University of Southampton
Cancer Sciences Division
School of Medicine

Genetic Profiling of Haematological Malignancies

Helen Parker

Thesis for the degree of Doctor of Philosophy

July 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

SCHOOL OF MEDICINE

Doctor of Philosophy

GENETIC PROFILING OF HAEMATOLOGICAL MALIGNANCIES

by Helen Parker

Acute lymphoblastic leukaemia (ALL) accounts for approximately 3% of all cancers, and leads to more than 4250 deaths a year. The different classes of leukaemia comprise numerous heterogeneous subgroups, which differ in their cellular and molecular characteristics, as well as their response to therapy. Accurate risk stratification is essential for tailoring of therapy, and achieving optimal outcome in contemporary treatment regimes. Based upon clinical features and cytogenetic and molecular diagnostics, the majority of ALL patients are assigned to one of the following prognostically significant subtypes having; *ETV6-RUNX1*, *BCR-ABL1* or *TCF3-PBX1* fusions, *MLL* rearrangements, high hyperdiploidy (HeH) with >50 chromosomes, hypodiploidy or T-ALL. Whilst these genetic abnormalities are important in leukaemogenesis, and provide diagnostic and prognostic markers, co-operating oncogenic aberrations are often required for the production of a full leukaemic phenotype. A number of additional aberrations have been identified within some cytogenetic subgroups, but the full complement of cooperating abnormalities, and their distribution within ALL subtypes remains to be defined.

In this study the genomic changes in a total of 94 ALL patients from the *ETV6-RUNX1* (n=34), iAMP21 (n=19) and unclassified (n=41) patient subgroups were characterised using cytogenetics, FISH, array comparative genomic hybridisation, molecular copy number counting, mutation analysis, qRT-PCR and targeted gene expression arrays. This approach allowed the identification of a range of large scale and submicroscopic aberrations, targeting multiple known and novel regions, including a number of genes. The elucidation of pathways and genes dysregulated in these malignant subgroups has provided further insight into the underlying cause of the disease phenotype. The *PDE9A* and *TBL1XR1* genes were implicated in the pathogenesis of iAMP21 and *ETV6-RUNX1* patients, respectively. Additionally, the potential prognostic significance of the *ADD3* gene in *ETV6-RUNX1* positive ALL was revealed.

Table of Contents

1: List of Figures

2: List of Tables

4: Acknowledgements

I would like to acknowledge the following people for their academic and personal support during the completion of my PhD:

Leukaemia Research for funding this work.

My primary supervisor, Professor Christine Harrison, my second supervisor, Professor Graham Packham and my co-supervisor, Dr Jon Strefford for their encouragement, guidance and advice in developing my research and writing skills.

All my colleagues at the LRCG for their guidance and training, specifically Dr Qian An.

Dr Fiona Ross, the haematology consultants at Southampton General Hospital, and all the regional Cytogenetic Laboratories for providing patient samples, without which, this project would not have been possible.

Dr Fiona Ross, Professor Nick Cross and Dr Ashutosh Wechaleker for their contribution to the design of the customized gene expression arrays.

Mr Richard Mitter for providing bioinformatics and biostatistics support.

Finally I would like to thank my husband and my family and friends for their ongoing support throughout the entirety of my PhD.

5: Abbreviations

aCGH	Array-based comparative genomic hybridisation
ADM-1	Aberration detection method-1
ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
ATM	Ataxia telangiectasia-mutated
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosomes
BFB	Breakage fusion bridge
bHLH	Basic helix-loop-helix
bHLHZIP	Basic helix-loop-helix zipper
BM	Bone marrow
bp	Base pairs
cAMP	Cyclic adenosine 3',5'-monophosphate
CCR	Continuous complete remission
cDNA	Complimentary DNA
cGMP	Cyclic guanosine 3',5'-monophosphate
CLL	Chronic lymphoblastic leukaemia
CML	Chronic myeloid leukaemia
CNA	Copy number aberrations
CNG	Copy number gain
CNL	Copy number loss
CNNLOH	Copy number neutral loss of heterozygosity
CNS	Central nervous system
CNV	Copy number variant
CR	Common reverse primer
CRA	Common region of amplification
CRD	Common region of deletion
cRNA	Complimentary RNA
Ct	Cycle threshold
CTP	Cytosine triphosphate
DBD	DNA binding domain
DLBCL	Diffuse large B-cell lymphoma
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSB	Double strand break
dsDNA	Double-stranded DNA
EBV	Epstein Barr virus

EF	External forward primer
EFS	Event free survival
EST	Expressed sequence tag
ETS	Erythroblast transformation specific
FAB	French-American-British
FISH	Fluorescence *In Situ* Hybridisation
FL	Follicular lymphoma
GF	Growth factor
GTP	Guanidine triphosphate
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HeH	High hyperdiploidy
HeL	Low hyperdiploidy
HIV	Human immunodeficiency virus
HLH	Helix-loop-helix
Ho	Hypodiploidy
HRE	Hormone response element
HRM	High resolution melt analysis
HSC	Haematopoietic stem cells
iAMP21	Intrachromosomal amplification of chromosome 21
ICL	Interstrand cross-linking agents
IF	Internal forward primer
IG	Immunoglobulin
IVT	*In vitro* transcription
kb	Kilobases
LB	Luria Bertani
LBD	Ligand binding domain
LDI-PCR	Long distance inverse-polymerase chain reaction
LRCG	Leukaemia Research Cytogenetics Group
MADGE	Microplate array diagonal gel electrophoresis
Mb	Megabases
MBCR	Major-breakpoint cluster region
mBCR	Minor-breakpoint cluster region
MCC	Multiple copy number counting
MDS	Myelodysplastic syndrome
MFISH	Multiplex Fluorescence *In Situ* Hybridisation
MM	Master mix
M-MLV	Moloney murine leukaemia virus reverse transcriptase
MRD	Minimal residual disease

mRNA	Messenger RNA
NCBI	National centre for biotechnology information
NCoR	Nuclear receptor co-repressor
NHEJ	Non-homologous end joining
NHL	Non-Hodgkin lymphoma
NHR	Nuclear hormone receptor
PB	Peripheral blood
PC	Plasma cell
PCM	Pericentriolar material
PCR	Polymerase chain reaction
PDE	Phosphodiesterases
Ps	Pseudodiploidy
qRT-PCR	Quantitative real time polymerase chain reaction
RA	Retinoic acid
RAR	Retinoic acid receptor
RE	Restriction enzyme
RNA	Ribonucleic acid
RPM	Revolutions per minute
RPMI medium	Roswell Park Memorial Institute medium
rRNA	Ribosomal RNA
RT	Reverse transcription
RT-PCR	Reverse transcription- polymerase chain reaction
RXR	Retinoic X receptor
siRNA	Small interfering RNA
SMRT	Silencing mediator of retinoic and thyroid hormone receptor
SNP	Single nucleotide polymorphisms
snRNA	Small nuclear RNA
T3R	Thyroid hormone receptor
TCR	T-Cell receptor
TF	Transcription factor
Tm	Melting temperature
tRNA	Transfer RNA
TSG	Tumour suppressor gene
TTP	Thymine triphosphate
UTP	Uracil triphosphate
UTR	Untranslated region
WHO	World health organisation

6: List of Genes

Gene Symbol	Gene Name
ABL1	*Abelson 1*
ADD3	*Adducin 3*
AFF1	*AF4/FMR2 family, member 1*
AIM1	*Absent in melanoma 1*
APP	*Amyloid beta (A4) precursor protein*
ATM	*Ataxia telangiectasia mutated*
ATP5O	*ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit*
B2M	*Beta-2-microglobulin*
BCL11B	*B-cell CLL/lymphoma 11B*
BCL2	*B-cell CLL/lymphoma 2*
BCL2L14	*BCL2-like 14 (apoptosis facilitator)*
BCR	*Breakpoint Cluster Region*
BLIMP1	*PR domain containing 1, with ZNF domain*
C21orf66	*Chromosome 21 open reading frame 66*
CASP2	*Caspase 2, apoptosis-related cysteine peptidase*
CBFA2T3	*Core-binding factor, runt domain, alpha subunit 2; translocated to, 3*
CBFB	*Core-binding factor, beta subunit*
CCL3L1	*Chemokine (C-C motif) ligand 3-like 1*
CCND1	*Cyclin D1*
CCND3	*Cyclin D3*
CD3D	*CD3d molecule, delta*
CD3G	*CD3g molecule, gamma*
CD44	*CD44 molecule*
CDKN1B	*Cyclin-dependent kinase inhibitor 1B*
CDKN2A/b	*Cyclin-dependent kinase inhibitor 2A*
CEBPA	*CCAAT/enhancer binding protein (C/EBP), alpha*
CLECSF6	*C-type lectin domain family 4, member A*
CRABP1	*Cellular retinoic acid binding protein 1*
CRABP2	*Cellular retinoic acid binding protein 2*
CSGlcA-T	*Chondroitin sulfate glucuronyltransferase*
DISC1	*Disrupted in schizophrenia 1*
DNAPTP6	*Viral DNA polymerase-transactivated protein 6*
EBF1	*Early B-cell Factor 1*
ELL	*Elongation factor RNA polymerase II*
ELN	*Elastin*

ERG	*V-ets erythroblastosis virus E26 oncogene homolog (avian)*
ETS2	*V-ets erythroblastosis virus E26 oncogene homolog 2 (avian)*
ETV6	*Ets variant gene 6*
EVL1	*Ecotropic viral integration site 1*
FGFR3	*Fibroblast growth factor receptor 3*
FLT3	*Fms-related tyrosine kinase 3*
FOXP1	*Forkhead box P1*
FYN	*FYN oncogene related to SRC, FGR, YES*
GAPDH	*Glyceraldehyde-3-phosphate dehydrogenase*
GPS2	*G protein pathway suppressor 2*
GRIK2	*Glutamate receptor, ionotropic, kainate 2*
GRP110	*G protein-coupled receptor 110*
GUSB	*Glucuronidase, beta*
HAP1	*Huntingtin-associated protein 1*
HIPK2	*Homeodomain interacting protein kinase 2*
HIST1H1E	*Histone cluster 1, H1e*
HIST1H2BH	*Histone cluster 1, H2bh*
HOX	*Homeobox*
HOXA5	*Homeobox A5*
HOXA9	*Homeobox A9*
HOXB5	*Homeobox B5*
IFNA1	*Interferon, alpha 1*
IGH	*Immunoglobulin heavy chain*
IGK	*Immunoglobulin kappa*
IGL	*Immunoglobulin light chain*
IKZF1	*IKAROS family zinc finger 1*
IL-3	*Interleukin 3*
JAK2	*Janus kinase 2*
KRAS	*V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*
LCK	*Lymphocyte-specific protein tyrosine kinase*
LMO1	*LIM domain only 1 (rhombotin 1)*
LMO2	*LIM domain only 2 (rhombotin-like 1)*
LYL1	*Lymphoblastomic leukemia*
MAD	*MAD homolog 5(Drosophila)*
MAF	*V-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)*
MAFB	*V-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)*
MAL	*T-cell differentiation protein*
MAX	*MYC associated factor X*

MEIS	*Meis homeobox 2*
MERTK	*C-mer proto-oncogene tyrosine kinase*
MLL	*Mixed Lineage Leukaemia*
MLLT1-10	*Mixed Lineage Leukaemia, translocated to 1-10*
MME	*Membrane metallo-endopeptidase*
MNT	*MAX binding protein*
MTAP	*Methylthioadenosine phosphorylase*
MUS81	*MUS81 endonuclease homolog (S. cerevisiae)*
MXI-1	*MAX interactor 1*
MYC	*V-myc myelocytomatosis viral oncogene homolog (avian)*
MYH11	*Myosin, heavy chain 11, smooth muscle*
NID2	*Nidogen2*
NRAS	*Neuroblastoma RAS viral (v-ras) oncogene homolog*
NUP98	*Nucleoporin 98kDa*
PAX5	*Paired box 5*
PBX1	*Pre-B-cell leukemia homeobox 1*
PDE9A	*Phosphodiesterase 9A*
PDGFRβ	*Platelet-derived growth factor receptor, beta polypeptide*
PLZF	*Promyelocytic leukemia zinc finger*
PML	*Promyelocytic leukemia*
PSCD2	*Pleckstrin homology, Sec7 and coiled-coil domains 2*
RAG1	*Recombination activating gene 1*
RAG2	*Recombination activating gene 2*
RARA	*Retinoic acid receptor, alpha*
RARβ	*Retinoic acid receptor, beta*
RARγ	*Retinoic acid receptor, gamma*
RB1	*Retinoblastoma 1*
RBM15	*RNA binding motif protein 15*
REXO1L1	*RNA exonuclease 1 homolog (S. cerevisiae)-like 1*
RUNX1	*Runt-related transcription factor 1*
RUNX1T1	*Runt-related transcription factor 1*
SH3GL1	*SH3-domain GRB2-like 1*
STIL	*SCL/TAL1 interrupting locus*
SPANXB	*SPANX family, member B2*
TAL1	*T-cell acute lymphocytic leukemia 1*
TAL2	*T-cell acute lymphocytic leukemia 2*
TBL1	*Transducin (beta)-like 1X-linked*
TBL1XR1	*Transducin (beta)-like 1 X-linked receptor 1*
TCBD	*Tubulin folding cofactor D*

TCF3	*Transcription factor 3*
TCF4	*Transcription factor 4*
TCR	*T-cell recpetor*
TCRA/D	*T-cell receptor α/δ*
TCRB	*T-cell receptor β*
TLX1	*T-cell leukemia homeobox 1*
TLX3	*T-cell leukemia homeobox 2*
TOX	*Thymocyte selection-associated high mobility group box*
TP53	*Tumour protein 53*
TRIM24	*Tripartite motif-containing 24*
VWF	*Von Willebrand factor*
WHSC1	*Wolf-Hirschhorn syndrome candidate 1*
XBP1	*X-box binding protein 1*
ZCCHC7	*Zinc finger, CCHC domain containing 7*

7: Introduction

7.1 The Human Genome

7.1.1 DNA, chromosomes and genes

Deoxyribose nucleic acid (DNA) is a double stranded helical molecule constructed from repeated nucleotides units. Each nucleotide is comprised of deoxyribose; a 5-carbon sugar, a phosphate molecule and one of four nitrogenous bases; adenine (A), cytosine (C), guanine (G) or thymine (T). Successive nucleotides are covalently linked via a 3',5'-phosphodiester bond. One end of the DNA strand terminates in a sugar residue with a 'free' 5' carbon and is therefore described as 5', whilst the opposite end is defined as 3'. The two DNA strands anneal via hydrogen bonds between laterally opposed bases. Adenine specifically binds to thymine (two hydrogen bonds), and cytosine specifically binds to guanine (three hydrogen bonds). The 5'→3' direction of one strand is opposite to that of its partner (1) (figure 1).

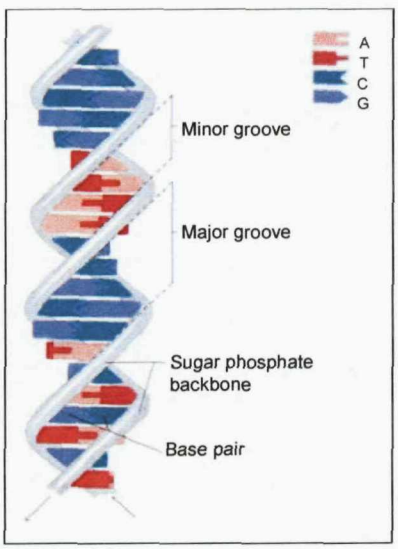


Figure 1. The structure of the DNA double helix. Adapted from www./chemistry.umeche. maine.edu/CHY431/Nucleic1.html

In humans, DNA is packaged into 46 chromosomes (diploid), consisting of 22 pairs of autosomes and two sex chromosomes (XX in females and XY in males). Chromosomes are necessarily highly condensed. During the first stage of packaging the DNA forms a nucleosome, a core of eight histone proteins around which 146 base pairs of DNA are coiled. A short length of spacer DNA separates adjacent nucleosomes, giving the DNA a 'string of beads' appearance. During interphase the nucleosomes are condensed into transcriptionally active chromatin fibres. As the cell cycle progresses toward metaphase, loops of chromatin fibres attach to a central protein scaffold, producing a transcriptionally inactive supercoiled structure (figure 2). Chromosomes are constricted at the centromere, which has an essential role during cell division, and are capped at either end by

telomeres. These are DNA–protein complexes which maintain the structural integrity of the chromosome, ensure replication of the extreme ends and help to position the chromosomes within the nucleus. Each chromosome has a short and long arm, termed the p and q arms, respectively. The alternating light and dark bands visualised using giemsa stain (refer to section 9.15) are numbered from the centromere, outwards toward the telomeres (2).

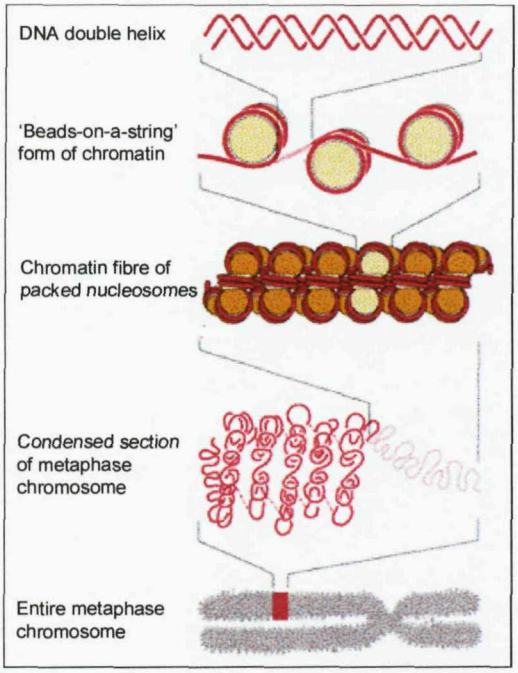


Figure 2. The packaging of DNA into metaphase chromosomes. Adapted from www.library.thinkquest.org/C004535/media/chromosome_packaging.gif

7.1.2 Cell cycle

The cell cycle consists of four stages. Under normal conditions the cell is maintained in stage G_1. When cell division is required for growth or repairs the cell enters S phase. Each of the 46 chromosomes consists of a single double helix or chromatid, which is replicated. By the end of S phase each chromosome consists of 2 chromatids. The process of DNA replication is semi-conservative. A helicase enzyme unwinds each double helix, and DNA polymerase catalyzes the synthesis of a new complementary daughter strand from each of the parent strands. After G_2, a gap phase to ensure replication is complete, the cell enters mitosis or M phase. Mitosis involves five stages;

1. Prophase. The chromosomes condense and the mitotic spindle is formed from tubulin-based microtubules and associated proteins.
2. Prometaphase. The nuclear membrane dissolves, allowing the chromosomes to migrate to the metaphase plate on the equatorial plane. Kinetochore fibres develop from some of the spindle microtubules and attach to the kinetochore; a large multi-protein complex attached to the centromeres. The remaining microtubules form polar microtubules.

3. Metaphase. The chromosomes are fully condensed and the 2 sister chromatids can be visualised. They are aligned on the metaphase plate via the kinetochore fibres.
4. Anaphase. As the cells progress to anaphase the centromeres split and the two chromatids of each chromosome are pulled to opposite poles of the cell along the polar microtubules.
5. Telophase. The chromatids reach the poles and begin to decondense. The nuclear membrane reforms and the cytoplasm divides. Finally, cytokinesis occurs; the cells cleave to give two identical daughter cells (figure 3a).

In the ovaries and testis specialised diploid cells divide by meiosis to give haploid gametes containing 22 autosomes and 1 sex chromosome. Meiosis is also preceded by one round of DNA replication, followed by two cell divisions; meiosis I and meiosis II. In meiosis I prophase involves five stages;

1. Leptotene. Chromosomes consist of 2 sister chromatids.
2. Zygotene. The maternal and paternal homologues of each chromosome join together in a synaptonemal complex to form a four-stranded structure called a bivalent.
3. Pachytene. The double helices of one maternal and one paternal chromatid break, cross over and recombine, forming a chiasma. Multi-protein complexes located at intervals along the bivalent mediate recombination. Chiasmata have a role analogous to the centromere in mitosis.
4. Diplotene. Homologues that are joined by chiasmata begin to separate.
5. Diakinesis. The separated homologues contract.

Following prophase I the chromosomes continue onto metaphase I and anaphase I as described above for mitosis. Meiosis II is identical to mitosis except that 23 chromosomes are involved instead of 46 (figure 3b).

Collectively, stages G_1 to G_2 of the cell cycle are referred to as interphase (2).

7.1.3 Transcription and translation

Each chromosome is subdivided into discrete units called genes. The human genome is thought to contain approximately 30,000 genes, which comprise approximately 25% of the genome. Only 10% of this is coding DNA, the rest is comprised of pseudogenes, introns, untranslated regions and non-coding repetitive DNA. The sequence of nucleotides within a gene encodes a specific amino acid sequence, which forms the basic primary structure of a protein. Each nucleotide triplet, or codon, encodes an amino acid. The genetic code is described as degenerate; there are 64 codons but only 20 amino acids, therefore most amino acids are encoded by multiple codons. The amino acid sequence of a protein is deciphered from the genetic code in the two-step process of transcription and translation.

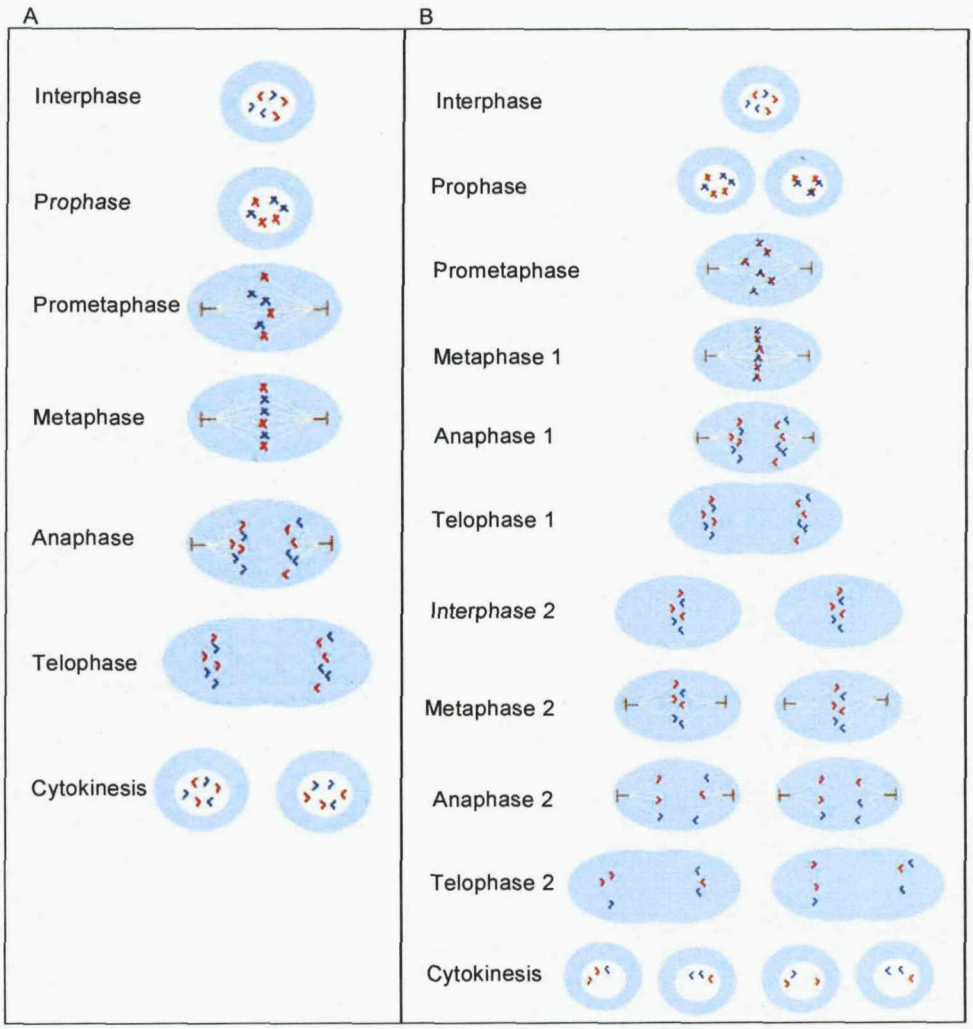


Figure 3. The processes of mitosis (A) and meiosis (B). Adapted from www.ba-education.demon.co.uk/for/science/dnabiology1.html

In transcription a single stranded ribonucleic acid (RNA) molecule is synthesised from the DNA template. Specific sequence elements, such as the TATA box or GC box, form a promoter region, typically 5' to the gene. Transcription factor proteins bind to these sequence elements, guide RNA polymerases to the DNA, and activate them. RNA polymerase unwinds the local region of the double helix, exposing the nucleotides. Moving along the template strand in a 5'→3' direction ribonucleoside triphosphates, which are complimentary to the template nucleotide, are added. The new RNA transcript has the same base sequence as the non-template strand (although thymine is replaced by uracil). The non-template and template strands are therefore referred to as the sense strand and anti-sense strand, respectively. As the RNA is synthesised the DNA releases it and the helix reforms. Ribosomal RNAs (rRNAs), transfer RNAs (tRNA) and small nuclear RNAs (snRNAs) are the functional product of the gene, and are not translated into proteins. RNA molecules that do code for proteins products are called messenger RNAs (mRNAs) (1).

Messenger RNA undergoes a number of post-transcriptional processes before translation occurs, including splicing, capping and polyadenylation. The coding sequence of a gene is interspersed with non-coding regions, termed exons and introns, respectively. Splicing removes the introns from the mRNA to leave a continuous coding sequence. During capping 7-methylguanosine is linked to the 5' nucleotide of the mRNA transcript. This protects the RNA from 5'→3' exonuclease attack, and facilitates splicing and transport of the molecule to the cytoplasm. Polyadenylation involves the cleavage of the 3' end of the mRNA and the addition of approximately 200 adenylate resides by poly(A) polymerase. The poly(A) tail facilitates transport to the cytoplasm and stabilizes the mRNA.
Following post-transcriptional processing the mRNA migrates from the nucleus to the cytoplasm, where it is bound by the ribosome; a large protein-RNA complex consisting of the 60S and 40S subunits. The ribosome scans the mRNA until it encounters the initiation sequence, AUG. Each codon from this point is bound by an anti-codon; a tRNA molecule with a complimentary sequence. Covalently bound to the tRNA molecule is a specific amino acid. Adjacent amino acids form peptide bonds via a condensation reaction catalysed by the RNA components of the ribosome. The tRNA molecules are released once their amino acids have bound to the growing polypeptide chain. This continues until the termination codon is encountered. The translated region is flanked by 5' and 3' untranslated region (UTR) which, together with the 5' cap and the poly(A)tail, is instrumental in recruiting mRNAs for translation and assists in binding and stabilization of the mRNA on the ribosome.
The finished polypeptide chain often undergoes post-translational processing such as phosphorylation, hydroxylation or the addition of carbohydrate or lipid groups (1).
Cancer is a disease caused by genetic aberrations, and an understanding of the structure and function of the human genome is essential to understanding the aetiology of cancer.

7.2 Cancer

Cancer refers to any one of a large number of diseases characterized by the development and uncontrolled proliferation of a malignant clone of cells. The stimulation of a proto-oncogene to a hyperactive oncogene, and/or the inactivation of an inhibitory tumour suppressor gene (TSG) results in deregulation of cellular differentiation, proliferation or apoptosis, leading to the establishment of the neoplastic cells. Proto-oncogenes encode proteins that function to control cell growth, differentiation, proliferation and death. They include growth factors (GF) and GF receptors, transcription factors (TF), signal transducers and cell death regulators (3). Tumour suppressor genes act as negative controls of cell growth. These stimulatory or inhibitory events are caused by: mutations or deletions of coding sequences, leading to deregulation of the mutated protein; gene amplifications, resulting in increased expression levels; or chromosomal rearrangements, which create fusion proteins with aberrant functions or juxtapose proto-

oncogenes with promoter elements. Factors that can initiate these changes include chemical carcinogens, radiation, viruses and spontaneous mutations caused by errors in DNA replication and repair. In 1971 Alfred Knudson proposed that cancer was the result of accumulated mutations to a cell's DNA (The Knudson Hypothesis) (4). It has since been found that carcinogenesis depends both upon the activation of an oncogene (the first 'hit') and the deactivation of a TSG (the second 'hit').

7.3 Haematopoiesis

Haematopoiesis is the formation and development of blood cells in the bone marrow from pluripotent haematopoietic stem cells (HSC). HSC have the ability to produce other undifferentiated progeny (self renewal) and progeny that develop into all blood cell types, through a series of committed progenitors. The HSC first differentiates into a myeloid or lymphoid progenitor cell and becomes committed to these lineages. Myeloid cells give rise to erythrocytes, granulocytes (neutrophils, eosinophils and basophils), macrophages and platelets. Lymphoid cells give rise to T and B lymphocytes and natural killer cells (figure 4). T- and B- lymphocytes mature in the thymus and bone marrow respectively. Mature unstimulated lymphocytes are maintained in the G_0 phase of the cell cycle. In response to antigen stimulation they proceed through the cell cycle to produce effector cells, which eliminate the antigen, or memory cells, which provide life long immunity (5). The B-lymphocytes synthesise and display membrane bound immunoglobulin molecules employed in the detection and elimination of antigen.

Self renewal, differentiation, proliferation and apoptosis are carefully regulated to maintain homeostasis. Deregulation of this balance leads to the development of leukaemia, of which there are many different subtypes (6).

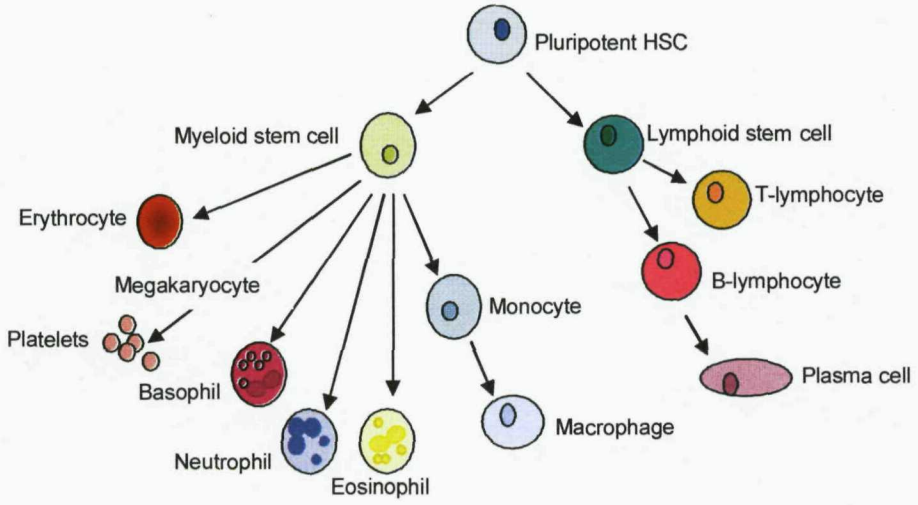


Figure 4. The differentiation of pluripotent HSC into mature blood cells, down either the lymphoid or myeloid cell lineage.

7.4 Acute lymphoblastic leukaemia (ALL)

ALL is a disease characterised by the clonal expansion of a transformed haematopoietic cell with differentiation arrested at an early stage in lymphoid cell development. It appears suddenly and progresses rapidly. The malignant immature lymphoid cells accumulate mainly in the bone marrow, but also in the peripheral blood of some patients, preventing the formation of adequate platelets, white and red blood cells. The initial transforming events occur in HSC (7). ALL accounts for 80% of childhood leukaemia. The incidence peaks at ages 3-4 years and 75% of cases present in patients less than 15 years old. It has a higher predominance in males compared to females, although this does reverse in infancy. It comprises a heterogeneous group of haematological malignancies composed of numerous subtypes that differ in their cellular and molecular characteristics, as well as their response to therapy (8).

7.4.1 Classification

Historically ALL was classified as subtype L1, L2 or L3 based solely on morphologic features, according to the French-American-British (FAB) classification system (9). However, this practice has largely been abandoned, in favour of the more advanced World Health Organisation (WHO) classification system, which considers the importance of immunophenotypic, cytogenetic and molecular features (10). The most important classification in ALL is now based upon cell lineage commitment and maturation (immunophenotype) of blasts. B-cell ALL is a malignancy of the lymphoblasts committed to the B-cell lineage, which accounts for 70-80% of cases (11). It is primarily a disease of childhood in which 75% of patients are under the age of six years. The most primitive cells in B-ALL are pre-pre-B cells, followed by early pre-B or common cells, and pre-B cells. B-cells are the most mature form. T-lineage ALL is a high risk malignancy of the thymocytes, which accounts for approximately 15% of childhood and 25% of adult ALL (12). It is sub-categorised as early, intermediate, late or mature. Each stage of cell maturation in both B- and T-ALL is determined according to the expression of specific cytoplasmic or surface markers (13).

Age, white cell count, morphology, immunophenotype and gender are all prognostic indicators, age being the most powerful. Infants and adolescents generally have a poor prognosis, whilst patients aged 3-5 years fare better. Karyotype is also an independent prognostic marker by which patients are classified. Since this discovery in 1978 (14), cytogenetic analysis has become a routine part of diagnosis and management in acute leukaemia (15). Contemporary treatment regimes for ALL achieve long term survival rates of approximately 80% in children and 40% in adults (16), by tailoring the intensity of the therapy to the patient's risk of relapse. Accurate diagnosis and sub-classification is therefore essential for appropriate risk stratification and optimal outcome (8).

7.4.2 Structural chromosomal aberrations in ALL

Structural aberrations commonly seen in ALL include translocations, inversions, deletions or duplications of whole or partial chromosomes and gene amplifications. Translocations and inversions frequently disrupt transcription factor (TF) genes, which bind to DNA and regulate the expression of cellular proteins. Leukaemia-specific translocations result in the deregulation of proteins involved in lymphoid cell growth, differentiation and survival by 1 of 2 mechanisms; in T and B-cell ALL TF genes are juxtaposed with transcriptionally active T-cell receptor (*TCR*) or immunoglobulin (*IG*) genes. In B-cell ALL translocations lead to illegitimate gene recombinations, creating chimeric fusion genes with oncogenic properties (7;17). Duplication or amplification of chromosomal material can lead to increased gene expression by increasing gene dosage, and deletion of genetic material typically results in the loss of tumour suppressor genes. Some examples are given below.

7.4.2.1 t(9;22)(q34;q11)

This translocation is characterised by the presence of a small chromosome known as the Philadelphia chromosome (Ph), after the city in which it was discovered. It was first identified in chronic myeloid leukaemia (CML) as the derived chromosome 22 (der(22)) in the reciprocal translocation t(9;22)(q34;q11) (18;19). The translocation produces the functional fusion gene *BCR-ABL1,* between the Abelson (*ABL1)* oncogene (9q34) and the breakpoint cluster region (*BCR)* gene (22q11) (figure 5). It accounts for 30% of adult ALL, 5% of paediatric ALL and 95% CML cases (17). It is associated with a poor outcome among all age groups in acute leukaemia (11;20).

ABL1 is a non-receptor tyrosine kinase found in the nucleus and cytoplasm of dividing cells, although the fusion protein is found only in the cytoplasm. Nuclear ABL1 is activated in association with ataxia telangiectasia-mutated (ATM) kinase to promote apoptosis (21). When located in the cytoplasm it is involved in signalling pathways regulating growth factor-induced proliferation (22). The BCR protein has serine/threonine kinase activity but little else is known about its function.

In acute leukaemia the breakpoint in *ABL1* can occur anywhere in the first 300kb of the gene, typically between exons 1b and 1a. In *BCR* it usually occurs in one of two regions; the minor breakpoint cluster region (m-BCR) between exons 1 and 2 (50-70% of cases), or within the major BCR (M-BCR) which spans exons 12 to 16, in the remaining cases. Fusions that occur in the m-BCR and M-BCR encode 190KDa proteins (p190), and 210KDa proteins (p210), respectively (20). A third BCR (μ-BCR) has been identified downstream of exon 19, which gives rise to a 230-kDa fusion protein in rare cases (23). All variants have enhanced tyrosine kinase activity, resulting in the activation of numerous signalling pathways, including the RAS, jun-kinase and PI-3 kinase pathway. Cell proliferation, differentiation, adhesion and apoptosis are dysregulated (17). BCR-ABL1 enhances DNA repair, control of cell cycle checkpoints and inhibits apoptosis,

resulting in resistance to cytostatic drugs and irradiation, and antagonizing the cytotoxic effects of DNA damaging agents, accounting for poor response of these patients to therapy. However, the tyrosine kinase inhibitor, Imatinib, has yielded promising results in terms of improved survival in CML (17;24), and trials are beginning to show similar promise in childhood ALL, in which the combination of Imatinib and intensive chemotherapy has resulted in a significant improvement in early event free survival (EFS) and reduction in early minimal residual disease (MRD) (25).

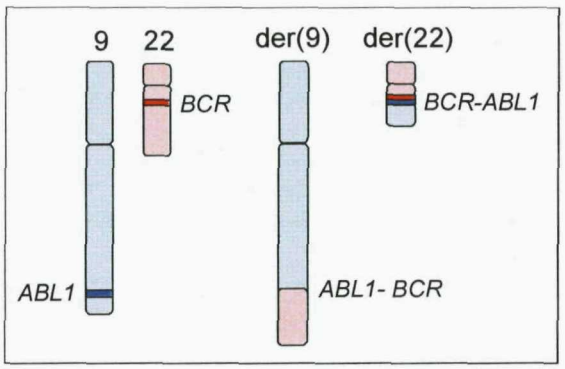


Figure 5. The positions of the ABL1 *and* BCR *genes on the normal chromosome 9 and 22, respectively, and the derivative chromosomes produced from the t(9;22)(q34;q11), with the* BCR-ABL1 *fusion gene on the der(22).*

7.4.2.2 t(12;21)(p13;q22)

The fusion gene, *ETV6-RUNX1,* is the molecular consequence of the cryptic t(12;21)(p13;q22) translocation, seen in approximately 25% of paediatric common or pre-B-ALL cases (26;27). In contrast to its high incidence among 3-6 year-olds (28), it has a frequency of less than 3% among adults, with most patients under 30 years old (29). For further details refer to section 10.2.

7.4.2.3 t(1;19)(q23;p13)

The translocation, t(1;19)(q23;p13), has an incidence of 6% in B-cell ALL and 25% in pre-B cell ALL. It juxtaposes the trans-activation domain of the *TCF3* gene (19p13) with the majority of the *PBX1* gene (1q23), to give the *TCF3-PBX1* fusion gene (although approximately 5-10% of patients with visible translocations, have no apparent molecular involvement of *TCF3* and *PBX1* (30).

TCF3 is a TF which plays a critical role in B-cell lymphopoiesis. It contains a basic helix-loop-helix (bHLH) domain which is responsible for DNA binding and protein dimerization, and two transcriptional transactivation domains (17). *PBX1* is a homeobox-containing gene thought to influence transcription by binding DNA co-operatively with other TF on target gene promoters (31).

TCF3-PBX1 is localized in the nucleus where it binds to DNA and acts as a transcriptional activator. The binding of PBX1 and TCF3-PBX1 to DNA is controlled by

the interaction of PBX1 with a subset of HOX proteins. Because these proteins direct TCF3-PBX1 to DNA sites recognised by HOX-PBX1 complexes, the fusion protein probably disrupts haematopoietic differentiation by interfering with genes that are usually regulated by HOX proteins. The haplo-insufficiency of *TCF3* as a result of the translocation may also contribute to leukaemogenesis.
The adverse prognostic implications of this fusion protein have been overcome with more aggressive chemotherapy (17).

7.4.2.4 11q23 (MLL) *rearrangements*

The chromosome band, 11q23, is a heritable fragile site. Structural rearrangements involving 11q23, the location of the *MLL* gene, occur in approximately 80% of infant ALL, and are also seen in AML and therapy related leukaemias.
The MLL protein, which is localized in the nucleus, has two central zinc finger domains and three N-terminal A-T hooks that appear to bind A-T rich sequences in the minor groove of DNA. TF activity is suggested by the homology of MLL to trithorax, a TF in *Drosophilia.* Knockout mouse models indicate that MLL plays a role in haematopoiesis and *HOX* gene regulation (11). Deregulation of *HOX* gene expression by MLL fusion proteins may contribute to leukemogenesis.
At least 51 partner chromosomal loci participate in 11q23 translocations including 1q32 (*MLLT5*), 6q27 (*MLLT4*), 9q22 (*MLLT3*), 10p12 (*MLLT10*), 17q21 (*MLLT6*), 19p13 (*MLLT1*) and Xq13 (*MLLT7*). Many of these are disease specific. The most common rearrangements in ALL are the t(4;11)(q21;q23), t(9;11)(p22;q23) and t(11;19)(q23;p13.3) which juxtapose *MLL* with *AFF1, MLLT3* and *MLLT1* respectively. The proteins encoded by these genes contain nuclear targeting sequences and may function as transcriptional activators. (32). The prenatal origins of *MLL-AFF1* rearrangements were confirmed in three patients by the presence of genomic *MLL-AFF1* sequences unique to the patients leukaemic clone in neonatal blood spots (33).
Although *MLL*-rearranged ALL is a high-risk subgroup, the prognosis is influenced by the partner gene involved, and the patient's age. Expression of the tyrosine kinase FLT3 is a frequent finding in association with these aberrations and FLT3 inhibitors have shown potential as therapeutic agents (11).

7.4.2.5 Mature B-ALL

Mature B-cell ALL is associated with the t(8;14)(q24;q32) translocation and the variants t(8;22)(q24;q11) and t(2;8)(p12;q24). 8q24 is the location of the *MYC* oncogene, which encodes a bHLH/leucine zipper (bHLHZip) TF. These translocations juxtapose *MYC* with the transcriptionally active *IG* heavy chain (*IGH*), *IG* lambda (*IGL*) and *IG* kappa (*IGK*) genes, respectively.
MYC is normally inactive in non-dividing cells but expression is up-regulated when cells

are stimulated to enter the cell cycle. MYC dimerizes with numerous proteins, including MAX, through its bHLHZip domain. The MYC:MAX heterodimer binds to target genes and activates transcription and cell proliferation. However, MAX also dimerizes with MAD, MXI-1 and MNT. These complexes inhibit transcription both directly and by competing with MYC for MAX. The t(8;14) translocation results in over-expression of MYC, increasing the MYC:MAX complexes relative to MAX:MAD, thus increasing transcriptional activation. MYC target genes are involved in cell growth, division, apoptosis, metabolism, adhesion and motility (17).

7.4.2.6 T-ALL

Recurring translocations in T-ALL involve the *TCRA/D* (14q11) and *TCRB* (7q34) genes. The t(1;14)(p32-34;q11) translocation, seen in 3% of T-ALL cases, juxtaposes the TF *TAL1* with *TCRA/D*, leading to dysregulation of *TAL1* and uncontrolled growth of primitive T-cells. Other rearrangements involving *TAL1* are seen in approximately 25% cases. For example, the combination of 5' *TAL1* to *STIL* in the *STIL-TAL1* fusion gene is the result of an interstitial deletion of 90kb of genomic DNA from chromosome 1p32 (34). A higher percentage of cases aberrantly express *TAL1*. TAL1 is a member of the class II family of bHLH proteins and dimerizes with class I bHLH proteins such as TCF3. Mouse models have shown that TAL1 is required for erythropoiesis and the early stages of lineage commitment of pluripotent embryonic HSC (17). Other translocations involving *TCRA/D* include t(8;14)(q24;q11), t(10;14)(q24;q11), t(11;14)(p13;q11) and t(11;14)(p15;q11), involving *MYC*, *TLX1, LMO2* and *LMO1,* respectively (34). The homeobox gene *TLX1* encodes a homeodomain TF, while *LMO1* and *LMO2* encode cysteine rich proteins involved in protein-protein interactions (17).

Translocations of 7q34 include t(1;7)(p34;q34), t(7;9)(q34;q32) and t(7;11)(q35;p13), which result in the fusion of *TCRB* to *LCK*, *TAL2* and *LMO2,* respectively.

The cryptic translocation, t(5;14)(q35;q32), results in the fusion of *TLX3* (5q35) to *BCL11B* (14q32), leading to up-regulation of *TLX3* (35). *TLX3* belongs to the class II homeobox gene family, whilst *BCL11B* encodes a transcription factor essential in T-cell development (36). Abnormalities of the short arm of chromosome 9 (9p) and 12p, as well as deletions of the long arm of chromosome 6 (del(6q)) are also common in, but not specific to T-ALL (34).

7.4.2.7 Intrachromosomal amplification of chromosome 21 (iAMP21)

iAMP21 is a recently defined recurrent abnormality, with a 1.5% incidence in paediatric B-cell ALL. It is a distinct subgroup in older children and is associated with a poor prognosis. It is defined as the intrachromosomal duplication of chromosome 21, with amplification of the *RUNX1* gene, located to 21q22 (37) (refer to section 10.3 for further details).

7.4.2.8 Other aberrations

In addition to the t(12;21)(p13;q22) translocation, other aberrations involving 12p are common in ALL. In a study of childhood leukaemia, Raimondi et al (38) identified a structural aberration of 12p11-p13 in 11.5% of patients. A total of 69 translocations, 20 different partner chromosomes and 43 distinct breakpoints were observed. These involved deletions, inversions, the formation of i(12)(q10) and translocations including dic(9;12)(p11;p12), dic(7;12)(p11;p11-12), dic(7;12)(q36;p13) and t(12;13)(p13;q14). These translocations were associated with additional chromosomal changes in 78% of cases (38).

Deletions of 6q occur in 4-13% of pediatric cases, with breakpoints typically localized to 6q15 and 6q21. The short arm of chromosome of 9 is deleted in 30% of cases and the genes *CDKN2A, PAX5, PSCD2* and *IFNA2* are frequently targeted. The recurrent aberration, dic(9;20)(p13;p11), mostly found in pre-B ALL, is also associated with a deletion of 9p (11).

7.4.3 Numerical chromosomal abnormalities in ALL

Numerical aberrations involve the gain or loss of entire chromosomes. Patients are classified by chromosome number as follows:

Clones harbouring between 51 and 65 chromosomes are defined as high hyperdiploid (HeH). The most frequently gained chromosomes are 21, 6, X, 14, 4, 18, 17, 10 and 8. Structural aberrations, such as duplications of 1q, deletions of 6q, and i(17)(q10), are also observed in approximately 50% of cases. This subgroup is seen in 25-30% of paediatric patients and is associated with a good prognosis, although it is linked with the occurrence of early relapse. Patients with 56-67 chromosomes fare better than those with 51-55 chromosomes. In rare cases, clones can also be near triploid (~69 chromosomes) or near tetraploid (~92 chromosomes) (11).

Clones with 47-50 chromosomes are described as low hyperdiploid (HeL), which occurs in 10-15% of childhood ALL, and is associated with an intermediate prognosis. Chromosomes 21, X, 8 and 10 are the most frequently gained. Structural aberrations are also common and involve chromosome arms 1q, 6q, 12p and 19p (11).

Clones with 46 chromosomes and an abnormal karyotype are described as pseudodiploidy (Ps). The incidence in children is approximately 40%, while it is higher in adults. This group was associated with a high mortality rate, although prognosis was improved by modern multi-agent chemotherapy. Many Ps clones harbour structural abnormalities, in particular the t(4;11) and t(9;22), which are associated with high drug resistance and may explain the poor outcome associated with Ps clones. Diploid cells appear to be cytogenetically normal. They have an incidence of approximately 30% in T-

cell ALL and 18% in B-ALL. Many of these cases harbour sub-microscopic genetic changes undetectable by cytogenetics (11).

Hypodiploid (Ho) clones have fewer then 46 chromosomes and are found in 5% of ALL. These cases may be divided into three distinct groups (39). In the 42-45 chromosome group, the chromosomal deficiency is usually the result of an unbalanced translocation, the loss of a whole chromosome or the formation of dicentric chromosomes. Sex chromosomes are commonly lost, and monosomy 7 is seen infrequently. The most common recurrent dicentric aberrations seen in this group are dic(9;12), dic(9;20) and dic(7;9). Low hypodiploidy (35-44 chromosomes), and near haploidy (24-34 chromosomes) are extremely rare and most frequently occur in adults and children, respectively. Near haploid clones usually retain both sex chromosomes with addition of chromosomes 14, 18 or 21 to the haploid chromosome set. Structural abnormalities are rare within this group. Patients with fewer then 44 chromosomes have a poor outcome when treated with conventional therapies (11;40).

To date, little is known about the mechanisms by which numerical abnormalities affect pathogenesis.

7.5 Acute Myeloid Leukaemia (AML)

AML differs from ALL in the lineage commitment of the transformed haematopoietic cells. The leukaemic population consists primarily of myeloid blasts. It is the commonest form of leukaemia in adults with a median age of 60 (41), but also accounts for approximately 20% of paediatric leukaemias (42). The incidence in children peaks between ages 2-5 years (43). AML is characterised by numerous heterogeneous subgroups with specific chromosomal aberrations that commonly target TF involved in lineage commitment, survival, proliferation and differentiation pathways. There are marked differences in cure rates between the genetic subtypes, but most are below 60% (44).

According to the FAB classification system, AML is assigned to one of eight subsets, M0-M7, according to lineage commitment along the granulocyte, monocyte, erythroid or megakaryocytic lineage, the degree of maturation and the presence of recurrent cytogenetic and molecular features (Table 1).

AML subtype	**Defining characteristics**
M0 (AML with minimal evidence of myeloid differentiation)	The presence of at least 30% blasts with minimal evidence of myeloid differentiation. It is associated with a higher frequency of *RUNX1* mutations and other unfavourable cytogenetic aberrations, such as monosomy 7, inv(3q) and del(9q), than other AML subtypes.
M1 (AML without maturation)	90% or more of the non-erythroid marrow cells are immature myeloid blasts.

M2 (AML with maturation)	The marrow cells are comprised of at least 30% myeloblasts and 10% of the non-erythroid cells are maturing granulocytes.
M3 (Acute promyelocytic leukaemia)	The predominance of promyelocytes and the t(15;17)(q22;q12) translocation.
M4 (Acute myelomonocytic leukaemia)	Cells manifests granulocytic and monocytic features with 20-79% monocytic marrow cells
M4Eo (M4 variant)	5% of the non-erythroid marrow is comprised of abnormal eosinophilic precursors
M5 (Acute monoblastic/monocytic leukaemia)	80% or more of the non-erythroid marrow cells are monocytic. These cases are further classified by the degree of monocytic differentiation; a predominance of monoblasts indicates M5a, and maturation to monocytes denotes M5b
M6 (AML with predominant erythroid differentiation)	Predominant erythroid differentiation, accounting for at least 50% of the marrow nucleated cell population. Myeloblasts account for 30% of the non-erythroid cells.
M7 (Acute megakaryoblastic leukaemia)	The marrow is infiltrated by megakaryoblasts. It is often associated with Down syndrome. Certain cytogenetic aberrations, such as t(1;22)(p13;q13) (*RBM15-MAL*) are observed in the paediatric age group. M7 is genetically distinct from other subgroups and is associated with a poor prognosis.

Table 1. The defining characteristics of the seven AML subtypes (10).

The main recurrent structural aberrations include;

- The t(15;17)(q22;q12) translocation (M3) results in the fusion gene *PML-RARα* on the der(15). *PML* (15q22) has roles in differentiation, senescence and apoptosis and may be a tumour suppressor. *RARα* (17q12) is a TF essential for normal haematopoiesis. *PML-RARα* is pathogenic because it is differentiation-blocking during the promyelocytic stage of myeloid maturation (45). It inhibits apoptosis in haematopoietic cells (46).
- The inv(16)(p13q22) and t(16;16)(p13;q22) (M4Eo) generates the fusion gene, *CBFβ-MYH11*. CBF*β* is a subunit of core binding factor, a TF in early haematopoietic stem cells. It associates with and regulates the RUNX1 subunit, which binds to CBF target genes. The fusion transcript is found in the cytoplasm where it dimerizes with RUNX1. MYH11 associates with the actin cytoskeleton which prevents RUNX1 entering the nucleus and activating target gene

promoters (11;32).

- The t(8;21)(q22;q22.3) translocation (mostly M2, M1 and M4) occurs in 12% of paediatric and 6% of adult AML cases (11). It results in the fusion of the 5' region of *RUNX1* (21q22.3) to the 3' region of *RUNX1T1* (8q22), creating a chimeric gene on the der(8). RUNX1-RUNX1T1 interacts with CBFβ and represses RUNX1 mediated transcriptional activation. It is thought that expression of this fusion transcript alone is not sufficient for leukaemogenesis (32).
- 11q23 aberrations (M4, M5) are the most common cytogenetic abnormality in childhood AML, particularly frequent in infants. At least 30 partner genes have been documented in AML. *MLLT3* on t(9;11)(p22;q23), *ELL* on t(11;19)(q23;p13.1) and *SH3GL1* on t(11;19)(q23;p13.3) are the most common. AML is often seen as a secondary leukaemia in patients initially treated for ALL with epipodophyllotoxins and etoposide. Approximately 85% of treatment related leukaemias have 11q23 aberrations, and their prognosis is extremely poor (11).

Other cytogenetic aberrations seen in AML involve 8p11.2, 11p15.5, 12p, 5q and numerous breakpoints in chromosomes 16, 21 and 22. The inv(3)(q21q26) and ins(3;3)(q21;q26) rearrangements juxtapose *EVI1* (3q26) with genes on 3q21 (11). Duplication of chromosome 21 is also seen in AML, typically associated with a complex karyotype (47). Complex karyotypes are defined as having more than three chromosomal abnormalities (48) and are associated with a poor outcome.

At least two cooperating genetic lesions are thought to be required for the development of AML; one that impedes haematopoietic cell differentiation and one that imparts a survival or proliferation advantage (type I and type II, respectively). Mutations of the tyrosine kinase gene, *FLT3*, result in constitutive tyrosine kinase activity and are the most common secondary abnormality in AML (32).

Numerical aberrations are unusual in AML and are typically associated with a recurrent structural abnormality. Trisomy 8 is the most frequent numerical aberration, followed by monosomy 7, including the associated partial deletion of 7q, and acquired trisomy 21 (11).

7.6 Lymphoma

Lymphomas are a heterogeneous group of malignancies that are broadly classified as Hodgkin or Non-Hodgkin lymphoma (NHL). Several cytogenetic abnormalities are observed in almost all types of lymphoma, including translocations involving chromosome 1, deletions of 6q21 and 6q27, trisomy 12 and translocations of 14q32. Lymphomas are often cytogenetically complex. The presence of complex clones is associated with a poor prognosis regardless of the particular abnormalities present.

Hodgkin lymphoma is the most common type of lymphoma, predominating in young males. It is not associated with any specific chromosome aberrations.
NHLs are subdivided according to the cell size, organ involvement, BM infiltration, how the cells are distributed and whether or not they are cleaved. Common subtypes include;

- Burkitt's lymphoma, an aggressive immature B-cell malignancy. Almost all cases have one of the three translocations involving *MYC*; t(8;14)(q24;q32), t(2;8)(p12;q24) or t(8;22)(q24;q11), as seen in mature B-ALL (section 7.4.2.5).
- Diffuse large B-cell lymphoma (DLBCL) accounts for 30% of lymphoid malignancy and targets a broad age range. The most common translocations associated with DLBCL are the t(3;22)(q27;q11), t(2;3)(p12;q27) and t(3;14)(q27;q32), involving *BCL6* (3q27).
- Follicular Lymphoma (FL) is the most common type of lymphoma in adults, accounting for 30% of all NHL. FL typically involves the t(14;18)(q32;q24), translocation which juxtaposes *IGH@* and *BCL2*. Less common translocations fuse *BCL2* to *IGL* or *IGK*
- Mantle cell Lymphoma is rare and refractory to treatment. Mutations of *IGH@* are seen in 25% of cases and it is strongly associated with the t(11;14)(q13;q32) translocation, which results in constitutive *CCND1* expression (49).

7.7 Chronic Lymphocytic Leukaemia (CLL)

CLL is the most common leukaemia in elderly adults and arises from an accumulation of non-functional B-cells in the peripheral blood (PB) and bone marrow (BM). Treatment may increase remission time but an allogenic stem cell transplant is the only therapy with curative potential (50). This disease is characterised by deletions, which most frequently target 13q14, 11q22-q23, 17p13 and 6q21. Gains of 12q (usually trisomy 12), 8q and 3q are also observed. Deletions of 17p13 and 11q22-q23 are thought to target the *TP53* and *ATM* genes, respectively. They are independent prognostic markers associated with rapid disease progression and short survival. Single abnormalities are most common, but multiple rearrangements can occur, often associated with disease progression (51).

7.8 Chronic Myeloid Leukaemia (CML)

CML is a BM stem cell disease characterised by the t(9;22)(q34;q11) translocation, as described for ALL (section 7.4.2.1). It accounts for 15-20% of adult leukaemias with a median age at diagnosis of 53 years. CML is associated with a very low mortality rate due to the availability and efficacy of tyrosine kinase inhibitors, such as Imatinib, which target the *BCR-ABL1* oncogene. However, the only curative therapy is allogenic BM transplant. The disease evolves through three stages; in chronic phase patients are most responsive to kinase inhibitor therapy. As they enter accelerated phase, patients begin to fail in their response to treatment. The number of peripheral blasts and basophils increases and in

50-80% of patients there is evidence of clonal evolution. Cytogenetic abnormalities, in addition to the t(9;22)(q34;q11), are observed, such as a second Ph or trisomy 8. Blast phase is characterised by a much higher percentage of blasts and transformation to acute leukaemia. In 60-70% of cases the leukaemia is myeloid in origin, and in the remaining 20-30% the blast crisis is lymphoid derived. In both the outcome is extremely poor (52).

7.9 Myeloma

Myeloma is an incurable malignancy characterized by the clonal expansion of transformed mature plasma cells (PC) at multiple sites in the BM. It accounts for approximately 10% of haematological neoplasia (53;54). Translocations involving *IGH@* at 14q32 occur in 60-70% of cases. To date five recurrent partners have been identified; *CCND1* (11q13), *CCND3* (6p21), *MAF* (16q23), *MAFB* (20q11) and *FGFR3* and *WHSC1* (4p16) (both genes are dysregulated by the t(4;14)(p16;q32)) (55). Most tumours directly or indirectly dysregulate at least one of the cyclin genes. Secondary events, which occur late in disease progression, include rearrangements of *MYC*, and mutations of *NRAS*, *KRAS, FGFR3 and TP53* (55). Patients are also classified according to ploidy. Hyperdiploid karyotypes (>47 chromosomes) are characterized by recurrent gains of 3, 5, 7, 9, 11, 15 and 19. Non-hyperdiploid karyotypes are sub-divided into hypodiploid, pseudodiploid or hypotetraploid karyotypes (56).

7.10 Array-based comparative genomic hybridisation (aCGH) of haematological malignancies

The detection of chromosomal abnormalities by cytogenetic and molecular techniques, such as G-banded karyotyping, fluorescence in situ hybridisation (FISH), and RT-PCR, is an essential component in the diagnosis, classification and risk stratification of leukaemia. However, G-banding has a limited resolution of 5-10 Mb, dividing cells are required, and a high percentage of patients lack clinically relevant cytogenetic aberrations, whilst FISH analysis and RT-PCR is restricted to the screening of previously identified aberrations. High resolution molecular techniques, such as comparative genomic hybridisation (CGH), and gene expression arrays, are continually evolving and have overcome some of these limitations. The analysis of the cancer genome with these techniques has revealed a plethora of previously unidentified chromosome aberrations, has highlighted pathways and cellular mechanisms that may be intrinsic to cancer development and progression, and revealed novel diagnostic and prognostic markers and potential therapeutic targets.

CGH, a technology developed in 1992 (57), allows detection of whole genome copy number changes, in a single experiment. Genomic DNA from test and reference cell populations are differentially labeled with fluorescent dyes, and competitively hybridised

to a glass slide containing a representation of the normal genome, known as the target (58;59). Until recently, normal metaphase spreads provided this target and copy number aberrations were mapped to a chromosomal location. This technique had a resolution similar to G-banding and was highly dependent on good quality metaphase spreads. Advances in our knowledge of the human genome, and the ability to spot DNA sequences onto microarray slides at high density have allowed the chromosome template to be superseded by array-based methods. Initially bacterial artificial chromosomes (BAC) or other large insert genomic clones were used to produce array elements approximately 200kb in length, providing an average resolution of 1Mb. The BAC clones could then be used for FISH confirmation of CGH data. The optimization of arrays using PCR products, cDNA and oligonucleotides has dramatically increased the sensitivity and resolution of the technique. Current platforms, such as the NimbleGen HD2 array, containing 2.1 million probes, can offer breakpoint mapping at gene-level resolution. Additionally, platforms such as the Affymetrix Genome-Wide SNP Array 6.0, which features over 1.8 million markers, include probes for the detection of copy number variation, and single nucleotide polymorphisms (SNP) which allows the detection of copy number neutral loss of heterozygosity (CNNLOH).

Following hybridisation, lasers scan the surface of the slide, exciting the dyes to fluoresce. The ratio of the fluorescence intensities gives a measure of the relative copy number variation between the reference and test cell populations (54;58). Normalization of the data sets the modal ratio for the experiment to a standard value (usually one) and accounts for systematic experimental variation, such as unequal dye incorporation, detection efficiencies and background fluorescence. The signal intensities are arbitrary values so techniques such as FISH or molecular copy number counting (MCC) are used to confirm results and generate an accurate DNA copy number.

Limitations of aCGH include its inability to identify CNNLOH and balanced rearrangements that do not result in CNA. They cannot provide information on the type of unbalanced structural rearrangements and detection of ploidy changes is difficult (60). Normalization is affected by the presence of a highly abnormal genome or contaminating normal cells, and aberrations present in a small subpopulation of cells (less then 40%) may not be detected (61).

This technique has been widely applied to the study of the malignant genome, including a range of haematological malignancies. Studies focusing on ALL have demonstrated a higher level of karyotype complexity then anticipated from cytogenetic analysis (62). Deletions most frequently target 9p21.3, 6q, 12p and 11q. Recurrent gains target 21q21.2-q22.2 and various loci on chromosome 17. This high resolution technique has also facilitated the identification of submicroscopic CNA, targeting single genes such as *ETV6*, *PAX5*, *EBF1*, *RB1*, *TCF3* and *IKZF1* (62-66).

Copy number changes can have a significant impact on the expression levels of genes in the targeted region and are frequently associated with disease (62;67). However, variation in DNA sequence between normal individuals is also a common occurrence which contributes to genetic diversity. Genetic variation in the human genome can take many forms including SNP and copy number variations (CNV) such as deletions, insertions and duplications, ranging from several kilobases to hundreds of kilobases in length (68-70). At least 1,447 regions of CNV have been identified, accounting for approximately 12% of the genome. Although preferentially located outside genes and ultra-conserved elements, they nevertheless encompass or flank thousands of putatively functional sequences, including known disease related genes (70). Although they occur in healthy individuals, many studies have noted an association between these sequence and genetic instability. Regions flanked by repetitive sequence are more susceptible to the formation of chromosomal rearrangements and have been implicated in more then 25 recurrent genomic disorders. Although the nature of the relationship between CNV and disease phenotype remains to be elucidated, it is also believed that by targeting important genes including oncogenes and tumour suppressor genes, CNV may directly influence disease initiation or progression, be linked to disease susceptibility, or be an important cause of variable response to drugs. So far they have been implicated in a number of diseases and disorders, including autism (71), some neurological disorders (68), prostate cancer (72) and susceptibility to lung cancer (73). Gonzalez *et al* (74) found that increased copy number of the immune system gene, *CCL3L1*, protected individuals against HIV demonstrating that CNV may also exert a positive influence.

7.11 Gene expression profiling of haematological malignancies

The carefully balanced regulation of gene expression confers unique properties and functions to different cell types, and is essential to the differentiation, development and maintenance of any organism. Deregulation of gene expression is often associated with disease.

Gene expression arrays enable the quantitative measurement of the expression of thousands of genes in a single experiment. Complementary DNA (cDNA) from test and reference cell populations are differentially labelled with fluorescent dyes and competitively hybridised to DNA representations of expressed sequences. Alternatively, a single labelled sample can be hybridised, and the data from separate arrays compared. The arrays are manufactured in 1 of 2 ways. The DNA sequences, which can be PCR products, plasmids or oligonucleotides, are mechanically spotted onto glass slides at a density of approximately 23000 features per slide. Photochemical oligonucleotide arrays are constructed by photolithography, which synthesizes the sequences one base at a time directly onto the slide (75;76). The optimization of this technique has allowed the development of arrays that represent all currently known or deduced exons (77).

Following hybridisation, lasers scan the surface of the array and excite the dyes to fluoresce. The ratio of the fluorescence intensities gives a measure of the relative gene expression between the two cell populations. Normalisation of the data sets the modal ratio for the experiment to a standard value and accounts for systematic experimental variation.

This technology can be used to deduce probable functions of new genes based on similarities in expression patterns to known genes, expand the size of existing gene families and reveal new ones. Increasing our understanding of how genes interact in functional pathways and how their expression is regulated in healthy systems will help to determine how altered gene expression contributes to disease. Arrays can aid the discovery of new drug targets and clarify mechanisms of drug action, be applied to disease diagnosis and classification, provide information on prognoses and enable patients to receive a more targeted therapy (75).

Whilst microarray technology is a high throughput, rapid method for investigating gene expression changes, limitations include the expense, the large quantity of RNA required and the complex analysis of huge volumes of data, requiring bioinformatics expertise.

This technique has been extensively applied to the study of haematological malignancies with the aim of enhancing risk-stratification and tailoring therapy appropriately, identifying candidate genes for novel therapies and enhancing our understanding of the disease mechanisms. In ALL distinct expression profiles for each of the prognostically significant subgroups: T-ALL, *ETV6-RUNX1*, *TCF3-PBX1*, *MLL* rearrangements, *BCR-ABL1* and HeH, and a novel subgroup, have been identified. Class assignment according to these profiles was, on average, 97% accurate. Expression signatures that accurately predict relapse have also been identified (8;78).

8: Research Aims

In recent years the application of increasingly high resolution techniques to the study of the leukaemic genome has revealed many previously unidentified aberrations in leukaemia patients, and advanced our understanding of the pathways and mechanisms involved in disease pathogenesis. However, much currently remains unknown. We hypothesized that the application of one of the highest resolution aCGH platforms available at the time, to study the genomic events occurring in a large cohort of leukaemia patients would allow us to detect previously unidentified genomic changes and cast more light on the underlying causes of leukaemogenesis. The aims of this project were;

- To characterize previously identified, and discover novel genomic changes in a range of acute leukaemia patients, using cytogenetics, FISH, aCGH and targeted gene expression arrays.
- Provide further insight into the underlying cause of the disease phenotype and identify novel diagnostic and prognostic markers, as well as potential therapeutic targets, using additional supporting techniques such as molecular copy number counting (MCC) and qRT-PCR.
- Compare and evaluate these techniques to provide information about their relative qualities, such as sensitivity, resolution and reproducibility, and their potential use in a clinical setting.

9: Materials and Methods

Instructions for the preparation of all solutions, and details of all kits used in these techniques can be found in Appendices 1 and 2, respectively.

9.1 Patient samples

Patients recruited to the Medical Research Council/National Cancer Research Institute childhood (UKALLXI, ALL97 and ALL2003) or adult (UKALLXII) treatment trials for ALL are entered onto the Leukaemia Research Cytogenetics Group (LRCG) Karyotype Database (79). Currently, clinical, cytogenetic and FISH data are stored for more than 20,000 patients. This resource provides information and material for research projects, and the possibility of comparing pre-treatment findings and patient outcome. Local ethical approval was granted to extract RNA and DNA from diagnostic bone marrow samples, from patients referred to Southampton General Hospital with ALL. Karyotype data was provided by Salisbury District Hospital. DNA and RNA/cDNA samples were also sent from Bristol Regional Cytogenetics Centre, West Midlands Regional Genetics Service, Merseyside and Cheshire Genetics Laboratory, Liverpool Women's Hospital, North Trent Cytogenetics Laboratory, The Northern Institute for Cancer Research and St Bartholomew's Cytogenetics department, through the LRCG Karyotype Database.

9.2 Tissue Culture

Tissue culture is a technique for the *in vitro* proliferation of animal cells in a nutrient medium, for extended periods of time. It provides an inexhaustible supply of heterogeneous material such as DNA, RNA and protein, and the cells can be easily manipulated i.e. by adding drugs to the media. Culture conditions mimic the cells' natural environment. The incubator maintains a constant temperature and oxygen level, and controls pH by adjusting carbon dioxide levels. The supplemented media contains all the necessary nutrients for the cells to grow: amino acids (including the essential amino acid L-glutamate), vitamins, inorganic salts, insulin, fibroblast growth factor, epidermal growth factor and antibiotics to prevent bacterial growth.

1. All culturing was carried out in a laminar flow hood using aseptic techniques. The hood and all instruments were cleaned with 70% ethanol prior to use.
2. 500ml Roswell Park Memorial Institute medium (RPMI) (Invitrogen) was supplemented with 10ml penicillin-streptomycin-glutamate (PSG) (Invitrogen) and 50ml foetal calf serum (FCS) (Invitrogen). This was warmed to 37°C in a water bath before use.

3. Liquid nitrogen stored cell pellets were defrosted quickly at room temperature and transferred to a 15ml falcon tube. 10ml RPMI was added and the cells were centrifuged at 260g for 3 minutes at room temperature.
4. The supernatant was removed and the cell pellet resuspended in 10ml fresh medium. This was transferred to a 10ml tissue culture flask with the lid loosely fastened. The flask was placed in the incubator at 37°C with CO_2 levels at 5%.
5. According to the specific requirements of different cell lines, the cells were supplemented with appropriate volumes of fresh medium at approximately three day intervals. The cell cultures were split and transferred into larger flasks as the volume increased.
6. To count the number of cells in a culture, 5ul was removed and mixed with 5µl 0.4% trypan blue solution (Sigma). This was pipetted into a 10 well disposable haematocytometer. Trypan blue is negatively charged and only interacts with damaged cell membranes. Viable cells exclude the dye whilst dead cells are stained blue. The numbers of cells in a 16 square area were counted and the total number of cells in the culture calculated as follows;

   Number of cells in a 16 square area x 2 (dilution factor of trypan blue) x 10,000 (to give cell number in 1ml of medium) x total volume of medium.

7. For long term storage of cell pellets, 10^7 cells were removed from culture and pelleted in the centrifuge (360g, 3 minutes). The supernatant was removed and the cells resuspended in 1ml freezing solution (50% FCS, 30% RPMI (un-supplemented) and 20% dimethyl sulfoxide (DMSO) (BDH)). The solution was transferred to a cryotube and placed at -80°C overnight. The following day, tubes were placed in liquid nitrogen. DMSO is an organic solvent that maintains the membrane as fluid, so that cells remain intact as they freeze.

9.3 Mononucleated cell separation by density gradient centrifugation

To investigate the effect that a disease has on its target cells, it is preferable to study material from a pure population. ALL affects the B and T lymphoblasts, so these cells are separated from as many other bone marrow components as possible. However, the abnormal clone is often only present in a subpopulation of cells. Previously published data suggest that accurate detection of CNA by aCGH requires a leukaemic population at least 40% (62), although the accuracy will depend upon the size of the aberration and the level of copy number change.

1. Bone marrow or peripheral blood samples were centrifuged in an appropriate sized tube at 300g for 10 minutes at room temperature. The transport medium was removed and the pellet loosened by gentle flicking.
2. 5ml phosphate buffered saline (PBS) (Sigma) was added.
3. 2 volumes of sample were carefully layered over 1 volume Ficoll-paque (GE Healthcare) at a 45° angle, and centrifuged at 360g for 30mins with the brake off.
4. Red blood cells are aggregated by Ficoll-paque, and they have the greatest cell density, as they are un-nucleated. They separated to the bottom of the tube. The mononucleated cells, such as monocytes and lymphocytes, are not dense enough to penetrate the Ficoll-paque layer, and collected as a concentrated band at the interface between the Ficoll-paque and the plasma. This layer was removed and the cells washed 3 times in 5-10ml PBS, at 420g for 10 minutes, to remove all Ficoll-paque.

9.4 DNA extraction using DNeasy kit (Qiagen)

DNA is used in many applications such as CGH, molecular copy number counting, PCR and LDI-PCR.

1. Buffers AW1 and AW2 were made up with 25µl and 30µl 100% ethanol, respectively.
2. Up to $5x10^6$ cells (per column) were centrifuged for 5 minutes at 300g and resuspended in 200µl PBS.
3. 200µl buffer AL and 20µl proteinase K (20mg/ml) were added, mixed by vortexing and incubated at 70°C for 10 minutes. Buffer AL is a cell lysis solution containing guanidinium chloride, which acts as a protein denaturant. It breaks open the cell and nuclear membranes, exposing the DNA to proteins in the sample tissue. Proteinase K is a non-specific serine protease used to digest these proteins and keep the DNA intact.
4. 200µl 100% ethanol was added and mixed by vortexing.
5. The reaction mixture was pipetted into a DNeasy mini spin column placed in a 2ml collection tube and centrifuged for 1 minute at 6000g. The flow through and collection tube were discarded.
6. The column was placed in a new collection tube. 500µl buffer AW1 was added to the column and centrifuged as above. The flow through and collection tube were discarded. Buffer AW1 contains guanidinium chloride and ethanol, and is used to wash the sample.
7. The column was placed in a new collection tube. 500µl buffer AW2 was added to the column and centrifuged for 3 minutes at 20,000g to dry the membrane. The column was carefully removed from the collection tube to prevent any ethanol

carryover. The flow through and collection tube were discarded. Buffer AW2 is also used to wash the sample. The components are confidential.

8. The column was placed in a new collection tube and 200µl buffer AE was added directly to the membrane. Buffer AE is an elution buffer that pulls the DNA through the column membrane. Its components are confidential. It was incubated at room temperature for 1 minute then centrifuged for 1 minute at 6000g. This step was repeated using the same eluant. The extracted DNA was stored at 4°C or -20°C

9.5 RNA extraction

9.5.1 RNA extraction using Trizol (Invitrogen)

RNA extracted from target cells may be used in RT-PCR and expression array studies to investigate the effect of disease on gene expression levels.

1. 1ml Trizol was added to $5x10^6$ pelleted cells and was homogenized thoroughly by pipetting. Trizol contains two protein denaturants; phenol and guanidine isothiocyanate. It maintains the integrity of the RNA, while disrupting cells and dissolving cell components. The sample was incubated at room temperature for 5 minutes to allow the complete dissociation of nucleoprotein complexes.
2. 200µl chloroform (Sigma) was added and mixed by shaking vigorously for 15 seconds. The samples were incubated for 2-3 minutes at room temperature then centrifuged at 11920g in a microfuge, for 15 minutes at 4°C. The chloroform separates the solution into a colourless aqueous phase and an organic phase.
3. The RNA remains exclusively in the aqueous phase. This layer was removed to a fresh eppendorf and 500µl isopropanol (Fisher Chemicals) was added to precipitate the RNA. The sample was incubated at room temperature for 10 minutes and centrifuged at 11920g for 10 minutes at 4°C.
4. The RNA formed a pellet at the bottom of the tube. The supernatant was removed and the pellet washed in 1ml 75% ethanol (Fisher Chemicals). The sample was vortexed and centrifuged at 5345g for 5 minutes at 4°C.
5. The pellet was air dried and redissolved in 50µl RNase free water (Qiagen). It was incubated in a water bath at 60°C for 10 minutes to ensure total resuspension. Extracted RNA was stored at -80°C.

9.5.2 RNA extraction using RNeasy kit (Qiagen)

1. The DNase stock solution was prepared by suspending the lyophilized DNase 1 (1500 units) in 550µl RNase free water and gently mixing. 10µl aliquots were stored at -20°C.
2. Buffer RLT was supplemented with 10µl β-mercaptoethanol (BDH) per 1ml, in the fume hood. This solution was stable for 1 month.

3. 4 volumes 100% ethanol were added to buffer RPE to obtain a working solution.
4. 350µl buffer RLT was added to $<5x10^6$ pelleted cells. 600µl buffer RLT was added to a maximum of 10^7 cells. Buffer RLT is a cell lysis solution that contains the protein denaturant, guanidinium thiocyanate. β -mercaptoethanol is added to destroy any RNases that may be present, which will degrade the RNA. It is a reducing agent that reduces the disulfide bonds of the RNases, thereby destroying the conformation and the functionality of the enzyme.
5. To completely homogenise the cells, they were passed through a 20-gauge needle fitted to an RNase free syringe, five times.
6. One volume 70% ethanol was added and mixed by pipetting. 700µl of the sample was applied to a column and centrifuged at 10,000g for 15 seconds at room temperature. The flow through was discarded. If the sample volume exceeded 700µl, successive volumes were applied to the column and centrifuged as above.
7. 350µl buffer RW1 was applied to the column and centrifuged as above. The flow through was discarded. RW1 is a wash buffer containing guanidinium thiocyanate.
8. 70µl buffer RDD was added to a 10µl aliquot of DNase 1 solution, and mixed by inverting the tube. This solution was applied directly onto the RNeasy silica-gel membrane and incubated for 15 minutes at room temperature. The components of buffer RDD are confidential.
9. 350µl buffer RW1 was added to the column and centrifuged as above. The column was then transferred to a new 2ml collection tube.
10. 500µl buffer RPE was applied to the column and centrifuged as above. The flow through was discarded. Another 500µl buffer RPE was added and centrifuged at 10,000g for 2 minutes to dry the membrane. The components of buffer RPE are confidential.
11. The column was transferred to a new 1.5ml collection tube and centrifuged at 10,000g for 1 minute.
12. 30µl RNase free water was added to the membrane and incubated at room temperature for 10 minutes. The column was spun at 10,000g for 1 minute. This step was repeated once. The RNA was stored at -80°C.

9.6. Quantification and quality assessment of extracted material

All RNA and DNA samples were assessed for quality and quantity before being used in experiments.

9.6.1 DNA quality assessment

Charged molecules placed in an electric field migrate toward the positive (anode) or negative (cathode) pole according to their charge. DNA has a negative charge imparted by the phosphate backbone. Semi-permeable agarose gel is placed in a tank of buffer, which provides the ions to carry a current, and the DNA is applied at the cathode. As it

migrates through the gel toward the anode the DNA fragments are separated according to size and electric charge. Molecules with a low molecular weight pass through the gel faster then those with a higher weight. A higher concentration of agarose will allow for the separation of smaller DNA fragments, whilst low concentrations allow for the resolution of larger DNA fragments. Ethidium bromide is incorporated into the gel to allow visualization of the DNA. It intercalates between nucleic acid bases and fluoresces when exposed to UV light. DNA comprised of different sized fragments will pass through the gel at different times creating a ladder like image. Genomic DNA is comprised of very large molecules that will not migrate far, creating a discrete band. A smear indicates DNA degradation or digestion (figure 6).

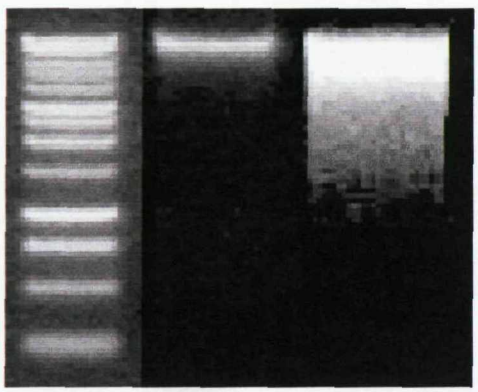


Figure 6. Agarose gel image of DNA ladder (lane 1), genomic DNA (lane 2) and degraded genomic DNA (lane 3)

1. 1g agarose powder (Promega) and 100ml 1xTAE were combined and heated in the microwave for 2 minutes. After cooling slightly, 10µl Ethidium bromide (Fisher Chemicals) was added and the gel poured into a casting tray. When set (approximately 20 minutes), the tray was placed into an electrophoresis tank filled with 1x TAE.
2. 2µl 6x DNA loading buffer (Promega) was added to 1-8µl DNA, depending on the expected concentration, and made up to 10µl with nuclease free water. The sample was loaded into one of the wells. Loading buffer contains ficoll which pulls the DNA to the bottom of the well, and dyes to allow the sample to be tracked down the gel. A 1kb DNA ladder was also loaded for reference (Promega).
3. 100 volts was applied to the gel for 30 minutes, after which a UV transilluminator was used to visualize the DNA.

9.6.2 DNA quantification using the NanoDrop-1000 (NanoDrop Technologies)

NanoDrop-1000 is a spectrophotometer which determines the concentration and purity of DNA by measuring the amount of light that a sample absorbs. The sample is pipetted onto a fiber optic cable (the receiving cable) and a second fiber optic cable (the source cable) is bought into contact with the sample, causing the liquid to bridge the gap

between the fiber optic ends. A pulse of light originating in the source cable is passed through the sample. When a photon encounters a DNA molecule it is absorbed and the intensity of light reaching the receiving cable is reduced and measured. DNA absorbs UV light at a wavelength of 260nm, proteins absorb light at 280nm and 230nm. Other contaminates such as carbohydrates also absorb at 230nm. Absorbance is measured at these three wavelengths, allowing the concentration and purity of the DNA sample to be determined (figure 7).

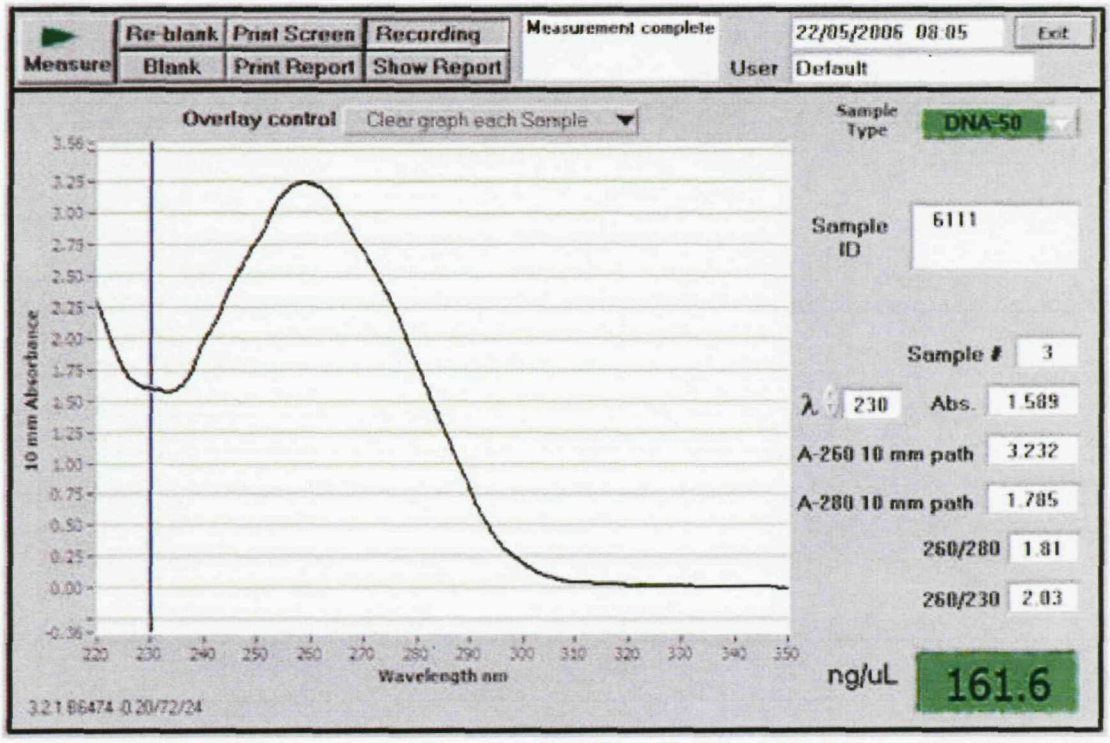


Figure 7. NanoDrop DNA profile showing the concentration and absorbance ratios.

1. With the sampling arm in the down position the NanoDrop-1000 software was started and the nucleic acid application selected.
2. NanoDrop-1000 was calibrated by applying 1.5µl nuclease free water to the lower measurement pedestal, lowering the sampling arm, and selecting blank.
3. The water was wiped from both pedestals. 1.5µl nuclease free water was measured and a flat base-line was returned. Both pedestals were wiped.
4. 1.5µl of each DNA sample was measured. Each profile was saved after measurement and the pedestals were wiped.
5. The software provided a concentration in ng/µl. This is calculated in the following way:

Optical density (absorbance reading at 260nm) x 50 (1 absorbance unit at 260nm= 50µg/µl DNA)

6. DNA purity is calculated by dividing the absorbance at 260nm by the absorbance at 280nm and at 230nm. This ratio should fall between 1.5 and 2.0. A ratio <1.5 indicates high contamination.

9.6.3 RNA 6000 nano or pico assay (Agilent bioanalyser 2100)

Nano and picochips use the principles of electrophoresis to assess RNA quality and quantity. The software generates a double peaked trace, representing the 18s and 28s ribosomal RNA (figure 8). The ratio of these two components indicates the quality of the sample. As degradation proceeds there is a decrease in the 18s to 28s rRNA band ratios. A good quality sample should have a ratio of 2:1 (28s:18s), whereas ratios of less then 1.5 indicate degradation. The concentration and size of the bands in the supplied ladder are known, which provides the concentration;area ratio. This is compared with the sum of the sample peak areas, to calculate the sample concentration. The nanochip has a qualitative range of 5-500ng RNA and a quantitative range of 25-500ng. The picochip has a qualitative range of 200-5000pg.

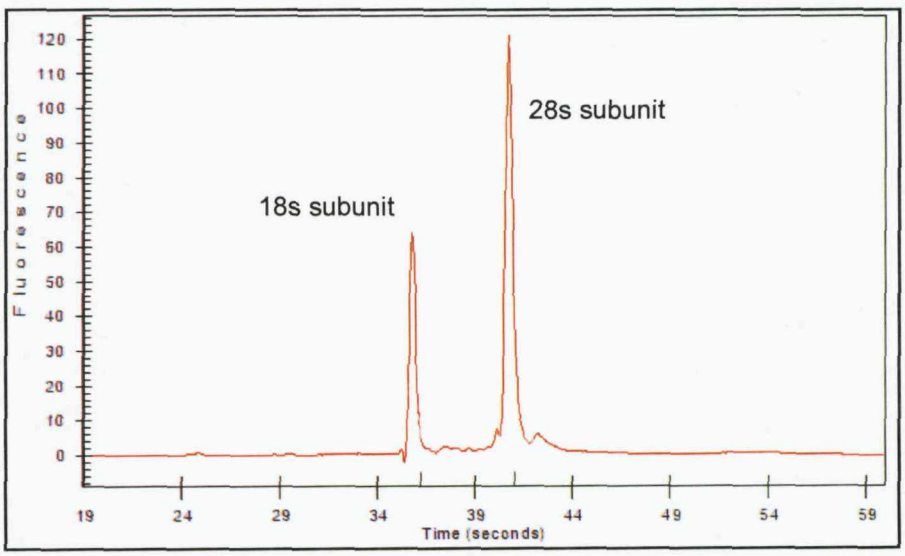


Figure 8. An RNA trace generated by the Nanochip software, showing the 28s and 18s rRNA peaks.

1. The electrodes were decontaminated by filling the wells of an electrode cleaner with 350µl RNAseZAP (Invitrogen) and placing it inside the bioanalyser for 1 minute. A second electrode cleaner filled with 350µl water was used to wash excess RNAseZAP from the electrodes. The lid was left open for 10 seconds to allow excess water to evaporate.
2. The reagents were equilibrated to room temperature for 30 minutes then 550µl gel matrix was applied to the top receptacle of a spin filter. The filter was centrifuged at 1520g for 10 minutes. The eluate was aliquotted in 65µl volumes.

3. The dye concentrate was vortexed and spun briefly to sediment any particles. 1µl was added to 65µl gel matrix and spun at 16000g for 10 minutes.
4. A RNA nano chip was placed in the chip priming station and 9µl gel-dye mix was pipetted to the bottom of the first gel well. The station was closed and the plunger depressed from 1ml until it was held by the syringe clip. This was held for exactly 30 seconds to force the gel into the microfluidics channels. The clip was then released.
5. After 5 seconds the station was opened and another 9µl gel-dye mix was pipetted into the remaining two gel wells.
6. 5µl Nano marker was pipetted into all the remaining wells. 6µl marker was added to any unused sample wells. The marker contains a 50bp DNA fragment and provides an internal standard which is used to align the ladder data with the sample data
7. 1µl sample and 1µl ladder were denatured at 70°C for 2 minutes to minimize secondary structure and then pipetted into the appropriate wells.
8. The chip was vortexed at 2400rpm for 1 minute and placed inside the bioanalyser.
9. In the 2100 Expert software the appropriate assay (nano or pico) was selected and the run started. When completed, the chip was removed and the electrodes decontaminated as described in step 1.

9.7. Quantitative Real Time-Polymerase Chain Reaction (qRT-PCR)

This technique is primarily used for determination of gene expression levels. It can also be used for RNA, DNA and cDNA quantitation. The Taqman system uses a primer and probe combination specific for the gene of interest. The probe contains a fluorescent reporter dye on the 5' end and a quencher molecule on the 3' end. While the probe is intact, the proximity of the quencher molecule greatly reduces the fluorescence emitted by the reporter dye. The probe anneals to a target sequence downstream of the primer site and as the primer is extended by Taq DNA polymerase its 5' nuclease activity cleaves the reporter dye from the quencher. This allows the fluorescence of the reporter to increase. Each cycle of denaturation, primer annealing and primer extension cleaves another probe. In this study FAM was used as the fluorescent reporter dye. Data collection occurs throughout the process and reactions are characterized by the point in time during cycling when the fluorescence exceeds the threshold. The cycle number needed to generate a defined amount of fluorescence when the PCR is in its linear phase is called the cycle threshold (Ct) (figure 9). The higher the Ct value the less RNA present in the sample. The comparative Ct method is used for quantitation of gene expression relative to an endogenous control gene, such as housekeeping gene GAPDH.

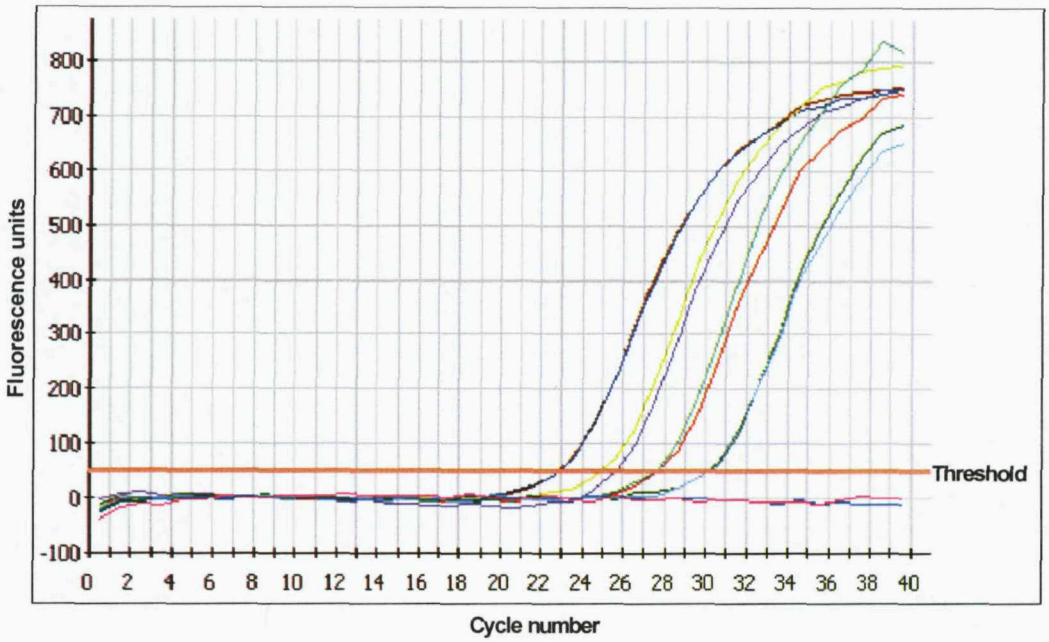


Figure 9. Graph displaying the increase in fluorescence in seven cDNA samples undergoing qRT-PCR cycling. The sample represented by the black line has the highest RNA concentration of the gene of interest. After 23 cycles the fluorescence levels exceed the threshold (orange line) and continue to increase, exponentially, until they plateau after 35 cycles.

9.7.1 cDNA Synthesis

1. 1µg RNA and 1µl oligo$(dt)_{15}$ primer (500µg/µl) (Promega) were combined and made up to 12.5µl with PCR-grade water, mixed and centrifuged to collect. Oligo$(dt)_{15}$ primer has a tail of 15 thymine residues which bind to the poly(A) tail of mRNA.
3. The reaction mixture was heated at 70°C for 5 minutes to melt the secondary structure of the RNA and allow the primer to bind. It was cooled on ice and centrifuged to collect.
4. 1µl M-MLV reverse transcription (RT) enzyme (200U/µl), 5µl 5X buffer, 1.25µl 10mM dNTP mix, 0.625µl RNAsin (40U/µl) (Promega) and 4.625µl water were added, mixed and centrifuged to collect.
5. The reaction was heated at 42°C for 60 minutes, followed by 95°C for 5 minutes. It was cooled on ice and centrifuged to collect.
6. The cDNA was diluted with 75µl water. The approximate concentration of the cDNA was calculated as 10ng/µl (Concentration of RNA / final volume of cDNA).
7. The cDNA was diluted to 1ng/µl and stored at –20°C.

9.7.2 qRT-PCR

1. 10µl 2x Taqman Universal Mastermix, 1µl assay-on-demand 20x primer and probe (Applied Biosystems) for the gene of interest, and 4µl water were added to

the appropriate wells of a 96/384 well plate (Applied Biosystems). 5µl of appropriate cDNA was added, in triplicate. For each gene a well containing no cDNA is included as a negative control.

2. The plate was sealed with an adhesive film (Applied Biosystems) and briefly centrifuged to collect the samples. It was placed in the ABI 7900HT Sequence Detection System under the following cycling conditions; 2 minutes at 50°C, 10 minutes at 95°C, 40 cycles of 15 seconds at 95°C (cDNA denaturation) and 1 minute at 60°C (Primer annealing and extension).
3. Results were analysed using the 7900 System software and exported into Excel for calculation of fold change.

9.8. Gene expression arrays (CodeLink, GE Healthcare)

Custom GE Healthcare CodeLink 16-assay bioarrays were designed for this study. The multi-assay chamber allows up to 16 samples to be processed in parallel on a single slide. A 3D aqueous gel matrix slide surface, spotted with 30bp oligonucleotide probes, provides an aqueous environment for optimal interaction between probe and target. Each array includes 1200 features, plus positive and negative controls. Feature extraction was performed using CodeLink Expression 4.2 software. Bioconductor 2.1 software (www.bioconductor.org) was used to analyze the data.

9.8.1 Preparation of solutions

1. The following solutions were prepared;

   1xTNT (Rinse storage vessel with 150ml isopropanol, followed by deionised water).

   0.75x TNT

   TNB

   Cy5- streptavidin

   0.1x SSC/0.05% Tween 20

   Bacterial controls

9.8.2 cDNA synthesis

1. 1µg total RNA, 0.5µl working solution of bacterial control mRNAs and 0.5µl T7 oligo(dT) primer were combined and made up to 6µl with nuclease free water.
2. The reaction mix was incubated at 70°C for 10 minutes, cooled on ice, and centrifuged to collect.
3. 1µl 10x first strand buffer, 2µl 5mM dNTP mix, 0.5µl RNase inhibitor and 0.5µl reverse transcriptase were added, on ice, incubated at 42°C for 2 hours and centrifuged to collect.
4. 31.5µl nuclease free water, 5µl 10x second strand buffer, 2µl 5mM dNTP mix, 1µl DNA polymerase mix and 0.5µl RNase H were added to the first strand reaction

mix. The tube was flicked to mix the reagents, centrifuged to collect then incubated at 16°C for 2 hours. The sample was centrifuged to collect, mixed, and placed on ice.

5. The cDNA was purified using the Qiagen QIAquick PCR purification kit. 500µl buffer PB was added to the cDNA, mixed and transferred to a spin column in a 2ml collection tube. Buffer PB contains guanidinium chloride and isopropanol, and acts to bind the cDNA to the column membrane.
6. The column was centrifuged at 10,000g for 30-60 seconds.
7. The flow through was discarded. 700µl buffer PE was added to the column and centrifuged as before. Buffer PE is used to wash the sample. Its components are confidential.
8. The flow through was discarded and the column centrifuged at 10,000g for 1 minute to dry the membrane.
9. The column was placed in a new collection tube and 30µl buffer EB was applied to the centre of the membrane. After 1 minute incubation at room temperature the column was centrifuged for 1 minute at 10,000g to elute the cDNA. This step was repeated. Buffer EB is a 10mM Tris-Cl solution used to elute the cDNA.

9.8.3 *In Vitro* Transcription (IVT)

1. The cDNA was dried in a speedvac concentrator and resuspended in 4.75µl nuclease free water.
2. An IVT master mix of 2µl 10x T7 reaction buffer, 2µl T7 ATP, 2µl T7 GTP, 2µl T7 CTP, 1.5µl T7 UTP, 3.75µl 10mM biotin-11-UTP and 2µl 10x T7 enzyme mix was added to the resuspended cDNA, mixed and incubated for 14 hours at 37°C.
3. Biotin labelled cRNA was recovered using the Qiagen RNeasy kit. 80µl nuclease free water and 350µl buffer RLT was added to the cRNA and mixed.
4. 250µl 100% ethanol was added and mixed. The sample was applied to an RNeasy column and centrifuged at 8,000g for 15 seconds.
5. The column was transferred to a new collection tube and 500µl buffer RPE added. It was centrifuged as above. The flow through was discarded and the step repeated.
6. The column was placed in a new collection tube and centrifuged at 8,000g for 2 minutes, to dry the membrane.
7. The column was placed in a new collection tube and 50µl nuclease free water was applied to the membrane and it was incubated for 10 minutes at room temperature. The column was centrifuged at 8,000g for 1 minute to elute the cRNA. This step was repeated.

8. The cRNA concentration was assessed using the Bioanalyser 2100 as described in section 9.6.3. The cRNA was concentrated in the speedvac if there was less then 0.5µg/µl. It was stored at -80°C.

9.8.4 Bioarray hybridisation

1. 1µg of cRNA was made up to 5.3µl with nuclease free water. 1.3µl 5x fragmentation buffer was added and the sample incubated at 94°C for 20 minutes, then cooled to 0°C for at least 5 minutes.
2. 19.8µl hybridization buffer component A, 33µl hybridization buffer component B and 6.6µl nuclease free water were added to the fragmented cRNA. It was vortexed for 5 seconds, incubated at 90°C for 5 minutes and cooled on ice for 5 minutes.
3. The bioarrays were placed in the 12-slide universal shaker tray.
4. The hybridization mixture was vortexed for 5 seconds, centrifuged briefly and placed back on ice. 60µl was expelled into the middle of the chamber by placing the pipette tip at an angle above the bioarray, resting on the edge of the chamber unit.
5. The chamber unit was sealed with sealing strips and a sealing tool. The shaker tray was fixed into the platform inside the shaker-incubator. The speed was set at 225rpm and the slides were incubated at 37°C for 18-24 hours.

9.8.5 Bioarray washes

1. The following day the sealing strip was removed by lifting the tab and slowly pulling it back at a 60° angle.
2. Using an 8-channel pipette each chamber was flushed three times with 250µl 0.75x TNT buffer. 250µl 0.75x TNT was added to each chamber and the entire unit sealed with a sealing strip. It was incubated at 46°C for one hour.
3. The sealing strip was removed as before. The 0.75x TNT buffer was aspirated, leaving a fluid layer over the bioarray.
4. 250µl Cy5-Streptavidin working solution was added to each chamber which was then stored in the dark for 30 minutes at room temperature.
5. The Cy5-Streptavidin was aspirated and each chamber was flushed three times with 250µl 1x TNT buffer. 250µl 1x TNT buffer was added to each chamber and incubated at room temperature for 20 minutes.
6. The 1xTNT buffer was aspirated and 250µl 0.1xSSC/0.05% Tween 20 was added to each chamber. The chamber unit was immediately inverted to decant the liquid.
7. To dry the bioarrays the chamber unit was inverted into a holder which was placed in a paper towel lined centrifuge bucket. They were centrifuged at 644g for 3 minutes with maximum acceleration and deceleration.

8. The unit was held with the chambers facing away and the unit clips were removed. The slide was peeled away from the chambers and placed in a light protected box until it was scanned, using the Agilent DNA Microarray scanner.

Note: During the early stages of developing this technique, a number of samples failed to hybridize correctly. These samples were reprocessed following ethanol precipitation before and after labeling.

9.9 Ethanol precipitation

Ethanol precipitation can be used to concentrate RNA and DNA samples, and to remove contaminants.

1. 2 volumes 100% ice cold ethanol, and 1/10 volume 3M sodium acetate (NaOAC (BDH)) were added to the samples.
2. They were stored at -20°C for at least 30 minutes then centrifuged for 30 minutes at 13,000rpm, at 4°C.
3. The pellet was washed with 100% ice cold ethanol and centrifuged for 10 minutes at 13,000rpm, at 4°C.
4. Step 3 was repeated.
5. The pellet was air dried and resuspended in an appropriate volume of nuclease-free water.

9.10 Array-based Comparative Genomic Hybridisation (aCGH) (Agilent)

The Agilent Human Genome CGH Microarray Kits 185A and 244A were selected for this study. Feature extraction was performed using Agilent Feature Extraction 9.1 software. Unless stated, the arrays were analysed using Agilent CGH Analytics 3.4.27, which is based upon the University of California at Santa Cruz (UCSC) May 2004 assembly (HG17). Aberrant regions were initially identified using Z-scoring. Each CGH ratio was converted to a log ratio and Z-normalized, setting the modal ratio of the array to 0. The moving average was set at 2Mb, meaning log ratios were averaged for each point in the data set using a 2Mb window size, centered on that point. Z-scores indicated statistically significant groups of probes that appeared to deviate from the typical distribution of values in that window. Outliers were classified using cut-offs of +/- 0.25. The extent of the deviation from 0 is not considered. Once aberrant regions had been identified in this way, the aberration detection method 1 (ADM-1) algorithm was used. In contrast to the z-score algorithm, adjacent probes were sampled rather then a set window size, to arrive at a robust estimation of the true range of the aberrant segment. In addition, ADM-1 takes into account the number of probes in the aberrant region and the extent of their deviation from 0 (figure 10). A minimum of five consecutive deviant probes were required for an aberrant call. In order to make reliable amplification or deletion calls, the Derivative Log Ratio

Spread (DLRS) algorithm is used. Noting that most pairs of probes adjacent to each other along a chromosome have the same true copy number, the large majority of differences between log ratios of adjacent probes are just noise. The DLRS metric calculates the spread of the log ratio differences between consecutive probes along all chromosomes, and hence the minimum log ratio difference required for an aberrant call. The quality of the array data is reflected by the DLRS score; excellent arrays have a score of <0.2, whilst good quality and poor arrays have scores of 0.2-0.3 and >0.3, respectively.

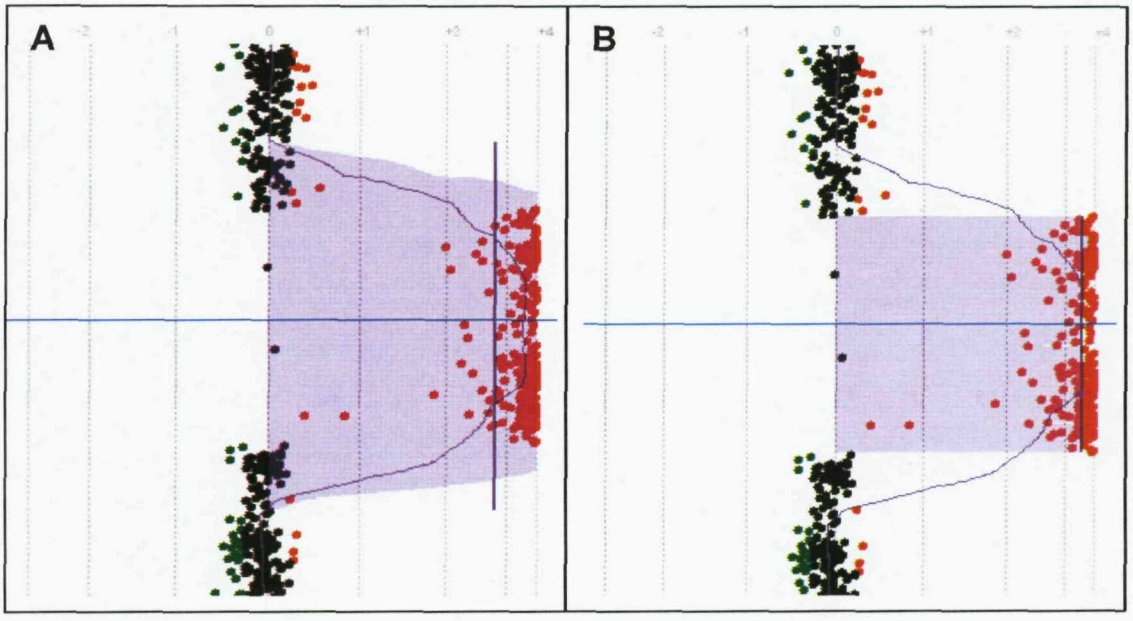


Figure 10. An aCGH profile of a copy number gain, calculated using (A) Z-scoring and (B) ADM-1. Green, black and red dots represent deleted, normal and gained probes, respectively. The moving average is represented by the purple line. The purple shaded area, and the purple sidebar indicate the extent of the aberration; ADM-1 has identified a more defined region.

9.10.1 Restriction digest of genomic DNA

1. 1.5µg genomic test DNA and reference DNA (Promega) was made up to 20.2µL with nuclease-free water.
2. 2µl nuclease free water, 2.6µl 10x buffer C (10mM Tris-HCL, 10mM $MgCl_2$, 50mM $NaCl_2$), 0.2µl acetylated BSA (10mg/ml) and 0.2µl *Alu I* (10U/µl) and *Rsa I* (10U/µl) (Promega) were added to both samples, on ice. They were incubated at 37°C for 2 hours. The enzymes were inactivated by incubation at 65° for 20 minutes

9.10.2 DNA labelling

1. 5µL Random Primers was added to each sample. They were incubated at 95°C for 3 minutes, and placed on ice for 5 minutes.
2. 10µL 5x buffer, 5µL 10x dNTP and 1µL exo-klenow (Agilent) was added to each sample. 3µL Cy3-dUTP was added to the reference sample and 3µL Cy5-dUTP

was added to the test sample. They were incubated at 37°C for 2 hours and the enzymes were inactivated by incubation at 65° for 10 minutes

9.10.3 Clean-up of labeled DNA

1. 430µL TE buffer was added to each sample. A Microcon YM-30 filter (Millipore) was placed in a collection tube and the sample loaded into the filter. It was centrifuged for 10 minutes at 8,000g at room temperature. The flow-through was discarded.
2. 480µL 1xTE was applied to the filter. It was centrifuged as before. The flow-through was discarded.
3. The filter was inverted into a fresh collection tube. It was centrifuged for 1 minute at 8,000g.
4. If the sample volume exceeded 80.5µL, the sample was returned to its filter and centrifuged for 1 minute at 8,000g. The flow-through was discarded. Step 3 was repeated.
5. The test and reference DNA were combined and the volume was brought to 161µL with nuclease-free water.

9.10.4 Hybridisation

1. The lyophilized 10x Blocking Agent was reconstituted in 1350µL nuclease-free water, for 60 minutes at room temperature.
2. 50µL Human Cot-1 DNA (1mg/ml) (Invitrogen), 52µL 10x blocking agent and 260µL 2x hybridisation buffer (Agilent) was added to the sample, mixed and centrifuged to collect.
3. The sample was incubated at 95°C for 3 minutes followed by 37°C for 30 minutes. It was centrifuged for 1 minute at 17,900g to collect
4. A clean gasket slide was loaded into the Agilent SureHyb chamber base with the label facing up.
5. 490µL hybridization sample was slowly dispensed into the gasket well. An array was placed onto the gasket slide, so the numeric barcode side was facing up and the "Agilent" barcode was facing down.
6. The chamber cover was placed onto the sandwiched slides and the clamp assembly slid into place and hand-tightened.
7. The assembled chamber was vertically rotated to wet the slides and assess the mobility of the bubbles. The assembly was tapped on a hard surface to move stationary bubbles.
8. The chamber was placed in rotator rack in a hybridization oven at 65°C. It was rotated at 20 rpm for 40 hours.

9.10.5 Washes and drying

1. The appropriate volume of wash buffer 2 was warmed to 37°C in a water bath, overnight.
2. To dissolve precipitates, the stabilization and drying solution was warmed to 40°C overnight, in a closed container with sufficient head space to allow for expansion. It was allowed to equilibrate to room temperature prior to use.
3. Slide-staining dish 1 was filled with Wash Buffer 1. A slide rack was placed into slide-staining dish 2 and a magnetic stir bar was added. It was filled with Wash Buffer 1 and placed on a magnetic stir plate.
4. Slide-staining dish 3 was placed on a magnetic stir plate and a magnetic stir bar was added.
5. Slide-staining dish 4 was filled approximately ¾ full with acetonitrile (Sigma). A magnetic stir bar was added and the dish placed on a magnetic stir plate.
6. Slide-staining dish 5 was filled approximately ¾ full with Stabilization and Drying Solution. A magnetic stir bar was added and the dish placed on a magnetic stir plate. Agilent Stabilization and Drying Solution is a proprietary solution that acts as a protectant comprising an ozone scavenger dissolved at near-saturation levels in acetonitrile. It prevents ozone-induced degradation of Cy-3 and Cy-5 dyes during microarray hybridization and processing steps.
7. The hybridization chamber was removed from the incubator and placed on a flat surface. The clamp was loosened and removed and the chamber cover removed. The array-gasket sandwich was lifted from the chamber base by the ends. It was quickly submerged in wash buffer 1 in slide-staining dish 1. The sandwich was prized open from the barcode end using the blunt ends of the provided forceps. The gasket slide was released and the microarray slide was placed in the slide rack in slide-staining dish 2, with minimum exposure to air.
8. Wash buffer 1 was stirred using setting 3 for 5 minutes. With 30 seconds remaining, wash buffer 2 was removed from the water bath and transferred to slide-staining dish 3.
9. The slide rack was transferred to slide-staining dish 3 and stirred using setting 3 for 1 minute.
10. The slide rack was removed from slide-staining dish 3 and tilted slightly to minimize wash buffer carry-over. It was immediately transferred to slide-staining dish 4, and stirred using setting 3 for 1 minute.
11. The slide rack was transferred to slide-staining dish 5 and stirred using setting 3 for 30 seconds. It was slowly removed over 5-10 seconds to minimize droplets on the slides.
12. The slide was scanned immediately using the Agilent DNA Microarray scanner.

9.11 Molecular copy number counting (MCC)

MCC uses PCR to directly count target sequences in a series of miniscule amounts of genomic DNA, and therefore delineate DNA content to within a few hundred base pairs (bp) of a genomic aberration. Aliquots of patient DNA diluted to contain less than one haploid genome of DNA (0.9pg) are used. Sets of markers external forward (EF), internal forward (IF) and common reverse (CR) primers are designed at regular intervals throughout the DNA sequence of interest. The 1st PCR is a multiplex amplification step for each DNA aliquot using all EF and CR primers, which amplifies all copies of any target sequence to some extent. The second PCR is semi-nested and carried out separately for each individual IF and CR primer, using the 1st PCR product as a template. The PCR products are then analyzed by gel electrophoresis and the proportion of aliquots that are positive for any marker reflects the relative copy number of that marker in the genome (figure 11). For example, in the case of a non-reciprocal translocation there will be 1*n* copy number of the marker for the distal portion of the chromosome and 2*n* copy number of the marker proximal to the translocation (80).

9.11.1 Primer design

1. The DNA sequence for the region of interest was obtained from www.ncbi.nlm.nih.gov.
2. Approximately 500bp lengths of sequence were selected at regular intervals throughout the sequence and copied into Editseq (LaserGene 6 software). The Editseq files were opened in PrimerSelect, and EF, IF and CR primers were designed according to the following criteria;
    I. Primer length should be 17-22 bp.
    II. A G nucleotide is preferred at the 5' and 3' end, two if possible. C nucleotides can be substituted if G is not available.
    III. The sequence should not contain the same nucleotide more then three times in a row.
    IV. The GC content should be 50-60%.
    V. The 5' and 3' end should differ by at least 4 nucleotides to prevent the formation of a ring.
    VI. The melting temperature (Tm (the temperature at which half of the DNA is denatured)) should be 52-60°C. The Tm is calculated as follows;

    $$Tm = (4x(G+C)) + (2x(A+T))$$

    Primers need to have the same Tm if possible as the PCR temperature is obtained by subtracting 4 or 5°C from the Tm.
    VII. The distance between the EF and IF primers should be 80-100bp, and the distance between the IF and CR primer should be <150bp.
3. Primers were ordered from MWG (www.mwg-biotech.com).

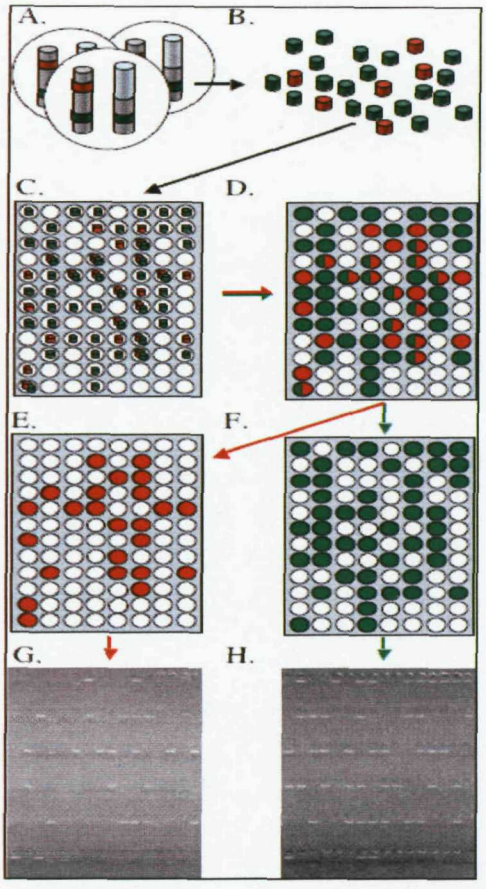


Figure 11. Overview of the MCC method. (A) Cells harbour a nonreciprocal translocation, and therefore one marker sequence (green) is present in twice as many copies as a second marker (red), which lies telomeric of the translocation breakpoint. (B) DNA is prepared from the target cells and diluted; for simplicity, only the markers are illustrated. (C) DNA is dispensed at less than one genome per aliquot into wells of a 96-well plate. (D) All markers are amplified in a multiplex PCR, using all EF and CR primers. (E, F) The multiplex reaction products are split into replica plates, and a second semi-nested PCR is carried out using the IF and CR primers specific for each marker. (G, H) Gel electrophoresis analysis of the semi-nested PCR products. In this example, the 'red' marker is found in 24 of the wells (G) and the 'green' marker in 46 of the wells (H), corresponding to a twofold increase in copy number. Adapted from Daser et al, *2006 (80).*

9.11.2 Primer QC

1. The desiccated primers were reconstituted with TE into a 100pmol/µl concentration. A 5pM working solution was prepared.
2. To confirm that the primers work the EF and CR, and IF and CR primers for each marker were combined and used to amplify normal control DNA. The following master mix (MM) was added to 0.6µl of each primer pair;

Reagents	Volume (µl)	Concentration
GoTaq Flexi 5x buffer (Promega)	4	1x
$MgCl_2$ (25mM) (Promega)	1.6	1.5mM
dNTP (10mM) (Promega)	0.8	200nM
DMSO	2	
DNA	1	100ng
GoTaq DNA polymerase (5U/µl) (Promega)	0.2	0.05 µM
Water	9.8	

3. PCR was performed using an MJ PTC-225 Thermal cycler (MJ Research) at 95°C for 2 minutes, then 30 cycles of 30 seconds at 95°C, 30 seconds at 50-56°C and 1 minute at 72°C, followed by a terminal extension at 72°C for 7 minutes.
4. The PCR products were run on a 1% agarose gel to confirm the presence of bands.

9.11.3 MCC

1. Patient DNA was diluted to give a total of 86.4pg in a maximum volume of 518.4µl.
2. A MCC MM, sufficient for 96 wells was prepared, containing the following;

Reagents	Volume/well (µl)
GoTaq Flexi 5x buffer	2
$MgCl_2$ (25mM)	0.8
dNTP (10mM)	0.4
EF and CR primers for each marker	0.3
DMSO	1
DNA	≤ 5.4
GoTaq DNA polymerase (5U/µl)	0.1
Water	Make up to 10

3. A 10µl aliquot of MM was dispensed into 88/96 wells. 10µl of a second MM, containing additional water instead of DNA, was added to the remaining 8 wells as a negative control.
4. The plates were sealed with FoilSeal (Alpha Laboratories) and centrifuged to collect.
5. PCR was performed as described in section 9.11.2, with 5 additional cycles.
6. Each reaction was diluted with 190µl water. 4µl of each diluted PCR product was transferred to a new 96 well plate. Each plate was used to amplify a single marker. A MM was prepared for 105 wells, containing the reagent volumes listed above with the following differences; 1µl of the relevant IF and CR primers (10µM) and

0.7µl water per well. 6µl aliquots were added to all wells. The plates were sealed, vortexed, and centrifuged to collect.

7. PCR was performed as described in step 5.

9.11.4 Electrophoresis of PCR products

1. A glass plate was cleaned with 70% ethanol. 500µl sticky silane solution was evenly applied using tissue and dried for several minutes.
2. A 7.5% polyacrylamide gel was prepared and poured into a 192 well MADGE (microplate array diagonal gel electrophoresis) former. This was covered with the glass plate, silane side down, and left to set for 15 minutes. The gel was removed from the MADGE former and soaked in 1xTBE (Natural Diagnostics) for several minutes.
3. 3µl of each PCR product was mixed with 2µl formamide dye mix, and 3µl was loaded onto the gel.
4. 200V were applied to the gel for 15 minutes to separate the PCR products by electrophoresis. The gel was then stained in a tank containing 100ml 1x TBE and 10µl 1x SYBR Gold fluorescent stain (Invitrogen) for 10-20 minutes, followed by a 20-30 minute wash in water. (If the SYBR Gold stain is being used on additional gels the length of staining should be increased by several minutes each time).
5. The gel was scanned on a Typhoon Trio Imager and analyzed in MADGE- specific gel image analysis software (Phoretix Ltd) (figure 12). The number of wells positive for each marker is compared.

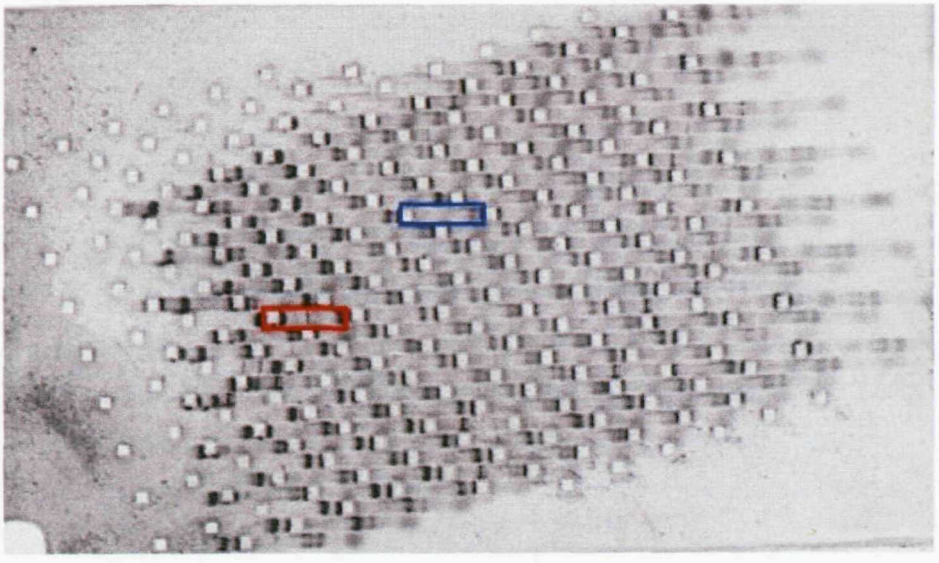


Figure 12. MADGE gel image of markers 7 and 9 for PDE9A, *in patient 7619. The red and blue boxes highlight positive and negative wells, respectively.*

6. If DNA copy number change is identified, additional markers can be designed for

the appropriate sequence, to further refine the breakpoint.

9.12 Long Distance inverse PCR (LDI PCR)

This technique is used to clone the breakpoints of translocations in which only one partner chromosome/gene is known. Genomic DNA is restriction endonuclease-digested, either side of the sequence of interest, then self-ligated to form a circular structure. Forward primers amplify the known sequence and reverse primers amplify any unknown sequence fused to it.

9.12.1 Primer design

1. Primers are designed according to the criteria described in section 9.11.1. A 3' G nucleotide is essential. Forward primers are designed proximal to the breakpoint region, and the CR primer is designed upstream of them. Primer QC is performed as described in section 9.11.2.

9.12.2 Restriction digest

1. Restriction endonucleases (RE) were selected to cleave the DNA upstream of the CR primer and downstream of the breakpoint region (www.neb.com).
2. 1µl of each RE (10U/µl) and 5µl of appropriate 1x enzyme buffer was added to at least 1µg of patient DNA and normal reference DNA. The reactions were made up to 50µl with nuclease-free water and incubated at 37°C for 4 hours.
3. The DNA was purified using the Qiagen QIAquick PCR purification kit, as described in section 9.8.2, and eluted in 50µl of buffer EB.

9.12.3 Ligation

1. To ligate the digested DNA into a circular structure 2µl T4 DNA Ligase (1-3U/µl) and 10µl 10x buffer (Promega) was added. The reaction was made up to 100µl with nuclease-free water and placed at 4°C overnight.
2. The DNA was purified using the Qiagen QIAquick PCR purification kit and eluted in 30µl EB buffer.

9.12.4 LDI PCR

1. The following MM was added to 5µl ligated DNA;

Reagents	Volume (µl)
GoTaq Flexi 5x buffer	5
$MgCl_2$ (25mM)	2
dNTP (10mM)	2
EF and CR primers	2
DMSO	2.5

GoTaq DNA polymerase (5U/µl)	0.3
Water	Make up to 25

2. PCR was performed at 95°C for 2 minutes, then 35 cycles of 40 seconds at 95°C, 40 seconds at 50-56°C and 5 minute at 72°C, followed by a terminal extension at 72°C for 10 minutes.
3. Steps 1 and 2 were repeated using the IF and CR primer, to amplify the PCR product.
4. The PCR products are run on a 0.8% agarose gel (refer to 9.6.1).

9.12.5 Gel Extraction (Qiagen QIAquick gel extraction kit)

Due to the presence of different sequence in the test and normal DNA, different sized DNA fragments are resolved by gel electrophoresis. The abnormal patient DNA band is excised from the gel and the DNA extracted from it.

1. Using a scalpel, the DNA fragment was excised from the gel and weighed.
2. 300µl buffer QG was added per 100mg gel, up to a maximum of 400mg per column. Buffer QG contains the protein denaturant guanidinium thiocyanate, and acts as a solubilization and binding solution
3. The gel was incubated at 50°C for 10 minutes, with vortexing every 2-3 minutes. DNA adsorption to the QIAquick membrane is only efficient at pH≤7.5. Buffer QG contains an indicator which is yellow at this pH. If the mixture was orange or violet 10µl 3M NaOAc was added to reduce the pH.
4. 100µl isopropanol was added per 100mg gel and mixed
5. The reaction was transferred to a QIAquick column and centrifuged at 17,900g for 1 minute. The flow-through was discarded.
6. 500µl buffer QG was applied to the column and centrifuged as above.
7. 750µl buffer PE was added to the column, left for 3 minutes, and centrifuged as above.
8. The flow-through was discarded and the column centrifuged as above, to dry the membrane.
9. To elute the DNA 30µl buffer EB was applied to the membrane, left for 1 minute and centrifuged as above.
10. The DNA was sent to GeneService (www.geneservice.co.uk) for sequencing. The sequences were copied into the web browser, BLAST (www.ncbi.nlm.nih.gov/blast/), which aligns homologous sequences in the human genome to the patient target.

9.13 Reverse Transcription-PCR (RT-PCR)

RT-PCR is used to determine if a particular sequence of DNA is being transcribed into mRNA, such as fusion gene products resulting from gene rearrangements. RNA samples are reversed transcribed into cDNA, which is subsequently amplified by a PCR reaction. The PCR products are run on an agarose gel. The presence of a band indicates that the sequence of interest has been transcribed.

1. Forward and reverse primers were designed for the sequences of interest, and the quality was checked as described in sections 9.11.1 and 9.11.2 respectively.
2. The following master mix was added to patient and normal control cDNA;

Reagents	Volume/well (µl)
GoTaq Flexi 5x buffer	4
$MgCl_2$ (25mM)	1.6
dNTP (10mM)	0.8
F and R primers	0.6
DMSO	2
GoTaq Flexi polymerase (5U/µl)	0.2
cDNA	1
Water	9.8

3. PCR was performed as described in section 9.12.4. The time required for terminal extension was determined by the size of the PCR product; 1 minute was sufficient for products <1kb in size. Up to 5 minutes was required for larger products.
4. Steps 2 and 3 were repeated using 0.5µl of PCR product and 10.3µl of water.
5. PCR products were run on an agarose gel, excised and set for sequencing as described in section 9.12.5.

9.14 High resolution melt analysis (HRM)

HRM analysis is used to detect mutations in genomic DNA sequence. PCR amplification of target sequences is performed in the presence of a saturating dye that fluoresces brightly when intercalated with dsDNA. The dsDNA is then exposed to increasing temperatures. As it dissociates into ssDNA, the dye molecules are released and the fluorescent signal diminishes. The observed Tm is characteristic of a particular DNA sample and is influenced by sequence length, GC content and DNA sequence complimentarity. HRM analysis can be used to identify single base changes, insertions, deletions and base pair substitutions. It can quantitatively detect a small proportion of variant DNA in a background of wild-type sequence, at sensitivities approaching 5%. Samples with abnormal populations of at least 30% were used in this study.

Melting curves are generated, which display the decrease in fluorescent signal with increasing temperature (figure 13a). The Tm of all the samples is then normalized to within 10°C by defining regions in the pre- and post-denaturation parts of the curve. The wildtype curve is selected as the base-line, and normalized and temperature-shifted difference plots are generated. These display the relative signal difference between the wildtype and test samples (figure 13b). A shift in Tm between wildtype and test samples is indicative of a homozygous mutation. A shift in Tm and a change to the shape of the curve indicates a heterozygous mutation (81).

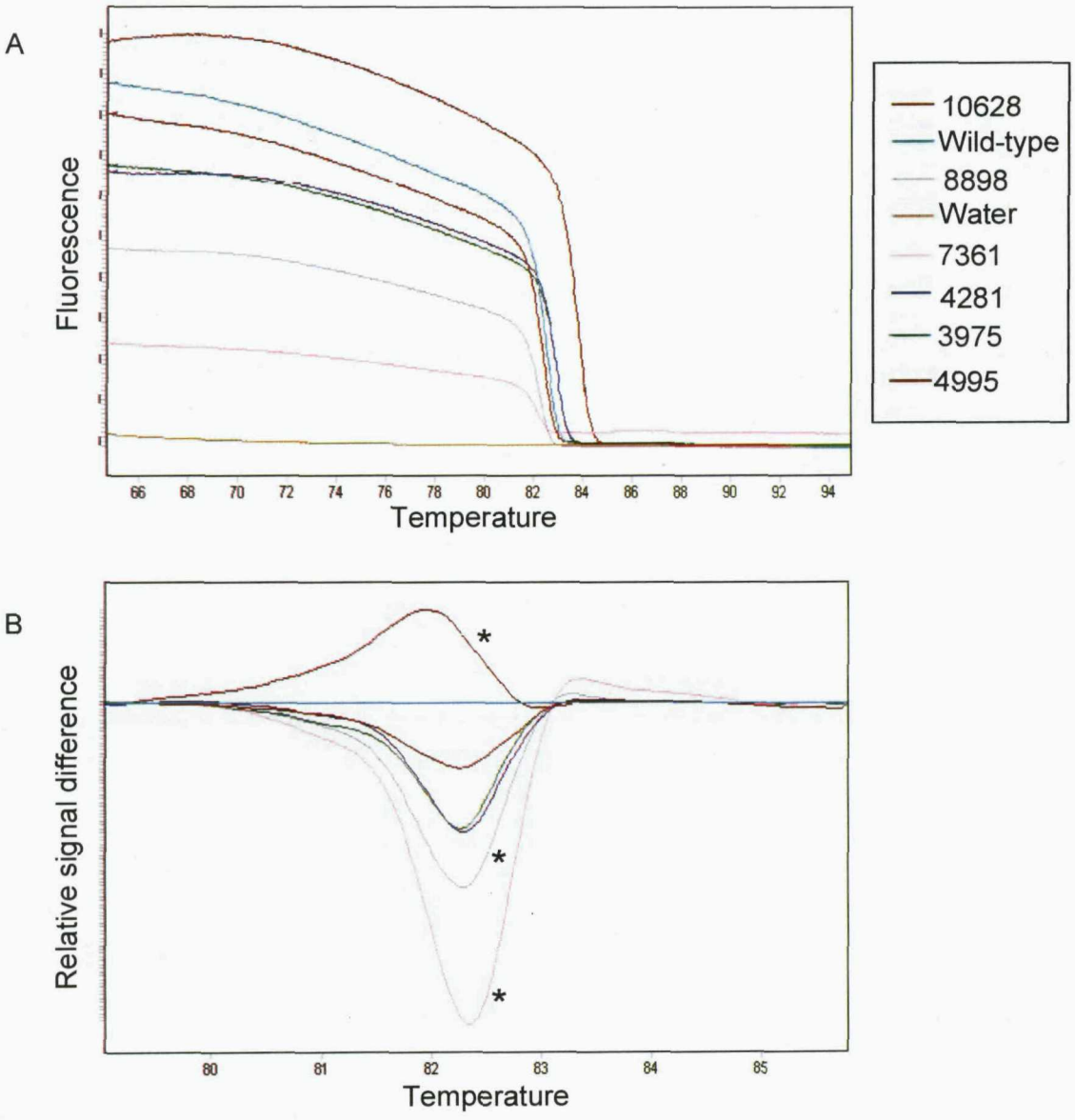


Figure 13. Melting curves (A) and normalized and temperature shifted difference plots (B) for exon 6 of the TBL1XR1 *gene, in six patient samples, a wildtype control and a water control. Traces marked with an asterisk (*) are indicative of a mutation. Patient 7361 and 8898 have homozygous mutations (temperature shift). Patient 4995 has a heterozygous mutation (temperature and shape shift).*

9.14.1 Primer Design

1. Forward and reverse primers are designed as described in section 9.11.1, to give an amplicon length of less then 150bp. Primer QC is performed as described in section 9.11.2.

9.14.2 HRM analysis

1. Test and normal reference DNA was diluted to a 5ng/µl concentration. 5mM SYTO 9 dye (Invitrogen) was diluted 1:100 to give a stock solution.
2. The following master mix was prepared for each sample and dispensed into a LightCycler 480 Instrument 96 well plate (Roche);

Reagents	Volume/well (µl)	Concentration
Platinum Taq 10x buffer (Invitrogen)	1	1x
5mM dNTPs	0.4	0.2mM
Platinum Taq DNA polymerase (5U/µl) (Invitrogen)	0.05	0.025U
50mM $MgCl_2$ (Invitrogen)	0.3	1.5mM
10µM F and R primers	0.5	0.5µM
DNA	2	
SYTO 9	1	
Water	4.75	

3. The plate was sealed and loaded into the LightCycler 480 System (Roche). PCR was performed with Taq activation at 95°C for 10 minutes, followed by 40 cycles of 15 seconds at 95°C, 30 seconds at 60°C and 30 seconds at 72°C.
4. The samples were heated to 95°C for 1 minute to denature the dsDNA, and cooled to 40°C for 1 minute to allow the complimentary single strands to reanneal.
5. The temperature was gradually increased from 65°C to 95°C, and the fluorescent signal measured 25 times per 1°C increase. Finally the plate was cooled to 40°C for 1 minute.
6. The data was analyzed with the LightCycler 480 software.

9.14.3 Sequencing

1. Forward and reverse primers were designed to amplify the mutated sequences, and the quality was checked as described in sections 9.11.1 and 9.11.2 respectively.
2. The following master mix was added to each sample and the PCR was performed as described in section 9.14.2.

Reagents	Volume/well (µl)	Concentration
Platinum Taq 10x buffer	5	1x
5mM dNTPs	2	0.2mM
Platinum Taq DNA polymerase (5U/µl)	0.25	0.025U
50mM $MgCl_2$	1.5	1.5mM
10µM F and R primers	2.5	0.5µM
DNA	10	
Water	28.75	

3. The amplified PCR products were sent to GeneService for sequencing. The mutated and wildtype sequences were aligned in LaserGene 6 software to locate and characterize the mutations.

9.15 G-Banding

G-banding is a staining technique which exploits the underlying structural and functional organization of DNA. Fixed metaphase chromosomes are trypsinised, *in situ* on a glass slide, to partially digest the histone proteins responsible for DNA packaging. This allows the positively charged thiazine component of Giemsa/Leishman's stain to bind to the negatively charged phosphate groups on DNA (82). This produces a pattern of alternating dark and light bands unique to each chromosome pair. The dark bands are relatively late replicating, AT rich regions containing few active genes (83). Structural rearrangements can be identified by a change to the expected pattern of bands on a chromosome (figure 14). G-banding of leukaemic samples has a resolution of approximately 5-10Mb (84).

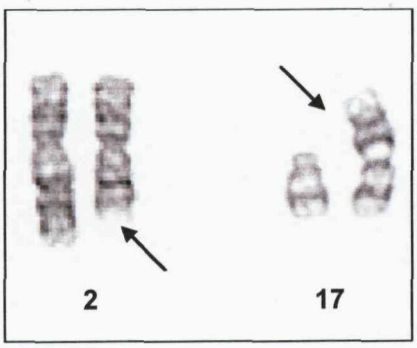


Figure 14. G-banded metaphase chromosomes. The normal chromosome 2 and 17 is shown on the left of each pair. Arrows highlight a deletion of the long arm of chromosome 2, and an unbalanced translocation involving chromosomes 9 and 17 (der(9)t(9;17)(q10;q11.2)).

9.15.1 Metaphase harvest

During metaphase, chromosomes condense and become visible down the microscope. By analysing metaphase cells, it is possible to identify chromosome pairs and visualise

structural and numerical chromosomal abnormalities. Cells are arrested in metaphase by the addition of colcemid to cell cultures, which in the study of leukaemia is bone marrow or peripheral blood. Colcemid binds to tubulin and prevents its polymerisation into spindle microtubules, whilst inhibiting their aggregation. The mitotic index, or the proportion of metaphase cells, will increase with longer exposure to colcemid but the chromosomes will also continue to condense. As longer chromosomes provide improved resolution of chromosome bands two cultures, with varying incubation times are set up.

1. Bone marrow aspirates diluted in culture medium, or peripheral blood samples were centrifuged at 300g for 10 minutes and the transport medium removed.
2. 5ml supplemented RPMI (section 9.2) was added to two flat sided culture tubes and 0.5ml of the cell pellet was added to each. 50µl 10µg/ml colcemid (Invitrogen) was added to one tube and both tubes were placed in an incubator at 37°C overnight.
3. The following day 100µl colcemid was added to the remaining tube and incubated for 1-2 hours.
4. The cultures were centrifuged for 5 minutes at 300g and the supernatant removed.
5. The tubes were filled with 0.75M potassium chloride (KCL) and incubated for 10 minutes at 37°C. KCL is a hypotonic solution which forces water to move into the cells by osmosis. As they swell the chromosomes spread out. The tubes were centrifuged as above.
6. The supernatant was removed and 5ml fixative (1 volume acetic acid: 3 volumes methanol) was added drop-wise to the pellet, with vortexing to prevent the formation of clumps. Methanol replaces the water in the cells and acetic acid softens the membranes.
7. The tubes were centrifuged as above and the pellets washed twice in fixative. The fixed cell suspension was stored at -20°C.

9.15.2 Slide preparation

1. Fresh fixative was used to clean the slides, to assist metaphase spreading.
2. The fixed cell suspension was centrifuged at 10,000g for 1 minute. The supernatant was removed and the pellet resuspended in sufficient fixative to give a solution that was slightly cloudy.
3. 3µl cell suspension was dropped onto the slide. The plasma membranes of the swollen cells erupt, releasing the cell contents. When Newton's rings were visible, two drops of fixative were added and allowed to spread for approximately 1 minute. As the fixative evaporates the cells flatten and the chromosomes spread further.

4. The slide quality was checked with a phase contrast microscope. The metaphases should be phase dark and non-overlapping.
5. Slides were aged overnight at room temperature before use.

9.15.3 Staining

1. 1-2ml trypsin solution (Worthington) was added to a coplin jar containing 50ml PBS. The slides were dipped in this solution for 20 seconds and immediately rinsed with PBS followed phosphate buffer. This time may be varied to optimize the chromosome banding.
2. The slide surfaces were flooded with Leishmans stain for 1 minute and immediately rinsed with phosphate buffer followed by distilled water.
3. A coverslip was placed on the wet slides and the banding quality checked using a light microscope. The coverslip was removed. The trypsin and staining steps may be repeated if required. Once satisfied with banding quality the slides were dried on a hotplate.
4. Working in the fume hood, the slides were mounted using a drop of DPX (BDH) applied to the slide surface, followed by a coverslip. Slides were analysed the following day.

9.16 Fluorescence *in situ* hybridisation (FISH)

FISH provides a way to visualize and map the genetic content of a cell. It is used to detect numerical and structural chromosomal aberrations. Vectors such as BAC, containing DNA sequences complementary to specific regions of the genome, are used to clone probes, which are labeled with fluorescent dyes. The probes are hybridized to fixed metaphase and interphase cells and the signals examined with regard to number and position, using fluorescence microscopy (figure 15). The resolution of FISH is determined by the size of the probe used. BAC probes are typically 100-200kb, but probes as small as 20kb are generated by fosmid vectors.

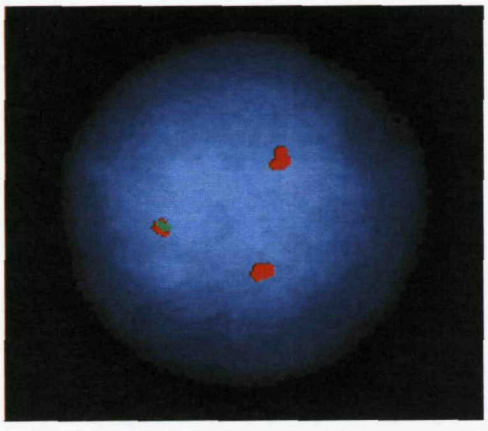


Figure 15. Interphase nuclei hybridised with BAC probes labeled with red and green fluorescent dyes.

9.16.1 Culturing BAC clones

1. The genome browser Ensembl (www.ensembl.org) was used to select appropriate clones. The clones were requested from the Sanger Institute (www.sanger.ac.uk), and were received as glycerol stabs that were stored at 4°C.
2. Agar plates were prepared containing either kanamycin (1µg/ml) or chloramphenicol (1µg/ml), depending on the resistance of the clone. A sterilized rod was used to streak cultures onto the plates, and they were incubated overnight at 37°C.
3. A single colony was removed from the plate and used to inoculate 50ml of Luria-Bertani (LB) broth. The cultures were incubated overnight in a shaking incubator at 210rpm at 37°C.

9.16.2 Midi kit DNA extraction from bacterial cells (Qiagen)

1. Prior to harvesting the bacteria, buffer P1 was supplemented with 2mg RNase and buffer P3 was placed on ice.
2. For long-term storage, 150µl glycerol was added to 850µl of culture, vortexed and placed in the -80°C freezer.
3. The remaining culture was transferred to culture bottles and centrifuged at 3,020g at 4°C for 5 minutes. The supernatant was discarded and the pellet resuspended in 4ml buffer P1. Buffer P1 contains 50mM Tris-Cl and 10mM EDTA. The RNase digests any cellular RNA.
4. Clumps were removed by vortexing, the cells were lysed in 4ml buffer P2 and inverted 5-6 times, then incubated for 5 minutes at room temperature. Buffer P2 contains 1% SDS and 200mM NaOH. These components solubilise the cell membranes by breaking down the lipids, releasing the cell contents, and denature the DNA and protein molecules, respectively. The bacterial proteins, DNA and cellular debris were then precipitated with 4ml buffer P3, which contains acetic acid, and inverted 5-6 times. The circular plasmid DNA remained in solution.
4. The reaction mix was applied to a Qiagen tip, following equilibration with 4ml buffer QBT. The columns were washed twice with 10ml QC buffer and the DNA was eluted with 5ml QF buffer. All three buffers contain isopropanol.
5. 3.5ml isopropanol was added and the sample was centrifuged for 30 minutes at 23,665g at 4°C. The supernatant was discarded and 2ml 70% ethanol was added. The sample was centrifuged under the same conditions for 20 minutes. The supernatant was discarded and the DNA left to air-dry for approximately 1 hour. It was resuspended in 200µl distilled water and left at room temperature overnight.
5. The DNA quality was assessed as described in section 9.6.1.

9.16.3 Nick translation probe labeling (Vysis)

During nick translation, labeled nucleotides are introduced into the probe sequence, and the DNA is reduced in size to fragments <500kb. Smaller probes give weak signals whilst larger probes produce increased background signals. DNase 1 introduces single strand breaks, at random sites in the DNA, exposing the 3' and 5' termini. DNA polymerase then introduces labeled nucleotides to the 3' end and removes them from the 5' end.

1. 50nM spectrum red and spectrum green dUTP were reconstituted in 50µl distilled water to give a 1mM solution. 10µl was added to 40µl nuclease free water to give a working solution of 0.2mM.
2. 11µl of 0.3mM dTTP was added to 22µl nuclease free water to give a 0.1mM solution. 22µl each of 0.3mM dATP, dGTP and dCTP were combined to give a 0.1mM dNTP mix.
3. Working on ice, 2.5µl spectrum red or green working solution, 5µl 0.1mM dTTP, 10µl 0.1mM dNTP and 5µl 10x nick translation buffer were combined and vortexed. 17.5µl of DNA was added and pipette mixed. 10µl DNase and polymerase enzyme combination was added and the reaction was incubated at 16°C for 75 minutes.
4. 3µl 0.5M EDTA was added to stop the reaction.
5. To remove unincorporated dUTP and reduce the background signals, the probe was applied to a G50 sephadex column (Amersham). The bottom was snapped off the column, which was placed in a collection tube and centrifuged at 1,520g for 1 minute to compact the sephadex.
6. The flow through was discarded, 50µl TE buffer was applied to the column and it was centrifuged as above.
7. The column was placed in a new collection tube, the probe was applied and centrifuged at 1,520g for 2 minutes.

9.16.4 Probe precipitation

1. 10µl Cot-1 DNA was added to the probes to binds to the repetitive sequences. 1/10th of the volume of NaOAc and 2.5x the volume of 100% ice cold ethanol was added.
2. The probes were stored at -20° for 1-2 hours to precipitate then centrifuged at 16,000g for 30 minutes at 4°C.
3. The supernatant was removed and the pellet was air dried for 1 hour. 14µl hybridisation buffer (Vysis) and 6µl nuclease free water was added. The probes were stored at -20°C. The hybridisation buffer contains 50% formamide, 2x SSC and 10% dextran sulphate. Formamide reduces the melting temperature of DNA and allows lower temperatures to be used for denaturation and hybridisation. SSC

regulates the ionic strength of the solutions used and helps stabilize the nucleic acid duplexes.

9.16.5 Hybridisation

1. 1.5µl of both red and green labeled FISH probe was added to a 22x22mm coverslip and the slide was inverted over it.
2. Any air bubbles were gently squeezed from under the coverslip and the edges were sealed with rubber cement.
3. The slides were placed in a temperature controlled hotplate (Hybrite) at 70°C for 5 minutes, then overnight at 37°C.

9.16.6 Post hybridisation washes and counterstaining

1. Wash 1 was heated to 72°C in a water bath.
2. The slide was removed from the Hybrite and the rubber cement seal removed using forceps. The slide was dipped in 2xSSC to remove the coverslip.
3. The slide was transferred to wash 1 for 2 minutes then to the wash 2 for 1 minute. Post hybridisation washes remove any unbound or loosely bound probes, reducing background signals.
4. 7µl Vectashield mounting medium containing 4',6-diamidino-2-phenylindole (DAPI) (Vector laboratories) was applied to a 22x50mm coverslip. The slide was inverted over it and any air bubbles were squeezed out. DAPI specifically binds to the A-T rich sequences of double-stranded DNA in cell nuclei, forming blue fluorescent complexes. Its fluorescence provides vivid contrast to green, yellow or red fluorescent probes.

9.16.7 Analysis

1. Fluorescent signals were visualized using an Axioscope fluorescence microscope (Zeiss) and Macprobe 4.3 software (Applied Imaging). This system uses a mercury lamp, with filter sets to control wavelengths of light. Signals are seen as red, green or yellow dots.
2. The number, colour and position of fluorescent signals were counted in 100-200 nuclei per case, using appropriate filters for the probes. Only discrete and intact nuclei were scored.

9.17 Multiplex FISH (MFISH)

MFISH uses a mixture of 24 whole chromosome paint probes (wcp), each labeled with different combinations of up to three of the following fluorochromes; spectrum gold, spectrum red, spectrum far red, spectrum aqua, and spectrum green. Each of the 22 autosomes and the two sex chromosomes can be simultaneously identified in 24 discrete colors. It allows for the detection of subtle and complex chromosomal rearrangements,

and the identification of marker chromosomes in fixed metaphase spreads (figure 16). MFISH has a resolution similar to G-banding.

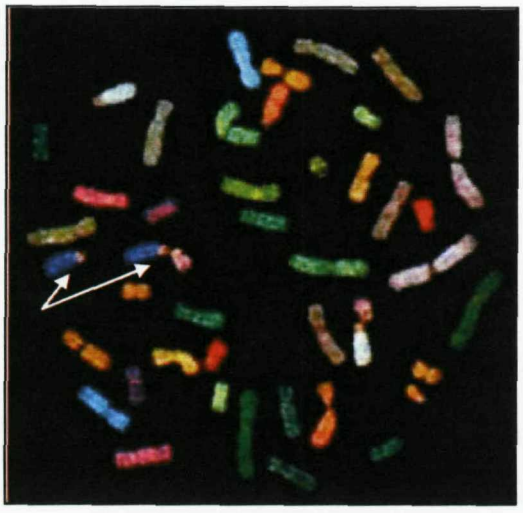


Figure 16. An MFISH labeled metaphase spread. Arrows indicate a chromosome pair involved in translocations.

9.17.1 Slide preparation

The slides were prepared as described in section 9.15.2. For MFISH it is of particular importance that metaphases are well spread to avoid overlapping of fluorescence signals. There should be no cytoplasm around the cells as this prevents probe hybridisation.

9.17.2 Hybridisation

1. The denaturation solution was heated to 73°C in a water bath. The slide was denatured in this solution for 2 minutes.
2. The slide was dehydrated in an ascending ice cold ethanol series at 75%, 90% and 100%, for 1 minute each. Washing in increasing concentrations of ethanol removes any water that might dilute the probes, and reduce hybridization, if left on the slide.
3. The MFISH probe solution (SpectraVysion, Vysis) was prepared containing 5µl MFISH probe and 5µl whole chromosome paint hybridisation buffer, and denatured at 73°C for 5 minutes.
4. 5µl of the probe was applied to a 22x22mm coverslip and the slide was inverted over it.
5. Any air bubbles were gently squeezed from under the coverslip and the edges were sealed with rubber cement.
6. The slide was placed in the Hybrite and hybridized for 48-72 hours at 37°C.

9.17.3 Post-Hybridisation washes

1. The post hybridisation washes were carried out as described in section 9.16.6 with the following amendment: the slides were air dried after wash 2.

2. The DAPI III (Vysis) was removed from the freezer and allowed to thaw. 10µl was applied to a 22x50mm coverslip. The slide was inverted over the coverslip and any air bubbles were squeezed out.
3. The slide was placed in the fridge for 10 minutes to allow the DAPI to spread out and bind to the DNA to improve the visualization of the metaphases.

9.17.4 Analysis

1. Metaphases were visualized as described in section 9.16.7. They were analysed using Powergene MFISH 1.1.1 software (Applied imaging).

10: Results

10.1 Cytogenetic and molecular characterisation of the cell line, ARH77.

10.1.1 Introduction

Developments in molecular cytogenetic techniques, such as MFISH, have greatly improved the accuracy of cytogenetic analysis, and aCGH has become the most widely used method for the genome wide detection of CNA. There are now a number of commercially available aCGH platforms, with variable levels of resolution and sensitivity. This study evaluates cytogenetic procedures, and compares the utility of three aCGH platforms, in the characterization of the cell line, ARH77. ARH77 was established from a 33 year old female with IgG plasma cell leukaemia (85), and is widely used in myeloma studies (86-88). However, more recent evidence has shown ARH77 to be an EBV-transformed B lymphoblastoid cell line (89). The complementary merits of these molecular cytogenetic approaches, and the value of aCGH in the discovery of new chromosomal abnormalities, were demonstrated during analysis of the cell line. On the basis of this pilot study, an appropriate aCGH platform for future large-scale projects was selected.

10.1.2 Material and methods

10.1.2.1 Cell Culture

The cell line, ARH77 (American Tissue Culture Collection (ATCC)), was cultured, DNA was extracted and cytogenetic preparations were made as described in sections 9.2, 9.4 and 9.15.1, respectively.

10.1.2.2 Cytogenetics

G-banding and MFISH were carried out on the same fixed cell suspensions as described in sections 9.15.2-3 and 9.17 respectively. The karyotype was described according to International System for Human Cytogenetic Nomenclature (84).
FISH confirmation of the aCGH data was undertaken using BAC clones (Appendix 3) from the genome locations as indicated by results from platform 1 (table 2). Clones were prepared, and FISH performed, as described in section 9.16.

10.1.2.3 aCGH

aCGH was performed on a DNA sample from the same cell culture passage using three independent platforms:

1. The Human BAC Array (Spectral Genomics), containing 2621 BAC probes at an average 1Mb resolution. The position of genes and BAC clones were determined using the National Centre for Biotechnology Information (NCBI) MapViewer for

Homo Sapiens, Build 35.1 (hg17). Dye-swap experiments were performed to give more confidence in the results.

2. The 44K Human Genome CGH Microarray (Agilent), consisting of approximately 43,000 60mer oligonucleotide probes, with a median probe spacing of 35kb. The content was sourced from The NCBI Homo Sapiens Build 35. Well characterized and cancer relevant genes were represented by at least 1 and 2 probes, respectively.
3. The 385K Human Whole-Genome Array (NimbleGen Systems Inc), which contains 385,000 oligonucleotide probes, spaced at approximately 6Kb throughout the genome. Probe lengths were adjusted (50mer-75mer) to equalize melting temperature across the entire set. The content was sourced from The NCBI Homo Sapiens Build 35.

The respective companies performed DNA labeling, hybridization and image capture (90-92). Spectral Genomics SpectralWare software was used to analyze Platform 1 data. The normalized fluorescent intensity ratios were computed for each clone from both dye-swap experiments, and were plotted together for each chromosome. DNA copy number gains were identified as the simultaneous deviation of the ratio plots for at least 2 clones by +/- 0.25 from a modal value of 1.0, with one ratio plot showing a positive deviation (upward) and the other showing a negative deviation (downward) at the same locus. DNA copy number losses show the opposite pattern (figure 18b). Platform 2 was analysed using Agilent CGH Analytics v3.2.32 as described in section 9.10 (figure 18c). NimbleGen Systems SignalMap1.8.0 software was used to analyse Platform 3 data. Each CGH ratio was converted to a log ratio and averaged based on a 60Kb window size. Copy number changes were identified as >5 consecutive data points deviating by +/- 0.25 from a normal value of 0 (figure 18d). In all platforms breakpoints were defined as the proximal and distal normal clones flanking the CNA.

10.1.3 Results

10.1.3.1 G-Banding and MFISH

G-banding and MFISH analysis of ARH77 (Appendix 4) revealed nine rearrangements involving chromosomes 2, 3, 5, 6, 8, 9, 11, 12 and 17, as shown in figure 17 and listed in table 2.

MFISH identified the chromosomal origin of several abnormalities which were ill-defined by G-banding alone, for example, add(2)(q?) and add(12)(q24) were more fully resolved as der(2)t(2;9) and der(12)t(8;12), respectively.

Chr	G-banding/MFISH result	aCGH results		
		Platform 1	Platform 2	Platform 3
2	der(2)t(2;9)(q?;?) der(3)t(2;3)(?;p14)	dim(2)(145.5[q22.3]-154.9[q24.1])	dim(2)(148.8[q23.3]-154.1[q24.1])	dim(2)(149[q23.3]-153.9[q24.1])
3	der(3)t(2;3)(?;p14) der(3;5)(q10;p10)	dim(3)(0.2[p26.3]-4.2[p26.2]) dim(3)(15.9[p25.1]-83.8[p12.2]) enh(3)(183.4[q27.2]-204.6[q29])	dim(3)(0.2[p26.3]-3.1[p26.2]) dim(3)(16.4[p25.1]-83.2[p12.2]) enh(3)(184.4[q27.3]-199.2[q29])	dim(3)(0[p26.3]-3.0[p26.2]) dim(3)(16.8[p25.1]-83.1[p12.2]) enh(3)(184.8[q27.3]-199.2[q29])
5	der(3;5)(q10;p10)	enh(5)(0[p15.33]-41.3[p12])	enh(5)(0.14[p15.33]-40.6[p12])	enh(5)(0[p15.33]-40.5[p12])
6	der(6)del(6)(p23)del(6)(q10)	dim(6)(0.1[p25.3]-25.5[p22.1])	dim(6)(0.2[p25.3]-33.9[p21.31]) **dim(6)(81.7[q14.3]-170.8[q27])**	dim(6)(0[p25.3]-33.5[p21.31]) **dim(6)(81.3[q14.3]-170.8[q27])**
8	der(9)t(8;9)(q22.1;p13) der(12)t(8;12)(q24.21;q24.3)	enh(8)(94.9[q22.1]-145.7[q24.3])	enh(8)(96.5[q22.1]-146.2[q24.3])	enh(8)(94.7[q22.1]-146.2[q24.3])
9	der(2)t(2;9)(q?;?) der(9)t(8;9)(q22.1;p13) der(9)t(9;17)(q10;q11.2)	enh(9)(35.4[p13.2]-64.9[q21.11]) enh(9)(84.7[q21.33]-134.1[q34.3])	enh(9)(35.3[p13.2]-38.7[p12]) enh(9)(88.3[q22.1]-138.2[q34.3])	enh(9)(35.4[p13.2]-39[p12]) enh(9)(88.3[q22.1]-138.4[q34.3])
11	del(11)(p13)	dim(11)(0.8[p15.5]-32.1[p13]) dim(11)(100.4[q22.1]-104.9[q22.3]) dim(11)(113.2[q23.3]-116.5[q23.3]	dim(11)(0.8[p15.5]-33.6[p13]) dim(11)(101.9[q22.2]-102.8[q22.3]) dim(11)(113.1[q23.3]-115.1[q23.3])	dim(11)(0.1[p15.5]-33.5[p13]) dim(11)(102[q22.2]-103[q22.3]) dim(11)(113.1[q23.3]-114.9[q23.3])
12	der(12)t(8;12)(q24.21;q24.3)	enh(12)(112.4[q24.21]-126.9[q24.32]) amp(12)(127.9[q24.32]-128.8[q24.33]) dim(12)(130.2[q24.33]-133.7[q24.33])	enh(12)(111.8[q24.21]-125[q24.31]) amp(12)(125.4[q24.32]-128.1[q24.33]) dim(12)(128.9[q24.33]-132.3[q24.33])	enh(12)(111.8[q24.21]-124.9[q24.31]) **dim(12)(124.9[q24.31]-125.3[q24.31])** amp(12)(126.6[q24.32]-128.1[q24.32]) dim(12)(128.1[q24.32]-132.4[q24.33])
17	der(9)t(9;17)(q10;q11.2)	dim(17)(0.1[p13.3]-20.2[p11.2])	dim(17)(0[p13.3]-21.8[p11.2])	dim(17)(0[p13.3]-22.5[p11.2)

Table 2. The copy number imbalances described by platforms 1-3, correlated with the cytogenetic results. CNA in bold were described by 1 or 2 platforms only.

10.1.3.2 aCGH analysis

Table 2 indicates the abnormalities defined by the three aCGH platforms. Platform 1 (BAC) revealed 16 CNA of which nine were deletions. Chromosomes 3, 9, 11 and 12

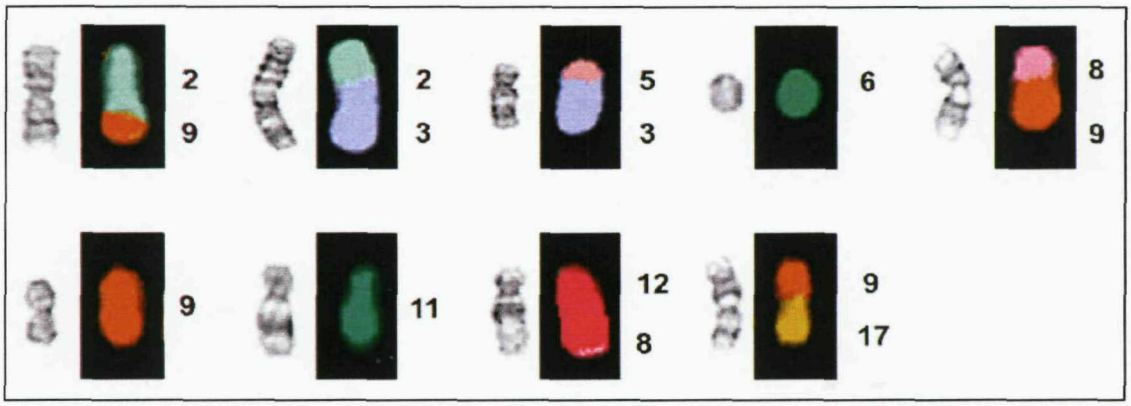


Figure 17. Partial G-banded and corresponding MFISH karyogram of the abnormal chromosomes identified in ARH77. The numbers indicate the chromosomes of origin for each abnormality.

harboured multiple abnormalities. The smallest aberration detected was a 3.3Mb deletion on chromosome arm 11q (dim(11)(113.2[q23.3]-116.5[q23.3]). This corresponded to the deletion of a single BAC clone. The largest CNA was a 67.9Mb deletion of 3p (dim(3)(15.9[p25.1]-83.8[p12.2])). The higher resolution available with platforms 2 and 3 provided further refinement of the breakpoints for these 16 aberrations. For example, all platforms exhibited complex CNA involving chromosome 3 (figure 18), while their position and extent was more precisely defined by platforms 2 and 3. In addition, a small deletion of 6q was observed with platforms 2 and 3 which was not seen using platform 1. Platform 3 also revealed a deleted region on 12q not detected by the other two platforms. The smallest aberration detected by the platforms 2 and 3 were dim(11)(101.9[q22.2]-102.8[q22.3]) (0.9Mb) and dim(12)(124.9[q24.31]-125.3[q24.31]) (0.4Mb), respectively (table 2). The largest CNA detected by platforms 2 and 3 were the 6q deletions dim(6)(81.7[q14.3]-170.8[q27]) (89.5Mb) and dim(6)(81.3[q14.3]-170.8[q27]) (89.1Mb) respectively).

10.1.3.3 FISH confirmation of aberrations detected by aCGH

BAC clones from platform 1 were used as FISH probes to confirm the copy number changes for chromosomes 2, 9, 11 and 12 (table 3).

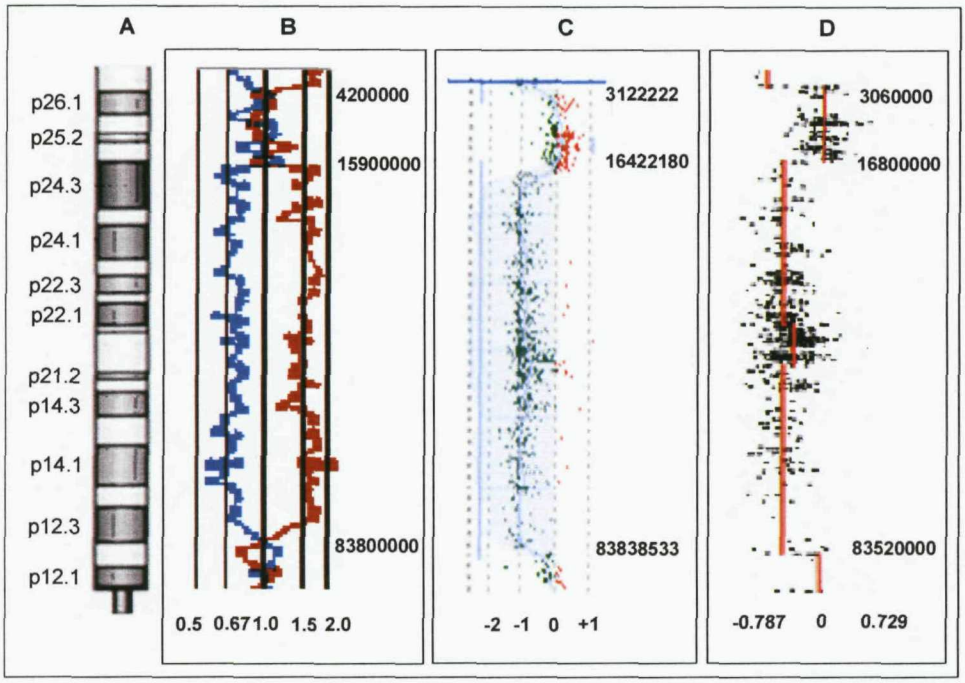


Figure 18. aCGH profiles of chromosome 3p abnormalities. To the right of each breakpoint, genomic positions are given. (A) Idiogram of chromosome 3p. (B-D) Profiles from platforms 1, 2 and 3, respectively

CNA	Corresponding BAC clone	Number of signals by FISH
dim(2)(145.5[q22.3]-154.9[q24.1])	RP11-357H16	1
dim(2)(145.5[q22.3]-154.9[q24.1])	RP11-79A11	1
enh(9)(85.5[p13.2]-134.1[q21.11])	RP11-89K14	3
dim(11)(0.8[p15.5]-32.1[p13])	RP11-44H16	1
dim(11)(100.4[q22.1]-104.9[q22.3])	RP11-88H18	1
dim(11)(100.4[q22.1]-104.9[q22.3])	RP11-33F6	1
enh(12)(112.4[q24.21]-126.9[q24.32])	RP11-91M21	2 (1 relatively large)
amp(12)(127.9[q24.32]-128.8[q24.33])	RP11-526P6	5 (1 relatively large)
amp(12)(127.9[q24.32]-128.8[q24.33])	RP11-91B1	3
dim(12)(130.2[q24.33]-133.7[q24.33])	RP11-81G12	1

Table 3. The CNA confirmed by FISH, the selected BAC clone targeted by the aberration, and the number of fluorescent signals observed in interphase FISH analysis of ARH77 cells.

Correlation between the aCGH and FISH of the 12q deletion is shown in figure 19. This complex CNA was shown to have >two copies of RP11-91M21, >five copies of RP11-

526P6, on different chromosomes, and one copy of RP11-81G12, confirming the gain, amplification and deletion defined by aCGH, respectively.

The final karyotype of ARH77, based on a combination of G-banding, MFISH, and aCGH is shown in table 4.

Karyotype
46,XX,<u>der(2)t(2;9)(**q23**;?)</u>,der(3)t(2;3)(**q23**;p14),der(3)t(3;5)(q10;p10), der(6)del(6)(p23)del(6)(q10),i(9)(p10),der(9)t(8;9)(q22.1;p13), der(9)t(9;17)(q10;q11.2),del(11)(p13),<u>der(12)t(8;12)(q24.21;q24.3)</u>[5]/46,idem,-6[5].

Table 4. The revised karyotype for ARH77. The underlined abnormalities were defined by MFISH, while the breakpoints in bold were identified by aCGH.

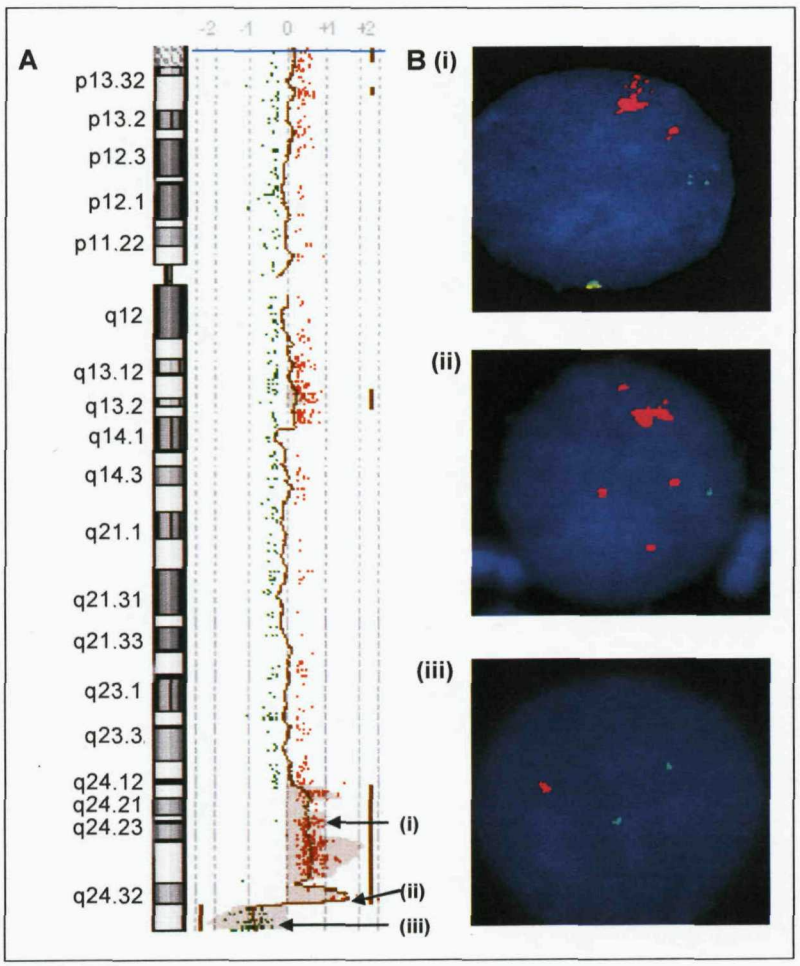


Figure 19. Array CGH and confirmatory FISH results for chromosome 12. (A) aCGH profile (platform 2) for chromosome arm 12q. The genomic location of the FISH probes is indicated. (B) Interphase FISH confirmation of each CNA. Probes targeting 12q and control loci were labelled in red and green respectively. (i-iii) Gain, amplification and deletion of 12q24.3 seen with probe RP11-9IM21, RP11-526P6 and RP11-81G12, respectively.

10.1.3.4 Comparison of cytogenetic and aCGH data

The CNA involving chromosomes 5, 8, 11 and 17 detected by aCGH corresponded to the der(3)t(3;5)(q10;p10), der(9)t(8;9)(q22.1;p13), der(12)t(8;12)(q24.21;q24.3), del(11)(p13) and der(9)t(9;17)(q10;q11.2) detected by G-banding and MFISH. The loss of 3p and 12q material was concordant with the der(3)t(2;3)(?;p14), and der(12)t(8;12)(q24.21;q24.3). Complex rearrangements of chromosome 9 resulted in copy number gain of 9q. The chromosome 2 rearrangements seen by cytogenetics did not result in large-scale copy number changes. Due to the complexity of the karyotype and the improved resolution offered by aCGH, it was not always possible to correlate cytogenetic and aCGH data. For example, gains of 3q and 12q were detected by all aCGH platforms, but were not cytogenetically visible.

10.1.4 Discussion

In this study, G-banding, MFISH and three aCGH platforms were used to characterize the karyotype of the cell line, ARH77, as fully as possible. Prior to this study, only a partial karyotype was available and the aberrations described were incompletely identified (table 5) (DSMZ, German collection of micro-organisms and cell cultures (www.dsmz.de)).

DSMZ karyotype
human near diploid karyotype with 8% polyploidy; 44~48,**XX**,del(1)(q23),add(2)(q21), add(3)(p11),**der(3)t(2;3)(q23;q26)**,**del(6)(p21)**,+9,**der(9;17)(q10;q10)** -sideline with der(X)t(X;1)(q23;p32),del(16)(p13.2) and unresolved der(3) and der(9)hsr markers

Table 5. The karyotype for ARH77 published on the DSMZ website. Aberrations in bold were confirmed and refined in this study.

We have confirmed the diploid chromosome complement and refined several chromosome rearrangements, such as the der(3)t(2;3)(q23;p14), der(6)del(6)(p23)del(6)(q10) and der(9)t(9;17)(q10;q11.2). In addition several novel rearrangements were highlighted. As this cell line is derived from an EBV-transformed B lymphoblastoid cell, these genomic aberrations probably result from the *in vitro* evolution of the karyotype. In light of this genomic variability, the cellular phenotype of ARH77 may vary, a fact worthy of consideration when performing novel studies or attempting to replicate previous work with this cell line.

All platforms detected the same 16 CNA, although the precise position of the chromosome breakpoints varied between the platforms due to increasing resolutions. The BAC array platform (platform 1) provided high signal intensities and allowed easy FISH

confirmation of aberrant probes using the same clones. However, the oligonucleotide platforms (platforms 2 and 3) provided greater resolution, and in this study detected additional CNA beyond the resolution of platform 1. Platform 3 had the greatest probe density and identified a 0.4Mb aberration (dim(12)(124.9[q24.31]-125.3[q24.31])) not detected by platforms 1 and 2. However, it also demonstrated the highest level of background noise and exact breakpoints were difficult to interpret. Platform 2 provided a specific, sensitive, high-resolution tool with limited background noise. These conclusions were supported in a recent publication (93) comparing Agilent, Affymetrix, and NimbleGen DNA copy number profiling platforms. Agilent CGH arrays were shown to provide the highest degree of reproducibility and the best sensitivity and specificity. Based upon this data, Agilent CGH arrays were selected for use in future studies.

10.1.5 Conclusions

In conclusion, the genomic instability inherent in many cell line models is also a feature of ARH77 and should be taken into consideration when used in biological studies. This study indicated that as far as possible, aCGH results should be interpreted in conjunction with data produced from other cytogenetic techniques. These are required for the detection of balanced chromosomal abnormalities and to determine the formation of individual rearrangements, for example whether they are derived from inversions, translocations, insertions etc., as well as the chromosomal origins of CNA. Variations between cells as a result of karyotypic evolution rely on cytogenetic analysis, which would pose technical challenges for aCGH. Thus, complementary use of techniques provide the most comprehensive interpretation of complex karyotypes for use in biological studies

The findings of this study have been published;

Parker H, Cheung KL, Robinson HM, Harrison CJ, Strefford JC. Cytogenetic and genomic characterization of cell line ARH77. Cancer Genet Cytogenet. 2008 Feb;181(1):40-5.

10.2 Global genomic profiling of *ETV6-RUNX1* positive ALL patients.

10.2.1 Introduction

The *RUNX1* gene, located at position 35.081975-35.343511Mb in chromosome band 21q22, encodes a transactivator widely expressed in haematological lineages. It has a DNA-binding domain (DBD) related to Drosophilia *RUNT* (figure 20) and mediates the expression of genes important in lymphopoiesis and B-lymphoid differentiation, such as *TCR* genes, B-cell receptors genes and *IL-3*. RUNX1 dimerizes with CBFß, via the *RUNT* domain, in order to effectively bind to its DNA target sites (94).
The organization of chromatin is fundamentally important in transcriptional control. The effect of chromatin structure on gene transcription is largely determined by the nature of the post-translational histone modifications, such as methylation and acetylation, of the regulatory region (95;96). Many transcriptional activators or repressors have been shown to recruit co-factors with intrinsic or associated histone acetyltransferases (HAT, coactivators) or histone deacetylases (HDAC, corepressors) (97). Transcriptional activation or repression by RUNX1 is thought to be dependant on its recruitment of p300 HAT and mSin3A/HDAC or groucho-related corepressors respectively (98-100). When associated with mSin3A/HDAC, the deacetylation of lysine residues in the amino terminal tails of the histone molecules allows the formation of compact chromatin, reducing the accessibility of promoters to transcriptional machinery (7). *RUNX1* was first identified in the fusion to the *RUNX1T1* gene in the t(8;21)(q22;q22) translocation in AML (101). More recently, *RUNX1* has been demonstrated to fuse with a plethora of other genes, such as the RUNX1T1-related *CBFA2TB* (102).

ETV6, located at 11.694055-11.939588Mb on chromosome 12p13, is an ETS (erythroblast transformation specific) family transcription factor gene and its importance as a regulator of embryonic development and hematopoiesis has been demonstrated in mouse models (17). ETV6 self associates via its helix-loop-helix domain (HLH), forming dimers or oligomers (103). The HLH domain can also interact with mSin3A, and a central repression domain located between HLH and the ETS DBD (figure 20), interacts with N-CoR and HDACs (98;104). It was first identified by its fusion with *PDGFRβ* in chronic myelomonocytic leukaemia with the translocation t(5;12)(q33;p13) (105), and has since been found rearranged with a number of tyrosine kinase genes, including *ABL1* and *JAK2* (106;107).

It is the fusion of these two promiscuous genes, *ETV6* and *RUNX1*, that is the molecular consequence of the t(12;21)(p13;q22) (26). The fusion consists of the *ETV6* sequence situated amino-terminal to the ETS-DBD and all known functional sequences of *RUNX1*,

including the RUNT DBD (figure 20). In this regard, the *ETV6-RUNX1* fusion is unique in that the *RUNX1* RUNT and the C-terminal regulatory sequences are preserved in the chimeric protein, and the fusion transcript is expressed from the *ETV6* promoter (7). Genomic sequence analysis of *ETV6-RUNX1* suggests that the predominant mechanism of chromosome translocation is double stranded DNA break followed by normal but error prone repair by nonhomologous end-joining (NHEJ) (108). In NHEJ, double stranded DNA breaks are repaired by direct ligation, without the need for a homologous template (109). Microhomologies (short homologous DNA sequences) in the single-stranded overhangs that are often present on the ends of double-strand breaks are used to guide repair. When these overhangs are compatible, NHEJ almost always repairs the break accurately, with no sequence loss, however, imprecise error-prone repair, leading to loss of nucleotides and translocations can also occur.

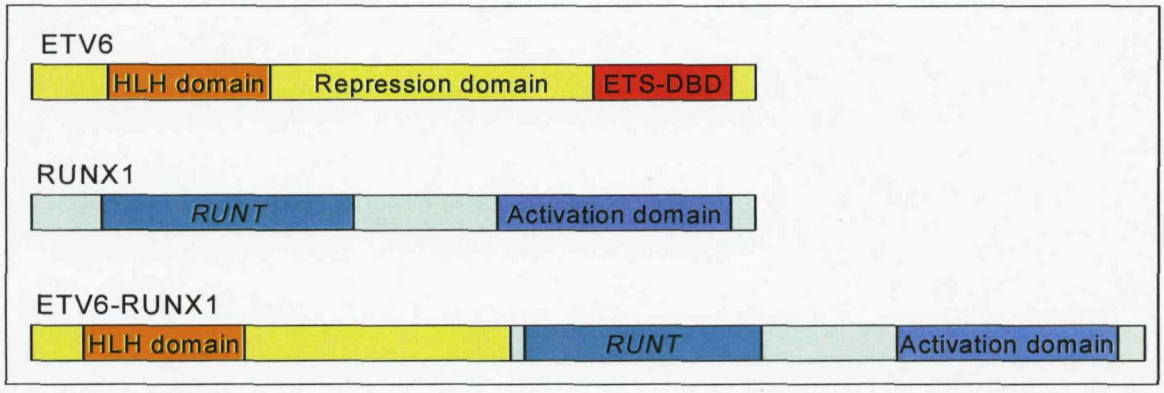


Figure 20. The organization of the protein domains in the ETV6, RUNX1 and ETV6-RUNX1 fusion proteins.

Nuclear co-repressors/HDAC complexes are recruited to the ETV6 moiety of the ETV6-RUNX1 fusion protein, converting RUNX1 to a constitutive repressor of transcription of hematopoietic genes (104). The interaction of ETV6-RUNX1 with N-CoR results in chromatin remodeling to an inactive configuration in promoters that are activated by normal RUNX1 (17). In mouse models, expression of ETV6-RUNX1 inhibits B-cell differentiation in early B-cell progenitors, leading to a deficit in mature B-cells (7). Therapeutically, HDAC inhibitors could be used, in combination with conventional therapies to relieve ETV6-RUNX1 mediated repression, allowing for dose reduction and lower therapeutic toxicities of currently used drugs (7).

Neonatal blood spots from 6/12 children and one set of twins presenting with *ETV6-RUNX1* fusions at diagnosis were PCR positive for the same fusions, confirming the prenatal initiation of this leukaemia (33). While translocation breakpoints show a degree of micro-clustering (110), only identical twins share the same clonotypic *ETV6-RUNX1*

genomic sequence, demonstrating a single-cell origin of leukaemia and spread of the clonal progeny from one twin to the other via the monochorionic placenta (7).

The protracted latency between the *in utero* formation of the fusion gene and disease presentation can be up to a decade or more (111). In addition, the concordance rate in monozygotic twins is approximately 10% (112) and the frequency of functional *ETV6-RUNX1* fusions in normal newborns is approximately 100x the rate of leukaemia (113). Also, *in vivo* functional studies of the *ETV6-RUNX1* fusion gene under control of the B-cell specific *IGH@* enhancer failed to cause leukaemia (114), while the introduction of additional genetic defects into the model, such as *CDKN2A/ CDKN2B* deletions, induce leukaemia at high frequencies (115). Collectively, these observations suggest that the development of *ETV6-RUNX1* positive ALL conforms to a two- or multi-step model of cancer development, and additional genetic changes are required for leukemogenesis. Deletion of the homologous *ETV6* allele and/or duplication of the fusion gene frequently accompany the translocation (116). Deletions are subclonal (117) and distinct in their genomic boundaries in twins (118), and in relapse versus diagnostic samples from the same individual (119), suggesting that these are postnatal secondary events. Other abnormalities of 12p, trisomy 21, monosomy X and hyperdiploidy with less then 52 chromosomes are also secondary aberrations associated with this subtype (17).

Several studies have investigated the prognostic value of the *ETV6-RUNX1* fusion, which in general is associated with a favorable outcome (120). Overall, the reported prognostic relevance of the fusion seems to depend on the intensity of the treatment regime (121). Within the *ETV6-RUNX1* group, significant heterogeneity in cellular *in vitro* drug sensitivity has been found, suggesting that additional genetic alterations may also be important in drug sensitivity and clinical outcome (122).

The aims of this study were;

- To investigate the frequency, size and genomic position of genomic changes in *ETV6-RUNX1* positive ALL patients using a combination of techniques.
- To identify copy number changes that may contribute towards leukaemogenesis, and identify genes that may represent diagnostic or prognostic markers, or therapeutic targets.

10.2.2 Materials and Methods

10.2.2.1 Patient samples

Diagnostic bone marrow samples from 34 patients positive for the *ETV6-RUNX1* fusion by FISH or RT-PCR were obtained, and DNA was extracted as described in section 9.4.

Chromosomal analysis was performed in the UK regional cytogenetics laboratories, and described according to the International System for Human Cytogenetic Nomenclature (84).

10.2.2.2 Array- based CGH

The samples were processed for aCGH analysis using the Agilent Human Genome CGH Microarray kit 244k, as described in section 9.10.

10.2.2.3 Quantitative RT-PCR

Where RNA was available, qRT-PCR was used to assess the expression of *TOX,* and *TBL1XR1* and its downstream target genes *RARB*, *RARA*, *CRABP1* and *CRABP2*, as described in section 9.7. Normal bone marrow and peripheral blood derived RNA samples were used as controls and tested in parallel. Based on data published on the Genehub-Gepis website (http://www.cgl.ucsf.edu/Research/genentech/genehub-gepis/index.html), salivary gland and thymus tissues were selected as positive and negative controls of *TBL1XR1* expression, respectively. However, in a series of pilot plates, with highly reproducible results, over- and under-expression of the gene could not be confirmed in these tissues. Therefore positive and negative controls were not included in the test plate. *GAPDH, GUSB* and B_2M were used as endogenous controls in the pilot plates. Comparison to *GAPDH* provided the most pronounced and consistent fold changes and it was therefore selected as the endogenous control for the test plates. For each patient the average Ct value for *GAPDH* was subtracted from the average Ct value for each test gene, generating the ΔCt value. For each test gene, the ΔCt values for each patient were then subtracted from the ΔCt for normal BM, and fold changes were calculated (2^-ΔΔCt) (see Appendix 6 for a worked example). Thymus and salivary gland tissues were selected from Genehub-Gepis as positive and negative controls for *TOX* expression, respectively. The over- and under- expression of *TOX* in these tissues was confirmed in pilot plates. *GAPDH* was used as an endogenous control, and patient expression levels were compared to normal BM as described above.

10.2.2.4 Mutation analysis

HRM analysis and DNA sequencing was performed for exons 4-14 of *TBL1XR1*, in 6 patients with deletions including or flanking the gene, as described in section 9.14. (See Appendix 5 for primer sequences)

10.2.2.5 FISH

Where fixed cells were available, selected CNA detected by aCGH were confirmed by FISH, as described in section 9.16 (Appendix 3).

10.2.3 Results

10.2.3.1 Array-based CGH

In total aCGH detected 326 CNA in 34 patients, with between two and 24 aberrations per case (Appendix 7 and figure 21). Twenty four aberrations involved whole chromosome losses (n=17) or gains (n=7), and of the remaining 302 CNA, deletions were far more frequent than gains (n=237 versus n=65). Deletions targeted chromosome arms 12p (n=34, two bi-allelic deletions), 6q (n=27), 9p (n=24, five bi-allelic deletions), 3q (n=23), 8q (n=16) and 11q (n=13). Recurrent gains affected chromosome arms Xq (n=11), 21 (n=9) and 12p (n=6). The smallest copy number change observed was an 11kb deletion on chromosome arm 9q (13.4547308-13.4558379 Mb). By consulting the database of genomic variants (www.projects.tcag.ca/variation/), an additional 325 copy number variants (CNV) were identified, which targeted 48 different loci (Appendix 8). No CNV were detected on chromosomes 12, 18, 19, 20, 21 and Y. Deletions at 2p11.1 (n=30), 7q34 (n=33), 14q11 (n=26), 14q32.33 (n=30) and 22q11.22 (n=37) target the *TCRA/D, TCRB, IGK@, IGH@* and *IGL@* loci, respectively, and most likely result from somatic recombinations. An additional 12 CNV were targeted in over 10% of patients, and between four and 26 CNV were observed in each patient.

The diagnostic cytogenetic and FISH data is provided in Appendix 7. Cytogenetic analysis was successful in 91% (31/34) of cases. Seven cases exhibited a normal karyotype and one (patient 11469) had a balanced translocation, (t(3;9)(q29;p21)), as the sole visible abnormality. In these cases, aCGH demonstrated a total of 60 CNA, while a further 18 were detected in three cases with a failed cytogenetics result. In total, cytogenetic analysis detected 65 visible abnormalities that would result in copy number imbalances. Array-CGH confirmed and refined 37 of these aberrations. A further 21 imbalances had been described as additional unidentified material or marker chromosomes. The origin of the material in 19 of these aberrations could be deduced from the aCGH data, although FISH would be required for definitive confirmation. Of the remaining nine aberrations unconfirmed by aCGH, eight were detected in fewer then nine metaphase cells by cytogenetics (range=1 to 9 cells), indicating their presence in a low proportion of cells in the sample. Patient 10086 had a deletion of 12p, which was not seen by aCGH. However, cytogenetics also identified additional material on chromosome 18 in this patient, suggesting this may be a balanced translocation.

Chromosome 12p

Deletions of the short arm of chromosome 12 ranged in size from 3.8 Mb to 28.3 Mb, and encompassed the *ETV6* gene in 16 patients (figure 22). Two patients harbored

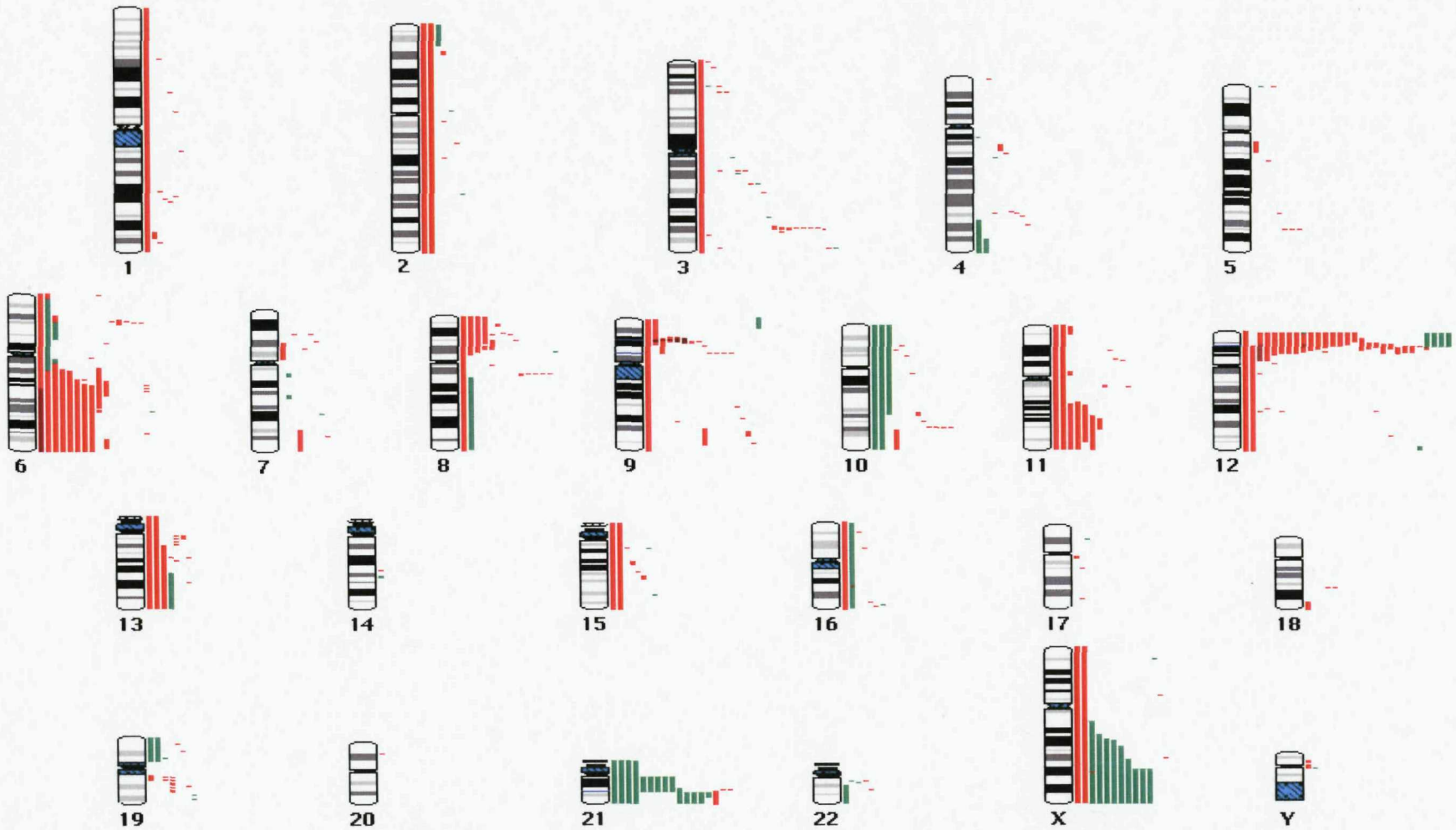


Figure 21. CNAs detected in 34 patients with ETV6-RUNX1 *positive ALL. To the right of each chromosome idiogram, red and green vertical lines show copy number losses and gains, respectively. Each vertical line is positioned to show its extent and genomic location and represents a single change in a single patient. Homozygous deletions are shown in dark red.*

heterogeneous bi-allelic deletions of this region (patients 7361 and 10073). A 3.24 Mb commonly deleted region (CRD) was identified between genomic position 11.466866 and 14.708804 Mb, encompassing 33 expressed sequences including *ETV6*, *BCL2L14* and *CDKN1B*. A further eight cases targeted *ETV6* directly with either a proximal (n=3) or distal (n=5) deletion breakpoint. These breakpoints occurred within intron 1 (n=2), and intron 5 (n=6). One case displayed both deletion breakpoints within intron 4 and exon 8 of *ETV6* respectively (patient 11469).

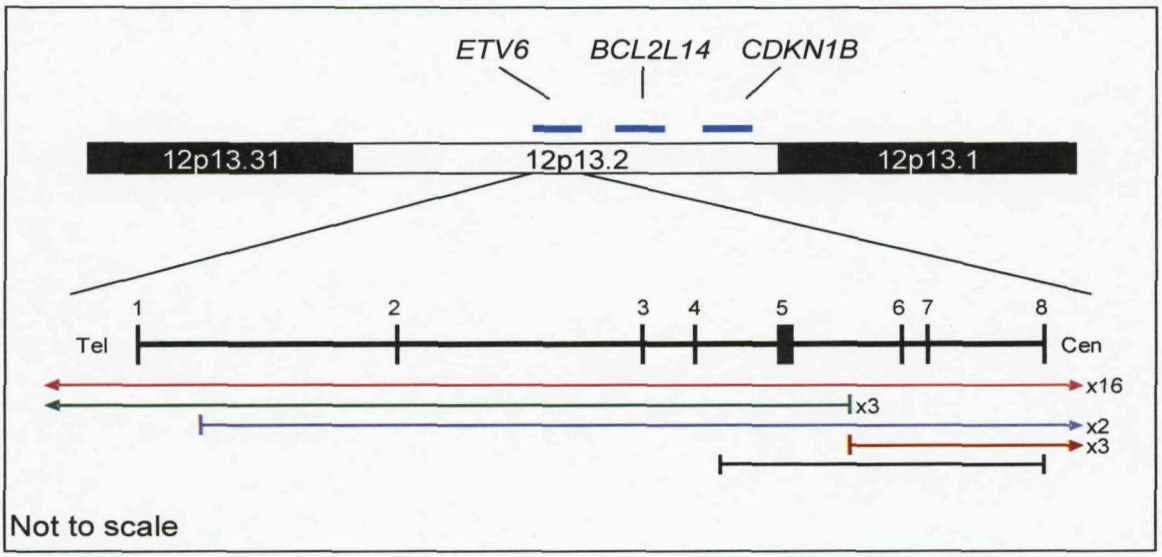


Figure 22. A partial ideogram of the 12p13.2 region, showing the location of the ETV6, BCL2L14 *and* CDKN1B *genes, and a schematic representation of the* ETV6 *gene, showing the positions of the exons (numbered 1-8), and their relative size (indicated by thickness of line). The location and extent of 25 deletions involving* ETV6 *are shown (xn indicates the number of patients a deletion was observed in). Arrows indicate that a deletion extends beyond the gene locus.*

Chromosome 12p and 21q

Four patients exhibited gains of both 12p and 21q (figure 23), which reflected the presence of an additional der(21)t(12;21), with a second copy of the *ETV6-RUNX1* fusion, confirmed by FISH. In a single case, a gain of 21q targeted *RUNX1* at the distal breakpoint, but this was not associated with a gain of 12p. A further four cases had gains of 21q which encompassed the entire gene (n=1) or disrupted *RUNX1* with the centromeric breakpoint (n=3). Of the three deletions involving 21q, two shared a common region of deletion (CRD) between 34.226926 and 34.419703 Mb.

Chromosome 6

Twenty seven deletions involving chromosome 6 were detected, with nine and 18 targeting the short and long arm, respectively. The 6p deletions involved two common

regions, each observed in four patients. These were genomic positions 26.253494 to 26.378998 Mb (containing 16 histone gene family members from *HIST1H1E* to *HIST1H2BH*) and 29.201891 to 29.260804 Mb. Two cases shared the distal breakpoint at 29.201891Mb. The 18 deletions of 6q were identified in 11 patients. Nine were large scale, ranging from 40.0 Mb to 92.0 Mb in size. All shared a CRD between 111.393615 and 112.432599 Mb.

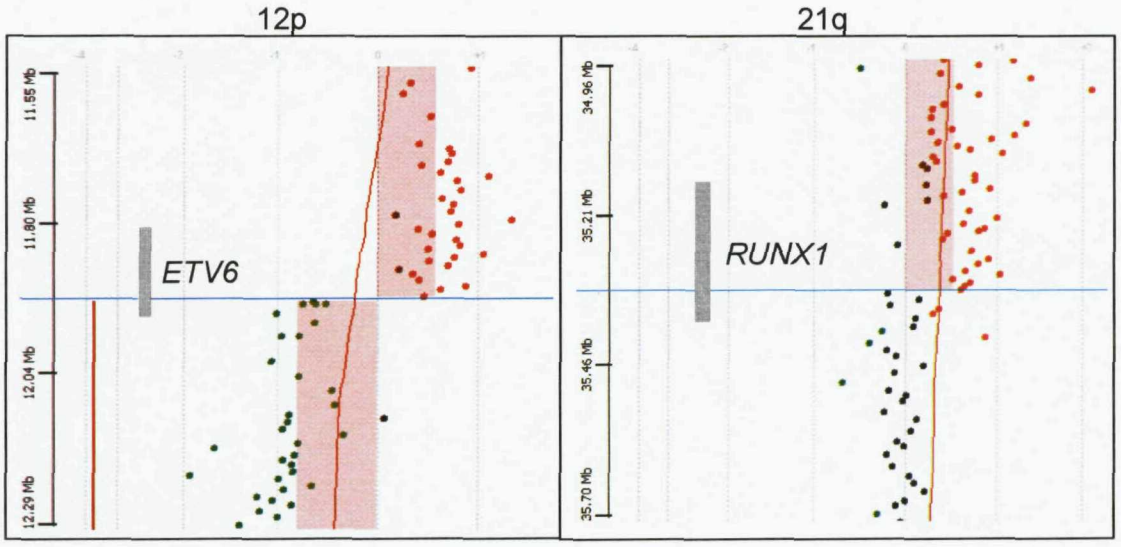


Figure 23. Array-CGH profiles for patient 3975 showing gain of 12p and 21q, with breakpoints located within the ETV6 *and* RUNX1 *genes (grey bars). Green, black and red dots represent deleted, normal and gained probes respectively. The moving average is represented by the red line. The pink shaded area indicates the extent of the aberration.*

Chromosome 11

Twelve deletions targeted chromosome 11 in 10 patients, affecting both the short and long arms (n=5 and 7, respectively). The only recurrent CNA on 11p resulted in targeted deletion of the *RAG1* and *RAG2* gene loci (n=2). Two CRD were identified on 11q; the first (n=2) was between 65.094963 and 65.151518 Mb containing six sequences and the second (n=5) spanned genomic position 102.510235 and 115.446921 Mb encompassing 30 genes/EST.

Chromosome 9p

Deletions of the short arm of chromosome 9 were observed in 12 patients, and targeted two genomic regions. Seven patients had deletions of regions including the *CDKN2A/B* locus, with a CRD between genomic positions 21.947735 and 21.993710 Mb, excluding the *MTAP* gene. Five of these cases harbored bi-allelic deletions of a common region between 21.030758 and 22.130685 Mb in addition to a larger deletion of 9p. In a single case (patient 2196), the 9p deletion encompassed both the *CDKN2A/B* and *PAX5* loci,

while in a further four cases, *PAX5* was targeted specifically. Although the CRD containing *PAX5* was between 36.996073 and 37.011079 Mb, the deletion in three cases included *ZCCHC7*, either resulting in partial (n=2) or complete (n=1) loss of the gene (figure 24).

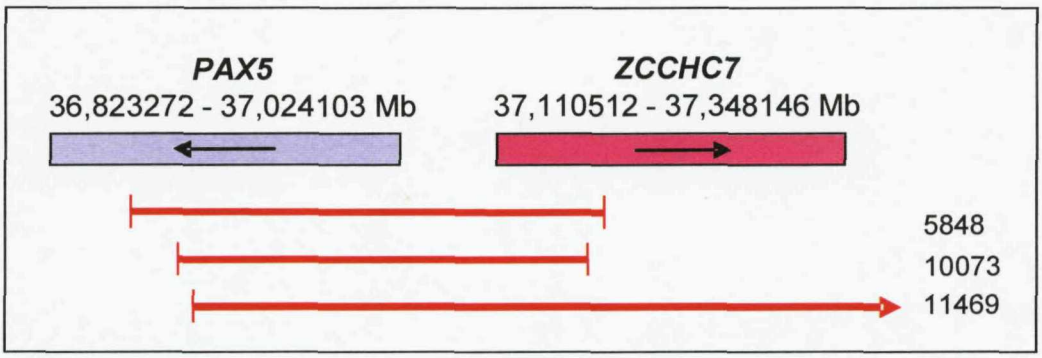


Figure 24. The CNL involving the PAX5 *and* ZCCHC7 *genes. Red bars represent the position and extent of the deletions. Black arrows indicate the direction of gene transcription.*

Chromosome 13q

Ten deletions of 13q occurred in five patients, of which three exhibited a common region between 43.764662 and 43.899260 Mb, and a single case showed a focal deletion of *RB1* (patient 3602).

Chromosome 8

Large scale and submicroscopic deletions of 8p were identified in eight patients (n=5 and 4, respectively). Four of six patients with deletions of the long arm of chromosome 8 shared a CRD between 60.199387 and 60.367552 Mb. The sequence 5' (telomeric) of the *TOX* gene is targeted in five cases. Two cases shared a distal breakpoint at 60.403052 Mb (figure 25).

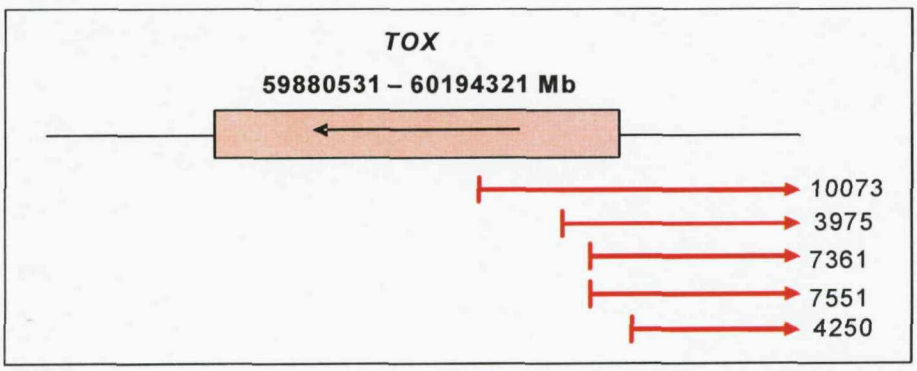


Figure 25. The CNL involving the TOX *gene. Red bars represent the position and extent of the deletions. The black arrow indicates the direction of gene transcription.*

Chromosome 3q

Eight patients had deletions of 3q that targeted the *TBL1XR1* gene and/or distal sequences, of which five deletions exhibited a common region between 178.265404 and 178.392151 Mb (figure 26).

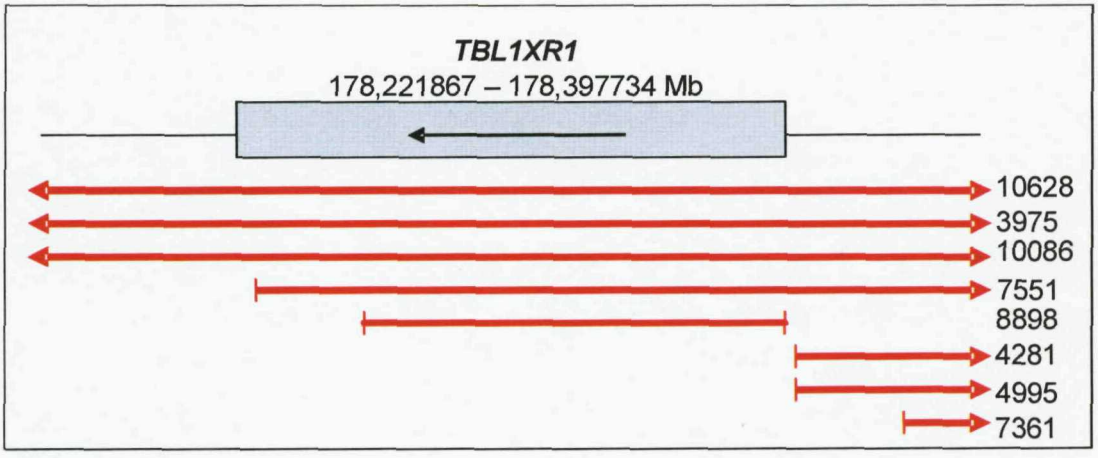


Figure 26. The CNL involving the TBL1XR1 *gene. Red bars represent the position and extent of the deletions. The black arrow indicates the direction of gene transcription.*

Chromosome Xq

Duplications of the telomeric part of Xq were observed in nine patients and, with one exception, all shared a common 22.64 Mb region of duplication between genomic position 122.949886 and 145.592983 Mb. In the remaining case the duplication was intersected by a 44.09kb deletion between positions 128.794606 Mb and 128.838697 Mb.

Other aberrations

Focal deletions of *ADD3* (111.746116-111.885313 Mb) occurred in three patients with 10q deletions, and three of the six deletions of 5q specifically targeted *EBF1* (158.058006-158.459347 Mb). Deletion of *CEBPA*, located at 38.482543-38.485460 Mb on 19q13.1 occurred in a single patient. A single case had a deletion encompassing the *IKZF1* (50.411725-50.440293 Mb) gene on chromosome 7p11.2

10.2.3.2 Quantitative RT-PCR

Quantitative RT-PCR was used to investigate the effect of copy number loss on the expression levels of *TBL1XR1,* and its downstream target genes *RARA*, *RARB*, *CRABP1* and *CRABP2*. RNA or cDNA was available for six patients with deletions of the gene or the surrounding sequences (patients 7551, 8898, 4281, 7361, 4995, 10086), and six *ETV6-RUNX1* positive patients with normal copy number at this locus (patients 4323, 5062, 5848, 3700, 4932, 4598).

Considered together, the cohort of patients with deletions of *TBL1XR1* had lower expression levels of the gene than patients with normal copy numbers (average fold change = 2.19 vs. 5.13 respectively) (table 6/figure 27). However, individually the fold changes were variable; three cases with deletions (8898, 4281 and 4995) had fold changes similar to those without deletions. Two cases with low expression levels (10086 and 7551) had deletions of the entire gene, whilst two cases with higher expression levels (4281 and 4995) had deletions of the downstream region only. The remaining two cases (8898 and 7361) did not correspond to this pattern.

On average the expression levels of the other genes were highest in the patients with *TBL1XR1* deletions (table 6/figure 27). However, as described above, the individual fold changes were variable and for each gene the fold change for several of the patients with deletions was similar to those without. The difference in expression levels between the two cohorts were statistically significant for *TBL1XR1* only (p=0.005). This experiment was repeated twice, and the results remained consistent.

	Patient Id	*TBL1XR1*	*RARB*	*RARA*	*CRABP1*	*CRABP2*
Patients with deletions	7551	0.65	1.13	8.34	0.19	2.2
	8898	3.61	1.74	3.73	1.18	9.65
	4281	3.16	0.07	1.67	0.98	0.34
	7361	0.56	8.06	0.6	0.3	2.43
	4995	4.72	0.02	1.95	0	1.16
	10086	0.44	0.01	0.4	0.01	0.39
	Mean	**2.19**	**1.83**	**2.78**	**0.44**	**5.69**
	StDev	**1.86**	**3.13**	**2.97**	**0.509**	**3.51**
Patients without deletions	4323	6.19	0.08	1.44	0.04	0.47
	5062	3.58	0.01	1.31	0.01	0.03
	5848	4.82	0.01	0.82	0.01	0.51
	3700	5.35	0.04	1.34	0.01	0.26
	4932	5.39	0.03	1.66	0	0.08
	4598	5.46	0.13	1.04	0.02	0.19
	Mean	**5.13**	**0.05**	**1.26**	**0.015**	**0.25**
	StDev	**0.877**	**0.047**	**0.297**	**0.013**	**0.198**
	P-value	**0.005**	**0.192**	**0.242**	**0.066**	**0.121**

Table 6. The expression levels of the genes TBL1XR1, RARB, RARA, CRABP1 *and* CRABP2 *in six patients with deletions of* TBL1XR1*, and six patients with normal copy number at this locus. The means, standard deviations and p-values are included.*

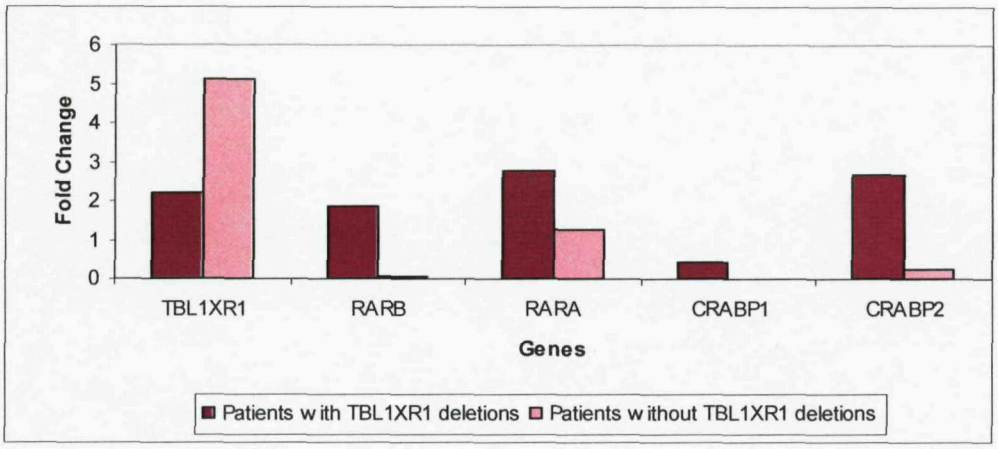


Figure 27. Histogram showing the expression levels of the five genes in the two cohorts of patients, relative to normal bone marrow.

The effects of CNL on expression levels of *TOX* were also investigated. The six *ETV6-RUNX1* positive patients with normal copy numbers of *TBL1XR1* were used as controls, as they also had normal copy numbers for *TOX*. RNA or cDNA was available on 3/5 patients harboring deletions of the sequence 5' of the *TOX* gene (10073, 7551 and 7361). In all three patients with deletions of *TOX*, expression levels of the gene were lower then in patients without a deletion at this locus (p=0.04) (table 7/figure 28).

	Patient Id	TOX expression level
Patients with deletions	10073	12.36
	7551	2.34
	7361	8.97
	Mean	**7.89**
	StDev	**5.09**
Patients without deletions	4323	13.15
	5062	14.99
	5848	13.84
	3700	25.96
	4932	32.11
	4598	33.96
	Mean	**22.33**
	StDev	**9.53**
	P-value	**0.04**
	Thymus (+)	15.45
	Salivary gland (-)	0.73

Table 7. Expression levels of TOX *in three patients with deletions of the gene, and six* ETV6-RUNX1 *positive patients with normal copy number at this locus. The means, standard deviations and p-values are included.*

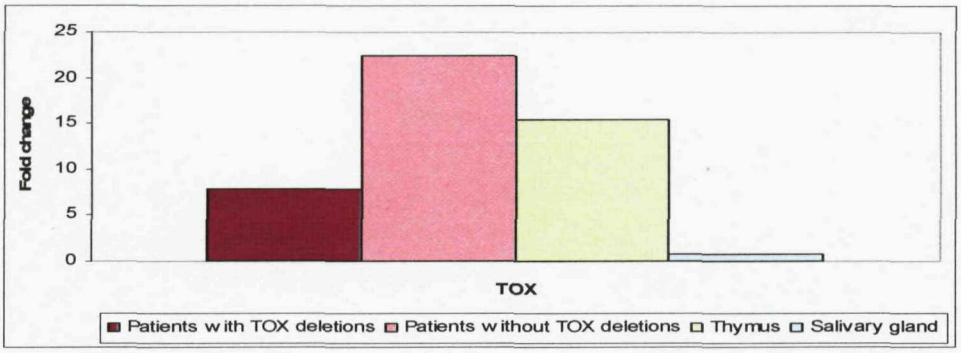


Figure 28. Histogram showing the expression levels of TOX *in the two cohorts of patients, and the positive and negative control (thymus and salivary gland respectively), relative to normal bone marrow.*

10.2.3.3 Mutation analysis

The author gratefully acknowledges Dr Qian An for providing much of the following data.

Exons 4-14 of the *TBL1XR1* gene code for protein sequence. HRM analysis was performed on these exons to determine whether sequence mutations were contributing to *TBL1XR1* inactivation. Six patients with deletions of the gene were analysed (3975, 4281, 4995, 7361, 8898 and 10628) and melting curve shifts, indicating sequence mutations, were observed in each exon, in at least one patient (see figure 29 as an example). However, sequencing failed to confirm any mutations.

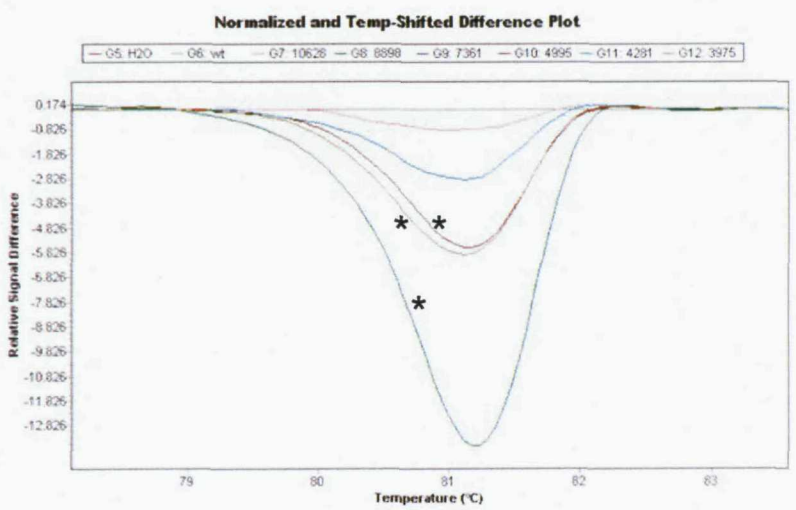


Figure 29. The normalized and temperature-shifted difference plot for exon 7 of TBL1XR1*, in six patients with deletions of the gene. Asterisks (*) indicate three patients (3975, 4995 and 8898) with significantly shifted curves.*

10.2.3.4 FISH

The author gratefully acknowledges the Leukaemia Research Cytogenetics Group, Southampton, for providing the following data.

Where fixed cell material was available deletions of *6p22.1, ADD3, EBF1, PAX5, TBL1XR1, TOX* and gains of Xq were confirmed, for the majority of cases, using specifically designed FISH probes. FISH probes were selected to provide the best possible coverage of the commonly gained or deleted regions. However, due to the heterogeneous nature of these recurrent copy number changes a small number of these cases could not be confirmed with the selected probes. (See Appendices 3 and 9 for details of probes and FISH results, respectively).

To investigate whether these recurrent CNA have prognostic significance, the diagnostic samples for up to 41 *ETV6-RUNX1* positive patients who had relapsed were screened with these probes, as well as break-apart probes for *ETV6* and *RUNX1* (material permitting). These break-apart probes were included because they provide more information on the nature of the *ETV6-RUNX1* rearrangements than the extra signal probe used in the diagnostic test. Based on the results of the confirmatory FISH screen above, CNA may have been missed in a subset of patients due to probe coverage limitations. Table 8 summarizes the results of this screen.

CNA	**Percentage of patients with CNA**
6p22.1 deletions	2.4% (n=1)
ADD3 deletions	19.5% (n=8)
EBF1 deletions	9.7% (n=4)
PAX5 deletions	10.5% (n=4/38)*
TBL1XR1 deletions	9.7% (n=4)
TOX deletions	13.1% (n=5/38)*
Gain of Xq	12.1% (n=5)
Loss of 2nd *ETV6* allele	68.3% (n=28)
+der(21)	14.6% (n=6)
Loss of 2nd *ETV6* allele and +der(21)	7.3% (n=3)

Table 8. The percentage of relapse patients harboring the selected recurrent copy number changes. Asterisks indicate where fewer then 41 patients were analysed, due to limited availability of fixed cells.

The most common CNA in these patients was loss of the homologous *ETV6* allele, seen in a proportion of cells in 68.3% of patients. A rearranged *ETV6* signal pattern and normal *RUNX1* signal pattern were seen in one case, indicating that a section of *ETV6*, including one of the probes, was excised from chromosome 12 and inserted into *RUNX1* between the 2 break-apart probes, to produce the *ETV6-RUNX1* fusion by an insertion rather than a fusion event. Deletions of *ADD3* and *TOX*, an additional copy of der(21), and gains of Xq were also frequent (19.5%, 13.1%, 14.6% and 12.1% respectively). In 13/38 cases, none of these secondary aberrations were detected.

Due to the increased frequency of *ADD3* deletions in relapsed patients, a retrospective screen of a further 200 diagnostic samples from *ETV6-RUNX1* positive patients was performed using the *ADD3* probe, and the *ETV6* and *RUNX1* break-apart probes. Table 9 summaries the results.

CNA	Percentage of patients with CNA
ADD3 deletions	10.5% (n=21)
Loss of 2nd *ETV6* allele	49% (n=98)
+der(21)	10.5% (n=21)
Loss of 2nd *ETV6* allele and +der(21)	11.5% (n=23)

Table 9. The percentage of patients harboring ADD3 *deletions, deletions of the second* ETV6 *allele, or extra copies of the der(21) in 200 ETV6-RUNX1 positive patients.*

10.2.4 Discussion

10.2.4.1 Array- based CGH

To investigate genomic abnormalities in ALL, numerous cytogenetic and CGH analyses have been conducted (62-66;123-125), all of which corroborate the findings of this study. Most, however, were performed at lower resolution, or on fewer patients than documented here.

Four of the patients in this study were also included in the 1 Mb BAC aCGH study performed by Strefford *et al* (62). The higher resolution array used in this investigation confirmed and refined all of the CNA documented by Strefford *et al* in these patients, as well as identifying an additional 26 aberrations.

The application of high resolution aCGH to study secondary abnormalities in *ETV6-RUNX1* positive ALL patients has revealed unexpected genomic complexity, with as many as 24 CNA seen in a single patient. At least two chromosomal imbalances were involved in all samples, further supporting the two- or multi-step model of cancer development. The use of FISH probes to confirm the CNA affecting 6p22.1, *ADD3*, *EBF1*, *PAX5*, *TBL1XR1*, *TOX* and Xq, and the strong correlation between the cytogenetic and aCGH data, demonstrated the accuracy of aCGH. Many of the 238 aberrations detected by aCGH alone were classified incompletely, or were below the resolution of cytogenetics. A total of 330 CNA targeted reported polymorphic regions. However, constitutional samples would be required to fully confirm the nature of these aberrations. Currently, the significance of these findings is unknown. Although the majority of these CNV will reflect normal variation, a subset may influence disease susceptibility, drug response, or be directly involved in leukaemogenesis (refer to section 7.10).

This investigation highlighted genomic regions harbouring significant genes, worthy of further study.

Chromosome 12p and 21

Loss of the non-translocated *ETV6* allele is well documented in *ETV6-RUNX1* positive ALL (62;64;65;126;127). In this study, deletions of 12p targeted *ETV6* in 73% of cases, in accordance with previously published data. In two patients the deletions appeared to be biallelic. As a partial copy of one *ETV6* allele is retained in the *ETV6-RUNX1* fusion, this finding is implausible. The presence of subclones, with different copy numbers of the gene may account for the overall variation in copy number. Loss of *ETV6* is not confined to *ETV6-RUNX1* positive ALL patients (see section 10.3 and 10.4). It is seen in other subtypes of leukaemia (62;66;128) and in a subset of solid tumors, for example prostate cancer (129). In a retrospective screen of 41 *ETV6-RUNX1* relapsed patients, loss of the non-rearranged *ETV6* allele was the most frequent secondary aberration.
Four cases (11%) had gains of both 12p and 21q, corresponding to an additional copy of the der(21)t(12;21), a recurrent secondary abnormality specific to *ETV6-RUNX1* positive ALL, readily visible by FISH. Whole chromosome 21 gains were also observed (n=3) in accordance with published data (65). ALL patients with trisomy 21 are thought to have a low risk of treatment failure (130). Gain of chromosome 21 is seen in other subtypes of ALL, such as high hyperdiploidy (131) and iAMP21 (partial gains) (37). It is also the constitutional abnormality in Down syndrome (132) and has been associated with an increased incidence of Alzheimer's disease (133).

Chromosome 9p

Deletions of 9p, targeting the *CDKN2A* gene are found in approximately 30% of childhood and adult ALL (134). *CDKN2A* encodes two proteins; $p16^{INK4a}$ and $p14^{ARF}$, which have different promoters and first exons (exons 1α and 1β respectively), but share exons 2 and 3 in alternate reading frames. Both proteins are cell cycle regulators and are thought to be non-redundant tumour suppressor genes. The high incidence of 9p21 deletions implicates inactivation of *CDKN2A* in both T- and B-ALL leukemogenesis. The deletions are sometimes associated with gene inactivation by hypermethylation of promoter regions. Additionally, $p14^{ARF}$ is a transcriptional target of *RUNX1*. *ETV6-RUNX1* is a dominant inhibitor of *RUNX1*, suggesting that $p14^{ARF}$ expression may be repressed in *ETV6-RUNX1* positive ALL (135). Mono- and bi-allelic deletions of various sizes are found, leading to variable co-inactivation of contiguous genes such as *CDKN2B* and *MTAP*. A mixture of both mono- and bi-allelic deletions may be present in a single patient. Deletions are found in a spectrum of ALL subtypes and are also reported, at a lower frequency, in AML, CML and MDS. However, promoter hypermethylation is more common in these diseases (134). Whilst *CDKN2A* deletions have no prognostic value in

B-ALL (135), they have been shown to functionally cooperate with *ETV6-RUNX1* fusions during leukemogenesis (115). They may also be linked to disease progression. In a comparison of 17 diagnostic and relapse ALL samples, Beverloo *et al* (134) found four cases with *CDKN2A* deletions at relapse, that were not present at diagnosis. The co-inactivation of *CDKN2B* and *MTAP* also has no influence on outcome, although *MTAP* is involved in purine and methionine salvage metabolism and *in vitro* studies have shown its deletion makes cancer cells more sensitive to drugs such as methotrexate (135).
Seven patients in this study had deletions of *CDKN2A/B* (20%), in accordance with published data (62;64;66), including bi-allelic deletions in five cases. *MTAP* was not encompassed by the minimally deleted region.

In a single case, a large scale 9p deletion encompassed both *CDKN2A* and *PAX5*. *PAX5* was targeted specifically in four additional cases. PAX5 is a member of the highly conserved paired box-domain family of transcription factors. It has two distinct promoters, resulting in two 5' exons- 1a and 1b. PAX5b is expressed in the CNS and testis, and both transcripts are expressed in B-lymphoid lineage cells. PAX5a, also known as B-cell specific activator protein (BSAP), can act as a transcriptional activator or repressor, and has a critical role in B-cell lineage commitment, differentiation, and maintenance. PAX5 is continually expressed from the early stages of B-cell development up to the formation of mature B-cells. It promotes the activity of B-cell specific genes and suppresses alternate lineages choices, and represses the induction of plasma cell (PC) formation (136). Blimp-1, a crucial plasma cell (PC) differentiation protein, down-regulates PAX5 expression following antigen-driven B-cell activation. This allows the induction of PC specific genes, such as *XBP1*, and the terminal differentiation of PC (137).
In a large scale screen of 242 cases of pediatric ALL, Mullighan *et al* (63) found both mono- and bi-allelic deletions of whole or partial *PAX5* in 29% of B-ALL cases (n=56). In their *ETV6-RUNX1* positive patients 28% had mono-allelic *PAX5* deletions, compared to 15% in this study. Three cases also had deletions of *EBF1*, a gene that functions upstream of *PAX5* (63). Numerous translocations involving *PAX5* have also been described, including the t(9;14)(p13;q32) (*PAX5-IGH*) (138), dic(9;12)(p13;p13) (*PAX5-ETV6*) (139), t(7;9)(q11;p13) (*PAX5-ELN*) (136) and t(3;9)(p13;p13) (*PAX5-FOXP1*) (63). Mullighan *et al* (63) also found point mutations of *PAX5* in 5.7% of patients. All mechanisms of *PAX5* inactivation result in loss of function or inhibition of wild-type *PAX5* (63), and contribute to the differentiation arrest that is the hallmark of leukemic cells.
The 3' transcription factor gene, *ZCCHC7*, was also targeted in three cases with *PAX5* deletions in this study, in accordance with previously published data (65). Although little is currently known about its function, elevated expression levels during early B-cell development have been reported (140) and it was identified as the fusion partner of *MYC* in a B-cell lymphoma with a t(8;9)(q24;p13) (141). Kuiper *et al* (66) also identified two ALL

translocation cases with breakpoints in intron 1 of *ZCCHC7*. In this study cytogenetic analysis failed in the two cases with partial deletions of *ZCCHC7*. To identify potential translocation partners in these cases FISH and/or LDI-PCR could be performed. These collected findings implicate *ZCCHC7* as a potentially important target gene in leukemogenesis.

Chromosome 6

Deletions of 6q are well documented in a variety of haematological malignancies and solid tumours (142), including ALL (62-66). Numerous genes have previously been postulated as functional candidate targets, including *FYN, AIM1, BLIMP1* and *GRIK2* (65;66;143). Fyn is a member of the Src family of protein tyrosine kinases, and is an important regulator of B-cell development (143). *AIM1* is a putative tumour suppressor gene, associated with the control of tumourgenecity in human malignant melanoma (144). *GRIK2*, a putative TSG, encodes a transmembrane glutamate receptor subunit involved in the transduction of pro-apoptotic signals (145), and as described above, Blimp1 is essential for PC differentiation (137). In this study, all four genes were deleted in the nine cases with large scale CNL of 6q. Considering the frequent occurrence of the deletion and the absence of more targeted aberrations, multiple genes in this region may contribute to leukaemogenesis.

Four patients had deletions of a 0.12 Mb genomic region containing 16 histone gene family members. The structure of chromatin plays a role in the regulation of gene expression by controlling the access of DNA binding TF to their binding sites. The deletion of histone genes may result in changes to the organization of the chromatin, and therefore alter gene expression. Mouse models lacking specific H1 subtypes have indicated that these molecules do play differential roles in the control of gene expression (146). Similar deletions were observed by Mullighan *et al* (63) in 6.7% of B-ALL cases. In a retrospective screen of 41 *ETV6-RUNX1* patients who went on to relapse, deletions of 6p22.1 were observed at a lower frequency then anticipated from the aCGH data, indicating that this aberration may be associated with a more favourable outcome. The study of a larger cohort of patients would confirm this preliminary observation.

Chromosome 11q

Deletions of 11q22-q23 have been previously identified in B-ALL patients by FISH, often in association with a *MLL* translocation (147). A recent publication has reported the co-existence of *ETV6-RUNX1* and *MLL* aberrations in 9.2% of B-ALL patients. Deletions accounted for 71% of these aberrations, and to date these patients are responding well to treatment (148). Large scale deletions that encompassed the *MLL* gene were observed in 11.7% of patients in this study.

Deletions of 11q22-q23 are also one of the most common aberrations in B-CLL and are associated with an adverse prognosis (149). The ATM (Ataxia- Telangiectasia Mutated) gene, which is involved in DNA repair, DNA recombination and cell cycle control, is an interesting candidate in this region. Mutations of the gene are responsible for Ataxia-telangiectasia, and predispose these patients to lymphoid malignancies. Mutations of ATM have also been identified in T-cell prolymphocytic leukaemia, mantle cell lymphoma and diffuse large B-cell lymphoma (150). Deletions of 11q, encompassing the ATM gene, have been previously reported in ALL (150), and were observed in five patients in this study. This data is suggestive of a pathogenic role for ATM in ALL, and further investigation into the mutation status of the retained allele in these patients would be informative.

Chromosome 13q

Deletions of 13q12~14 have been described in a wide range of haematological malignancies (63;66). In accordance with published data (63), a single focal deletion of the tumor suppressor and cell cycle regulator, *RB1* was detected in this study. *RB1* is frequently inactivated in solid tumours but it is postulated that a second tumour suppressor gene, immediately telomeric to *RB1,* is the target in ALL. This putative gene is yet to be identified (151). The deletion in case 3602 terminates 33kb telomeric of *RB1*, lending support to this hypothesis.

Chromosome 8

TOX is a member of the High-motility group family of DNA binding proteins. It is thought to regulate gene expression by altering local chromosomal structure. TOX is highly expressed in the thymus and is involved in the regulation of T-cell development (152). Five cases in this study had deletions of the sequence immediately 5' (telomeric) of the gene, and exon 1 was also targeted in four cases. Mullighan *et al* (63) identified similar deletions in seven B-cell ALL cases, four of which were *ETV6-RUNX1* positive. Quantitative RT-PCR was used to investigate the expression levels of *TOX* in patients with deletions of this locus. In all three cases, expression of the gene was reduced compared to *ETV6-RUNX1* patients with normal copy number at the same locus as well as the positive control. *TOX* was selected for further investigation because this aberration has not been previously described, and the function of the protein as a regulator of gene expression highlights it as an interesting target. Further research is required into the effect that this decrease in expression has on downstream targets of *TOX.*

Chromosome 10q

Focal deletions of *ADD3* were observed in three cases. Similar deletions were recently described by Mullighan *et al* (63) in *ETV6-RUNX1* (4%), HeH (2.6%), *BCR-ABL1* (33%)

and unclassified (8.3%) patients. *ADD3* is highly conserved and encodes the γ subunit of the heterodimeric cytoskeletal protein, adducin. Adducin is involved in the formation of an actin-spectrin lattice, actin polymerization and cell signal transduction. It is more typically associated with hypertension (153), but has been identified as the fusion partner of *NUP98* in a T-ALL patient (154).

In a retrospective FISH screen of 41 relapsed *ETV6-RUNX1* patients, deletions of *ADD3* were seen at a much higher frequency than anticipated from the aCGH data (19.5% versus 8.8%). The increased frequency of *ADD3* deletions in these patients highlights this aberration as a putative poor prognostic indicator. FISH was used to screen the diagnostic samples from an additional 200 *ETV6-RUNX1* positive patients for the presence of this deletion. It was observed in only 10.5% of these cases, in accordance with the aCGH data, lending further weight to this hypothesis. Additionally, Mullighan *et al* (63) observed this deletion in a high proportion of *BCR-ABL1* patients, a subgroup associated with a very poor prognosis.

Chromosome 5q

Three patients in this study had specific deletions of the early B-cell factor gene, *EBF1*, supporting recently published data (63). In conjunction with TCF3, EBF1 initiates the B-cell specific gene expression programme. However, B-lineage commitment is not ensured unless *PAX5*, a downstream target of EBF1, is expressed. Mouse models have shown that a bi-allelic deletion of *EBF1* leads to differentiation arrest, and haploinsufficiency results in a partial block (63).

Chromosome 19q

The transcription factor *CEBPA* is crucial in proliferation and differentiation, and has a central role in the pathogenesis of various forms of myeloid malignancy. Recently *CEBPA* was implicated in B-ALL for the first time, in nine patients with t(14;19)(q32;q13) (*IGH@-CEBPA*) (155). Deletions of *CEPBA*, as described in one case in this study, have not been previously reported. The function of *CEBPA* in the development of normal lymphocytes and lymphoid malignancies remains largely unknown.

Chromosome 7p

IKZF1 encodes a family of early haematopoietic and lymphocyte restricted TF, which promote differentiation of pluripotent HSC. Georgopoulos *et al* (156) demonstrated that mice with mutations or deletions of this gene lack T- and B-lymphocytes, natural killer cells and their earliest progenitors, and are predisposed to the development of T-cell malignancies (156;157). Mullighan *et al* (63) and Kuiper *et al* (66) have recently identified deletion of this gene in 8% and 7% of B-ALL cases, respectively. A single case included

in this study had a deletion of this locus (2.9%), a lower incidence then expected from the published data.

Chromosome Xq

Recurrent duplication of Xq in *ETV6-RUNX1* positive ALL was recently described by Lilljebjörn *et al* (123). In an aCGH study of 17 patients they found gain of Xq in 35% (n=6) of cases, making it the most common aberration detected in their series. All patients with dup(Xq) were male and unbalanced translocations were involved in three cases. Gene expression profiling of the 6 cases with dup(Xq) revealed distinct over-expression of genes in the duplicated segment. The high and uniform over-expression of the testis-specific gene *SPANXB* (Xq27.1) suggested an important pathogenic role for the gene. Mullighan *et al* (63) also documented gain of Xq in 21% (n=10) of *ETV6-RUNX1* positive patients. In accordance with these findings, 26% (n=9) of patients in this study harbored duplications of Xq. Conversely, 55% of these patients were female. The greater number of females included in this study compared to the one of Lilljebjörn *et al* (6 versus 17) may account for this discrepancy.

In a retrospective screen of 41 relapsed *ETV6-RUNX1* patients, gains of Xq were seen at a lower frequency than expected from the aCGH data, indicating that this aberration may be associated with a favourable prognosis.

Disruption of nuclear co-repressor complexes in *ETV6-RUNX1* positive ALL

The *TBL1XR1* gene, located at 3q26 comprises 18 exons and exists as at least two isoforms (TB1XR1α and TB1XR1β), that are identical through to exon 14 but differ at their carboxyl end. It is a transcriptional regulator that interacts with the co-repressors of nuclear hormone receptor (NHR) activity. The NHR superfamily is a family of ligand-inducible transcription factors, including thyroid hormone receptor (T_3R) and retinoic acid receptor (RAR). They play critical roles in the regulation of reproduction, growth, differentiation, and homeostasis. The structural features of NHR include a conserved DBD that targets the receptor to hormone response element (HRE) sequences, and a ligand-binding domain (LBD). When ligand is bound, NHR dimers recruit co-activators, such as CBP-p300, and bind to the HRE within their target genes, stimulating transcription. In the absence of ligand, T_3R and RAR associate with the nuclear receptor co-repressor (N-CoR) and silencing mediator of retinoic and thyroid hormone receptor (SMRT) complexes (158). N-CoR and SMRT exist in large protein complexes and mediate repression through their ability to recruit enzymes such as GPS2 (159) and HDAC3, which produces a more compact chromatin structure that is inaccessible to transcriptional activators. Although the composition of the repressor complex is not fully understood, both TB1XR1 and a highly homologous protein, TBL1, are present (96) (figure 30). TBL1XR1 contains six WD-40 repeat elements, which allow association with

the N-CoR/ SMRT complex and directs interaction with histone molecules (preferentially histones H2B and H4). The interaction between TBL1XR1 and the histones stabilizes the association of N-CoR/SMRT complexes with the chromatin (158;160). Changes in repression correlate to changes in the levels of N-CoR and SMRT. These levels are regulated by the rate of both synthesis and degradation of the co-repressor. Over expression of TBL1XR1 has been linked to increased expression of NHR co-repressors, and the F-box domain of TBL1XR1 plays an essential role in the recruitment of ubiquitin/ proteasome machinery (160). Interestingly, TBL1XR1 and TBL1 are also required for transcriptional activation of T_3R, but not RAR, due to their ability to mediate the dismissal of the co-repressor and its subsequent degradation (158).

TBL1XR1 is widely expressed and one mRNA splice variant is particularly abundant in haematopoietic tissues. Additionally, *TBL1XR1* is over-expressed in early haematopoietic cells. Although both co-repressors are expressed widely, extensive hematological abnormalities, including blocks in erythrocyte and T-cell development, follow targeted deletion of N-CoR (158). Deletions of N-CoR are also observed in AML patients (161). In this study, deletions targeted this gene (n=5), and the distal sequence (n=3) in 23% of cases, the latter of which may target regulatory or promoter sequences 5' of *TBL1XR1*. These data suggest that loss of this gene may be contributing to pathogenesis. Focal deletions of *TBL1XR1* have recently been reported in two studies, at a frequency of 3% (64) and 12% (63). The deletions were only seen in 9.7% of the relapsed patients screened using FISH, indicating that they may be associated with a favourable outcome. From this study it was postulated that loss of *TBL1XR1* would result in a decrease in gene expression. Without TBL1XR1, the N-CoR/SMRT complexes would be unable to efficiently bind to the NHR and the histone molecules in the HRE of their target genes. As a result, these target genes could freely bind co-activator complexes, and expression levels would increase. To test this theory the expression levels of *TBL1XR1* and four downstream genes; *RARα, RARβ, CRABP1* and *CRABP2* were investigated using qRT-PCR.

RARα, *RARβ* and *RARγ* are the three genes that encode RAR. Each isotype exerts a mixture of overlapping and specific actions in target gene regulation (162). *RARβ* has four isoforms (RARβ1-RARβ4), with RARβ2 being the most abundant and the major RA inducible form. *RARβ2* is a target of RARα; activation of the promoter is mediated by RAR and RXR (retinoic X receptor) dimers (163;164). *RARβ2* is a tumour suppressor gene that is silenced in a number of malignant tumours including AML (165), where the frequently observed PML-RARα fusion protein acts as a constitutive repressor of RARα target genes, through its aberrant association with NCor/SMRT (166).

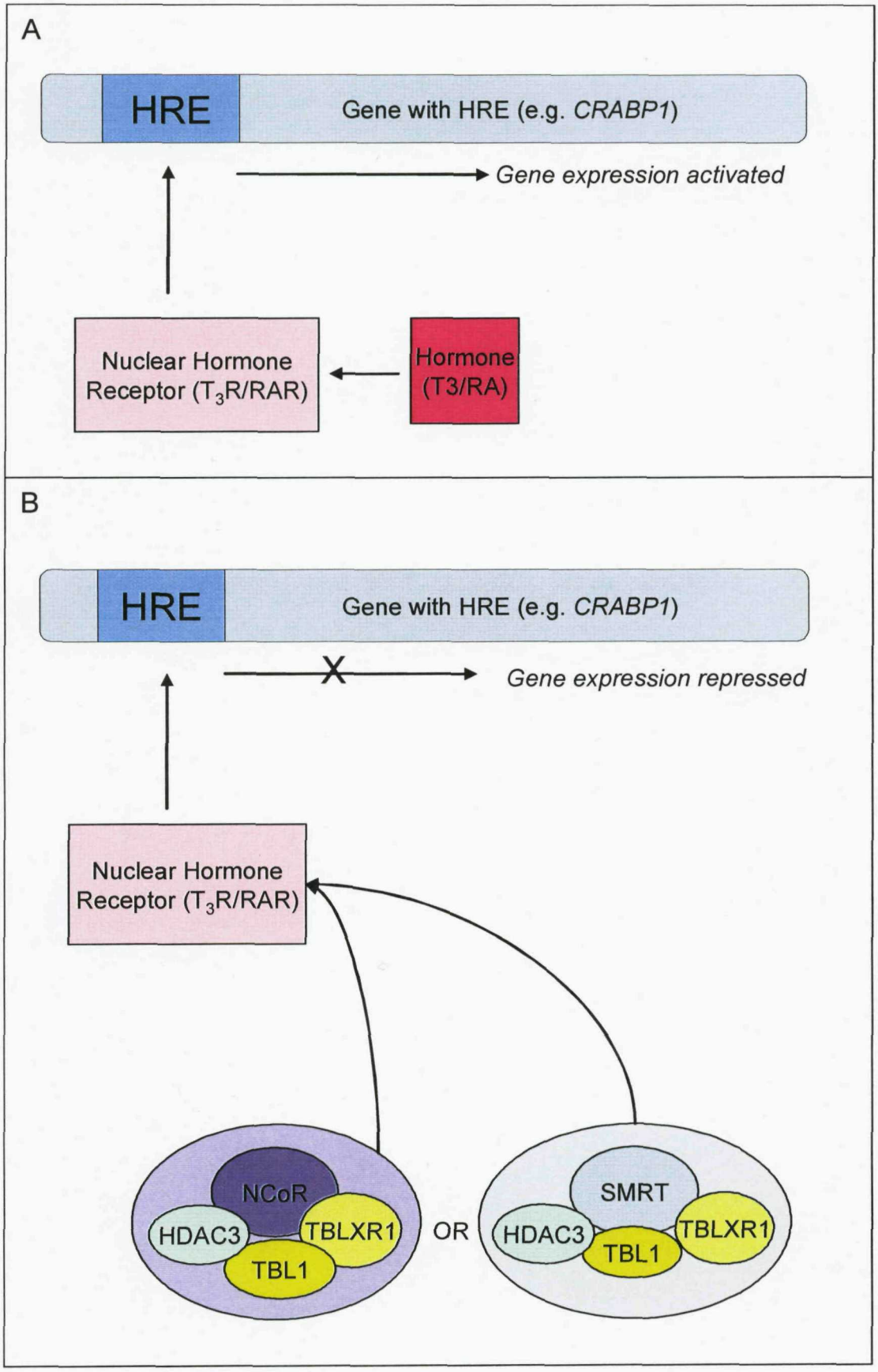


Figure 30. A. The activation of gene expression in response to the binding of ligand bound nuclear hormone receptor to the hormone response element. B. The repression of gene expression in response to the binding of corepressor bound nuclear hormone receptor to the hormone response element.

The cellular retinoic acid binding protein genes, *CRABP1* and *CRABP2* are also downstream targets of *RARα*. The CRABPs bind RA and function in part to direct the metabolism of retinol. There is also evidence that they may lower the active intracellular RA concentrations, protecting the cells from the differentiation-inducing effects of RA. They may transport RA through the cytosol to the nucleus where it can interact with RAR and initiate gene transcription. However, the precise function of the CRABPs remains unknown and the two proteins may function differently (167). Both *CRABP1* and *CRABP2* are targets for aberrant methylation in a variety of tumour types (168;169) and the *RARA-PLZF* fusion (t(11;17)(q23;q21)) in acute promyelocytic leukaemia induces promoter hypomethylation and *CRABP1* over-expression (170). *CRABP2* is also expressed in a wide variety of human cancers and is thought to play an important role in cancer development (171).

Overall, the cohort of patients with deletions of *TBL1XR1* had lower expression levels of the gene than patients with normal copy numbers, and up regulation of all four target genes was observed in response to the decreased expression of *TBL1XR1*. This supports the hypothesis that loss of *TBL1XR1* destabilizes the co-repressor complexes, disrupting hormone mediated gene expression. However, in individual patients the evidence is not compelling, as the fold changes observed in the two cohorts are very similar for some cases. Whilst mutation sequencing failed to confirm mutations in exons 4-14 of the *TBL1XR1* gene in the six patients with deletions at this locus, mutations may be involved in gene inactivation in patients with normal copy numbers of the gene. Additionally, although unconfirmed by sequencing, mutations of the gene were detected in the six patients by HRM analysis. For most exons the primers designed for HRM analysis either spanned or extended into an intron. Mutations of the intronic sequence may explain the discrepancy between HRM analysis and sequencing data. Additionally, HRM analysis can detect variant DNA in a background of wild-type sequence at sensitivities approaching 5%. Mutated subclones present at such low levels would not be detectable by sequencing. Patients with the lowest expression levels of TBL1XR1 may harbour a deletion of one allele, and mutations of specific exons in the remaining allele. Lastly, there were many possible choices for the downstream target genes. The four selected for this study may not provide the best representation of the effects of TBL1XR1 down-regulation. Other pathways and cellular processes may impact upon their expression levels. Time limitations have prevented further research in this area.

Conclusions

The aCGH approach used in this study provided further insight into the size, genomic position and gene content of CNA affecting *ETV6-RUNX1* positive patients. Numerous recurrent large scale and submicroscopic aberrations were described, targeting both previously identified and novel regions. The vast majority of genes involved in these

aberrations are tumour suppressors, transcription factors or regulators of B-cell differentiation and development. Loss of *ADD3* was identified as a potential poor risk marker. The effect of the deletions on gene expression was shown for *TOX*, and for *TBL1XR1* and its downstream targets, providing preliminary data toward further studies. *ETV6-RUNX1* positive ALL is a heterogeneous disease, with no two patients having the same combination of genomic alterations. Many possible avenues for further research remain, which may lead to improved diagnosis and risk stratification, and may facilitate the identification of novel therapeutic targets.

The findings of this study have been published in abstract form;

Parker H, An Q, Case M, Barber K, Konn ZJ, Stewart ARM, Wright SL, Moorman AV, Hall AG, Harrison CJ, Irving JA, Strefford JC (2008) Deletion of *TBL1XR1* in *ETV6-RUNX1* positive acute lymphoblastic leukaemia induces expression of retinoic acid receptor response genes through potential disruption of nuclear co-repressor complexes. *Proc Am Ass Cancer Res 49, 4241.*

10.3 Global genomic profiling of ALL patients with intrachromosomal amplification of chromosome 21 (iAMP21).

10.3.1 Introduction

Genomic aberrations of the *RUNX1* gene are frequently observed in leukaemia, as a result of gain or loss of the entire chromosome 21, or structural rearrangements involving the long arm, for example; t(12;21)(p13;q22) in ALL, t(8;21)(q22;q22) in AML and t(3;21)(q26;q22) in MDS and CML (172). In 2001 Busson-Le Coniat *et al* (173) described three cases of B-ALL with multiple copies of the *RUNX1* gene on a marker chromosome resulting from complex rearrangements and uneven duplications of different regions of 21q. Two later studies (174;175) identified this aberration in a significant number of patients and described it as a novel recurrent abnormality in B-ALL, involving intrachromosomal amplification of chromosome 21 (iAMP21) and *RUNX1* amplification. It is now clear that iAMP21 represents a distinct genetic subtype of ALL effecting older children, with a median age of 9 years at diagnosis. In a prospective screen of 1630 patients treated on ALL97, iAMP21 was detected in 2% of cases (n=28). The prognosis was very poor, with a 5 year event free survival (EFS) of 29% and overall survival at 5 years of 71%, versus 78% and 87%, respectively, in other ALL patients (176).

Harewood *et al* (174) first described iAMP21 as a distinct ALL subgroup in 2003. During routine FISH screening of paediatric ALL patients for the presence of the *ETV6-RUNX1* fusion, they identified 20 patients that were negative for the fusion, but had multiple copies of the *RUNX1* signal. These were arranged in clusters in interphase cells and in tandem duplication on a single abnormal chromosome 21 in metaphase cells (174) (figure 31). Exon-specific probes to *RUNX1* were also applied, verifying that the entire gene was amplified. Several studies have shown that the number of copies of *RUNX1* is highly variable between patients as well as from cell to cell within the same patient. Although it is often difficult to accurately enumerate the signals due to their close apposition within the nuclei, between 4 and 15 *RUNX1* signals, of varying size and intensity, are frequently observed within nuclei of iAMP21 patients (37;172;174;177).

IAMP21 cannot be diagnosed by conventional cytogenetics as the abnormal 21 can adopt a range of different morphologies in G-banded cells. These include large or small acrocentric, metacentric, sub-metacentric, or ring chromosomes (figure 32), and the morphology of the dup(21) can differ between cells of the same patient (172;174). Therefore, the dup(21) is often misclassified as a marker chromosome of unidentified origin during routine cytogenetics. The duplication of 21q is often the sole cytogenetically visible abnormality (174), suggesting that this aberration is the primary chromosomal change which contributes to the malignant phenotype in these patients. Cells are usually

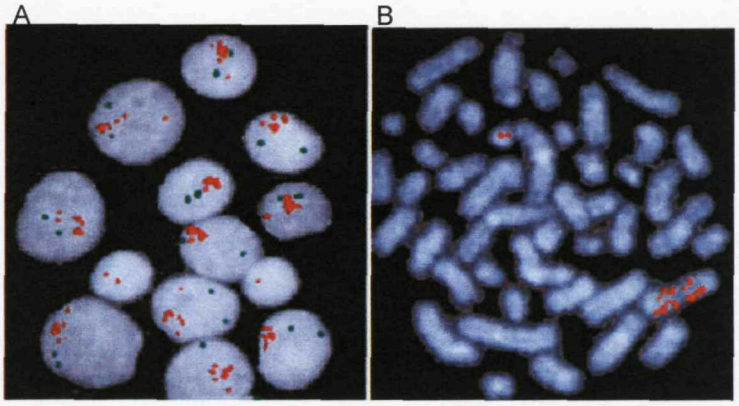


Figure 31. A. Clusters of RUNX1 *FISH signals (red) in interphase cells. Two normal copies of* ETV6 *are present (green signals) (*TEL-AML1 *ES probe, Vysis). B.* RUNX1 *exon specific FISH signals in tandem duplication on a single abnormal chromosome 21 in a metaphase cell. Adapted from Harewood* et al *(174).*

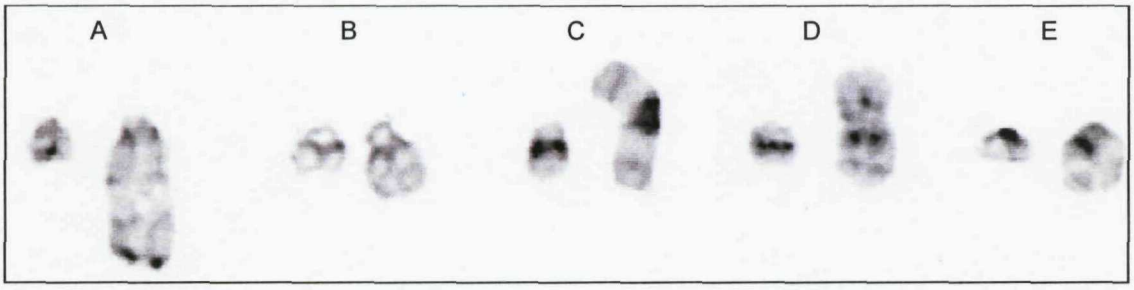


Figure 32. Partial G-banded karyograms showing examples of the different forms of dup(21), compared to normal chromosome 21 (left hand image). A-E show large acrocentric, small acrocentric, metacentric, submetacentric and ring chromosomes, respectively. Adapted from Harewood et al *(174).*

near-diploid or pseudodiploid. With the exception of gain of chromosome X, which has been reported in 19-25% of cases (172;174) all other associated abnormalities are non-recurrent. To date a diagnosis of iAMP21 is based on the presence of three or more copies of *RUNX1* detected by FISH on a single abnormal chromosome 21. However, it is not possible to determine the genomic location of the FISH signals in the absence of metaphase cells. As patients with i(21)(q10) or hidden hyperdiploid clones including several copies of chromosome 21 have additional *RUNX1* signals, this diagnostic test may be misleading. Accurate diagnosis of iAMP21 is essential as these high risk patients require intensive treatment (37).

Attempts have been made to further characterise iAMP21 by the application of aCGH and gene expression profiling. In a recent study, Strefford *et al* (37) used genome-wide 1 Mb

BAC aCGH in the analyses of 10 iAMP21 patients. Complex patterns of genomic gain, amplification and deletion along chromosome 21, unique to each patient, were revealed. Although the amplicon size and degree of amplification varied considerably between individuals, a common 8.6 Mb region of amplification (CRA) was identified at position 31.5-40.1 Mb in all cases, encompassing the *RUNX1* gene. This was accompanied by a 4 Mb common region of deletion (CRD) close to the telomere in 70% of patients. FISH analysis of 26 loci along chromosome 21q confirmed the variations in copy number and identified the same CRA and CRD. Tiling path Oligo aCGH analysis of five patients furthur refined the CRA and CRD to 6.5-6.6 Mb (33.192-39.796 Mb) and 3.54 Mb, respectively (37). In a study of six iAMP21 patients, using 33K BAC aCGH, Kuchinskaya *et al* (177) further refined the telomeric deletion to a 0.6 Mb region.

Global gene expression profiling, using the Affymetrix U133A oligonucleotide array, revealed distinct gene expression profiles in eight iAMP21 patients, compared with *ETV6-RUNX1* positive patients and patients with HeH including gain of at least one copy of chromosome 21 (37). Fourteen of the top 150 significantly over-expressed genes in the iAMP21 patients were located to the CRA, including *ERG* and *ETS2*. Eleven genes were uniquely expressed in iAMP21 patients, only two of which were located in the CRA; C21orf66 and ATP50. Legumain (chromosome band 14q32.1), a lysosomal cysteine protease involved in antigen processing, was also over expressed. This may contribute to the poor prognosis as it is over-expressed in a number of aggressive cancers and may provide a survival advantage to the cells, as it enables them to invade extravascular spaces. In addition, a large proportion of the genes in the CRD were under-expressed, implicating copy number variation as a mechanism for gene dysregulation in these patients. Interestingly, *RUNX1* expression was variable among iAMP21 patients, irrespective of the copy number. This reflected its expression in other ALL patients, suggesting that *RUNX1* may not be the target gene of iAMP21 (37).

Similar amplifications of chromosome 21 are a recurrent finding in AML patients with complex karyotypes. The *RUNX1* also appears not to be the target of this abnormality, as normal *RUNX1* copy number is seen in some cases. The most frequently amplified region in AML contains the *ERG* gene (38.7-39.1 Mb), located close to the *ETS2* locus. This region overlaps the CRA seen in iAMP21. The highest degree of amplification is seen between positions 26.7-28.7 Mb, which contains the *APP* gene. All three genes are consistently over-expressed in these patients (47).

The inter- and intra-individual differences in the degree of amplification and the size of the amplified and deleted regions implies high level genomic instability and telomere dysfunction. This was supported by MFISH, M-banded analysis and metaphase FISH of

the duplicated chromosome 21 (dup(21)), which revealed variable morphology of the dup(21) between patients, and FISH signals representing the different regions of chromosome 21 located to unexpected positions in relation to each other. These findings indicated a series of complex intrachromosomal rearrangements within this chromosome, including deletions, duplications, amplifications and inversions (172). Breakage-fusion-bridge (BFB) cycles were proposed as the mechanism of iAMP21 formation. Following an initiating event, potentially telomeric loss, a double-stranded DNA break (DSB) occurs. A chromosome with two broken sister chromatids is formed. Following replication they fuse to produce a dicentric chromosome. During mitotic segregation of the centromeres to opposite poles at anaphase, a breakage occurs at a site close to the original one, leading to generation of a chromosome with an inverted repeat. Multiple cycles of BFB produce a chromosome with accumulated additional copies of intrachromosomal regions. The BFB cycle usually ends when the unstable chromosome acquires a new telomere, through non-reciprocal translocations, formation of a dicentric or ring chromosome, or addition by telomerase (figure 33) (172).

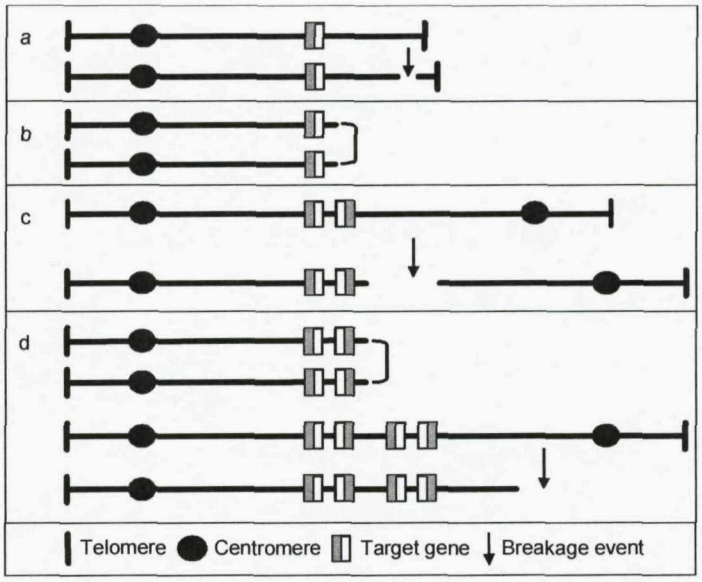


Figure 33. Diagrammatic representation of the BFB cycle. (a) DSB occurs, leading to loss of the telomere; (b) Two broken sister chromatids form a dicentric chromosome; (c) Breakage leads to a chromosome with an inverted repeat; (d) Multiple BFB cycles produce a series of inverted repeats. (Adapted from Robinson et al *(172))*

The aims of this study were;

- To fully characterise the genomic changes in iAMP21 patients, and identify novel markers that may facilitate accurate diagnosis of this subgroup.
- To identify recurrent copy number changes that may contribute towards leukemogenesis in iAMP21 patients.

- To identify genes that may represent prognostic markers or therapeutic targets in these patients.

10.3.2 Materials and methods

10.3.2.1 Patient samples

DNA from diagnostic bone marrow samples from 19 patients with iAMP21, and five patients with chromosome 21 abnormalities, was obtained from the MRD Childhood Leukaemia Cell Bank. Chromosomal and diagnostic FISH analysis was performed in the UK regional cytogenetics laboratories, and karyotypes were described according to the International System for Human Cytogenetic Nomenclature (84).

10.3.2.2 Array- based CGH

The samples were processed for aCGH analysis using the Agilent Human Genome CGH Microarray kit 185k, as described in section 9.10. Five consecutive probes with fluorescence ratios deviating from a normalized log ratio value of 0 by at least +1, were defined as copy number amplifications.

10.3.2.3 MCC, LDI-PCR and RT-PCR

Following the identification of a deletion within intron 1 of the *PDE9A* gene by aCGH, MCC was used to confirm the findings and define the breakpoints with greater accuracy, as described in section 9.11. Initially, six markers (M1-6) were designed at ~2000bp intervals throughout intron 1. A further three markers (M7-9) were designed at ~300bp intervals between M2 and M3 (see Appendix 10 for the location of markers within the *PDE9A* gene sequence). Using the MCC refined breakpoints, LDI-PCR (section 9.12) and sequence analysis were used to determine whether the retained *PDE9A* sequence was involved in a gene fusion. Where RNA was available, RT-PCR was used to ascertain whether the chimeric transcript was expressed (section 9.13) (see Appendix 5 for marker/primer sequences, restriction enzymes and sequencing data).

10.3.2.4 FISH

Five fosmid clones were selected and differentially labeled to construct a break-apart FISH probe targeted to *PDE9A*, as described in section 9.16. Figure 34a illustrates the probe design and Appendix 3 details the selected fosmids. Fosmids were selected for this analysis as they are cloning vectors which provide single-copy cloning, while allowing greater clone stability. This probe was used to screen for deletions of the 3' (telomeric) end of *PDE9A* on fixed cell suspensions available from 70 iAMP21 patients. A FISH probe was also designed to target the *PDE9A-MUS81* fusion gene identified by sequencing in patient 4279 (figure 34b).

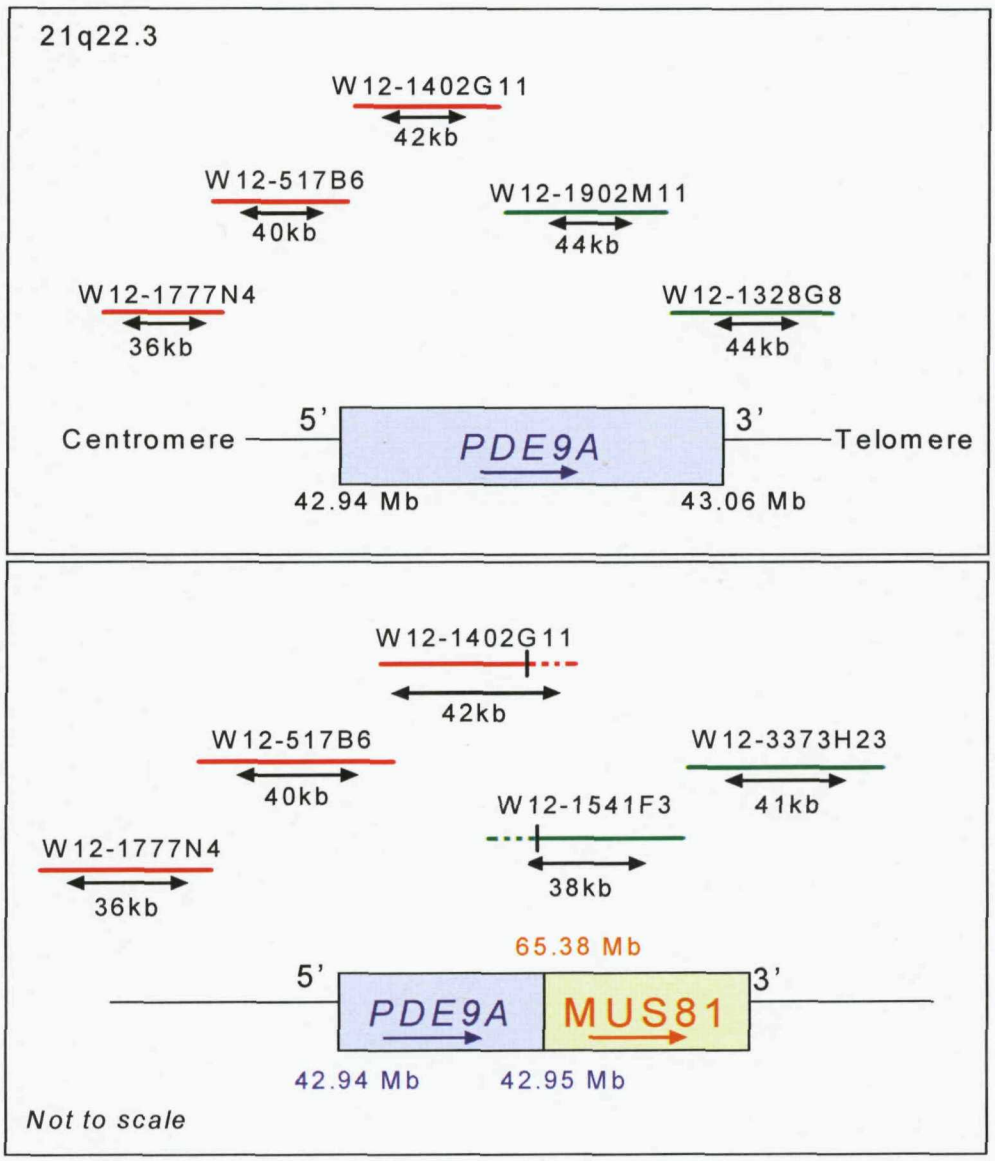


Figure 34. The name, genomic positions, size and labeling colour (red or green) of the five fosmids comprising the PDE9A *break-apart FISH probe (A) and the* PDE9A-MUS81 *fusion FISH probe (B). Dotted lines represent the section of the fosmid probe that extends beyond the end of the gene when it is involved in the fusion*

10.3.3 Results

10.3.3.1 Array-based CGH

All 19 iAMP21 patients had an abnormal aCGH profile. In total 178 CNA were detected, with between three and 22 aberrations per case (Appendix 11 and figure 35). Seven aberrations involved whole chromosome gain (n=6) or loss (n=1), and of the remaining 172 CNA, deletions were more frequent than gains or amplifications (n=96 versus n=56 and 27, respectively). Although copy number changes affecting chromosome 21 accounted for 42.4% of all aberrations detected (n=76), no patient had aberrations of this

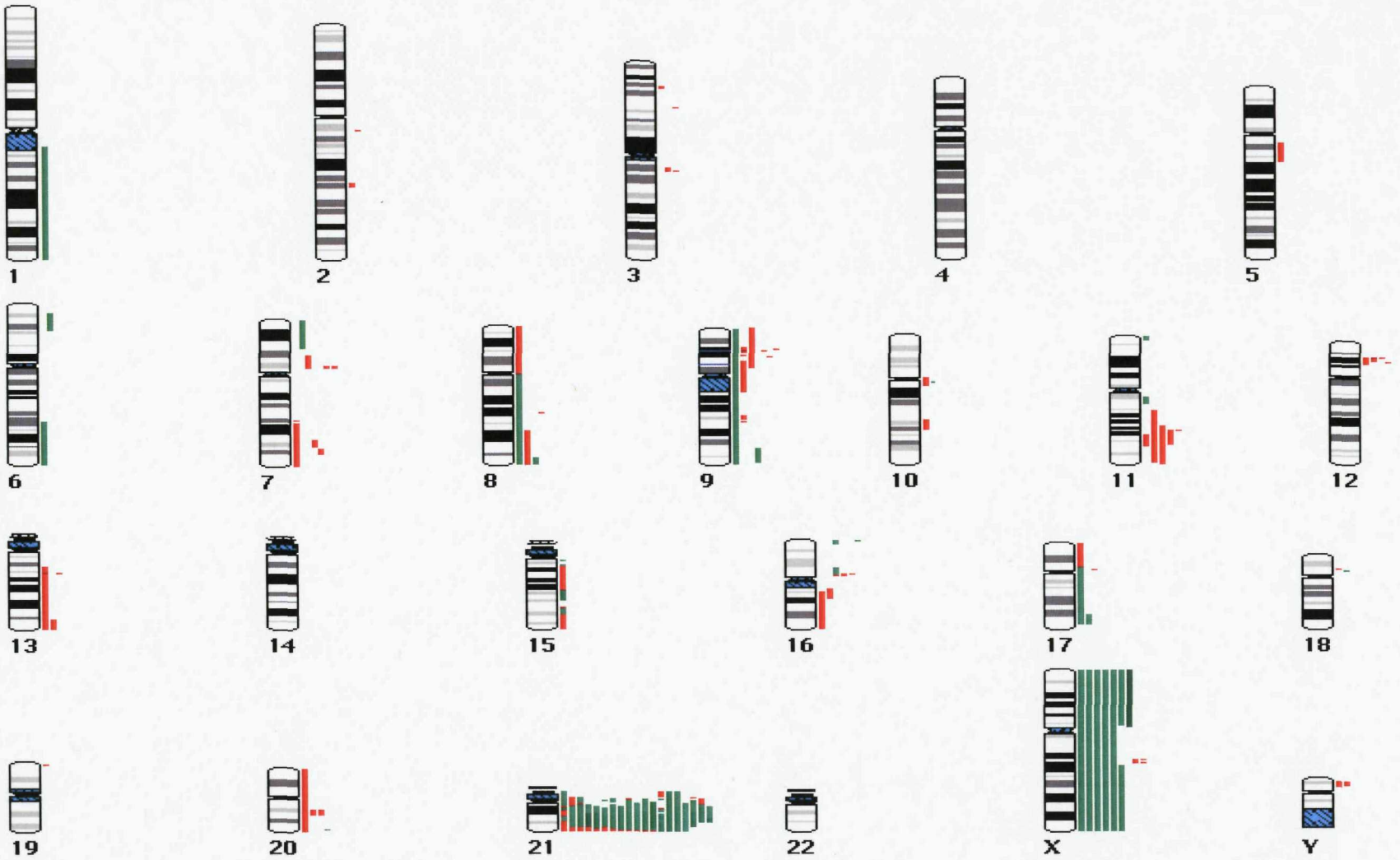


Figure 35. CNAs detected in 19 iAMP21 ALL patients. To the right of each chromosome idiogram, red, dark red, green and dark green vertical lines show copy number losses, homozygous deletions, gain and amplifications, respectively. Each vertical line is positioned to show the extent and genomic location of a single change in a single patient.

chromosome as the sole abnormality. Gains, amplifications and deletions of chromosome 21 were observed in approximately equal measure (n= 28, 25 and 23 respectively). Deletions also targeted chromosome arms 9p (n= 12, two biallelic deletions), 7q (n=5), 11q (n=5), 12p (n=5), and 13q (n=5). Recurrent gains affected chromosome X (n=8, including 5 whole chromosome gains). A complex pattern of eight copy number changes targeted chromosome 15 in a single patient. The smallest CNA observed was a 13.7kb gain of chromosome 21 material (46.646783-46.660527 Mb). Additionally, by consulting the database of genomic variants (www.projects.tcag.ca/variation/), 94 CNV were identified, which targeted 26 different loci (Appendix 8). No CNV were detected on chromosomes 10, 12, 13, 16, 18, 19, 20, 21, X and Y. Similar to the *ETV6-RUNX1* positive patients described in section 10.2, deletions at 2p11.1 (n=15), 7q34 (n=5), 14q11 (n=7), 14q32.33 (n=16) and 22q11.22 (n=4) were frequent. Eleven additional CNV were targeted in at least 10% of patients. A single patient had no CNV, while as many as 11 were targeted in remaining cases.

The diagnostic G-banding and FISH data for the 19 iAMP21 patients is provided in Appendix 11. The results of a FISH screen using six probes targeted to different chromosome 21 loci is included (screen performed by Hazel Robinson, LRCG). G-banded analysis failed in 26% of cases (5/19), and the karyotypes of two patients were cytogenetically normal (7828 and 8983). In these cases aCGH described 49 (failed cytogenetics) and 15 (normal karyotype) CNA respectively. All remaining patients had cytogenetic aberrations of other chromosomes in addition to the abnormal 21. Of the 40 cytogenetically visible aberrations that would result in copy number imbalance, 27 were confirmed and refined by aCGH. A further six were described as additional material or marker chromosomes. The origin of the chromosomal material in five of these, and two unconfirmed aberrations, could be deduced from the aCGH or cytogenetic data, although FISH would be required for definitive confirmation. The remaining cytogenetic aberrations unconfirmed by aCGH were observed in fewer then eight metaphase cells by cytogenetics (range=1-8), indicating their presence in a low proportion of cells in the sample, likely below the detection level of aCGH. In the 14 cases with successful cytogenetic analyses, aCGH detected nine large scale aberrations (>10Mb), excluding chromosome 21 abnormalities, that were misclassified by cytogenetic analysis.

Chromosome 9p

Twelve deletions of the short arm of chromosome 9 were observed in six patients. The deletions showed considerable heterogeneity in extent, ranging from 0.1 Mb to 39.0 Mb in size. Four patients had deletions encompassing the *CDKN2A*, *CDKN2B* and *MTAP* genes with a CRD between genomic positions 21.892294 and 21.993710 Mb. In addition to a larger deletion of 9p, case 7045 harboured a bi-allelic deletion of all three loci, whilst

the bi-allelic deletion in case 7583 targeted *CDKN2A* only. In two cases (8983 and 7045) the 9p deletion encompassed the *CDKN2A/B, PAX5* and *ZCCHC7* loci.

Chromosome 7

Chromosome arm 7q was targeted by four deletions in three patients. Two common regions of deletion (CRD) were identified between 132.045690 and 139.596420 Mb, and 141.843315 and 146.031777 Mb in two patients each. Three patients (15.7%) had a deletion of 7p, encompassing the *IKZF1* gene.

Chromosome 11q

Five patients had deletions of 11q and two CRD were identified; the first (n=2) was between 102.802893 and 102.976602 Mb and the second (n=2) spanned genomic position 106.603155 and 114.420246 Mb encompassing 55 genes/EST, including the *ATM* gene.

Chromosome 12p

Five deletions of chromosome 12p occurred in four patients. Three targeted the *ETV6* gene, including a bi-allelic deletion in case 7255. The commonly deleted region encompassed introns 1 to 2 of *ETV6*.

Chromosome 13q

Five deletions of 13q were described in three patients and two CRD were identified; a 6.4 Mb deletion between 107.653460 and 114.118329 Mb (n=2) and a 0.06 Mb deletion between 47.884042 and 47.954006 Mb (n=2, including a bi-allelic deletion). The 3' end of the *RB1* gene was targeted in 3 cases.

Chromosome 20

Deletions of 20q were observed in three patients, including a single whole chromosome loss. The commonly deleted region was between genomic positions 43.655971 and 47260714 Mb.

Chromosome 17q

A 10.8 Mb gain of 17q was detected in two patients, between genomic positions 67.647121 and 78.507610 Mb, encompassing the *TBCD* gene.

Chromosome X

Gain of whole chromosome X was observed in five patients. Duplication of the short arm was seen in an additional two cases, one of which also displayed a 60 Mb duplication of Xq.

Chromosome 21

All patients had copy number changes affecting chromosome 21 (see Appendix 12), with between two and eight alternating regions of copy number gain, loss and amplification observed (figure 36). Unique patterns of imbalance were identified in all cases.
In 19 iAMP21 patients, 28 gains of chromosome 21 material were observed, ranging from 0.013 to 16.9 Mb in size. In 13 patients the most centromeric aberration was a CNG (the remaining six patients had centromeric deletions (n=5) and amplifications (n=1). Gains most frequently involved genomic position 27.822473 to 27.847876 Mb. Twenty five amplifications were observed, which varied considerably in extent between patients (1.2 to 27.71 Mb). With one exception, a 3.4 Mb CRA, which completely encompassed *RUNX1,* was detected in all cases between positions 32.858116 to 36.314355 Mb. The remaining patient had a lower level gain of this region. Sub-telomeric deletions were observed in 68.4% of patients. They ranged from 0.25 to 8.97 Mb in size. All shared a common region between positions 46.660527 - 46.915341 Mb. Two patients (4279 and 6111) shared the start position of the deletion (42.954890Mb), which targeted the *PDE9A* gene, located at position 42.946931 to 43.068685 Mb. The FISH screen performed by Hazel Robinson (172) confirmed the variation in copy number along 21q (Appendix 11).

Non-iAMP21 patients

As described above, iAMP21 is difficult to diagnose on the basis of cytogenetics and FISH analysis, using *RUNX1* FISH probes only. To determine whether aCGH can detect CNA unique to iAMP21 patients, that may be exploited to facilitate accurate diagnosis of this subtype, five patients with cytogenetically detected chromosome 21 abnormalities, other than iAMP21 (for example;. i(21)(q10), rea(21), der(9)t(9;21) and del(21)(q?)), were analysed by aCGH and their profiles compared with those of iAMP21 patients (Appendix 14). In total 45 CNA were detected among the five patients, including 12 gains, amplifications and deletions on chromosome 21 (n=3, 4 and 5, respectively) (figure 36). Two aberrations involved whole chromosome gain and overall, deletions were more frequent than gains or amplifications (n=31 versus n=11 and 4, respectively)

In the single case that failed G-banded analysis, aCGH detected eight CNA. Among the successful cases, 11 of the 13 cytogenetically visible aberrations that should result in copy number imbalances were confirmed and their breakpoints were refined by aCGH. A single large scale aberration (>10Mb) was not detected by cytogenetics in one case.

Recurrent deletions of chromosome 12 were identified in three patients, encompassing genomic position 110.315771 to 110.389223 Mb. Two patients had deletions of the

CDKN2A/B and *MTAP* loci, including one bi-allelic deletion, and *PAX5* was deleted in a single case. Additionally, gain of whole chromosome X was observed in one patient.

All five patients had between one and four chromosome 21 CNA. A single case had a sub-telomeric deletion, and the CRA defined in the iAMP21 patients was amplified or gained in all cases (n=3 and n=2, respectively) (figure 36).

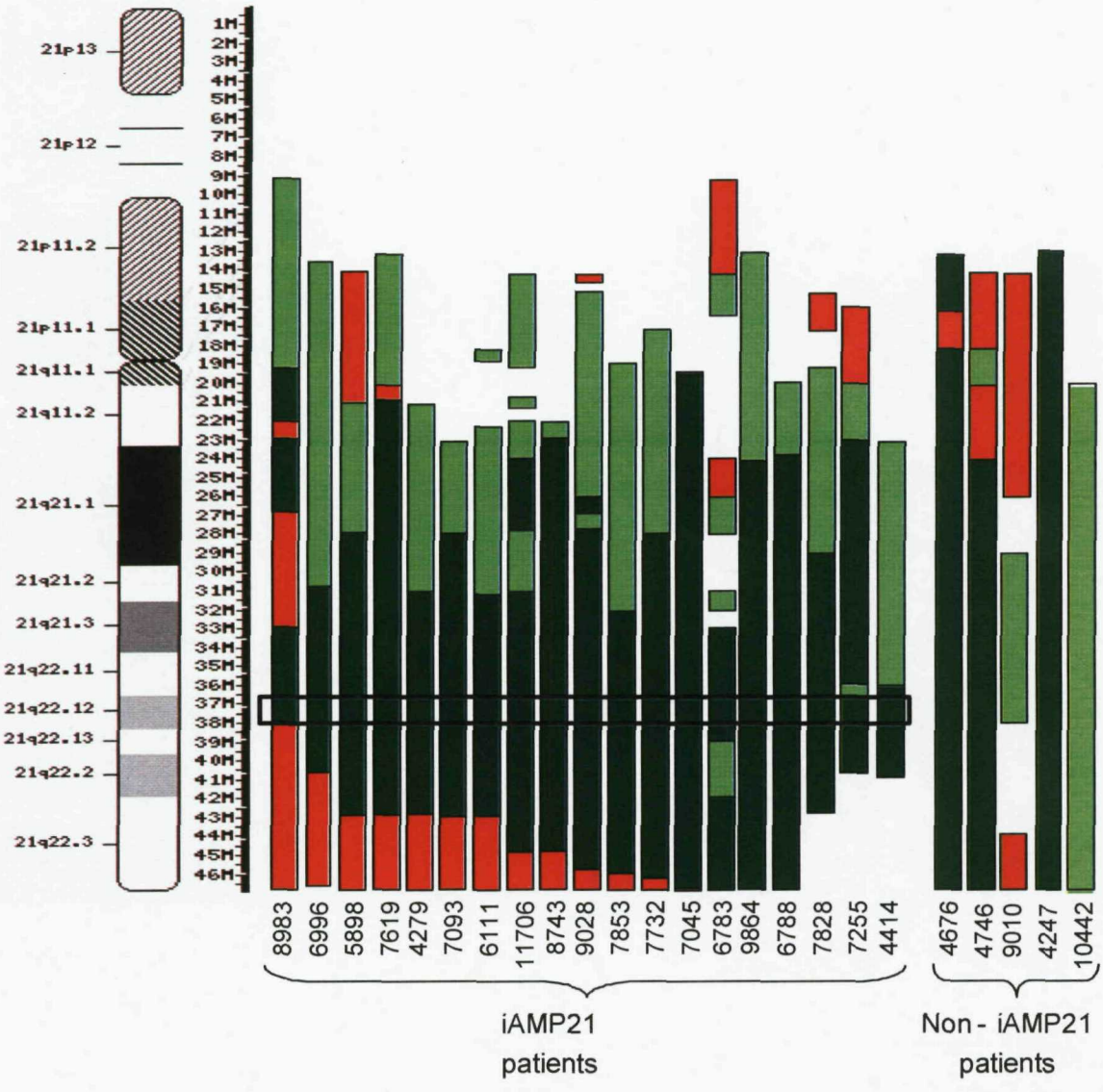


Figure 36. The CNA affecting chromosome arm 21q. To the right of the chromosome 21 idiogram, red, dark green and green vertical lines represent copy number deletions, amplifications and gains, respectively. Each vertical line is positioned to show the extent and genomic location of a single patient's CNA. The CRA is demonstrated by the black box.

10.3.3.2 MCC, LDI-PCR and RT-PCR

Two iAMP21 patients with deletion breakpoints within the *PDE9A* gene (4279 and 6111), and three patients with deletions of the entire gene (6996, 7619 and 8983) were selected for MCC analysis. In patients 4279 and 6111, copy number changes were observed between markers M2 and M7, and M9 and M3 respectively (table 10 and figure 37). This confirmed the copy number changes observed using aCGH, and refined the breakpoints from a 13kb region to a 280bp and 392bp region, respectively. In the three remaining patients copy number changes were also detected within *PDE9A* by MCC, despite the observed deletions of the entire gene by aCGH. Breakpoints were identified between M8 and M9, and M2 and M7 in patients 6996 and 7619, respectively. Patient 8983 had breakpoints between both M2 and M7, and M9 and M3, suggesting an interstitial deletion (table 10 and figure 37). All breakpoints were positioned within 2079bp of each other, in intron 1.

Patient	M1	M2	M7	M8	M9	M3	M4
4279	64	58	29	30	32	30	34
6111			48	52	53	27	
6996			56	68	38	32	
7619	42	39	13	12	14	12	15
8983		31	11	12	12	36	29

Table 10. The number of positive wells scored for the PDE9A *markers, M1-M4, in five iAMP21 patients. The copy number change is indicated by the shift in the number of positive wells (MCC was not performed for every marker in every patient).*

LDI-PCR was performed on these five patients and the products were sent for sequencing. In patients 4279, 6996 and 8983, intron 1 of *PDE9A* was fused to intron 12 of the *MUS81* gene on chromosome 11q13, intron 3 of the *VWF* gene on chromosome 12p13.3, and intron 2 of the *REX01L1* gene on chromosome 8q21.2, respectively (figure 39). LDI-PCR failed for patients 6111 and 7619, due to the poor quality of the DNA.

RT-PCR was performed on RNA from patient 4279 and revealed that the fusion gene, *PDE9A-MUS81*, was transcribed (figure 38).

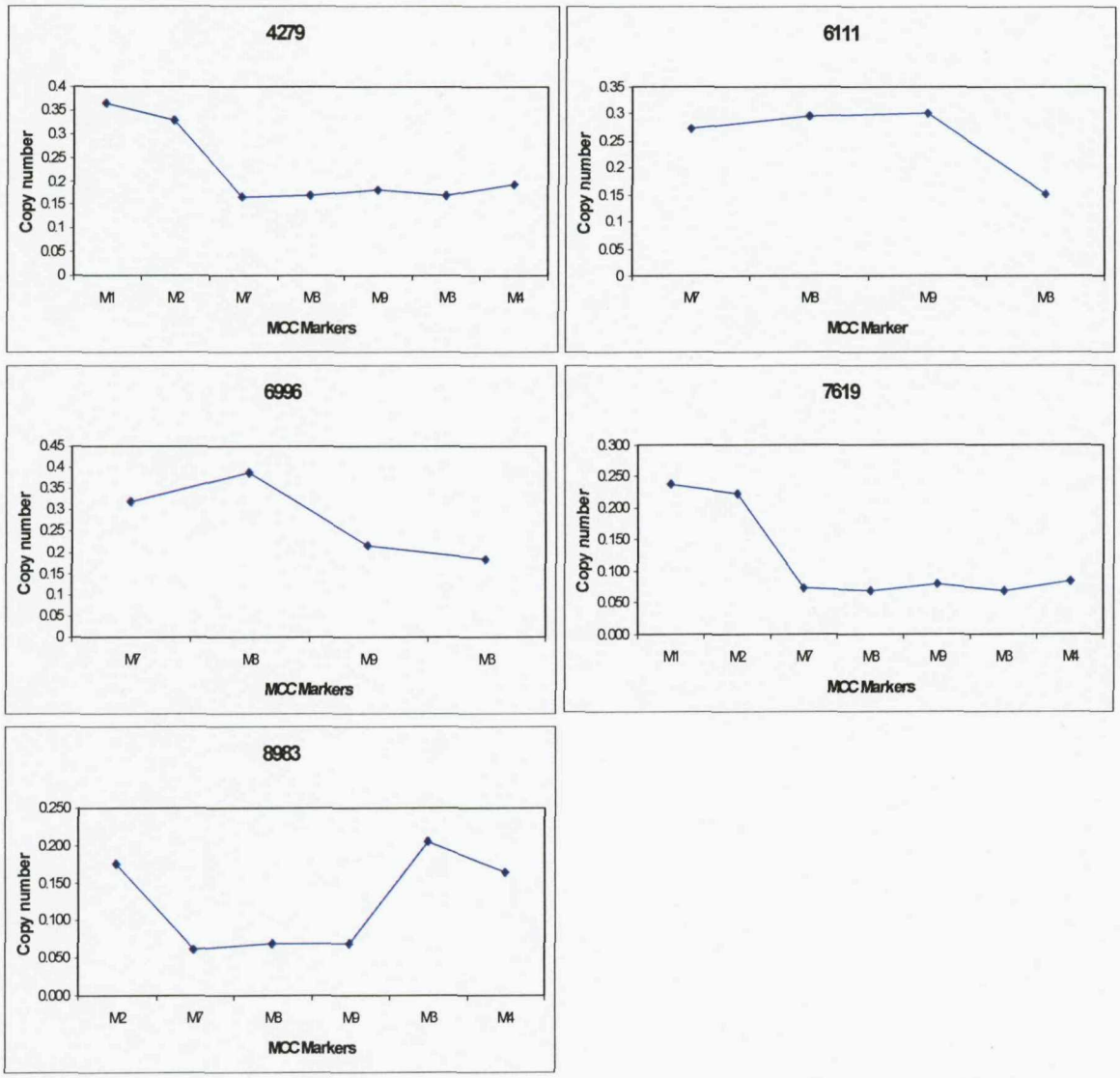


Figure 37. Graphs illustrating the copy number changes observed between markers M1-M4 in the five iAMP21 patients. Copy number is calculated as; (n (number of positive wells) x 0.5 (value assigned to a haploid genome)) / 88 (total number of wells counted).

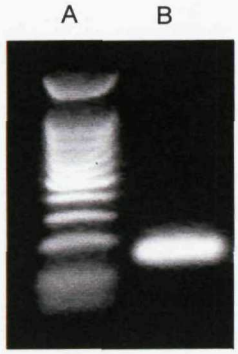


Figure 38. Agarose gel image of (A) DNA ladder and (B) the PDE9A-MUS81 *fusion transcript (150bp in size)*

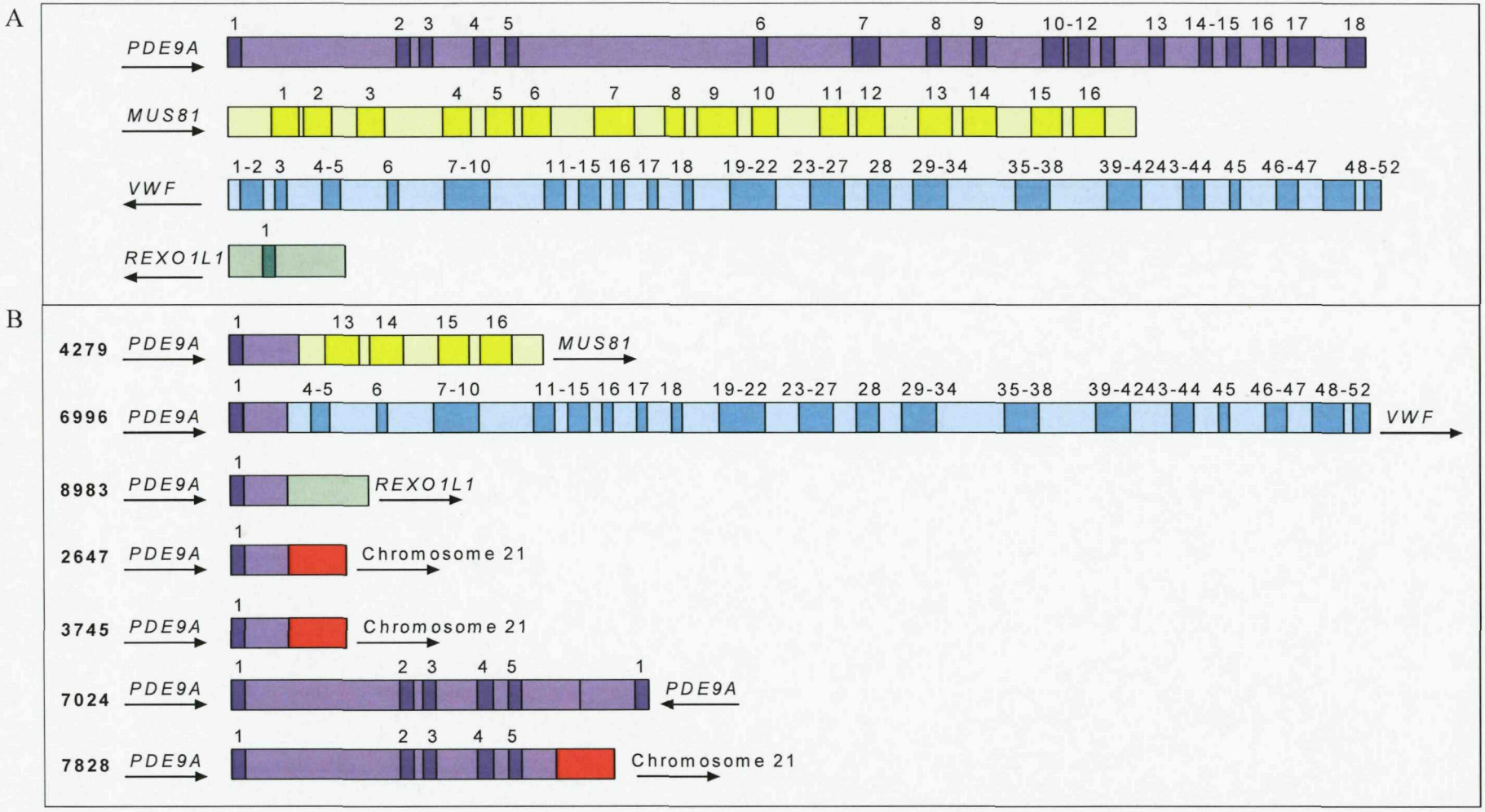


Figure 39. A. Schematic representation of the PDE9A, MUS81, VWF *and* REXO1L1 *genes, showing the number and relative positions of the exons. Arrows indicate the orientation of the gene in the chromosome, from centromere to telomere. B. Schematic representation of the fusions involving* PDE9A *in seven iAMP21 patients, illustrating the exons involved. Arrows indicate the orientation of the genes in relation to each other.*

10.3.3.3 FISH screen and further molecular analysis

The results of the *PDE9A* FISH screen are presented in Appendix 13. Fixed cells were available on 16 of the patients with aCGH profiles. Break-apart FISH probes targeted to *PDE9A* confirmed the aCGH findings for the same loci in 11 patients (4279, 5898, 6111, 6783, 6788, 7045, 7583, 7732, 8743, 9864, 11706). Patients 4279 (figure 40a) and 6111 (figure 40b) indicated a breakpoint within *PDE9A* with amplified 5' sequences. In four of the remaining five patients the aCGH and FISH data did not concur. For patient 8983, aCGH revealed deletion of *PDE9A* which was confirmed by FISH in 55% of cells (figure 40c). However, MCC showed a copy number change within the gene, which was seen in 26% of cells by FISH, indicting the presence of two different populations in this patient. Array-CGH for patient 7619 revealed deletion of the entire gene. MCC shows a copy number change within the gene, which was confirmed by FISH (figure 40d). In addition, FISH revealed a partial deletion of *PDE9A* in patient 7828 (figure 40e), in contrast to the normal aCGH profile. MCC was performed on this patient, confirming the FISH result.

A further 54 iAMP21 patients were FISH screened for *PDE9A* aberrations. Eleven patients had two normal copies of the gene, 16 patients had a single copy, indicating deletion of the entire gene, and in 24 cases the entire gene was amplified. Extra copies of the centromeric region of *PDE9A* were observed in four cases (2647, 3745, 7024 (figure 40f) and 7828). Array-CGH was not performed on patients 2647, 3745 and 7024 due to lack of available material. LDI-PCR and sequencing of patient 7024 revealed the fusion of intron 5 of *PDE9A* with intron 1 of a second, inverted copy of *PDE9A*. In patient 7828 intron 5 of *PDE9A* was fused to non-coding chromosome 21 material, 5' of the gene, and in patients 2647 and 3745 intron 1 of *PDE9A* was fused to non-coding chromosome 21 material, 5' and 3' of the gene, respectively (figure 40).

FISH probes targeted to *PDE9A* and *MUS81* were unable to detect a fusion between these two genes in patient 4279 (figure 40g).

Table 11 summarises the results of these different techniques for each patient.

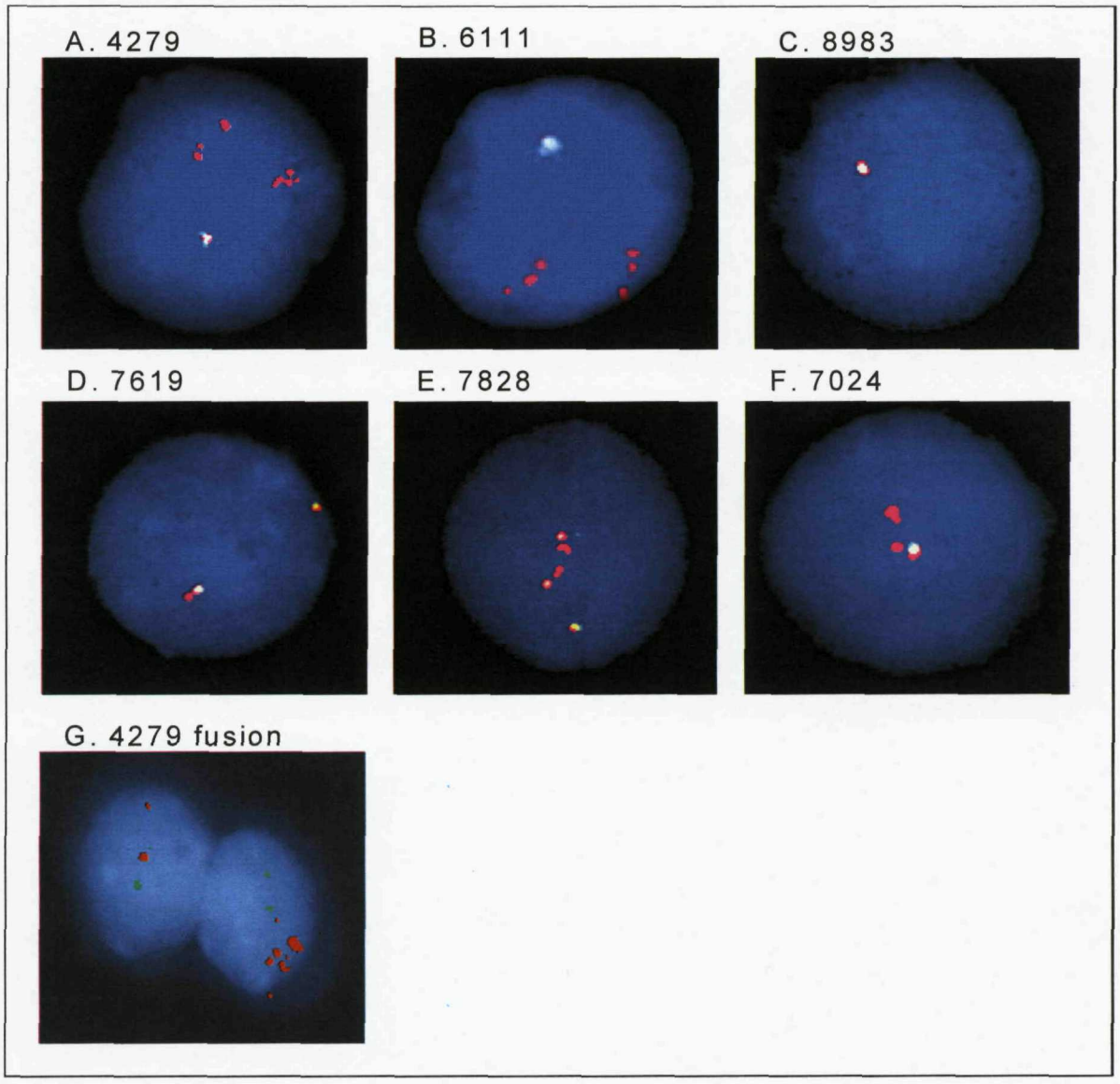


Figure 40. A-F. Interphase FISH images for patients 4279, 6111, 8983, 7619, 7828 and 7024, respectively, using the PDE9A *break-apart probe. Exon 1 and intron 1 were labeled in red, and the remaining sequence was labeled in green. Yellow fusion signals represent a normal copy of the entire gene. Red signals indicate loss of the telomeric region of* PDE9A. *G. Interphase FISH image for patient 4279, using the* PDE9A-MUS81 *fusion probe. The probes for* MUS81 *and* PDE9A *were labeled in green and red, respectively. PDE9A is amplified and MUS81 shows two normal copies, with no evidence of a fusion.*

Patient	aCGH profile for PDE9A	FISH signal pattern	MCC for PDE9A	LDI-PCR
4279	Partially deleted (from 42.954890 Mb)	ampR0G1F (66%)	Deleted from M7	*PDE9A* (intron 1)-*MUS81* (intron 12)
6111	Partially deleted (from 42.954890 Mb)	ampR0G1F (82%)	Deleted from M3	Failed
8983	Deleted (from 37.941425 Mb)	0R0G1F (55%) 1-2R0G1F (25%)	Deleted between M7 and M9	*PDE9A* (intron 1)-*REXO1L1* (intron 2)
7619	Deleted (from 42.590005 Mb)	0R0G1F (39%) 1R0G1F (31%)	Deleted from M7	Failed
7828	Normal	ampR0G1F (69%) 0R0G1F (16%)	Not done	*PDE9A* (intron 5)- Chr 21 (5' to *PDE9A*)
6996	Deleted (from 40.324891 Mb)	Not done	Deleted from M8	*PDE9A* (intron 1)-*VWF* (intron 3)
7024	Not done	2R0G1F (76%)	Not done	*PDE9A* (intron 5)-*PDE9A* (intron 1)
2647	Not done	2R0GampF (70%)	Not done	*PDE9A* (intron 1)- Chr 21 (5' to *PDE9A*)
3745	Not done	0R0GampF (66%) 2R0G1F (20%)	Not done	*PDE9A* (intron 1)- Chr 21 (3' to *PDE9A*)

Table 11. A summary of the results from aCGH, FISH, MCC and LDI-PCR, in nine iAMP21 patients

Discussion

10.3.3.4 Array-based CGH

In an attempt to better characterize the iAMP21 chromosomal abnormality, a number of cytogenetic and aCGH analyses have been conducted (37;172;174;177). Most were performed at lower resolutions or on fewer patients than presented here. Six patients from this investigation were included in both studies performed by Strefford *et al* (37;62) and Robinson *et al* (172). One additional case was included in the former study only, whilst six additional cases were included in the latter study only. The higher resolution array used in this investigation confirmed and refined all CNA documented by Strefford *et al* in these patients, as well as identifying an additional 21 CNA. A single discrepancy was detected; aCGH for patient 7828 revealed no copy number change at the 21q telomere. However, FISH performed by Robinson *et al* (172) showed a telomeric deletion, which was confirmed in this study using the *PDE9A* FISH probe. The DLRS values (0.6) suggested that the aCGH data for this patient was poor, explaining why this deletion was not detected (refer to section 9.10).

The application of high resolution aCGH to the genomic characterization of iAMP21 patients has revealed a complexity not anticipated from G-banded analysis. In contrast to the 40 cytogenetically visible aberrations that would result in copy number changes, aCGH detected 179 CNA, including 76 targeting chromosome 21.

The identification of nine large scale aberrations that were not detected, or misinterpreted by cytogenetics demonstrated the advantages of analyzing the genomic content of large cell numbers by aCGH.

Chromosome 9p

As previously described (section 10.2.4.1), deletions of 9p targeting the *CDKN2A* locus are a frequent finding in ALL. Although no known established chromosomal aberrations have been associated with iAMP21, a single study has reported loss of 9p in 3/7 iAMP21 patients studied (177). Four patients (21%) in this study had deletions targeting *CDKN2A*. In three of these cases the deletions were below the resolution of G-banding. Therefore, the lower resolution techniques employed in previous studies may not have detected similar CNL. *PAX5* and *ZCCHC7* deletions were observed at a lower frequency then previously reported in B-ALL (10% vs. 29%) (63).

Chromosome 7

In this study, deletions of chromosome arm 7q were observed in three patients. The first CRD encompassed 38 genes, including *TRIM24* and *HIPK2*. TRIM24 is involved in mediating transcriptional control by interacting with nuclear receptors, such as the

estrogen, retinoic acid, and vitamin D3 receptors. HIPK2 is a conserved serine/threonine nuclear kinase that interacts with homeodomain transcription factors. Forty one genes including the apoptosis related *CASP2* were encompassed by the second CRD. Kuchinskaya *et al* (177) have also described recurrent 7q deletions in iAMP21 patients (3/7 patients), but the genomic positions were not revealed, preventing direct comparison between studies. Deletions of 7q are recurrent events in a number of other haematological malignancies and have been correlated with adverse outcomes in children and adults with myeloid malignancies (178).

Three patients had a deletion of 7p, encompassing the *IKZF1* gene. As discussed in section 10.2.4.1, this deletion has previously been described in B-ALL (63;66), although at a lower frequency than observed in this study (15.7% versus 7-8%)

Chromosome 11q and 12p

Deletions of 11q have been previously reported in ALL, as described in section 10.2.4.1. Interestingly, the commonly deleted 11q regions in the iAMP21 patients (102.802893 to 102.976602 Mb and 106.603155 to 114.420246 Mb) overlap with the CRD observed in *ETV6-RUNX1* patients (102.510235 to 115.446921 Mb), and encompass the *ATM* gene. This provides further evidence for the pathogenic role of *ATM* deletions in ALL. Similarly, 12p deletions targeting the *ETV6* gene are a frequent finding in both leukaemia and solid tumours (refer to section 10.2.4.1). In accordance with this data, 15.7% of iAMP21 patients harboured 12p deletions that were not visible by FISH, using the commercial *TEL-AML1* ES probe.

Chromosome 13q

Copy number losses targeting 13q were identified in 26% of cases, consistent with deletions observed in a range of haematological malignancies, as described in section 10.2.4.1. The deletions targeting the *RB1* gene, seen in 15.7% of cases, also included telomeric sequences; the location of a second putative tumour suppressor gene. *RB1* deletions were observed at a higher frequency in these patients, than have been reported in other ALL subtypes (62;63).

Chromosome 20q

Deletions of chromosome 20q occurred in 15.7% of iAMP21 patients. Chromosome 20q deletions are seen in approximately 6% of myeloid disorders, but are rare in lymphoid malignancies. However, previously published data suggested that the deletion may arise in an undifferentiated progenitor that has not yet committed to a myeloid or lymphoid lineage. A 2.6-2.7 Mb CRD at 20q11 has been identified in myeloid disorders and is postulated to be the location of an unidentified TSG (179). A 3.6 Mb CRD on 20q13.12

was observed in the patients in this study. The prognostic significance of this aberration remains unclear.

Chromosome 17

Gain of 17q is well documented in a wide range of haematological malignancies, including CML, AML and ALL, and is most often associated with i(17)(q10). The incidence of this isochromosome in haematological malignancies, and the fact that it is often found as the sole karyotypic abnormality suggests an important pathogenic role in leukemogenesis (180). However, a number of recent publications (62;64;66) have identified partial gain of 17q, not associated with i(17)(q10), as a recurrent event in ALL. Strefford *et al* (62) described gain of 17q material in five ALL patients, including one with iAMP21. The 8 Mb common region of gain, between genomic positions 72.8 to 80.8 Mb overlaps with the 10.8 Mb CRG observed in this study, between genomic positions 67.647121 and 78.507610 Mb. This large region encompasses many genes, of which, *TBCD* is of particular interest. Gene expression profiling of eight iAMP21 patients revealed significant over-expression of this gene (37). *TBCD* encodes the tubulin folding cofactor D protein, one of four proteins involved in the pathway leading to correctly folded beta-tubulin from folding intermediates. Beta-tubulin represents one of the main subunits of microtubules, which are responsible for chromosome segregation during anaphase. In mammals, microtubules are organised by the centrosome, which is comprised of two centrioles (composed of α/β tubulin heterodimers) and a protein network called the pericentriolar material (PCM). Initiation of new microtubule growth is regulated by the recruitment of γ-tubulin ring complexes to the PCM. A recent study showed that over-expression of *TBCD* leads to loss of γ-tubulin at the centrosomes and production of cells that fail to initiate the growth of microtubules from centrosomes. This results in a block in mitosis (181).

Chromosome X

Gain of chromosome X is the only previously reported recurrent aberration in iAMP21 patients. Harewood *et al* (174) and Robinson *et al* (172) both described gain of X in approximately 25% of cases. In accordance with these data, gain of chromosome X was observed in 26% of patients in this study and gains specifically of Xp were identified in a further 10% of cases. Gain of whole chromosome X is a common aberration in HeH, and in Down patients who develop ALL (182). Gain of Xq is also a recurrent aberration in *ETV6-RUNX1* positive patients (section 10.2.4.1).

Chromosome 21

All iAMP21 patients had unique, complex abnormalities targeting chromosome 21. However, they were not seen as the sole abnormality in any of the patients analysed,

suggesting a multi-step model of cancer development. The genomic positions of the common regions of gain, amplification and deletion along chromosome 21 were consistent with previously published data (37;177). However, from the study of additional cases using high resolution aCGH, the extent of these common regions was greatly reduced. The 5.7, 6.5 and 0.6 Mb regions of gain, amplification and deletion were refined to 0.025, 3.4 and 0.25 Mb, respectively. The CRA remained the only consistent abnormality detected in all patients. The refinement of this region to within genomic positions 32.8-36.31 Mb may facilitate the discovery of a gene involved in the initiation of iAMP21. This newly defined region contains five protein coding sequences and two genes, *TCP10L* and *SYNJ1*, not previously encompassed by the CRA. Additionally, the *ERG* and *ETS2* genes are excluded from this CRA. The CRD contains a single gene, *DIP2A*, about which little is currently known.

To determine whether high resolution aCGH could identify a profile unique to iAMP21 patients, by which it might be distinguished, five non-iAMP21 patients with cytogenetically detected chromosome 21 abnormalities were included for comparison purposes. Although all five cases had extra copies of *RUNX1* by FISH (Appendix 14), they were not arranged on a single abnormal copy of chromosome 21. Gain of chromosome X, the only previously reported recurrent aberration in iAMP21 patients, was observed in a single case. In addition, one patient had a gain of 17q, encompassing the CRG observed in the iAMP21 patients. In three cases, the chromosome 21 abnormalities were complex and the sub-telomeric deletions and CRA, characteristic of iAMP21 patients, were observed. These findings emphasise the difficulties in accurate diagnosis of iAMP21, and it is apparent that aCGH alone cannot reliably be used to differentiate between these two subsets of patients. However, as the iAMP21 and non-iAMP21 aCGH profiles were sometimes similar, these patients may in fact belong to the same biological subgroup.

10.3.3.5 Disruption of the *PDE9A* gene is a recurrent event in iAMP21 patients

Array-CGH identified deletions of 3' *PDE9A* in two patients, highlighting this gene as a target of interest. Using a combination of aCGH, FISH, MCC and LDI-PCR the recurrent involvement of this gene in iAMP21 patients was confirmed;

Patient 4279

Array-CGH revealed a deletion from intron 1 of *PDE9A* in this patient. FISH and MCC confirmed this deletion, and sequence analysis showed a fusion of *PDE9A* to the *MUS81* gene on chromosome 11q. RT-PCR was able to confirm the expression of this fusion transcript, although FISH probes were unable to detect the fusion. It is possible that the fragment of *MUS81* involved in the fusion is too small to be detected by FISH. To determine the extent of *MUS81* is involvement, LDI-PCR is currently being performed

from various points within this gene. Mus81 together with its binding partner Eme1, forms an endonuclease complex responsible for the cleavage of a variety of DNA structures. It plays an essential role in DNA repair during replication (183). *MUS81* deficiency leads to hypersensitivity to interstrand cross-linking (ICL) agents, and elevated levels of genomic instability. ICL, such as mitomycin C, are widely used as anti-tumour agents and disrupt replication by generating stalled replication forks. These can lead to DSB and chromosome rearrangements (184). McPherson *et al* (185) have shown that $Mus81^{+/-}$ and $Mus81^{-/-}$ mice spontaneously develop tumors, thus establishing *MUS81* as a novel haploinsufficient tumor suppressor gene. The disruption of *MUS81* by this fusion may have played a role in leukemogenesis in this patient.

Patient 6111

Array-CGH, FISH and MCC all confirmed a deletion from intron 1 of *PDE9A* in this patient. However, LDI-PCR repeatedly failed due to poor DNA quality and a fusion could not be confirmed.

Patient 6996

Although aCGH revealed deletion of the entire *PDE9A* gene in this patient, MCC demonstrated a deletion from intron 1. FISH confirmation was not possible as no fixed cells were available. In addition LDI-PCR and sequencing identified a fusion between *PDE9A* and *VWF,* on chromosome 12p. The failure of aCGH to detect the intragenic breakpoint of this deletion may be due to the high background noise which was observed in this sample, or the presence of different sub-clones as observed in patient 8983 (see below). VWF is a large multimeric glycoprotein which is synthesised in endothelial cells and megakaryocytes and secreted into the plasma. VWF attaches to subendothelial collagen and to platelets via the glycoprotein Ib receptor, promoting formation of a platelet plug at the site of small vessel injury. It also binds, transports and protects coagulation factor VIII. It is crucial to the haemostasis process. Mutations in this gene or under-expression of the protein result in von Willebrand's disease, the most common inherited bleeding disorder in humans (186). It is unclear how disruption of this gene may be linked to the disease phenotype.

Patient 8983

In this patient deletion of the entire *PDE9A* gene was observed by aCGH. However, MCC revealed an interstitial deletion of approximately 1407bp between genomic positions 42957596-42959003 Mb. The arrays do not provide probe coverage of these sequences. FISH analysis detected the presence of two dominant clones. In 55% of cells a single copy of the gene was deleted, supporting the aCGH data. In 26% of cells additional copies of the centromeric part of *PDE9A* were seen. FISH performed by Hazel Robinson

also revealed telomeric deletion of chromosome 21. This suggests that subclones with different deletions, beyond that identified by MCC, may be present in this patient. This is currently under further investigation. Different sub-clones present at low levels are unlikely to be detected by aCGH. LDI-PCR revealed a fusion between *PDE9A* and the *REXO1L1* gene, an RNA exonuclease, on chromosome 8q. Little is currently known about this gene. No material was available to test for expression of the fusion transcript.

Patient 7619

In contrast to the deletion of the entire gene, as seen by aCGH, MCC revealed a deletion of *PDE9A* from intron 1. FISH analysis revealed the reason for this discrepancy; 39% of cells have deletion of the whole gene, whilst 31% have extra copies of the centromeric part of the gene. As for case 8983, these FISH results suggest that different cell populations may harbour different deletions. However, LDI-PCR failed due to poor DNA quality, and a fusion could not be confirmed.

Other patients

PDE9A FISH was performed on a further 65 iAMP21 patients and deletions of the telomeric part of the gene were identified in four cases (2647, 3745, 7024 and 7828). Patient 7828 had a normal aCGH profile, but the quality of this data was poor. MCC confirmed the FISH result for this case. LDI-PCR and sequencing was performed on these four cases. In three patients, *PDE9A* was fused to non-coding chromosome 21 material 5' (7828 and 2647), and 3' (3745) of the gene. In the remaining case (7024) it was fused to a second inverted copy of *PDE9A*.

In total, deletions of the telomeric part of *PDE9A* were identified in 12.8% of cases (n=9/70), at least in a sub-population of cells, by a combination of aCGH, MCC and FISH. Seven of the nine cases had gain or amplification of the retained centromeric sequence, and LDI-PCR identified fusions in seven cases.

PDE9A

Cyclic adenosine 3',5'-monophosphate (cAMP) and cyclic guanosine 3',5'-monophosphate (cGMP) are ubiquitous intracellular second messengers that transduce the effect of a variety of extracellular signals (187;188). The biological effects of cyclic nucleotides are dependant on their intracellular concentration, rate of formation and rate of hydrolysis. These factors are regulated by a complex signalling system involving adenylyl and guanylyl cyclases (which manufacture cAMP and cGMP), protein kinases and cyclic nucleotide phosphodiesterases (PDEs). PDEs hydrolyse cAMP and cGMP into their corresponding monophosphates (188) and play an important role in signal transduction by regulating the intracellular concentration of the cyclic nucleotides.

Twenty one *PDE* genes have so far been identified in mammals (187). They are classified into 11 families based on their expression patterns, cellular localization, regulation of catalytic activity, affinity for cAMP or cGMP and sensitivity to inhibitors (189). Within each family there are multiple isozymes, and multiple splice variants of each isozyme, encoding more then 60 isoforms in total (190). All PDEs have a highly conserved C-terminal catalytic domain (~270 amino acids in length), and an N-terminal regulatory domain responsible for intracellular targeting and binding of cofactors such as divalent cations (191).

PDE9 is a recently identified novel family with only one member so far; *PDE9A*, which is located on chromosome 21q22.3. It has the highest affinity for cGMP of all known *PDE* and is therefore likely to be important in regulating intracellular cGMP levels and in the activation of cGMP-dependant signalling pathways (192). It may function at lower cGMP levels then other PDE in the same cell. Several potential phosphorylation sites exist in the structure, indicating that PDE9A may be regulated by cAMP and cGMP dependant kinases and/or tyrosine kinase and protein kinase C (193).

PDE9A has at least 20 exons and currently 28 transcript variants, with differential tissue distribution and sub-cellular localization, have been identified. In at least 20 transcript variants splicing involves exons 1 to 8, and two distinct start codons in exon 1 give rise to splice variants with different N-terminal sequences. The entire catalytic domain, encoded by exons 12-18, is retained in all variants. There are no clear changes in the protein sequences to indicate a defined function for each variant (190).

Most *PDE9A* splice variants are expressed to varying degrees in all tissues, and are most abundant in the spleen, brain and immune system tissues and cells (188;191). However, the leukaemia cell lines HL-60 and MOLT-4, and the Burkitt's lymphoma cell lines Raji and Daudi, showed no expression of *PDE9A* (190).

According to Wang *et al* (191), PDE9A1, which contains all 20 exons, is localized solely in the nucleus. Mutations in the nuclear localization signal, starting at residue 105, prevent PDE9A1 from entering the nucleus, where it may have a potential role in the regulation of nuclear events (191). However, Rentero *et al* (194) found PDE9A1 predominantly in the membrane ruffles at the cell margin.

It is postulated that *PDE9A* is a possible candidate for genetic disease mapped to 21q22.3, such as bipolar affective disorder, and its over-expression might be implicated in the phenotypes of Down syndrome (189). PDE inhibitors are used as vasodilators, antidepressants, antithrombotics and agents for improving cognitive function (195). The potent and selective PDE9 inhibitor, BAY-73-6691, is under preclinical development for the treatment of Alzheimer's Disease (187).

With one exception (patient 8983), all the deletions targeting *PDE9A* resulted in loss of the entire catalytic domain (figure 41), preventing the normal function of the gene. These

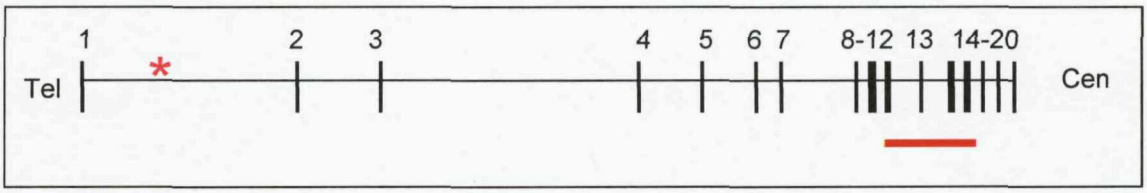


Figure 41. A schematic representation of the PDE9A *gene, showing the positions of the exons (numbered 1-20). The red line indicates the exons encoding the catalytic domain, and the asterisk indicates the approximate position of the deletion breakpoints identified in the iAMP21 patients.*

data implicate a role for *PDE9A* in the pathogenesis of iAMP21, in patients with the deletion. Its under-expression in leukaemia and lymphoma cell lines also supports this hypothesis (190). In patients with intact *PDE9A*, the gene may be inactivated by a different mechanism, such as mutation or promoter methylation. There are two CpG islands in the 5'UTR (genomic position 42967705-42968147 Mb), and coding sequence (genomic position 43000285-43000711 Mb) of *PDE9A*. The first is unmethylated, and the second is completely methylated in normal individuals (196).

In four cases *PDE9A* was fused to different regions of chromosome 21. This reflects the BFB cycle as the mechanism of iAMP21 formation, particularly in the case with inversion of chromosome 21 sequence (7024). The remaining three cases had fusions involving different chromosomes (8q, 11p and 12q). The partner chromosomes had normal aCGH profiles for the loci involved. It is not known whether these fusions are located on the abnormal chromosome 21. Further research into the role of *PDE9A* in iAMP21 pathogenesis is planned, including mutation analysis and investigation into the methylation status of the CpG islands and the effect of gene disruption of mRNA expression levels.

10.3.4 Conclusions

The high resolution aCGH approach used in this study has established the highly complex and heterogeneous nature of iAMP21 and provided insight into the size, genomic position and gene content of CNA affecting these patients. Novel recurrent large scale and submicroscopic aberrations were observed in these patients, and the previously reported common regions of amplification and deletion on chromosome 21 have been further refined. All abnormalities have been previously described in leukaemia. The previous identification of these aberrations in iAMP21 has been prohibited by the low resolution of other studies, and the smaller patient cohorts investigated. One aim of this study was to identify distinguishing characteristics in iAMP21 patients that would aid in accurate diagnosis. Based upon the chromosome 21 aCGH profiles it may be possible to design a panel of FISH probes targeted to regions of recurrent gain, amplification and

deletion, which would help differentiate between patients with iAMP21, HeH or isochromosome 21. Array-CGH has identified numerous markers that may contribute to leukemogenesis, impact on prognosis, or become future therapeutic targets, of which, *PDE9A* is significant. However, aCGH failed to identify copy number changes within this gene in four patients, highlighting the limitations of this technique, such as its inability to detect aberrations in sub-populations of cells. Additionally, it was not possible to differentiate between iAMP21 and non-iAMP21 patients with chromosome 21 aberrations based on aCGH data alone. Currently, a multi-disciplinary approach remains the most effective method of diagnosis.

10.4 Global genomic profiling of ALL patients lacking clinically relevant chromosome abnormalities.

10.4.1 Introduction

Cytogenetics is important for the classification and risk stratification of ALL patients. The presence of specific chromosomal abnormalities provides the single most important prognostic indicator in patients with this disease (16;197;198). However, established chromosomal changes with prognostic significance only account for approximately 70% of paediatric, and 50% of adult ALL. The remaining cases, including those with a normal karyotype, are unclassified. Unclassified karyotypes in B-ALL are defined as patients with 40-50 chromosomes, excluding patients with the defined cytogenetic subgroups of HeH, near-haploidy and low hypodiploidy, as well as those with the established genetic/chromosomal rearrangements; *BCR-ABL1*, *ETV6-RUNX1*, *MLL* rearrangements, *E2A-PBX1* and T-ALL. Unclassified karyotypes are observed in all age ranges and in a preliminary study it was discovered that this group includes patients with aberrations such as deletions of chromosome arms 6q, 7q and 9p, and trisomies of chromosome 19, 21 and X (62). Several recent studies (62;63;199) using high resolution array technology have identified a number of recurrent sub-microscopic deletions in these patients, targeting single genes, including *ADD3* and *ETV6*. Additionally, in a study of 200 adult ALL patients, including 54 with unclassified karyotypes, Pullarkat *et al* (16) determined that patients with a normal karyotype or deletion of 9p had a favourable outcome, whilst all other unclassified karyotypes were associated with an intermediate prognosis. However, there are currently wide gaps in our understanding of the mechanisms involved in disease initiation and progression in this relatively under-studied class of leukaemia.

The aims of this study were;

- To further characterise the genomes of unclassified ALL patients.
- To identify recurrent copy number changes and target genes that may contribute towards leukaemogenesis, or represent prognostic markers or therapeutic targets in these patients.

10.4.2 Materials and Methods

10.4.2.1 Patient samples

DNA was extracted from 41 diagnostic bone marrow samples from ALL patients (aged 1 to 29 years) lacking clinically relevant abnormalities, as described in section 9.4 (including the five 'non-iAMP21' patients as discussed in section 10.3.3.1). Chromosomal analysis was performed in the UK regional cytogenetics laboratories, and described according to the International System for Human Cytogenetic Nomenclature (84).

10.4.2.2 Array- based CGH

The samples were processed for aCGH analysis using the Agilent Human Genome CGH Microarray kit 244k, as described in section 9.10. The arrays were analysed using Agilent CGH Analytics 3.5.14, based upon the UCSC March 2006 assembly (HG18). Five consecutive probes with fluorescence ratios deviating from a normalized log ratio value of 0 by at least +1 were defined as copy number amplifications.

10.4.3 Results

In total, 228 CNA were detected by aCGH, in 37 patients, with between one and 23 aberrations per case (Appendix 14 and figure 42). No copy number changes were observed in four patients. Ten aberrations involved whole chromosome gains, and of the remaining 218 CNA, deletions (n=178) were far more frequent than gains or amplifications (n=47 and n=4, respectively). Deletions targeted chromosome arms 9p (n=38, 11 bi-allelic deletions), 4q (n=9), 7p (n=9), 8q (n=8), 12p (n=8) and 21q (n=8). Recurrent gains or amplifications affected chromosome 21 (n=22). The smallest copy number change observed was a 59bp deletion on chromosome arm 2p (1.492810-1.492869 Mb). Additionally, by consulting the database of genomic variants (www.projects.tcag.ca/variation/), 263 copy number variants (CNV) were identified, which targeted 51 different loci (Appendix 8). No CNV were found on chromosomes 12, 13, 20, 21, X and Y. As seen in the *ETV6-RUNX1* positive and iAMP21 patients (sections 10.2 and 10.3 respectively), deletions at 2p11.1 (n=24), 7q34 (n=19), 14q11 (n=24), 14q32.33 (n=29) and 22q11.22 (n=11) were frequent. An additional 10 CNV were targeted in at least 10% of patients, and between two and 20 CNV were observed in each patient. With one exception (patient 7322), all patients without CNA carried at least two CNV.

The diagnostic cytogenetic and FISH analysis for established aberrations is provided in Appendix 14. Two of the ten cases exhibiting a normal karyotype by cytogenetics had no CNA. In the remaining eight normal cases, aCGH revealed a total of 46 CNA, while a further 14 were detected in two cases with a failed cytogenetic result. No CNA where detected in patients 4051 and 7322, although both had abnormal karyotypes by G-banded analysis (46,XX,del(9)(p?)[3] and 46,XY,add(7)(p1?3),dup(21)(q22q22)[15], respectively). Three copies of *RUNX1* were observed in 92% of cells during FISH analysis of patient 7322, confirming the duplication of chromosome 21q seen by G-banding.

In total, cytogenetic analysis detected 81 visible abnormalities that would result in copy number imbalances. Array-CGH confirmed and refined 39 of these. A further 21 were

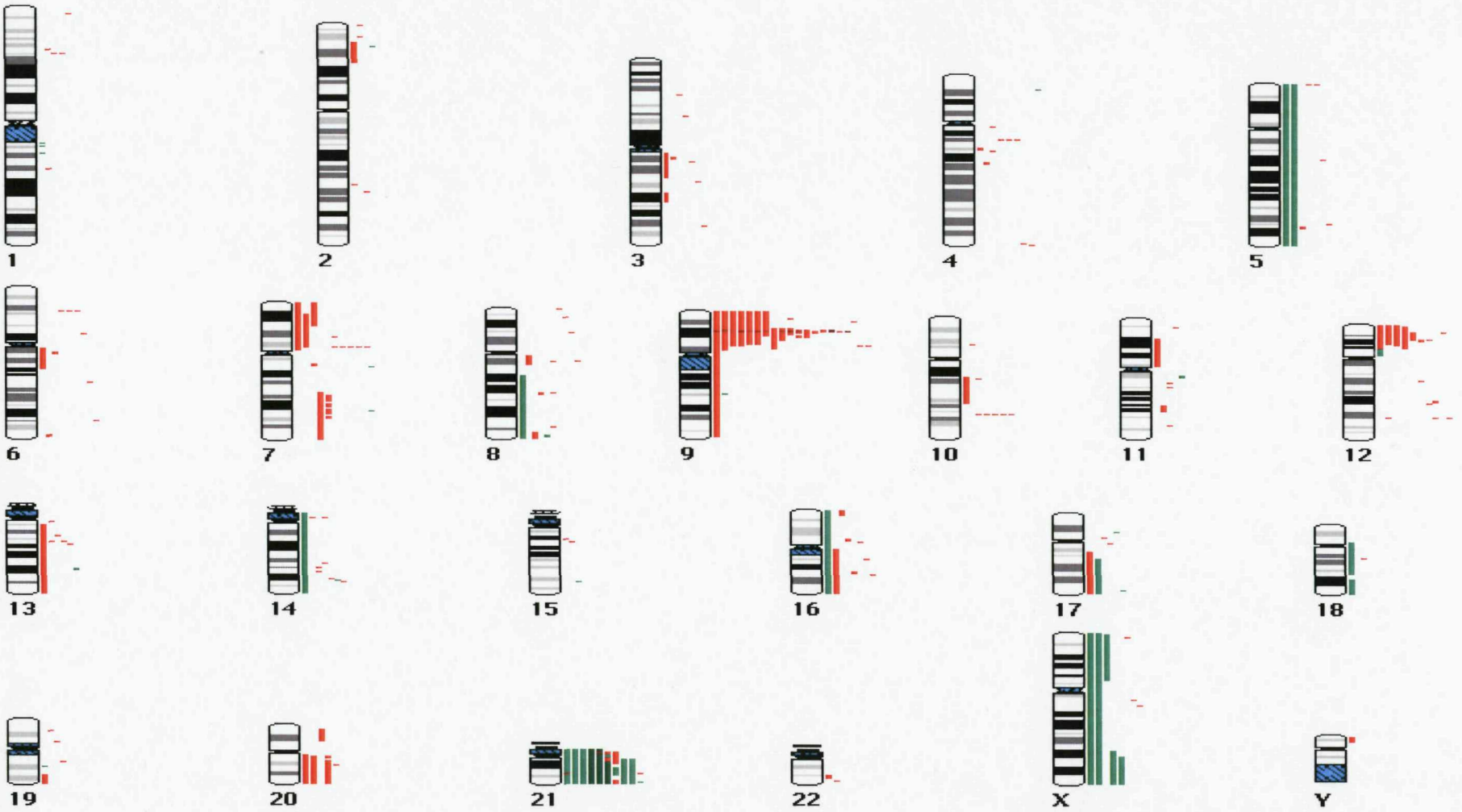


Figure 42. CNAs detected in 41 ALL patients lacking clinically relevant abnormalities. To the right of each chromosome idiogram, red, green and dark red vertical lines show copy number losses, gains and homozygous deletions respectively. Each vertical line is positioned to show the extent and genomic location of a single change in a single patient.

described as additional unidentified material or marker chromosomes, and the origin of chromosomal material in 14 of these aberrations could be deduced from the aCGH or cytogenetic data, although FISH would be required for definitive confirmation. Of the remaining aberrations unconfirmed by aCGH, 18 were present in fewer than ten metaphase cells by cytogenetics (range=1 to 10 cells). Therefore the abnormal population may be too small to be detected by aCGH. The missing chromosome 20 in patient 12377 is incorporated into the marker chromosome, and the four missing chromosomes in patient 772 are incorporated into the two markers and additional material. As discussed above, the aCGH and cytogenetic analysis for patient 7322 did not correlate.

Chromosome 21

Eighteen CNA targeting chromosome 21 were detected in nine patients. Gain or amplification of the entire chromosome was seen in three cases (n=2 and 1, respectively). Whole chromosome gain and amplification was also seen in patients 3043 and 4676, with the exception of 0.1 Mb (38.686491- 38.793256 Mb), and 1.6 Mb (15.957693-17.628955 Mb) deletions, respectively. In total seven deletions were observed, and in all cases the region between genomic positions 29.086997 and 38.049606 Mb was gained or amplified.

Chromosome 9p

Deletions of chromosome arm 9p were observed in 41% patients (n=17) and targeted two genomic regions. Seven of the eight patients with large scale deletions (>10Mb) harboured a bi-allelic deletion targeting *CDKN2A/B*. Bi-allelic deletions were also observed in four of the eight patients with focal deletions of this gene (figure 43). All patients shared a CRD between genomic positions 21.968405 and 21.980581 Mb, excluding the *MTAP* gene. In six patients the deletion involved both the *CDKN2A/B* and *PAX5* loci, and in a single case *PAX5* was targeted specifically, without *CDKN2A/B* involvement. In two cases the deletion breakpoint was located within the *PAX5* gene (proximal and distal breakpoint, n=1 and 1, respectively). The remaining cases had large scale deletions that encompassed the entire gene, including a deletion of the whole of chromosome 9 (patient 8451). *ZCCHC7* was also targeted in four cases.

Chromosome 12p

Eight deletions involving chromosome 12p were detected, five of which were large. In six cases the deletions encompassed the entire *ETV6* gene and in a further case the distal deletion breakpoint occurred within intron 1 of this gene.

Chromosome X

Gain of chromosome X was observed in five patients, including whole chromosome gain (n=2), and duplication of both the short (n=1) and long arms (n=2). In four of these patients the region between genomic positions 130.296688 and 154.340481 Mb was gained.

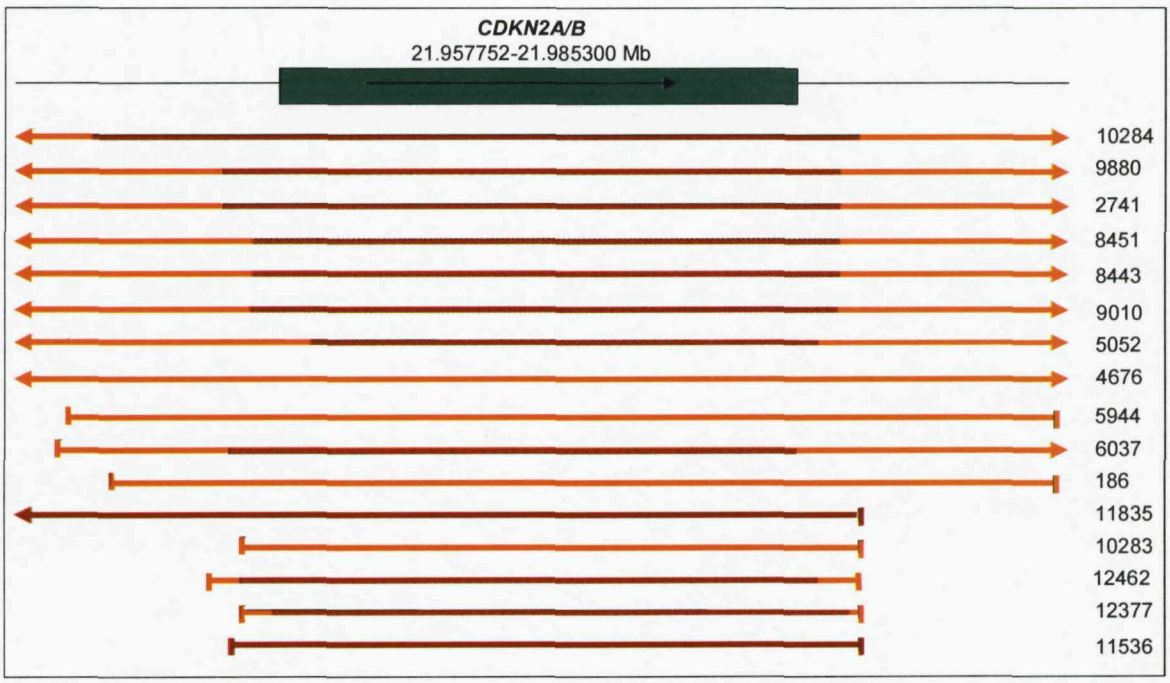


Figure 43. The CNL involving the CDKN2A/B gene, in 16 unclassified patients. Light and dark red bars represent the position and extent of mono- and bi-allelic deletions, respectively.

Other aberrations

Seven of the nine deletions of 7p shared a CRD between genomic positions 50.194152 and 50.256483 Mb, and focal deletion of the *ADD3* gene occurred in five of the seven patients with 10q deletions.

The deletion between genomic positions 26.245517 and 26.363050 Mb in chromosome band 6p22.22, in 3 cases, encompassed 17 histone gene family members. Three patients shared a 0.2 Mb CRD of 8q, targeting the sequence immediately 5' of the *TOX* gene. In one case the deletion encompassed the entire gene. Chromosome 20q was also targeted by three large deletions, with a CRD of 0.12 Mb between genomic positions 34.722945 and 60.220009 Mb.

Two of the deletions of chromosome 13 specifically targeted *RB1* and the downstream region, Deletions of *EBF1* on 5q, and of genomic position 178.399892 to 178.831991 Mb on chromosome 3, downstream of the *TBL1XR1* gene, were deleted in a single case each (5846 and 2741, respectively).

10.4.4 Discussion

ALL patients lacking clinically relevant chromosomal/genetic aberrations represent a poorly understood subtype of this disease. Despite the heterogeneity of the cytogenetic and clinical features of the patients included in this study, the application of high resolution aCGH to their characterisation has revealed numerous recurrent genomic abnormalities. The findings correlate well with data produced from several recently conducted array-based analyses (62;63;199).

Five of the unclassified patients here were also included in the 1 Mb BAC aCGH study performed by Strefford *et al* (62). An additional six CNA were detected in these patients by the higher resolution platform employed in this investigation. In a single case (3141), Strefford *et al* observed a deletion of 6q that was not confirmed in this study. This inconsistency may be attributed to the high level of background noise experienced for this case with the Agilent arrays.

No copy number changes were detected in four patients; two with normal karyotypes, and two with cytogenetically visible aberrations. The deletion of 9p in patient 4051 was seen in only 3 metaphase cells, which indicates that it was probably below the detection limit of the arrays in terms of the percentage of abnormal cells in the sample. However, the duplication of 21q22 seen in patient 7322 was confirmed in 92% of cells using FISH. The reason for this discrepancy between the cytogenetic and aCGH data is unclear. As no CNV were detected in this case, it is likely that the aCGH for this patient was sub-standard and should be regarded as a fail. In the normal cases with no CNA detected it cannot be ruled out that point mutations, copy number neutral loss of heterozygosity (CNNLOH), aberrations below the resolution of this aCGH platform or CNV may be responsible for the disease phenotype. Strefford *et al* (62) and Kuchinskaya *et al* (199) also identified normal aCGH profiles in 12% and 10.7% of patients, respectively.

The concordance between the cytogenetics and aCGH data was lower for these unclassified patients than was seen in the *ETV6-RUNX1* positive or iAMP21 patients (section 10.2.3.1 and 10.3.3.1). However, a large proportion of aberrations were detected in fewer then 10 metaphase cells, possible below the detection level of aCGH. Karyotypes containing additional material or marker chromosomes were frequent, indicating the chromosomal complexity of these cases. Rearrangements may have been balanced in nature, or misinterpreted.

Chromosome 21

Copy number changes, particularly gains, targeting chromosome 21 were the most common aberration detected in this cohort of patients. In a similar aCGH study of unclassified ALL patients, Strefford *et al* (62) also observed a high frequency of

chromosome 21 gains. However, five patients in this study (4247, 4676, 4746, 9010, 10442) had been specifically selected for array analysis because of their chromosome 21 involvement (refer to section 10.3). A common region of gain was seen in all unclassified cases with chromosome 21 imbalances, encompassing the *RUNX1* gene. Mullighan *et al* (63) did not identify any unclassified cases with this aberration.

Chromosome 9p

As discussed in section 10.2.4.1, chromosome 9p deletions, targeting *CDKN2A/B,* are frequently associated with ALL. However, they were found at a higher frequency in the unclassified patients in this study, than is normally observed in ALL (41% compared to 30%, respectively) (134). In agreement with this finding, Mullighan *et al* (63) and Strefford *et al* (62) also identified 9p deletion in a high percentage of unclassified patients (36% and 50%, respectively). Although the deletion targeting *CDKN2A/B* was detected in only 26% of the patients by Strefford *et al*, this was likely due to the inadequate probe coverage of the BAC arrays. In contrast, the frequencies of *PAX5* deletions were lower in this study than Mullighan *et al* (63) reported in both ALL in general (29%), and in unclassified patients (33%). The absence of *PAX5* deletions in the Strefford *et al* (62) study was also likely due to inadequate probe coverage.

Chromosome 12p

ETV6 plays an established and critical role in leukaemogenesis and deletions of this gene are frequent in a number of ALL subtypes (62;66;124), typically as secondary abnormalities associated with *ETV6-RUNX1* fusions. In accordance with previously published data (62;63;66), *ETV6* deletions were also observed in 15% of unclassified patients in this study.

Chromosome 7p

Deletions of 7p were observed in 21.9% of patients, and shared a CRD between genomic positions 50.194152 and 50.256483 Mb in 17% of cases. Although there are currently no significant known genes located at this position, *IKZF1* (refer to section 10.2.4.1) is located 0.5 Mb centromeric of this loci. These deletions may disrupt promoter sequences 5' of the gene.

Chromosome 10q

Focal deletions of the *ADD3* gene were detected in 12% of the unclassified patients here. Mullighan *et al* (63) also identified this deletion in 8.3% of unclassified cases, and Strefford *et al* (62) observed a single case with a large scale deletion of 10q, encompassing this gene. This data correlates with the *ADD3* deletions identified in the *ETV6-RUNX1* positive patients (refer to section 10.2.3.1), indicating its role in the

pathogenesis of ALL. As discussed in section 10.2.4.1, deletion of *ADD3* is a putative poor prognostic indicator. One of the three unclassified patients harbouring this aberration had relapsed. These findings indicate that the prognostic association of *ADD3* in ALL subtypes other than *ETV6-RUNX1* requires further investigation.

Chromosome 20q

Deletions of chromosome 20q occurred in 7.3% of patients, and were also identified in 8.8% of the unclassified patients in the Strefford *et al* study (62). As described in section 10.3.4.1, chromosome 20q deletions are rare in lymphoid malignancies and the prognostic significance of this aberration remains unclear. A single case (4746) had a cytogenetically visible aberration of 20q; der(20)t(20;21)(q13;q?). An almost identical translocation (der(20)t(20;21)(q11;q11)) was described in an ALL patient by Brezinova *et al* (179). In addition to the 20q deletion, patient 9880 had a deletion of 9p, indicating a possible translocation resulting in a dicentric chromosome (dic(9;20)).

Chromosome X

Gains involving chromosome X were observed in 12.1% of unclassified patients. Trisomy X occurred in two cases. As described in section 10.3.4.1, this aberration has been associated with iAMP21, albeit at a higher frequency. A single case with trisomy X also had a complex chromosome 21 profile, as seen in iAMP21. A further two cases had gains of Xq, an aberration recently described, at a higher frequency, in *ETV6-RUNX1* positive ALL (63;123) (refer to section 10.2.4.1).

Other aberrations

Deletions of *EBF1* (5q33.3), *RB1* (13q14.2), *TOX* (8q12.1) and 6p22.22 (observed in 2.4 %, 4.8%, 7.3% and 7.3% of patients, respectively), correlated with deletions seen in other subtypes of ALL (62;63). A single deletion of the sequence downstream of *TBL1XR1* was identified. This deletion had previously been reported in *ETV6-RUNX1* and HeH patients (n=1) only (63;64) (refer to section 10.2.4.1).

10.4.5 Conclusions

The preliminary data presented here confirmed the presence of complex and heterogeneous genetic abnormalities throughout the genome of patients with the unclassified ALL subtype. This study has provided insight into the size, genomic position and gene content of CNA affecting these patients. Numerous recurrent large and sub-microscopic aberrations were described, targeting both previously identified and novel regions, involving tumour suppressors, transcription factors or regulators of B-cell differentiation and development. The frequency of 6q and 11q deletions was lower in these patients than other subtypes of ALL (refer to section 10.2.4.1). In contrast, 7.3% of

patients had 20q deletions, a rare finding in ALL. Three of the patients included in this study relapsed. All have gain or amplification of chromosome 21, and two patients also had 9p deletions. However, loss of 9p has not been associated with high risk, and although iAMP21 patients have a poor prognosis, other ALL subtypes which commonly carry gains of chromosome 21 (HeH and *ETV6-RUNX1*) do not. In conclusion, this study has highlighted a number of genes which may be involved in the pathogenesis of ALL in patients with no established chromosomal abnormalities. If time had permitted, a number of the novel findings would be worthy of further study.

10.5 Gene expression profiling of ALL using customized gene expression arrays.

10.5.1 Introduction

Accurate risk stratification in ALL is essential for tailoring the intensity of therapy, and achieving optimal outcome in contemporary treatment regimes (8). Based upon clinical features and cytogenetic and molecular diagnostics, the majority of paediatric ALL patients are assigned to one of the following prognostically significant subtypes having; *ETV6-RUNX1*, *BCR-ABL1* or *TCF3-PBX1* fusions, *MLL* rearrangements, HeH and T-ALL. Previously published data has shown that global gene expression profiling of pediatric ALL patients identifies distinct expression profiles for these subtypes. It was claimed that they provided an overall diagnostic accuracy comparable to current multidisciplinary approaches (78). T-ALL and *ETV6-RUNX1* positive patients were characterized with 100% accuracy, and the remaining subtypes with between 96% and 99% accuracy in one study (78). A distinct expression signature was also identified for a subset of patients with normal, pseudodiploid and HeH karyotypes, and one group lacking any recurrent cytogenetic abnormality. The number of class discriminating genes which distinguished the groups from each other varied from 201 (*BCR-ABL1*) to 2063 (T-ALL) (8) . However, the authors claimed that a single gene was sufficient to give 100% diagnostic accuracy for T-ALL (*CD3D*) and *TCF3-PBX1* ALL (*PBX1*), whilst between 7 and 20 genes were required to make accurate predications for the remaining subgroups.

Comparison of diagnostic samples from patients who relapsed, remained in continuous complete remission, or developed therapy-induced AML, to samples collected at the time of relapse, identified distinct expression patterns that predicted relapse within individual subgroups. For T-ALL and HeH, relapse was predicated with 97% and 100% accuracy, respectively. In addition, a signature of 20 genes was identified that could predict, with 100% accuracy, those *ETV6-RUNX1* positive patients that would develop secondary AML (78).

These data supported the use of microarray based expression profiling as a single platform for the diagnosis of these cytogenetic subtypes, as a replacement for other 'gold standard' techniques. Since the publication of these data, the MILE (Microarray Innovations in LEukemia) study (200) has been initiated in 11 research centers. The first stage was designed to assess the ability of the Affymetrix HG U133 Plus 2.0 gene expression array to classify leukemias into 17 subgroups, as compared to conventional diagnostic assays. The subgroups include: ALL with *ETV6-RUNX1*, *BCR-ABL1*, *MYC-IGH* or *TCF3-PBX1* fusions, *MLL* rearrangements, HeH, unclassified karyotypes and T-ALL, AML with *PML-RARA*, *RUNX1-RUNX1T1* or *CBFβ-MYH11* fusions, *MLL* rearrangements, normal or complex karyotypes, CML, CLL, MDS and non-leukaemic samples. The cross validation accuracy of 1892 MILE Stage I samples is 95.4% concordant with the "Gold Standard" diagnosis, further confirming that standardized

microarray-based gene expression profiling may accurately classify leukemia samples into known diagnostic and prognostic sub-categories with a low technical failure rate and a high accuracy. However, data analysis, data storage issues, and the expense of using whole-genome expression arrays as a diagnostic tool are prohibitive. The second stage of the MILE study (200) involves prospectively processing 1000 samples on a custom designed microarray using 1,449 probe sets only (Roche). The results have not yet been published.

The aims of this study were;

- To attempt to validate the statement that gene expression profiling alone provides accurate diagnosis of genetic subtypes of ALL.
- To evaluate customized expression arrays, targeting the known class discriminatory genes, regarding their potential use in ALL sub-classification in routine clinical practice.

10.5.2 Materials and methods

10.5.2.1 Array design

Based upon previously published data (8;44;78;201-203), and collaboration with a number of investigators, 1200 relevant genes, which had previously been reported to discriminate the important genetic subtypes of ALL, were selected for printing onto the GE Healthcare 16-assay customized gene expression array (see Appendix 15 for list of genes). These included the top 100 dysregulated genes reported for each of the prognostically significant subtypes of ALL and AML, genes predictive of relapse in ALL, and over 100 genes aberrantly expressed in myeloma. Although no relapse, AML or myeloma samples were analysed in this study, relevant genes were included to potentially provide a comprehensive test for a range of haematological malignancies. GE Healthcare manufactured 25 arrays to this design. The CodeLink 16-assay arrays enable expression profiling of 16 patients in a single experiment, reducing the workload and reagent costs incurred from whole genome array studies, and potentially reducing data analysis.

10.5.2.2 Patient samples

Mononucleated cells were separated from 58 ALL patient bone marrow samples, as described in section 9.3. The disease sub-types were represented as follows: *ETV6-RUNX1* (n=9), *BCR-ABL1* (n=2), *TCF3-PBX1* (n=3), *MLL* rearrangement (n=1), HeH (n=16), T-ALL (n=6), Other (n=9) and Non-classified (NC) (n=12)
Patients classified as 'other' had non-recurrent cytogenetic aberrations. The disease subtype in NC patients is unknown due to failed cytogenetic analysis. Details of all patients are provided in Appendix 16.

RNA was extracted from the cells and the arrays were processed as described in sections 9.5 and 9.8, respectively. Each patient was processed in duplicate, on the same array.

10.5.2.3 Data analysis

The author gratefully acknowledges Richard Mitter, of the Cancer Research UK Bioinformatics and Biostatistics service for performing data analysis.

The following statistical analyses were performed, in keeping with the original studies (8;78). The raw fluorescent intensity values and quality flags generated by CodeLink Expression 4.2 software were entered into Bioconductor 2.1 software (www.bioconductor.org). A threshold was set for the raw signals so that any below background levels were set to 0.5, preventing problems when converting the values to a log_2 scale. Quantile normalization was applied to the data, normalizing the signals from each array so that their distribution was equal across all arrays (figure 44).
Before ranking genes according to their expression levels in a leukaemic subgroup, the data was corrected for batch effects using LIMMA (Linear models for microarray data) (www.bioconductor.org) (figure 45). A linear model is fitted to the expression data for each probe. The coefficients of the fitted models describe the differences between the RNA sources hybridized to the arrays. For these data the coefficients were disease group, patient identification and batch data.
An empirical Bayes t-test, which compares the mean signal intensities for each subgroup to all other subgroups, was then used to identify the top 100 differentially expressed genes in a disease subgroup and Hierarchical clustering of the samples was performed using a 1-Pearson distance metric and average linkage.

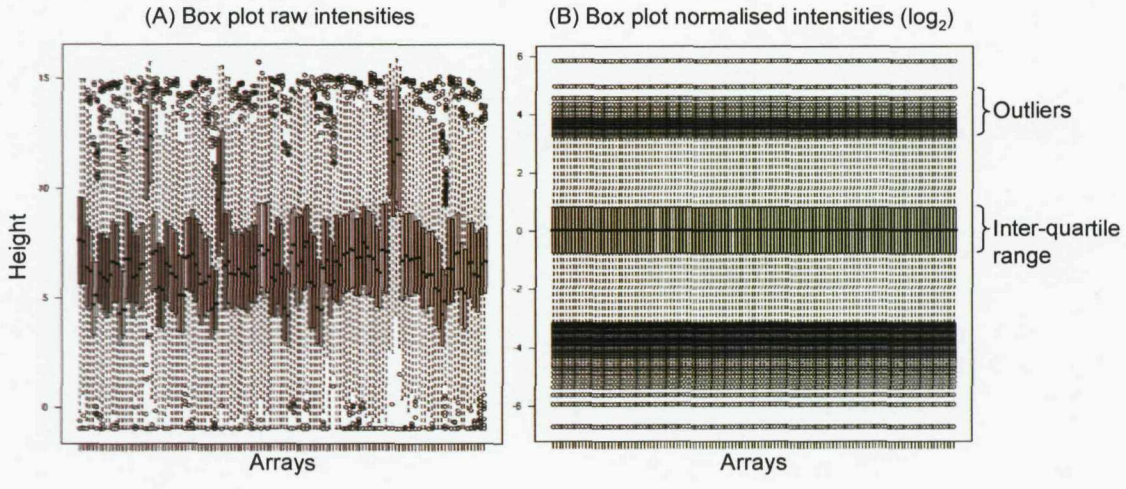


Figure 44. Box plots displaying the distribution of the fluorescence intensity signals before (A) and after (B) quantile normalization.

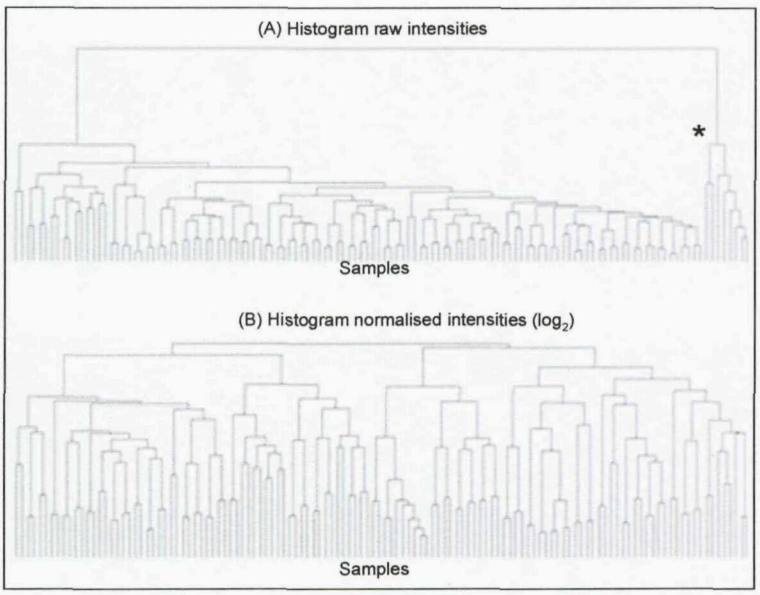


Figure 45. (A) Histogram of the raw fluorescence intensity signals, displaying the degree of similarity between the arrays. All of the arrays in the smallest cluster (*) were processed on the same day, indicating a strong batch effect. (B) Histogram of the normalised fluorescence intensities following correction for batch effects, using LIMMA.

10.5.3 Results

Comparison with published data

The top 100 differentially expressed genes, which were revealed by the above analyses for each of the eight subtypes, are presented in Appendix 17. Of these 800 genes, 284 appeared in the top 100 list for more than one subgroup, resulting in a unique set of 516 genes. These 284 genes may be commonly dysregulated in leukaemia.

A significant proportion, at least 25%, of the genes found in each list had been previously identified as discriminating genes for that subtype (table 12). For all subtypes, with the exception of T-ALL, this included the previously reported top ranking discriminatory genes, indicating consistency between platforms and different approaches, and providing positive controls. Up to four of the top five published genes for each subtype appeared in the corresponding list described here. The highest and lowest concordance was observed within the *ETV6-RUNX1* (56%), and the HeH (25%) subgroups, respectively. Nineteen of the top 20 HeH discriminatory genes had not been previously associated with this subtype, and the top 15 genes were not unique identifiers for HeH. In the previously published gene list only 10% of genes in the HeH group were under-expressed. In comparison, 67% of the top 100 genes identified from this analysis were down-regulated. Across all subtypes the percentage of under-expressed genes was much higher then anticipated from the published data. Additionally, the fold changes were generally much

higher in the published data, for example, the largest fold change observed in the published data was 7685.7 compared to 96.97 in this study.

Subgroup	Percentage of top 100 genes previously identified	P-value
ETV6-RUNX1	56	1.96E-21
BCR-ABL1	27	7.93E-07
TCF3-PBX1	38	6.10E-12
MLL rearranged	29	2.94E-07
HeH	25	1.66E-07
T-ALL	51	2.00E-21

Table 12. The percentage of genes in each of the top 100 lists that had previously been identified as discriminatory for that subgroup.

Hierarchical clustering

The expression profiles for each patient were clustered according to the similarity in the pattern of gene expression of the top 50 unique genes for each subgroup (50 genes provided better clustering then 100 genes) (figure 46). It was not possible to accurately cluster all patients from one subgroup together. In many cases the processing date had a greater influence on the expression profiles then the biological factors. Table 13 displays the processing date for each sample. All samples processed on 3.8.06 were subject to batch effects, and clustered regardless of the heterogeneous nature of the samples. Similarly, all samples processed on 30.4.07 were assigned to one of two clusters. Two sets of four NC patients were grouped according to process date (10.8.06 and 23.3.06). As the leukaemic subtype is unknown in these cases it is possible that they share the same genetic lesion.

Four cases (6845 (*ETV6-RUNX1*), 10073 (*ETV6-RUNX1*), 11210 (NC) and 11213 (NC)) were processed twice, on different days. The replicates clustered for a single case only (11213). Patient 6845 and 11210 were subject to batch effects and clustered with other samples that had been processed on the same array. Patient 10073 was grouped with another *ETV6-RUNX1* patient, and two 'other' patients. These four patients may all have common secondary abnormalities that have influenced clustering.

Four HeH and one 'other' patient, and both *BCR-ABL1* patients formed clusters. As they were processed on the same arrays it is not possible to discern if this is a batch or biological effect.

However, one NC and seven *ETV6-RUNX1* patients processed on four different dates were clustered together. Similarly, six HeH and five T-ALL patients formed subtype

specific clusters despite being processed on four different arrays. Two *TCF3-PBX1* patients, and one NC patient, all from different batches were also grouped together.

23.3.06	12.4.06	3.8.06	10.8.06	23.8.06	29.9.06	17.11.06	30.4.07
10073	6845	6845	2179	2058	8161	5972	6971
12883	7328	2586	5991	3043	10952	5957	7072
12884	10086	3043	10073		6948	8756	11234
12885	1333	3047	2328		10189	9497	11210
12886	2064		11208		10440	6074	11213
10028	6118		11210		10940	11739	7046
7152			11213		8431		7344
			11225		10186		7204
			11226		10817		6867
			11227		8988		
			11229		10283		
			11231		11050		
			11216		10430		
					10545		
					10655		

ETV6-RUNX1 *BCR-ABL1* *TCF3-PBX1* *MLL* rearranged
HeH T-ALL Other Non-classified

Table 13. The date of process for each of the patient samples. The cytogenetic subgroup of each patient is indicated by the colour of the cell.

BCR-ABL1 patients

The most discriminating gene for the *BCR-ABL1* patients was *DNAPTP6*, which was up-regulated 22.65 fold on average. However, this gene also appeared in the top 100 list for the T-ALL patients. Therefore, the top ranking gene unique to *BCR-ABL1* was *GRP110* (up-regulated 8.8 fold). Both genes appeared in the published top 100 discriminating genes for this subtype, at position 15 and 5 respectively. *ABL1*, together with various growth factor receptors (GFR), transcription factors (TF), cell surface marker and major histocompatibility (MHC) associated genes were dysregulated.

TCF3-PBX1 patients

In the *TCF3-PBX1* patients, *NID2* was the top ranking discriminatory gene (up regulated 30.03 fold). *PBX1* and *MERTK*, a tyrosine kinase that may be involved in leukemogenesis (8), were also over-expressed. All three genes had been previously identified as

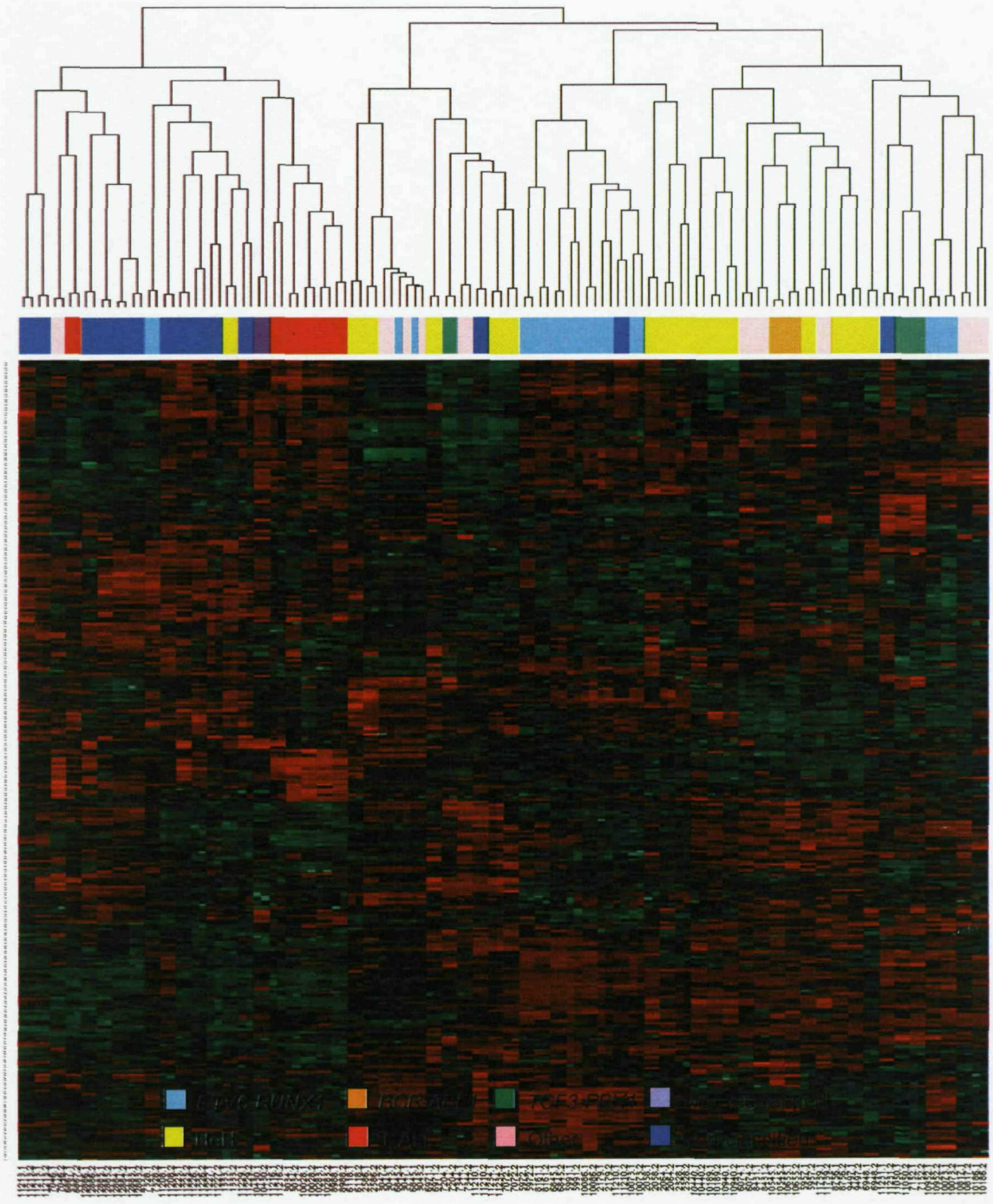


Figure 46. A 2-dimensional hierarchical cluster showing the expression levels of the top 50 ranked genes for each of the eight subgroups (rows) versus 58 ALL samples (columns). The normalized signal values for each gene were mean centered and divided by their standard deviations, and the values are represented by a colour. Red and green represent deviation above and below the mean respectively. The intensity of the colour corresponds to the magnitude of the deviation from the mean. The colored horizontal bar at the top of the heatmap indicates which samples belong to which subgroup.

discriminatory for this subtype. Other dysregulated genes included tyrosine kinase (TK) and TK receptors, GFRs and TFs.

MLL-rearranged patients

A single *MLL* rearranged case, harbouring a t(4;11) was analysed. This translocation results in the fusion of *MLL* and *AFF1*. Neither gene appeared in the top 100 list for this subtype. The most discriminatory gene was *MEIS* (up-regulated 96.97 fold), which was also the 4th highest ranking gene in the published gene list. In addition, three *HOX* genes (*HOXA5, HOXA9* and *HOXB5)* were over-expressed in this patient, although only *HOXA9* appeared in the published list for this subtype. *FLT3* was not up-regulated in this patient, although it did appear in the top 100 lists for the HeH and T-ALL patients.

HeH patients

The top 15 discriminatory genes for the HeH patients also appeared in the top 100 lists for other subgroups. The top ranking unique identifier was therefore *CSGlcA-T,* which was down-regulated 0.69 fold. This gene was similarly positioned in the published gene list for HeH patients. Yeoh *et al* (78) found that 70% of the class discriminating genes for this subtype were located on chromosomes X or 21, probably because these chromosomes are frequently gained in HeH. Although 14 HeH patients in this study had extra copies of chromosome 21, and 13 out of 15 patients had gains of X, only 2 of the top 100 genes were located on chromosome 21, and 15 were located on chromosome X.

T-ALL patients

The top ranking unique discriminator for the T-ALL patients was *CD3G* (position 3), which was up-regulated 9.79 fold. One patient had a *STIL-TAL1* fusion by FISH, and the *STIL* gene appeared in the top 100 list. Although over-expression of *TAL1* is associated with this fusion, neither gene was previously published in the top 100 list for these patients. None of the other genes that were rearranged in the T-ALL patients were in the top 100 list, although numerous cell surface markers, TFs, TKs and MHC associated genes were dysregulated.

ETV6-RUNX1 patients and comparison with aCGH data

The most discriminating gene for the *ETV6-RUNX1* patients was *HAP1*, which was up-regulated 32.44 fold. This gene appeared on the published list for this subtype at position 21.

The genomic locations of the top 100 discriminating genes, and of the most recurrent copy number alterations identified in *ETV6-RUNX1* positive patients by aCGH (refer to section 10.2.3.1), were compared. The top 100 genes were mapped to a chromosomal location, and copy number changes were looked for at these locations. A single down-

regulated gene (*CLECSF6*) was located on chromosome arm 12p, and two up-regulated genes (*APP* and *ERG*) were located on 21q, regions commonly deleted and gained in these patients, respectively. *EBF1,* the target of recurrent deletions in *ETV6-RUNX1* patients, was over-expressed 3.87 fold. It was also over-expressed in *TCF3-PBX1* patients, and down regulated in T-ALL and NC patients. Patients 10073 and 10086 had both aCGH and expression profiles. No correlation was observed between gene copy number and expression for patient 10073. In patient 10086 a single down-regulated gene on chromosome 1 (*DISC1*) was located within a deleted region, and the copy number gain on chromosome 21 encompassed the *APP* gene.

Two 'other' patients shared an additional chromosome 21 as the only cytogenetically visible abnormality (3043 and 10817). Array-based CGH was performed on patient 3043 (refer to section 10.4), confirming gain of chromosome 21 as the sole CNA. However, only a single gene in the top 100 list for this subtype was located on chromosome 21.

10.5.4 Discussion

Significant advances in the treatment of childhood ALL have been made in the past few decades. Improvements in classification, and risk-stratified treatment, guided by prognostic factors, have been central to these advances. With the development of high-throughput expression analysis using DNA-based microarrays in recent years, a number of studies have taken a genome wide expression approach to leukaemia classification. In 1999 Golub *et al* (204) used an oligonucleotide array approach to define a subset of 50 genes that could differentiate between ALL and AML patients. Subsequent studies have identified class discriminating genes that accurately distinguish between prognostically significant subtypes of ALL and AML, as well as predicting therapeutic outcome (8;44;78;201;205-207). In this study, customized expression arrays were designed, incorporating the most discriminatory genes published for each subtype, to assess the validity of using targeted arrays in the sub-classification of ALL.

Comparison with published dataset

A high degree of correlation was observed between the published top 100 gene lists for each subtype, and those described in this study, verifying that the choice of genes had been appropriate for the design of these customized arrays. The correlation may, in part, be affected by the number of patients analyzed from each subgroup; high concordance was observed between the *ETV6-RUNX1* lists (patients=9), compared to a much lower concordance between the *BCR-ABL1* lists (patients=2). However, the subgroup with the lowest concordance (HeH), had the largest patient cohort (n=16). The reason for the disparity between the HeH lists is unclear, although the batch effects observed during clustering may have played a significant role. Additionally, this subgroup may be more

heterogeneous then expected. Although specific chromosomes are commonly gained, different patients will have different combinations of chromosomal gains.
The much higher fold changes observed by Ross *et al* (8), across all subtypes, may result from a greater probe specificity and sensitivity of the Affymetrix arrays, leading to more efficient hybridization.

Hierarchical clustering

Hierarchical clustering, based on the top 50 unique genes for each subtype, had limited success. Overall, the batch effects were more influential then the molecular effect of the leukaemic subgroup, and technical replicates exhibited poor reproducibility. However, despite the low correlation between the top 100 gene lists for HeH patients, subtype specific clustering was observed for a subset of these patients, along with a subset of T-ALL, *TCF3-PBX1* and *ETV6-RUNX1* patients. Additionally two NC patients clustered with the *ETV6-RUNX1* and *TCF3-PBX1* patients, regardless of processing date. It is therefore possible that these NC patients were correctly sub-classified by their gene expression profiles, or at least have secondary abnormalities in common with the other patients in the same cluster. Ross *et al* (8) included 79 patients in their study, which they classified as hyperdiploidy with 47-50 chromosomes, hypodiploidy, normal or psuedodiploidy. Fourteen of these cases (which were not revealed) clustered together as a genetically distinct novel subtype. The remaining 65 cases (20% of the total number analyzed) had heterogeneous patterns of gene expression and did not cluster. As at least nine of the cases in this study could be classified in a similar way (the 'other' patients), their failure to cluster is unsurprising.
To test the accuracy with which patients could be sub-classified on the basis of their gene expression profile, two *ETV6-RUNX1* and HeH patients were removed from the training set and classification attempted. However, removing patients from the training set resulted in changes to the top 100 discriminatory genes. As the clustering capabilities observed using expression profiles generated from the maximum number of available patients were so poor, it seemed unlikely that these expression profiles could be used in this way, and this branch of the experiment was not pursued.

ETV6-RUNX1 patients and comparison with aCGH data

HAP1, the most discriminatory gene for the *ETV6-RUNX1* patients encodes a Huntingtin-associated protein involved in vesicular trafficking. Although this gene has no reported links to haematological malignancy, numerous other leukemia associated genes appeared in the top 100 list, including *EBF1*, *HOXA5*, *MYC*, *MME* and *CD44*.
Very little correlation was observed between the commonly gained and deleted regions in *ETV6-RUNX1* patients, and the genomic location of the top 100 genes. Similarly, the gain of chromosome 21 seen in two of the nine 'other' patients did not result in over-

expression of genes on chromosome 21. This may be expected, however, as only a subset of the patients included in this study would harbour the same recurrent CNA, and the biological effect of these aberrations may be diluted by the presence of other patients with normal copy number at these loci. Additionally, these copy number changes may not exert an effect on gene expression.

TCF3-PBX1 patients

The top ranking discriminatory gene for the *TCF3-PBX1* patients was *NID2*. The protein encoded by this gene, Nidogen-2, is a ubiquitous stabilizing component of the basement membrane. Loss of *NID2* expression, through aberrant DNA methylation, has been associated with a number of human cancers (208). However, the gene is over-expressed in association with this leukaemic subgroup.

Yeoh *et al* (78) were able to accurately diagnose *TCF3-PBX1* ALL based on the expression levels of *PBX1* alone. In the follow up study (8), this gene was over-expressed 3549.2 fold, compared to only 19 fold in this study. The raw data for the three patients was checked to determine if an anomalous value was skewing the results. The values were consistent across all cases. Up to four different probes for each gene are included on the Affymetrix arrays, compared to only one on the CodeLink arrays. A possible explanation for this discrepancy is that the selected CodeLink *PBX1* probe had a lower complementarity to the fusion transcript then the corresponding Affymetrix probes. To assist in the accurate sub-classification of leukaemias characterized by chimeric fusion genes, expression arrays would need to include probes for all known fusion transcript variants.

MLL- rearranged patients

In 2002, Armstrong *et al* (206) demonstrated that *MLL*-rearranged ALL patients have a gene expression profile distinct from both non *MLL*-rearranged ALL and AML cases. In a later study, Yeoh *et al* (78) distinguished *MLL*-rearranged ALL cases from six other subtypes of ALL. There was a high degree of correlation between the discriminating gene lists from these two studies. The subtle differences were attributed to the larger patient cohort used by Yeoh *et al* (78). In accordance with the published data, *MEIS* and *HOXA9* were exclusively expressed in the *MLL*- rearranged patient investigated in this study. MLL positively regulates *HOX* gene expression during development, and deregulation of *HOX* expression by MLL fusion proteins is thought to contribute to leukemogenesis (11). *HOXA9* is highly expressed in AML and its over-expression in mouse models induces leukaemia. It is also associated with a poor prognosis (206). MEIS is a co-factor for HOX proteins and can accelerate *HOXA9*-dependant leukemia (206).

An overall concordance of 29% was observed between the top 100 gene list identified in this study and the published data, including dysregulation of other *MLL* associated genes

such as *MME*, *HOXA5*, *CD44* and *TCF4*. Only a single *MLL*-rearranged case was available, accounting for the discrepancies with the published data. For example, although expression of *FLT3* is a frequent finding in association with *MLL*-rearrangements it was not seen in this one case.

T-ALL patients

In 2002, Ferrando *et al* (205) identified three molecularly distinct subtypes of T-ALL, based on expression patterns associated with the aberrant expression of the *TLX1, TAL1* or *LYL1* oncogenes. Yeoh *et al* (78) also observed heterogeneity within this subtype, as the T-ALL patients clustered into two groups. Subtle differences in the discriminating gene lists from these two studies were attributed to the higher density arrays used by Yeoh *et al* (78). Despite this heterogeneity, Yeoh *et al* (78) were able to accurately classify T-ALL cases on the basis of *CD3D* expression alone (over-expressed 35.7 fold), although it did not have the highest expression levels in the top 100, and 14 other genes shared the same chi-square value. In this study, *CD3D* was over-expressed 14.6 fold, and was the second highest ranking gene, whilst the top ranking unique gene was *CD3G*. Both genes are located within 9kb of each other on 11q23, and encode the CD3-gamma and –delta polypeptides, respectively. Both proteins, together with CD3-epsilon, and -zeta, form a complex with the TCR alpha/beta heterodimers (209). This complex plays an important role in coupling antigen recognition to several intracellular signal-transduction pathways.

Little is currently known about the most discriminatory genes for the *BCR-ABL1* and HeH subgroups (*GPR110* and *CSGlcA-T*, respectively).

Overall, the results of this study were interesting. Although the correlation between the published top 100 gene lists and those described here was highly significant in all cases, it was not possible to replicate the findings of previous studies (8;78), and differentiate between leukaemic subtypes on the basis of gene expression. Several factors may have contributed to this poor performance;

- Expression analysis was performed on significantly fewer patients in this investigation, compared to the study by Yeoh *et al* (58 versus 327) (78). In addition, nine of the patients here were cytogenetically diverse and did not belong to a defined subgroup. As no information was available on a further 12 patients, this number may be higher. Although it was not possible to further define these patients based on their expression profiles, Yeoh *et al* (71) also failed to classify 20% of their 'other' patients.
- A large degree of heterogeneity exists within the subtypes. Ross *et al* (8) found that *BCR-ABL1*, *TCF3-PBX1* and *MLL*-rearranged patients clustered in multiple

branches using 2D hierarchical clustering, and Ferrando *et al* (205) also demonstrated three molecularly distinct subtypes of T-ALL. In addition, each subgroup represents a genetically diverse population, in terms of age, sex and other features. These differences will result in variability in gene expression. To identify gene expression changes that are a consequence of the disease, comparison to constitutional samples from the same patients would be required.

- Technical difficulties were experienced during array processing. On a single array, processed using the same reaction conditions and reagents, some samples failed to produce fluorescent signals, whilst others generated only sporadic signals. Sample contamination was explored. Where possible, failed samples were reprocessed following ethanol precipitation before and after cRNA labeling. Successful labeling and hybridization was then achieved for some, but not all, samples. Limited material prevented further investigation of this problem. As described, in many cases the processing date had a greater effect on gene expression than the leukaemic subtype. As all arrays were processed by the same operator, using the same equipment and reagents, the reason for this is ambiguous, and reflects the unreliability of the technique for routine use, and the difficulties of working with RNA in general.
- Ross *et al* (8) suggested that defined expression profiles vary in their ability to sub-classify leukaemia in a bone marrow aspirate that is composed of less then 75% blasts, especially if the degree of expression change is low. Blast counts were available for 14 patients in this study, and were lower then 70% in two patients (one T-ALL and one HeH). The contaminating normal cells in these patients may have influenced their gene expression profiles.
- There is evidence that gene expression profiling is influenced by technical variability. For example, several other studies discussed here have attempted very similar analyses (8;78;205;206) and in all cases an Affymetrix platform was used. Whilst there is a high correlation between studies, differences were observed. Factors such as variations in sample preparation, the number of genes on the array, the number of samples investigated and the methods of analysis may all produce different results (78;204). Previous cross-validation studies have shown mixed results. Fioretos *et al* (210) concluded that cross-platform classification of childhood ALL and adult AML is feasible, and that diagnostic accuracy can be retained, whereas Stec *et al* (211) found that many genes with class-discriminating value on one profiling platform lose some of their discriminating value when measured by another profiling method. As this study differed from the original investigations in a number of key ways, including the choice of platform, the low correlation observed is perhaps unsurprising.

10.5.5 Conclusions

In conclusion, these data do not support the use of customized microarray based expression profiling as a single platform for the diagnosis and sub-classification of leukaemia. Although previously published studies were highly reproducible, and accurate sub-classification was achieved for patients with prognostically significant subtypes of leukaemia, a high proportion of patients did not belong to an established subgroup and thus could not be classified. As gene expression profiling is not applicable on an individual patient basis, accurate classification and therefore risk stratification cannot be guaranteed. The application of expression profiling in a diagnostic setting would not be cost or time effective. Additionally, to prevent technical variability influencing gene expression profiling, experimental approaches would need to be standardized across all laboratories, and local bioinformatics expertise would be essential. Yeoh *et al* (78) claimed that a single gene was sufficient for 100% diagnostic accuracy in T-ALL and *TCF3-PBX1* patients. Whilst this may be true for their limited cohort of patients, it is unrealistic to assume all cases will behave in the same way. If the expression levels of a single gene could be exploited to provide accurate diagnosis of leukaemic subtypes, gene expression profiling would no longer be needed, as more simple techniques such as qRT-PCR would be more appropriate. Despite these downfalls, expression arrays have provided insight into the molecular biology underlining different disease subtypes, and may prove useful in the prediction of relapse. In future they may provide a useful complementary tool, but for now standard diagnostic methods will not be replaced.

11: General Discussion

Paediatric ALL comprises numerous heterogeneous subgroups such as B-lineage ALL with the translocations, t(9;22)(*BCR–ABL1)*, t(1;19)(*TCF3–PBX1)* and t(12;21)(*ETV6–RUNX1)*, rearrangements of *MLL*, iAMP21, high hyperdiploid karyotypes, and T-lineage leukaemia (T-ALL). These genetically distinct subtypes differ in their cellular and molecular characteristics, as well as their response to therapy. Whilst these genetic aberrations are important in leukaemogenesis, and provide diagnostic and prognostic markers, co-operating oncogenic lesions are often required for a full leukaemic phenotype. A number of additional aberrations have been identified in association with some subgroups by cytogenetic analysis. However, it is the recent application of high resolution techniques to the study of the malignant genome that has significantly advanced our understanding of the mechanisms involved in disease initiation and progression. They have allowed the elucidation of pathways and genes dysregulated in this malignancy. Despite these advances, the full complement of cooperating lesions and their distribution within these subtypes of ALL remains to be defined (63).

In this study high resolution aCGH was applied to the study of *ETV6-RUNX1* positive, iAMP21 and unclassified acute leukaemia patients. *ETV6-RUNX1* positive ALL is one of the commonest forms of paediatric leukaemia and is associated with a good prognosis. It is well established that at least one secondary event is required to trigger the onset of leukemia in patients harbouring this translocation, but the nature of this additional aberration is currently undetermined. The genetic alterations responsible for the poor clinical outcome in a subset of these patients are also unknown. Intrachromosomal amplification of chromosome 21 is an emerging subgroup of leukaemia, seen in only 2% of B-ALL cases. Very few studies have attempted to characterise the genomes of these patients, and none have been performed at the resolution provided by the aCGH platform used in this investigation. Similarly, unclassified patients, which represent a heterogeneous subgroup lacking clinically relevant chromosomal abnormalities, are relatively understudied. This project aimed to provide further insight into the underlying genetic mechanisms of these disease phenotypes, and identify novel putative diagnostic and prognostic markers, and potential therapeutic targets. It also allowed an appraisal of the different techniques currently used in the diagnosis and characterization of leukaemic patients. Additionally, to attempt to validate the statement that gene expression profiling alone provides accurate diagnosis of genetic subtypes of ALL, customized expression arrays, targeting the known class discriminatory genes were designed. They were evaluated regarding their potential use in ALL sub-classification in routine clinical practice, and where possible correlation was made between copy number changes observed with aCGH, and gene expression data. Although this was a limited study, the

results indicated that gene expression profiling, using targeted arrays is not an appropriate method for leukaemia diagnosis and sub-classification, and that copy number changes do not necessarily equate to changes in gene expression levels. To date, no publications have validated the use of customized gene expression arrays in disease diagnosis and sub-classification, although this is currently being investigated in stage 2 of the MILE study (200).

In contrast, the aCGH studies were successful. Across all three patient cohorts, aCGH revealed a high level of genomic complexity, and identified a range of large and submicroscopic aberrations targeting multiple known and novel regions, highlighting a number of genes. The findings of this study correlated well with previously published data, confirming and refining the location, extent and gene targets of recurrent CNA in these patients. Some aberrations were observed in association with a particular subtype, whilst the majority were detected across all subtypes.

Common aberrations (table 14)

- Deletions of 9p, targeting *CDKN2A/B*, were observed in all subgroups. The frequency of this deletion was particularly high in the unclassified patients; 39%, compared to 31.5% and 20% in the iAMP21 and *ETV6-RUNX1* patients, respectively. This is higher than the reported frequency in ALL (134). The incidence of deletions involving *PAX5* and *ZCCHC7* were comparable across the subgroups.
- As expected, deletions of chromosome 12p, involving *ETV6* were observed at a very high frequency in *ETV6-RUNX1* positive patients (73.5%). However, 15.7% and 17% of iAMP21 and unclassified patients, respectively, also had similar deletions, highlighting the critical role of this gene in leukaemogenesis. In a number of cases the deletion breakpoints were within the gene. This raises the possibilities of novel *ETV6* fusion partners. Wiemels *et al* (212) have reported a non functional fusion between *ETV6* and non-coding chromosome 5 sequence in a single case.
- Aberrations involving chromosome 21 were observed in all subtypes. However, the origin of the observed copy number changes, for example, duplications and deletions, intrachromosomal rearrangements or translocations, varied between the subgroups. In total, gains of whole or partial chromosome 21 were observed in 26.4% and 17% of *ETV6-RUNX1* positive and unclassified patients, respectively, whereas all iAMP21 patients had complex chromosome 21 abnormalities, often involving gains, amplifications and deletions. Similar profiles were also observed in 7.3% of unclassified patients, suggesting that these two patient cohorts may belong to the same biological subgroup.

- Gain of whole or partial chromosome X was seen in a high frequency in both *ETV6-RUNX1* and iAMP21 patients (25.4% and 26.3%, respectively). This aberration has only recently been described in association with these subtypes (123;172;174) and has not been observed at high frequencies in other subgroups. Only 4.7% of unclassified patients had aberrations involving this chromosome.
- Deletions of chromosome 20 were detected in both iAMP21 and unclassified patients (15.7% and 10%, respectively). This aberration is rare in lymphoid malignancies and has not been previously associated with these genetic subtypes.
- An increased frequency of *IKZF1* deletions were identified in the iAMP21 patients (15.7%) compared with the *ETV6-RUNX1* (2.9%) patients. As the incidence of this deletion is high in adult ALL patients (18%) (213), this increase may be associated with patient age. IAMP21 is typically seen in older children, whereas the median age of *ETV6-RUNX1* patients in 3-4 years. The chromosome 7p deletions in 17% of unclassified patients were 0.5 Mb distal to this gene.
- Deletions targeting the histone gene family members on chromosome arm 6p, and the *ADD3* gene on chromosome arm 10q were observed, at similar frequencies, in *ETV6-RUNX1* positive and unclassified ALL patients only, whereas large scale deletions of 11q were detected in all but the unclassified subgroup. Deletions of *TOX* were not observed in iAMP21 patients, and the frequency of this deletion in *ETV6-RUNX1* positive patients was twice that in unclassified patients (14.7% versus 7.3%,respectively)

<u>*ETV6-RUNX1* patients (table 14)</u>

Deletions of *TBL1XR1* were observed in 23.5% of *ETV6-RUNX1* positive patients. Two previous studies have reported deletions of this gene in these patients, although at a lower frequency (3% (63) and 12% (64)). Loss of *TBL1XR1* has also been described in one HeH (63) and one unclassified case (section 10.4.4). Therefore, whilst this aberration is not unique to *ETV6-RUNX1* patients, it is highly associated with this subgroup. Deletions of *EBF1* were also more frequent in this group of patients (8.8% compared to 2.4% of unclassified patients and no iAMP21 patients). In addition deletions of 6q and *CEBPA* were restricted to these patients. The potential prognostic significance of the *ADD3* gene in *ETV6-RUNX1* patients was revealed. The frequency of this deletion was almost doubled in patients who relapsed, compared to those that remained in CCR (19.5% and 10.5%, respectively), indicating that this gene may be linked to a poor outcome in these patients.

<u>iAMP21 patients (table 14)</u>

Whilst this study revealed a number of chromosomal aberrations in these patients that had not previously been associated with this subtype (for example *CDKN2A/B* deletions), no novel recurrent CNA were identified. In addition, the hallmark complex CNA involving chromosome 21 were also observed in a subset of unclassified patients, suggesting an association between these patient cohorts or a number of unidentified iAMP21 patients within this group. The *PDE9A* gene emerged as potentially important in the pathogenesis of iAMP21 ALL. A high frequency of both *RB1* and *IKZF1* deletions were observed (15.7% (both genes)), compared with the *ETV6-RUNX1* (2.9% (both genes)) and unclassified patients (4.8% (*RB1*) and 0% (*IKZF1*)). Chromosome 20 deletions were also detected in a high proportion of cases. A number of aberrations targeting single genes were not observed in the iAMP21 patients, such as focal deletions of *TOX*, *ADD3*, *EBF1* and *CEBPA*. However, the array platform used for these patients had a lower probe density, which may account for the lack of detection.

Unclassified patients (table 14)

No distinguishing CNA were identified in this patient cohort, although the incidence of 9p deletion was much higher than typically observed in ALL.

Genomic aberration	**Patient subgroup**		
	***ETV6-RUNX1* positive**	**iAMP21**	**Unclassified**
Chr. 20		15.7%	10%
Chr. 21	26.4%	100%	17%
Chr. X	25.4%	26.3%	4.7%
6p deletion	11.7%		7.3%
6q deletion	32.3%		
9p deletion	20%	31.5%	39%
11q deletion	14.7%	26.3%	
12p deletion	73.5%	15.7%	17%
TBL1XR1 deletion	23.5%		2.4%
EBF1 deletion	8.8%		2.4%
IKZF1 deletion	2.9%	15.7%	
TOX deletion	14.7%		7.3%
ADD3 deletion	8.8%		12%
RB1 deletion	2.9%	15.7%	4.8%
CEBPA deletion	2.9%		

Table 14. A summary of the percentage of patients within each subgroup harbouring the recurrent genomic aberrations.

Copy number variants

As discussed in section 7.10, CNV are common in the normal population, and are important in genetic diversity (68;69). However, they may also be associated with disease susceptibility, response to treatment, or be directly linked to oncogenesis. As 1,447

regions of CNV have already been identified in normal individuals (70) the elucidation of these pathogenic loci will be a lengthy and difficult process.

A high number of CNV were detected in all patient subgroups, targeting 76 loci in total, including the *TCRA/D*, *TCRB*, *IGH@*, *IGK@* and *IGL@* gene loci (Appendix 8). An average of 10.4, 5 and 6.4 CNV were identified per patient in the *ETV6-RUNX1*, iAMP21 and unclassified groups, respectively. It is likely that the iAMP21 patients showed the lowest number of CNV because the array platform used in these cases had the lowest resolution. In a single iAMP21 patient and two unclassified patients, no CNV were detected. No CNV were found on chromosomes 12, 20, 21 or Y, in any subgroup. Four CNV; 4q (69.2-69.7 Mb), 7p (37.7-38.1 Mb), 15p (18.3-21.1 Mb) and 22q11.23 (22.5-22.9 Mb), were seen in at least 10% of patients from each group. A total of 17 genes are encompassed by these locations, and encode a number of transmembrane proteins, G-protein receptors, olfactory proteins and proteins associated with Prader-Willi Syndrome. The CNV located to chromosome 4q encompasses the *UGT2B17* gene, which is expressed in the prostate. UDP-glucuronosyltransferases (UGT) are a superfamily of detoxification enzymes that catalyze the glucuronidation of a variety of compounds, including steroid hormones, bilirubin, pharmaceuticals, and carcinogens. Deletion polymorphisms of this gene have previously been associated with prostate cancer, although the data from different studies is conflicting (72;214). They may also increase susceptibility to tobacco related cancers (215). Deletions of this locus were observed in a total of 23 patients, highlighting it as potentially important in the ALL disease phenotype. Although the CNV which were detected in patients from one subgroup only (*ETV6-RUNX1*; n=19, iAMP21; n=7, unclassified; n=19), were typically observed in a single patient, certain CNV were seen at much higher frequencies in particular patient groups. For example, the CNV at 8p12 (39.2-39.5 Mb) and 11q12.1 (54.9-55.5 Mb) were seen in 64.7% and 29.4% of *ETV6-RUNX1* patients, respectively, compared to only 5.2% of iAMP21 patients (both loci) and 19.5% (8q12) and 7.3% (11q12.1) of unclassified patients. No significant genes are encompassed by these CNV. Further studies are needed to determine whether CNV are correlated to particular genetic phenotypes and/or predisposition different leukaemic subtypes.

11.1 Future studies

In addition to those aberrations that targeted genes with an established and critical role in leukaemogenesis (for example *ETV6* and *RUNX1*), a large proportion of patients included in these studies showed alterations of genes that are key regulators of B-cell development and differentiation, (*EBF1*, *PAX5*, *IKZF1* (63;66)), or modulate G1/S cell cycle progression (*CDKN2A/B* and *RB1* (66)). The recurrent deletions of *TBL1XR1* in *ETV6-RUNX1* positive patients, and *TRIM24* in iAMP21 patients suggested that deregulation of key components of ligand inducible TF may also play a significant role in

leukaemogenesis. However, the role of other recurrently targeted genes, such as *ADD3*, *TOX* and the histone genes, in leukaemogenesis is not clear. The pathogenic effect of these various secondary lesions requires further investigation, which may help reveal novel therapeutic approaches for paediatric B-ALL.

Quantitative RT-PCR for the expression of *TBL1XR1* and its downstream genes in *ETV6-RUNX1* patients was inconclusive. Expanding this investigation with additional patients and different downstream genes, or using siRNA knockdown of *TBL1XR1* in a B-ALL cell line may be more informative, and highlight the role of this gene in the various biological pathways, such as apoptosis, differentiation or the cell cycle. This approach could also be applied to all the genes of interest, including *TOX.*

Deletions of *ADD3* were frequent in both *ETV6-RUNX1* positive patients and unclassified patients. Quantitative RT-PCR is required to determine the expression pattern of this gene, and its downstream targets. In addition, the potential prognostic importance of *ADD3* in the *ETV6-RUNX1* positive patients was revealed by FISH. To determine if deletions of this, and other highlighted genes, have clinical significance in other patient subgroups, similar FISH based studies of these patients should be performed.

To ascertain the clinical importance and pathogenic effect of *PDE9A* disruption in iAMP21 patients, a larger cohort of patients should be screened for this aberration using FISH, including relapsed patients to determine any association with prognosis. The value of FISH in revealing the presence of abnormalities involving this gene in subpopulations has been demonstrated. Novel disruption events could be identified with long-distance inverse PCR (LDI-PCR) and sequencing, including the presence of new *PDE9A* fusion partners. The effect of gene disruption on mRNA expression levels should be determined using qRT-PCR.
In addition, mutation analysis of conserved exons, and the methylation status of CpG islands in all genes of interest could be ascertained, to identify alternative mechanisms of gene inactivation. It would be of interest to determine the effect of gene disruption on protein expression using western blot analysis.

11.2 Conclusions

Both aCGH platforms used in these studies (Agilent 185k (iAMP21) and 244k (*ETV6-RUNX1* and unclassified)) provided high resolution maps of the genomic anomalies in these patients, although the slightly lower resolution array used in the study of the iAMP21 patients may have prevented the detection of focal deletions of single genes such as *ADD3* and *EBF1*. In comparison to the 1Mb BAC arrays previously used to investigate a subset of these patients (*ETV6-RUNX1*; n=4, iAMP21; n=7, unclassified;

n=6) (37;62), the Agilent platforms detected a total of 53 additional CNA, demonstrating the advances in this technology in just a few years, and the investigative power of the higher resolution arrays. This approach has provided further insight into the size, genomic position and gene content of CNA affecting *ETV6-RUNX1* positive, iAMP21 and unclassified ALL patients, and revealed an unexpectedly high level of genomic complexity. A number a novel targets and a potential prognostic indicator have been described, and many avenues for further study have been identified. As discussed in section 7.10, arrays with even higher probe densities are now available which will provide an even greater understanding of the mechanisms of disease pathogenesis. These arrays may eventually be superseded by the novel technique of high resolution sequencing, which is able to provide information on chromosomal rearrangements, mutations, SNP, copy number loss and copy number neutral loss of heterozygosity (CNNLOH) in a single experiment. However, this technique is still in the developmental stages and is prohibitively expensive. No single technique is currently able to fully characterize the plethora of aberrations that constitute the genomes of leukaemic patients. Therefore further elucidation of the importance of these aberrations, improvements in diagnosis and risk stratification, and the identification of novel therapeutic targets, will require a combination of both molecular and cytogenetic techniques.

12: References

1. Strachan T, Read A. DNA structure and gene expression. In: Strachan T, Read A. Human Molecular Genetics, Second ed. Oxford: BIOS Scientific Publishers LTD, 1999:1-26.

2. Strachan T, Read A. Structure and function of chromosomes. In: Strachan T, Read A. Human Molecular Genetics, Second ed. Oxford: BIOS Scientific Publishers LTD, 1999:27-53.

3. Pierotti M, Sozzi G, Croce C. Oncogenes. Cancer Medicine, sixth ed. Frei: BC Becker, 2003.

4. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc.Natl.Acad.Sci.U.S.A* 1971;68:820-3.

5. Goldsby R, Kindt T, Osbourne B. Cells and Organs of the Immune System. Immunology, Fourth ed. New York: W H Freeman and Company, 2000:27-59.

6. Eaves C, Eaves A. Anatomy and physiology of hematopoiesis. In: Pui CH. Childhood leukemias, Second ed. Cambridge University Press, 2006:69-105.

7. Zelent A, Greaves M, Enver T. Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia. *Oncogene* 2004;23:4275-83.

8. Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* 2003;102:2951-9.

9. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR et al. Proposals for the classification of the acute leukaemias. French- American-British (FAB) co-operative group. *British Journal of Haematology* 1976;33:451-8.

10. Onciu M, Pui CH. Diagnosis and classification. In: Pui CH. Childhood leukemias, Second ed. Cambridge University Press, 2006:21-47.

11. Raimondi SC. Cytogenetics of acute leukemias. In: Pui CH. Childhood leukemias, Second ed. Cambridge University Press, 2006:235-71.

12. Pui CH, Relling MV, Downing JR. Acute lymphoblastic leukemia. *N.Engl.J.Med.* 2004;350:1535-48.

13. Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A et al. Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). *Leukemia* 1995;9:1783-6.

14. Secker-Walker LM, Lawler SD, Hardisty RM. Prognostic implications of chromosomal findings in acute lymphoblastic leukaemia at diagnosis. *Br.Med.J.* 1978;2:1529-30.

15. Secker-Walker L. Clinical and biological features of acute lymphoblastic leukemia. Chromosomes and genes in acute lymphoblastic leukemia. R.G.Landes Company, 1997:1-11.

16. Pullarkat V, Slovak ML, Kopecky KJ, Forman SJ, Appelbaum FR. Impact of cytogenetics on the outcome of adult acute lymphoblastic leukemia: results of Southwest Oncology Group 9400 study. *Blood* 2008;111:2563-72.

17. Ferrando A, Rubnitz J, Look A. Molecular genetics of acute lymphoblastic leukaemia. In: Ching-Hon P. Childhood leukemias, Second ed. Cambridge University Press, 2006:272-97.

18. NOWELL PC, HUNGERFORD DA. Chromosome studies on normal and leukemic human leukocytes. *J.Natl.Cancer Inst.* 1960;25:85-109.

19. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 1973;243:290-3.

20. Secker-Walker L. Philadelphia positive acute leukemia t(9;22)(q34;q11). Chromosomes and genes in acute lymphoblastic leukemia. R.G.Landes Company, 1997:79-96.

21. Blume-Jensen P, Hunter T. Oncogenic kinase signalling. *Nature* 2001;411:355-65.

22. Saglio G, Cilloni D. Abl: the prototype of oncogenic fusion proteins. *Cell Mol Life Sci.* 2004;61:2897-911.

23. Specchia G, Albano F, Anelli L, Storlazzi CT, Zagaria A, Mancini M et al. Deletions on der(9) chromosome in adult Ph-positive acute lymphoblastic leukemia occur with a frequency similar to that observed in chronic myeloid leukemia. *Leukemia* 2003;17:528-31.

24. de LA, Rousselot P, Huguet-Rigal F, Delabesse E, Witz F, Maury S et al. Imatinib combined with induction or consolidation chemotherapy in patients with de novo Philadelphia chromosome-positive acute lymphoblastic leukemia - Results of the GRAAPH-2003 study. *Blood* 2006.

25. Schultz K, Bowman P, Slayton W, Aledo A, Devidas M, Sather H et al. Improved early event free survival (EFS) in children with Philadelphia chromosome-positive acute lymphoblastic leukaemia (ALL) with intensive Imatinib in combination with high dose chemotherapy: Children's Oncology Group (COG) study AALL0031. *Blood* 2007.

26. Romana SP, Mauchauffe M, Le CM, Chumakov I, Le PD, Berger R et al. The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood* 1995;85:3662-70.

27. Fears S, Vignon C, Bohlander SK, Smith S, Rowley JD, Nucifora G. Correlation between the ETV6/CBFA2 (TEL/AML1) fusion gene and karyotypic abnormalities in children with B-cell precursor acute lymphoblastic leukemia. *Genes Chromosomes.Cancer* 1996;17:127-35.

28. Rubnitz JE, Downing JR, Pui CH, Shurtleff SA, Raimondi SC, Evans WE et al. TEL gene rearrangement in acute lymphoblastic leukemia: a new genetic marker with prognostic significance. *J.Clin.Oncol.* 1997;15:1150-7.

29. Jabber Al-Obaidi MS, Martineau M, Bennett CF, Franklin IM, Goldstone AH, Harewood L et al. ETV6/AML1 fusion by FISH in adult acute lymphoblastic leukemia. *Leukemia* 2002;16:669-74.

30. Privitera E, Kamps MP, Hayashi Y, Inaba T, Shapiro LH, Raimondi SC et al. Different molecular consequences of the 1;19 chromosomal translocation in childhood B-cell precursor acute lymphoblastic leukemia. *Blood* 1992;79:1781-8.

31. Lu Q, Kamps MP. Selective repression of transcriptional activators by Pbx1 does not require the homeodomain. *Proc.Natl.Acad.Sci.U.S.A* 1996;93:470-4.

32. Lorsbach R, Downing JR. Molecular genetics of acute myeloid leukemia. In: Pui CH. Childhood leukemias, Second ed. Cambridge University Press, 2006:298-338.

33. Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G et al. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* 1999;354:1499-503.

34. Graux C, Cools J, Michaux L, Vandenberghe P, Hagemeijer A. Cytogenetics and molecular genetics of T-cell acute lymphoblastic leukemia: from thymocyte to lymphoblast. *Leukemia* 2006;20:1496-510.

35. Su XY, la-Valle V, ndre-Schmutz I, Lemercier C, Radford-Weiss I, Ballerini P et al. HOX11L2/TLX3 is transcriptionally activated through T-cell regulatory elements downstream of BCL11B as a result of the t(5;14)(q35;q32). *Blood* 2006;108:4198-201.

36. Wakabayashi Y, Watanabe H, Inoue J, Takeda N, Sakata J, Mishima Y et al. Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. *Nat.Immunol.* 2003;4:533-9.

37. Strefford JC, Van Delft FW, Robinson HM, Worley H, Yiannikouris O, Selzer R et al. Complex genomic alterations and gene expression in acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21. *Proc.Natl.Acad.Sci.U.S.A* 2006;103:8167-72.

38. Raimondi SC, Shurtleff SA, Downing JR, Rubnitz J, Mathew S, Hancock M et al. 12p abnormalities and the TEL gene (ETV6) in childhood acute lymphoblastic leukemia. *Blood* 1997;90:4559-66.

39. Harrison CJ, Moorman AV, Broadfield ZJ, Cheung KL, Harris RL, Reza JG et al. Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *Br.J.Haematol.* 2004;125:552-9.

40. Nachman JB, Heerema NA, Sather H, Camitta B, Forestier E, Harrison CJ et al. Outcome of treatment in children with hypodiploid acute lymphoblastic leukemia. *Blood* 2007;110:1112-5.

41. Buchner T, Hiddemann W, Berdel W, Wormann B, Schoch C, Loffler H et al. Acute myeloid leukemia: treatment over 60. *Rev.Clin.Exp.Hematol.* 2002;6:46-59.

42. Ebb DH, Weinstein HJ. Diagnosis and treatment of childhood acute myelogenous leukemia. *Pediatr.Clin North Am.* 1997;44:847-62.

43. Spector L, Ross J, Robison L, Bhatia S. Epidemiology and etiology. In: Pui C. Childhood leukaemias, Second ed. 2006:48-66.

44. Ross ME, Mahfouz R, Onciu M, Liu HC, Zhou X, Song G et al. Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* 2004;104:3679-87.

45. Chauffaille ML, Figueiredo MS, Beltrani R, Antunes SV, Yamamoto M, Kerbauy J. Acute promyelocytic leukemia: the study of t(15;17) translocation by fluorescent in situ hybridization, reverse transcriptase-polymerase chain reaction and cytogenetic techniques. *Braz.J.Med.Biol.Res* 2001;34:735-43.

46. Lo CF, Diverio D, Falini B, Biondi A, Nervi C, Pelicci PG. Genetic diagnosis and molecular monitoring in the management of acute promyelocytic leukemia. *Blood* 1999;94:12-22.

47. Baldus CD, Liyanarachchi S, Mrozek K, Auer H, Tanner SM, Guimond M et al. Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: Amplification discloses overexpression of APP, ETS2, and ERG genes. *Proc.Natl.Acad.Sci.U.S.A* 2004;101:3915-20.

48. Mrozek K, Heinonen K, Theil KS, Bloomfield CD. Spectral karyotyping in patients with acute myeloid leukemia and a complex karyotype shows hidden aberrations, including recurrent overrepresentation of 21q, 11q, and 22q. *Genes Chromosomes Cancer* 2002;34:137-53.

49. Swansbury J. Lymphoid disorders other that common acute lymphoblastic leukemia. In: Swansbury J. Cancer Cytogenetics. Methods and Protocols. Humana Press, 2003:93-110.

50. Boyiadzis M, Foon KA, Pavletic S. Hematopoietic stem cell transplantation for chronic lymphocytic leukemia: potential cure for an incurable disease. *Expert.Opin.Biol.Ther.* 2007;7:1789-97.

51. Stilgenbauer s, Lichter P, Dohner H. Genetic features of B-cell chronic lymphocytic leukemia. *Rev Clin Exp Hematol* 2000;4:48-72.

52. Sessions J. Chronic myeloid leukemia in 2007. *J.Manag.Care Pharm.* 2007;13:4-7.

53. Dispenzieri A, Kyle RA. Multiple myeloma: clinical features and indications for therapy. *Best.Pract.Res Clin.Haematol.* 2005;18:553-68.

54. Gutierrez NC, Garcia JL, Hernandez JM, Lumbreras E, Castellanos M, Rasillo A et al. Prognostic and biologic significance of chromosomal imbalances assessed by comparative genomic hybridization in multiple myeloma. *Blood* 2004;104:2661-6.

55. Hideshima T, Bergsagel PL, Kuehl WM, Anderson KC. Advances in biology of multiple myeloma: clinical applications. *Blood* 2004;104:607-18.

56. Smadja NV, Bastard C, Brigaudeau C, Leroux D, Fruchart C. Hypodiploidy is a major prognostic factor in multiple myeloma. *Blood* 2001;98:2229-38.

57. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992;258:818-21.

58. Pinkel D, Albertson DG. Comparative genomic hybridization. *annu rev genomics hum genet* 2005;6:331-54.

59. Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer. *Nat.Genet* 2005;37 Suppl:S11-S17.

60. Rosenberg C, Schut TB, Mostert MC, Tanke HJ, Raap AK, Oosterhuis JW et al. Comparative genomic hybridization in hypotriploid/hyperdiploid tumors. *Cytometry* 1997;29:113-21.

61. Kallioniemi A, Visakorpi T, Karhu R, Pinkel D, Kallioniemi OP. Gene Copy Number Analysis by Fluorescence in Situ Hybridization and Comparative Genomic Hybridization. *Methods* 1996;9:113-21.

62. Strefford JC, Worley H, Barber K, Wright S, Stewart ARM, Robinson H. Genome complexity in acute lymphoblastic leukaemia is revealed by array-based comparative genomic hybridization. *Oncogene* 2006.

63. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007;446:758-64.

64. Kawamata N, Ogawa S, Zimmermann M, Kato M, Sanada M, Hemminki K et al. Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* 2008;111:776-84.

65. Tsuzuki S, Karnan S, Horibe K, Matsumoto K, Kato K, Inukai T et al. Genetic abnormalities involved in t(12;21) TEL-AML1 acute lymphoblastic leukemia: analysis by means of array-based comparative genomic hybridization. *Cancer Sci.* 2007;98:698-706.

66. Kuiper RP, Schoenmakers EF, van Reijmersdal SV, Hehir-Kwa JY, van Kessel AG, van Leeuwen FN et al. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia* 2007;21:1258-66.

67. Shaffer LG, Bejjani BA. A cytogeneticist's perspective on genomic microarrays. *Hum.Reprod.Update.* 2004;10:221-6.

68. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P et al. Large-scale copy number polymorphism in the human genome. *Science* 2004;305:525-8.

69. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y et al. Detection of large-scale variation in the human genome. *Nat.Genet.* 2004;36:949-51.

70. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD et al. Global variation in copy number in the human genome. *Nature* 2006;444:444-54.

71. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T et al. Strong association of de novo copy number mutations with autism. *Science* 2007;316:445-9.

72. Gallagher CJ, Kadlubar FF, Muscat JE, Ambrosone CB, Lang NP, Lazarus P. The UGT2B17 gene deletion polymorphism and risk of prostate cancer: A case-control study in Caucasians. *Cancer Detect.Prev.* 2007;31:310-5.

73. Bouchardy C, Benhamou S, Jourenkova N, Dayer P, Hirvonen A. Metabolic genetic polymorphisms and susceptibility to lung cancer. *Lung Cancer* 2001;32:109-12.

74. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 2005;307:1434-40.

75. Nees M, Woodworth CD. Microarrays: spotlight on gene function and pharmacogenomics. *Curr.Cancer Drug Targets.* 2001;1:155-75.

76. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat.Genet.* 1999;21:20-4.

77. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engele P, McDonagh PD et al. Experimental annotation of the human genome using microarray technology. *Nature* 2001;409:922-7.

78. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002;1:133-43.

79. Harrison CJ, Martineau M, Secker-Walker LM. The Leukaemia Research Fund/United Kingdom Cancer Cytogenetics Group Karyotype Database in acute lymphoblastic leukaemia: a valuable resource for patient management. *Br.J.Haematol.* 2001;113:3-10.

80. Daser A, Thangavelu M, Pannell R, Forster A, Sparrow L, Chung G et al. Interrogation of genomes by molecular copy-number counting (MCC). *Nat.Methods* 2006;3:447-53.

81. Corbett Research. HRM Assay Design and Analysis. 1-9-2006. Ref Type: Generic

82. Holmquist GP, Comings DE. Histones and G banding of chromosomes. *Science* 1976;193:599-602.

83. Benn P, Perle M. Chromosome staining and banding techniqes. In: Rooney D, Czepulkowski B. Human cytogenetics I. Constitutional analysis, Second ed. Oxford University Press, 1992:91-118.

84. Shaffer L, Tommerup Ne. An International System for Human Cytogenetic Nomenclature. Basel: S.Karger, 2005.

85. Burk KH, Drewinko B, Turjillo JM, Ahearn MJ. Establishment of a human plasma cell line in vitro. *cancer res* 1978;38:2508-13.

86. Drucker L, Tohami T, Tartakover-Matalon S, Zismanov V, Shapiro H, Radnay J et al. Promoter hypermethylation of tetraspanin members contributes to their silencing in myeloma cell lines. *carcinogenesis* 2006;27:197-204.

87. Iyer R, Ding L, Batchu RB, Naugler S, Shammas MA, Munshi NC. Antisense p53 transduction leads to overexpression of bcl-2 and dexamethasone resistance in multiple myeloma. *Leuk.Res* 2003;27:73-8.

88. Gronich N, Drucker L, Shapiro H, Radnay J, Yarkoni S, Lishner M. Simvastatin induces death of multiple myeloma cell lines. *J Investig.Med* 2004;52:335-44.

89. Drexler HG, Dirks WG, MacLeod RA. False human hematopoietic cell lines: cross-contaminations and misinterpretations. *Leukemia* 1999;13:1601-7.

90. Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R et al. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc.Natl.Acad.Sci.U.S.A* 2004;101:17765-70.

91. Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P et al. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer* 2005;44:305-19.

92. Gunn SR, Mohammed M, Reveles XT, Viskochil DH, Palumbos JC, Johnson-Pais TL et al. Molecular characterization of a patient with central nervous system dysmyelination and cryptic unbalanced translocation between chromosomes 4q and 18q. *Am.J.Med.Genet.A* 2003;120:127-35.

93. Greshock J, Feng B, Nogueira C, Ivanova E, Perna I, Nathanson K et al. A comparison of DNA copy number profiling platforms. *Cancer Res.* 2007;67:10173-80.

94. Kurokawa M, Hirai H. Role of AML1/Runx1 in the pathogenesis of hematological malignancies. *Cancer Sci.* 2003;94:841-6.

95. Fischle W, Wang Y, Allis CD. Histone and chromatin cross-talk. *Curr.Opin.Cell Biol.* 2003;15:172-83.

96. Yoon HG, Chan DW, Huang ZQ, Li J, Fondell JD, Qin J et al. Purification and functional characterization of the human N-CoR complex: the roles of HDAC3, TBL1 and TBLR1. *EMBO J.* 2003;22:1336-46.

97. Grunstein M. Histone acetylation in chromatin structure and transcription. *Nature* 1997;389:349-52.

98. Fenrick R, Amann JM, Lutterbach B, Wang L, Westendorf JJ, Downing JR et al. Both TEL and AML-1 contribute repression domains to the t(12;21) fusion protein. *Mol.Cell Biol.* 1999;19:6566-74.

99. Imai Y, Kurokawa M, Tanaka K, Friedman AD, Ogawa S, Mitani K et al. TLE, the human homolog of groucho, interacts with AML1 and acts as a repressor of AML1-induced transactivation. *Biochem.Biophys.Res.Commun.* 1998;252:582-9.

100. Kitabayashi I, Yokoyama A, Shimizu K, Ohki M. Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *EMBO J.* 1998;17:2994-3004.

101. Erickson P, Gao J, Chang KS, Look T, Whisenant E, Raimondi S et al. Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood* 1992;80:1825-31.

102. Gamou T, Kitamura E, Hosoda F, Shimizu K, Shinohara K, Hayashi Y et al. The partner gene of AML1 in t(16;21) myeloid malignancies is a novel member of the MTG8(ETO) family. *Blood* 1998;91:4028-37.

103. Jousset C, Carron C, Boureux A, Quang CT, Oury C, Dusanter-Fourt I et al. A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerization interface essential to the mitogenic properties of the TEL-PDGFR beta oncoprotein. *EMBO J.* 1997;16:69-82.

104. Guidez F, Petrie K, Ford AM, Lu H, Bennett CA, MacGregor A et al. Recruitment of the nuclear receptor corepressor N-CoR by the TEL moiety of the childhood leukemia-associated TEL-AML1 oncoprotein. *Blood* 2000;96:2557-61.

105. Golub TR, Barker GF, Lovett M, Gilliland DG. Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* 1994;77:307-16.

106. Papadopoulos P, Ridge SA, Boucher CA, Stocking C, Wiedemann LM. The novel activation of ABL by fusion to an ets-related gene, TEL. *Cancer Res.* 1995;55:34-8.

107. Lacronique V, Boureux A, Valle VD, Poirel H, Quang CT, Mauchauffe M et al. A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science* 1997;278:1309-12.

108. Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. *Nat.Rev.Cancer* 2003;3:639-49.

109. Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol.Cell Biol.* 1996;16:2164-73.

110. Wiemels JL, Alexander FE, Cazzaniga G, Biondi A, Mayer SP, Greaves M. Microclustering of TEL-AML1 translocation breakpoints in childhood acute lymphoblastic leukemia. *Genes Chromosomes.Cancer* 2000;29:219-28.

111. Wiemels JL, Ford AM, Van Wering ER, Postma A, Greaves M. Protracted and variable latency of acute lymphoblastic leukemia after TEL-AML1 gene fusion in utero. *Blood* 1999;94:1057-62.

112. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. *Blood* 2003;102:2321-33.

113. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc.Natl.Acad.Sci.U.S.A* 2002;99:8242-7.

114. Andreasson P, Schwaller J, Anastasiadou E, Aster J, Gilliland DG. The expression of ETV6/CBFA2 (TEL/AML1) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. *Cancer Genet.Cytogenet.* 2001;130:93-104.

115. Bernardin F, Yang Y, Cleaves R, Zahurak M, Cheng L, Civin CI et al. TEL-AML1, expressed from t(12;21) in human acute lymphocytic leukemia, induces acute leukemia in mice. *Cancer Res.* 2002;62:3904-8.

116. Alvarez Y, Coll MD, Ortega JJ, Bastida P, Dastugue N, Robert A et al. Genetic abnormalities associated with the t(12;21) and their impact in the outcome of 56 patients with B-precursor acute lymphoblastic leukemia. *Cancer Genet.Cytogenet.* 2005;162:21-9.

117. Romana SP, Le CM, Poirel H, Marynen P, Bernard O, Berger R. Deletion of the short arm of chromosome 12 is a secondary event in acute lymphoblastic leukemia with t(12;21). *Leukemia* 1996;10:167-70.

118. Maia AT, Ford AM, Jalali GR, Harrison CJ, Taylor GM, Eden OB et al. Molecular tracking of leukemogenesis in a triplet pregnancy. *Blood* 2001;98:478-82.

119. Ford AM, Fasching K, Panzer-Grumayer ER, Koenig M, Haas OA, Greaves MF. Origins of "late" relapse in childhood acute lymphoblastic leukemia with TEL-AML1 fusion genes. *Blood* 2001;98:558-64.

120. Loh ML, Rubnitz JE. TEL/AML1-positive pediatric leukemia: prognostic significance and therapeutic approaches. *Curr.Opin.Hematol.* 2002;9:345-52.

121. Ramakers-Van Woerden NL, Pieters R, Loonen AH, Hubeek I, van DE, Beverloo HB et al. TEL/AML1 gene fusion is related to in vitro drug sensitivity for L-asparaginase in childhood acute lymphoblastic leukemia. *Blood* 2000;96:1094-9.

122. Stams WA, Beverloo HB, den Boer ML, de Menezes RX, Stigter RL, van DE et al. Incidence of additional genetic changes in the TEL and AML1 genes in DCOG and COALL-treated t(12;21)-positive pediatric ALL, and their relation with drug sensitivity and clinical outcome. *Leukemia* 2006;20:410-6.

123. Lilljebjorn H, Heidenblad M, Nilsson B, Lassen C, Horvat A, Heldrup J et al. Combined high-resolution array-based comparative genomic hybridization and expression profiling of ETV6/RUNX1-positive acute lymphoblastic leukemias

reveal a high incidence of cryptic Xq duplications and identify several putative target genes within the commonly gained region. *Leukemia* 2007;21:2137-44.

124. Ma SK, Wan TS, Cheuk AT, Fung LF, Chan GC, Chan SY et al. Characterization of additional genetic events in childhood acute lymphoblastic leukemia with TEL/AML1 gene fusion: a molecular cytogenetics study. *Leukemia* 2001;15:1442-7.

125. Raynaud SD, Dastugue N, Zoccola D, Shurtleff SA, Mathew S, Raimondi SC. Cytogenetic abnormalities associated with the t(12;21): a collaborative study of 169 children with t(12;21)-positive acute lymphoblastic leukemia. *Leukemia* 1999;13:1325-30.

126. Raynaud S, Cave H, Baens M, Bastard C, Cacheux V, Grosgeorge J et al. The 12;21 translocation involving TEL and deletion of the other TEL allele: two frequently associated alterations found in childhood acute lymphoblastic leukemia. *Blood* 1996;87:2891-9.

127. Cave H, Cacheux V, Raynaud S, Brunie G, Bakkus M, Cochaux P et al. ETV6 is the target of chromosome 12p deletions in t(12;21) childhood acute lymphocytic leukemia. *Leukemia* 1997;11:1459-64.

128. Takeuchi S, Seriu T, Bartram CR, Golub TR, Reiter A, Miyoshi I et al. TEL is one of the targets for deletion on 12p in many cases of childhood B-lineage acute lymphoblastic leukemia. *Leukemia* 1997;11:1220-3.

129. Latil A, Guerard M, Berthon P, Cussenot O. 12p12-13 deletion in prostate tumors and quantitative expression of CDKN1B and ETV6 candidate genes. *Genes Chromosomes.Cancer* 2001;31:199-200.

130. Zhang L, Parkhurst JB, Kern WF, Scott KV, Niccum D, Mulvihill JJ et al. Chromosomal changes detected by fluorescence in situ hybridization in patients with acute lymphoblastic leukemia. *Chin Med.J.(Engl.)* 2003;116:1298-303.

131. Heerema NA, Raimondi SC, Anderson JR, Biegel J, Camitta BM, Cooley LD et al. Specific extra chromosomes occur in a modal number dependent pattern in pediatric acute lymphoblastic leukemia. *Genes Chromosomes.Cancer* 2007;46:684-93.

132. Cheon MS, Dierssen M, Kim SH, Lubec G. Protein expression of BACE1, BACE2 and APP in Down syndrome brains. *Amino.Acids* 2007.

133. Tanzi RE, Bertram L. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell* 2005;120:545-55.

134. van Zutven LJ, van DE, de Bont JM, Wattel MM, den Boer ML, Pieters R et al. CDKN2 deletions have no prognostic value in childhood precursor-B acute lymphoblastic leukaemia. *Leukemia* 2005;19:1281-4.

135. Mirebeau D, Acquaviva C, Suciu S, Bertin R, Dastugue N, Robert A et al. The prognostic significance of CDKN2A, CDKN2B and MTAP inactivation in B-lineage acute lymphoblastic leukemia of childhood. Results of the EORTC studies 58881 and 58951. *Haematologica* 2006;91:881-5.

136. Bousquet M, Broccardo C, Quelen C, Meggetto F, Kuhlein E, Delsol G et al. A novel PAX5-ELN fusion protein identified in B-cell acute lymphoblastic leukemia acts as a dominant negative on wild-type PAX5. *Blood* 2007;109:3417-23.

137. Nera KP, Lassila O. Pax5--a critical inhibitor of plasma cell fate. *Scand.J.Immunol.* 2006;64:190-9.

138. Busslinger M, Klix N, Pfeffer P, Graninger PG, Kozmik Z. Deregulation of PAX-5 by translocation of the Emu enhancer of the IgH locus adjacent to two alternative PAX-5 promoters in a diffuse large-cell lymphoma. *Proc.Natl.Acad.Sci.U.S.A* 1996;93:6129-34.

139. Strehl S, Konig M, Dworzak MN, Kalwak K, Haas OA. PAX5/ETV6 fusion defines cytogenetic entity dic(9;12)(p13;p13). *Leukemia* 2003;17:1121-3.

140. Hystad ME, Myklebust JH, Bo TH, Sivertsen EA, Rian E, Forfang L et al. Characterization of early stages of human B cell development by gene expression profiling. *J.Immunol.* 2007;179:3662-71.

141. Bertrand P, Bastard C, Maingonnat C, Jardin F, Maisonneuve C, Courel MN et al. Mapping of MYC breakpoints in 8q24 rearrangements involving non-immunoglobulin partners in B-cell lymphomas. *Leukemia* 2007;21:515-23.

142. Jackson A, Carrara P, Duke V, Sinclair PB, Papaioannou M, Harrison CJ et al. Deletion of 6q16-21 in human lymophoid malignancies: a mapping and deletion analysis. *cancer res* 2000;60:2775-9.

143. Saijo K, Schmedt C, Su IH, Karasuyama H, Lowell CA, Reth M et al. Essential role of Src-family protein tyrosine kinases in NF-kappaB activation during B cell development. *Nat.Immunol.* 2003;4:274-9.

144. Ray ME, Wistow G, Su YA, Meltzer PS, Trent JM. AIM1, a novel non-lens member of the betagamma-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. *Proc.Natl.Acad.Sci.U.S.A* 1997;94:3229-34.

145. Sinclair PB, Sorour A, Martineau M, Harrison CJ, Mitchell WA, O'Neill E et al. A fluorescence in situ hybridization map of 6q deletions in acute lymphocytic leukemia: identification and analysis of a candidate tumor suppressor gene. *cancer res* 2004;64:4089-98.

146. Alami R, Fan Y, Pack S, Sonbuchner TM, Besse A, Lin Q et al. Mammalian linker-histone subtypes differentially affect gene expression in vivo. *Proc.Natl.Acad.Sci.U.S.A* 2003;100:5920-5.

147. Barber KE, Ford AM, Harris RL, Harrison CJ, Moorman AV. MLL translocations with concurrent 3' deletions: interpretation of FISH results. *Genes Chromosomes.Cancer* 2004;41:266-71.

148. mare Kadam PS, Raje GC, Pais AP, Banavali S. Coexistence of ETV6/RUNX1 and MLL aberrations in B-cell precursor acute lymphoblastic leukemia discloses a small subclass of BCP-ALL. *Cancer Genet Cytogenet.* 2008;182:27-32.

149. Schaffner C, Stilgenbauer S, Rappold GA, Dohner H, Lichter P. Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. *Blood* 1999;94:748-53.

150. Gumy PF, Wacker P, Maillet P, Betts D, Sappino AP. ATM gene alterations in childhood acute lymphoblastic leukemias. *Hum.Mutat.* 2003;21:554.

151. Kovacs BZ, Niggli FK, Betts DR. Aberrations involving 13q12 approximately q14 are frequent secondary events in childhood acute lymphoblastic leukemia. *Cancer Genet.Cytogenet.* 2004;151:157-61.

152. Wilkinson B, Chen JY, Han P, Rufner KM, Goularte OD, Kaye J. TOX: an HMG box protein implicated in the regulation of thymocyte selection. *Nat.Immunol.* 2002;3:272-80.

153. Manunta P, Bianchi G. Pharmacogenomics and pharmacogenetics of hypertension: update and perspectives--the adducin paradigm. *J.Am.Soc.Nephrol.* 2006;17:S30-S35.

154. Lahortiga I, Vizmanos JL, Agirre X, Vazquez I, Cigudosa JC, Larrayoz MJ et al. NUP98 is fused to adducin 3 in a patient with T-cell acute lymphoblastic leukemia and myeloid markers, with a new translocation t(10;11)(q25;p15). *Cancer Res.* 2003;63:3079-83.

155. Akasaka T, Balasas T, Russell LJ, Sugimoto KJ, Majid A, Walewska R et al. Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). *Blood* 2007;109:3451-61.

156. Georgopoulos K, Bigby M, Wang JH, Molnar A, Wu P, Winandy S et al. The Ikaros gene is required for the development of all lymphoid lineages. *Cell* 1994;79:143-56.

157. Winandy S, Wu P, Georgopoulos K. A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma. *Cell* 1995;83:289-99.

158. Zhang XM, Chang Q, Zeng L, Gu J, Brown S, Basch RS. TBLR1 regulates the expression of nuclear hormone receptor co-repressors. *BMC.Cell Biol.* 2006;7:31.

159. Zhang J, Kalkum M, Chait BT, Roeder RG. The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2. *Mol.Cell* 2002;9:611-23.

160. Perissi V, Aggarwal A, Glass CK, Rose DW, Rosenfeld MG. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell* 2004;116:511-26.

161. Stacey MW, Wang J, Byrd RL, Liu JM, Kearns WG. Nuclear receptor co-repressor gene localizes to 17p11.2, a frequently deleted band in malignant disorders. *Genes Chromosomes.Cancer* 1999;25:191-3.

162. Hauksdottir H, Farboud B, Privalsky ML. Retinoic acid receptors beta and gamma do not repress, but instead activate target gene transcription in both the absence and presence of hormone ligand. *Mol.Endocrinol.* 2003;17:373-85.

163. Zhuang Y, Faria TN, Chambon P, Gudas LJ. Identification and characterization of retinoic acid receptor beta2 target genes in F9 teratocarcinoma cells. *Mol.Cancer Res.* 2003;1:619-30.

164. Taneja R, Bouillet P, Boylan JF, Gaub MP, Roy B, Gudas LJ et al. Reexpression of retinoic acid receptor (RAR) gamma or overexpression of RAR alpha or RAR beta in RAR gamma-null F9 cells reveals a partial functional redundancy between the three RAR types. *Proc.Natl.Acad.Sci.U.S.A* 1995;92:7854-8.

165. Fazi F, Zardo G, Gelmetti V, Travaglini L, Ciolfi A, Di CL et al. Heterochromatic gene repression of the retinoic acid pathway in acute myeloid leukemia. *Blood* 2007;109:4432-40.

166. Reineke EL, Liu H, Lam M, Liu Y, Kao HY. Aberrant association of promyelocytic leukemia protein-retinoic acid receptor-alpha with coactivators contributes to its ability to regulate gene expression. *J.Biol.Chem.* 2007;282:18584-96.

167. Lane MA, Xu J, Wilen EW, Sylvester R, Derguini F, Gudas LJ. LIF removal increases CRABPI and CRABPII transcripts in embryonic stem cells cultured in retinol or 4-oxoretinol. *Mol.Cell Endocrinol.* 2008;280:63-74.

168. Tanaka K, Imoto I, Inoue J, Kozaki K, Tsuda H, Shimada Y et al. Frequent methylation-associated silencing of a candidate tumor-suppressor, CRABP1, in esophageal squamous-cell carcinoma. *Oncogene* 2007;26:6456-68.

169. Park J, Chae Y, Son C. Epigenetic silencing of human T (brachyury homologue) gene in non-small-cell lung cancer. *Biochem Biophys Res Commun* 2008;365:221-6.

170. Guidez F, Parks S, Wong H, Jovanovic JV, Mays A, Gilkes AF et al. RARalpha-PLZF overcomes PLZF-mediated repression of CRABPI, contributing to retinoid resistance in t(11;17) acute promyelocytic leukemia. *Proc.Natl.Acad.Sci.U.S.A* 2007;104:18694-9.

171. Gupta A, Williams BR, Hanash SM, Rawwas J. Cellular retinoic acid-binding protein II is a direct transcriptional target of MycN in neuroblastoma. *Cancer Res.* 2006;66:8100-8.

172. Robinson HM, Harrison CJ, Moorman AV, Chudoba I, Strefford JC. Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle. *Genes Chromosomes Cancer* 2007;46:318-26.

173. Busson-Le CM, Nguyen KF, Daniel MT, Bernard OA, Berger R. Chromosome 21 abnormalities with AML1 amplification in acute lymphoblastic leukemia. *Genes Chromosomes.Cancer* 2001;32:244-9.

174. Harewood L, Robinson H, Harris R, Al-Obaidi MJ, Jalali GR, Martineau M et al. Amplification of AML1 on a duplicated chromosome 21 in acute lymphoblastic leukemia: a study of 20 cases. *Leukemia* 2003;17:547-53.

175. Soulier J, Trakhtenbrot L, Najfeld V, Lipton JM, Mathew S, vet-Loiseau H et al. Amplification of band q22 of chromosome 21, including AML1, in older children with acute lymphoblastic leukemia: an emerging molecular cytogenetic subgroup. *Leukemia* 2003;17:1679-82.

176. Moorman AV, Richards SM, Robinson HM, Strefford JC, Gibson BE, Kinsey SE et al. Prognosis of children with acute lymphoblastic leukemia (ALL) and intrachromosomal amplification of chromosome 21 (iAMP21). *Blood* 2007;109:2327-30.

177. Kuchinskaya E, Nordgren A, Heyman M, Schoumans J, Corcoran M, Staaf J et al. Tiling-resolution array-CGH reveals the pattern of DNA copy number alterations in acute lymphoblastic leukemia with 21q amplification: the result of telomere dysfunction and breakage/fusion/breakage cycles? *Leukemia* 2007;21:1327-30.

178. Curtiss NP, Bonifas JM, Lauchle JO, Balkman JD, Kratz CP, Emerling BM et al. Isolation and analysis of candidate myeloid tumor suppressor genes from a commonly deleted segment of 7q22. *Genomics* 2005;85:600-7.

179. Brezinova J, Zemanova Z, Ransdorfova S, Sindelarova L, Siskova M, Neuwirtova R et al. Prognostic significance of del(20q) in patients with hematological malignancies. *Cancer Genet Cytogenet.* 2005;160:188-92.

180. Fioretos T, Strombeck B, Sandberg T, Johansson B, Billstrom R, Borg A et al. Isochromosome 17q in blast crisis of chronic myeloid leukemia and in other hematologic malignancies is the result of clustered breakpoints in 17p11 and is not associated with coding TP53 mutations. *Blood* 1999;94:225-32.

181. Cunningham LA, Kahn RA. Cofactor D functions as a centrosomal protein and is required for the recruitment of the gamma -tubulin ring complex at centrosomes and organization of the mitotic spindle. *J Biol.Chem.* 2008.

182. Forestier E, Izraeli S, Beverloo B, Haas O, Pession A, Michalova K et al. Cytogenetic features of acute lymphoblastic and myeloid leukemias in pediatric patients with Down syndrome: an iBFM-SG study. *Blood* 2008;111:1575-83.

183. Nomura Y, Adachi N, Koyama H. Human Mus81 and FANCB independently contribute to repair of DNA damage during replication. *Genes Cells* 2007;12:1111-22.

184. Pamidi A, Cardoso R, Hakem A, Matysiak-Zablocki E, Poonepalli A, Tamblyn L et al. Functional interplay of p53 and Mus81 in DNA damage responses and cancer. *cancer res* 2007;67:8527-35.

185. McPherson JP, Lemmers B, Chahwan R, Pamidi A, Migon E, Matysiak-Zablocki E et al. Involvement of mammalian Mus81 in genome integrity and tumor suppression. *Science* 2004;304:1822-6.

186. Lillicrap D. Von Willebrand disease - phenotype versus genotype: deficiency versus disease. *Thromb.Res* 2007;120 Suppl 1:S11-S16.

187. Wunder F, Tersteegen A, Rebmann A, Erb C, Fahrig T, Hendrix M. Characterization of the first potent and selective PDE9 inhibitor using a cGMP reporter cell line. *Mol.Pharmacol.* 2005;68:1775-81.

188. Fisher DA, Smith JF, Pillar JS, St Denis SH, Cheng JB. Isolation and characterization of PDE9A, a novel human cGMP-specific phosphodiesterase. *J.Biol.Chem.* 1998;273:15559-64.

189. Guipponi M, Scott HS, Kudoh J, Kawasaki K, Shibuya K, Shintani A et al. Identification and characterization of a novel cyclic nucleotide phosphodiesterase gene (PDE9A) that maps to 21q22.3: alternative splicing of mRNA transcripts, genomic structure and sequence. *Hum.Genet.* 1998;103:386-92.

190. Rentero C, Monfort A, Puigdomenech P. Identification and distribution of different mRNA variants produced by differential splicing in the human phosphodiesterase 9A gene. *Biochem.Biophys.Res Commun.* 2003;301:686-92.

191. Wang P, Wu P, Egan RW, Billah MM. Identification and characterization of a new human type 9 cGMP-specific phosphodiesterase splice variant (PDE9A5). Differential tissue distribution and subcellular localization of PDE9A variants. *Gene* 2003;314:15-27.

192. Andreeva SG, Dikkes P, Epstein PM, Rosenberg PA. Expression of cGMP-specific phosphodiesterase 9A mRNA in the rat brain. *J.Neurosci.* 2001;21:9068-76.

193. Staveren W, Glick j, Markerink-van Ittersum M, Shimizu M, Beavo J, Steinbusch H et al. Cloning and localization of the cGMP-specific phosphodiesterase type 9 in the rat brain. *J Neurocytol* 2002;31:729-41.

194. Rentero C, Puigdomenech P. Specific use of start codons and cellular localization of splice variants of human phosphodiesterase 9A gene. *BMC.Mol.Biol.* 2006;7:39.

195. Huai Q, Wang H, Zhang W, Colman RW, Robinson H, Ke H. Crystal structure of phosphodiesterase 9 shows orientation variation of inhibitor 3-isobutyl-1-methylxanthine binding. *Proc.Natl.Acad.Sci.U.S.A* 2004;101:9624-9.

196. Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T et al. A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res.* 2004;14:247-66.

197. Mancini M, Scappaticci D, Cimino G, Nanni M, Derme V, Elia L et al. A comprehensive genetic classification of adult acute lymphoblastic leukemia (ALL): analysis of the GIMEMA 0496 protocol. *Blood* 2005;105:3434-41.

198. Moorman AV, Harrison CJ, Buck G, Richards S, Secker-Walker LM, Martineau M et al. Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. *Blood* 2007;109:3189-97.

199. Kuchinskaya E, Heyman M, Nordgren A, Schoumans J, Staaf J, Borg A et al. Array-CGH reveals hidden gene dose changes in children with acute lymphoblastic leukaemia and a normal or failed karyotype by G-banding. *Br.J Haematol.* 2008;140:572-7.

200. Haferlach, T., Kohlmann, A, Basso, G, Bene, M, Downing, J. R., Shurtleff, S. A., Hernandez, J., Hofmann, W, Kipps, T, Kronnie, T, Lui, W, MacIntyre, E, Mills, K, Preudhomme, C, Chiaretti, S, Rassenti, L, Vos, J, yeoh, A, Koay, E, Papenhausen, P, Williams, M, Wieczorek, L, and Foa, R. **An International Multi-Center Study To Define the Application of Microarray-Based Gene Expression Profiling in the Diagnosis and Sub-Classification of Leukemia (MILE Study): Analysis of Completed Stage I with 2030 Patients Achieved a 95.4% Accuracy.** Blood 110. 2007.
Ref Type: Abstract

201. Van Delft FW, Bellotti T, Luo Z, Jones LK, Patel N, Yiannikouris O et al. Prospective gene expression analysis accurately subtypes acute leukaemia in children and establishes a commonality between hyperdiploidy and t(12;21) in acute lymphoblastic leukaemia. *Br.J.Haematol.* 2005;130:26-35.

202. Claudio JO, Masih-Khan E, Tang H, Goncalves J, Voralia M, Li ZH et al. A molecular compendium of genes expressed in multiple myeloma. *Blood* 2002;100:2175-86.

203. Zhan F, Hardin J, Kordsmeier B, Bumm K, Zheng M, Tian E et al. Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. *Blood* 2002;99:1745-57.

204. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531-7.

205. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 2002;1:75-87.

206. Armstrong SA, Staunton JE, Silverman LB, Pieters P, den Boer ML, Minden MD et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 2002;30:41-7.

207. Fine B, Stanulla M, Schrappe M, Viehmann S, Ho M, Harbott J et al. Gene expression patterns associated with recurrent chromosomal translocations in acute lymphoblastic leukemia. *Blood* 2004;103.

208. Ulazzi L, Sabbioni S, Miotto E, Veronese A, Angusti A, Gafa R et al. Nidogen 1 and 2 gene promoters are aberrantly methylated in human gastrointestinal cancer. *Mol.Cancer* 2007;6:17.

209. Pan Q, Gollapudi AS, Dave VP. Biochemical evidence for the presence of a single CD3delta and CD3gamma chain in the surface T cell receptor/CD3 complex. *J Biol.Chem.* 2004;279:51068-74.

210. Nilsson B, Andersson A, Johansson M, Fioretos T. Cross-platform classification in microarray-based leukemia diagnostics. *Haematologica* 2006;91:821-4.

211. Stec J, Wang J, Coombes K, Ayers M, Hoersch S, Gold DL et al. Comparison of the predictive accuracy of DNA array-based multigene classifiers across cDNA arrays and Affymetrix GeneChips. *J Mol.Diagn.* 2005;7:357-67.

212. Wiemels, J., Hofmann, j, Kang, M, Selzer, R., Green, R, Zhang, L, Smith, M, Buffler, P, and Yeh, R. Timing and formation of Chromosome 12p deletions in TEL-AML1 childhood acute lymphoblastic leukaemia. Proceeding of the AACR annual meeting 2008 49, 383. 2008.
Ref Type: Abstract

213. Paulsson, K, Cazier, J, Stevens, J, Chaplin, T, MacDougall, F, Lillington, D, Lister, A, and Young, B. Genome-Wide 500K SNP Array Analyses Reveal High Frequencies of Cryptic Genetic Abnormalities and Identify Recurrent Microdeletions Targeting Novel Genes in Adult Acute Lymphoblastic Leukemia. Session Type: Oral Session. Blood 110(11). 2007.
Ref Type: Abstract

214. Park J, Chen L, Ratnashinge L, Sellers TA, Tanner JP, Lee JH et al. Deletion polymorphism of UDP-glucuronosyltransferase 2B17 and risk of prostate cancer in African American and Caucasian men. *Cancer Epidemiol.Biomarkers Prev.* 2006;15:1473-8.

215. Lazarus P, Zheng Y, Aaron RE, Muscat JE, Wiener D. Genotype-phenotype correlation between the polymorphic UGT2B17 gene deletion and NNAL glucuronidation activities in human liver microsomes. *Pharmacogenet.Genomics* 2005;15:769-78.

13: Appendices

Appendix 1

Recipes for solutions

Agar for culture plates (sufficient for approximately 25 plates)

Prepare Luria Bertani (LB) broth as described below and add 15g agar. Sterilise at 115°C for 15 minutes, and allow to cool to 60 °C. Prepare a stock solution of kanamycin/chloramphenicol by diluting 5mg into 50ml water, and filter sterilising. Dispense 1ml of the stock solution into 1L LB broth to give a final concentration of 1µg/ml. Pour the plates and allow the agar to set. Store at 4°C for up to 1 month.

Bacterial mRNA control stock solution

Combine the following to give a 16.7ng/µl solution:

5µl 0.1 µg/µl *araB* mRNA

5µl 0.1 µg/µl *entF* mRNA

5µl 0.1 µg/µl *fixB* mRNA

5µl 0.1 µg/µl *gnd* mRNA

5µl 0.1 µg/µl *hisB* mRNA

5µl 0.1 µg/µl *leuB* mRNA

Store in 3µl aliquots at -80°C for 1 month

Prepare a 50.2pg/µl solution by diluting a 3µl aliquot into 997µl. Store in aliquots at -80°C for 1 month.

Bacterial mRNA control working solution

Dilute 2ul 50.2pg/µl stock solution with 998µl nuclease free water.

Cy-5 Streptavidin stock solution

Add 1ml nuclease free water to 1mg lyophilized Cy5 streptavidin. Pipette mix without allowing foam to form then place on ice for 5 minutes. Repeat 3 additional times. Protect from light at all times. Store in 90µl aliquots at -20°C.

Cy-5 Streptavidin working solution

Centrifuge the thawed stock at 8,000g for 1 minute. Add 6.8µl of stock solution to 3393.2µl thawed TNB buffer, for each array to be processed. Mix gently by inversion and use within 15minutes of preparation.

0.5M EDTA (pH8)

Add 46.53g EDTA disodium salt to 150ml distilled water

Add concentrated sodium hydroxide (NaOH) until the solution is at pH8.

Make volume up to 250ml with water

FISH wash 1

Combine the following:

2ml 20x SSC

98ml water

0.3 ml Igepal (NP-40)

FISH wash 2

Combine the following:

10ml 20x SSC

90ml water

0.1 ml Igepal (NP-40)

MADGE gel loading dye

Combine the following in a fume hood;

9.8ml 98% deionised formamide

200µl 10mM EDTA (pH8.0, 0.5M)

1.5mg xylene cyanol FF

MFISH denaturation buffer

Combine the following in a fume hood:

25ml formamide

10ml sterile water

5ml 20x SSC

Luria Bertani (LB) broth

Combine the following in a fume hood:

10g tryptone

10g NaCl

5g yeast extract

Make volume up to 1L with water

Leishmans stain stock solution

Gradually add 0.3g powdered Leishman's stain to 200ml methanol. Stir overnight on a magnetic stirrer.

Leishmans stain solution

Dilute the stock solution 1:5 with the phosphate buffer.

PBS

Add 1 tablet to 200ml water. Adjust to pH7.4 or pH6.8 appropriately

Polyacrylamide gel

Each gel requires 40ml gel mix

Combine the following;

75ml 40% acrylamide:bis

40ml 10x TBE

285ml water

Store 40ml aliquots at 4°C

To promote polymerization add 120µl TEMED and 120µl 20% ammonium persulphate (APS).

0.1x SSC/0.05% Tween 20

Combine the following:

994.5ml deionised water

0.5ml Tween 20

5ml 20x SSC buffer

Use within 2 weeks

2x SSC (sodium chloride/sodium citrate)

Combine the following

100ml 20x SSC

100ml water

Sticky silane

Combine the following;

500µl glacial acetic acid

500µl methacryloxypropyltrimethoxysilane

99ml 100% ethanol

50x TAE buffer

Add the following to 900ml distilled H_2O:

242g Tris base.

57.1ml Glacial Acetic Acid.

18.6g EDTA

Adjust volume to 1L with additional distilled H_2O.

TE buffer

Add the following to 500ml distilled H_2O:

0.727g Tris

1.2ml 0.5M EDTA

pH to 7.8 with approximately 4ml 1M HCL

Adjust volume to 600ml with additional distilled H_2O

TNB buffer

Combine the following in a flask with a stir bar:

522ml nuclease free water

60ml 1M Tris-HCL (pH7.6)

18ml 5M NaCl

On a stir plate heated to 60°C add 3g of NEN blocking reagent in 0.5g increments until completely dissolved. Turn off the heat and continue mixing for 30 minutes. Filter through a 0.88µm filter. Store in 50ml aliquots at -20°C, for up to 12 weeks.

1xTNT buffer

Combine the following:

100ml 1M Tris-HCL (pH7.6)

30ml 5M NaCl

0.5ml Tween 20 (Sigma)

870ml deionised water

Filter through a 0.2µm filter. Use with 1 month

Appendix 2

List of Suppliers and details of kit contents

Agilent

44K/185K/244K Human Genome CGH Microarray

Genomic DNA Labeling kit plus: *Random primers, 5x buffer, 10x dNTP, Cy-3 dUTP, Cy-5 dUTP, Klenow Fragment,* H_2O.

Gasket slides

OligoCGH hybridisation kit: *lyophilized 10x blocking agent, 2x hybridisation buffer.*

RNA 6000 nano kit: *gel matrix, dye concentrate, nano marker*

RNA 6000 pico kit: *gel matrix, dye concentrate, nano marker, conditioning solution*

Stabilization and drying solution

Wash buffer 1 (aCGH)

Wash buffer 2 (aCGH)

Alpha Laboratories

Foilseal

Ambion

RNA 6000 molecular weight ladder

Applied Biosystems

Assay-on-demand primers and probes

Taqman universal PCR mastermix

96/384 well clear optical reaction plates

Optical Adhesive covers

BDH

B-mercaptoethanol

DMSO

DPX

EDTA

Glycerol

HCL

KCL

Leishman's stain

NaOAc

NaOH

SSC

Dakocytomation

DAKO ETV6 break-apart FISH probe

Fisher Chemicals

Acetic acid

Acrylamide:bis

Agar

Chloramphenicol

Ethanol

Ethidium bromide

Formamide

Isopropanol

Kanamycin

Methanol

NaCl

TEMED

Tris base

Tryptone

Yeast extract

GE healthcare

CodeLink 16-assay bioarrays

CodeLink expression assay reagent kit; *T7 oligo(dT) primer, 10x first-strand buffer, 5mM dNTP mix, reverse transcriptase, RNase inhibitor, 10x second-strand buffer, DNA polymerase, RNase H, nuclease free water, bacterial control mRNAs* (araB, entF, fixB, gnd, hisB, leuB), *10x T7 reaction buffer, 10xT7 enzyme mix, T7 ATP solution, T7 GTP solution, T7 CTP solution, T7 UTP solution, 5x fragmentation buffer, hybridization buffer component A, hybridization buffer component B.*

CodeLink 20K human arrays

Ficoll-paque

Genomiphi DNA amplification kit: *sample buffer, reaction buffer, enzyme mix.*

G50 sephadex columns

Invitrogen

BioPrime array CGH genomic labeling system: *2.5x random primers solution 10x dUTP, exo-klenow fragment, stop buffer.*

Colcemid

BAC clones

Fetal calf serum
Human Cot-1 DNA
Penicillin-streptomycin-glutamate
Platinum Taq DNA polymerase: *Platinum Taq DNA polymerase, 10x platinum Taq buffer, 50mM MgCl,*
RNAseZAP
RPMI medium
SYBR Gold
SYTO 9
Trizol

Millipore
Microcon YM-30 filters

Miltenyi biotec
CD138+ magnetic beads

MWG
Primers

Natural Diagnostics
10X TBE

New England Biosystems
BbvCI
NdeI
NcoI
HindIII

NimbleGen
385K Human Whole-Genome Array

Perkin Elmer
10mM biotin-11-UTP
Cy5-streptavidin
Cy3-dUTP
Cy5-dUTP
NEN blocking reagent

Promega

Agarose

Alu1

Ammonium Persulphate

1Kb DNA Ladder

10mM dNTP mix

GoTaq Flexi DNA polymerase: *5x colourless/green GoTaq Flexi buffer, $MgCl_2$, GoTaq DNA polymerase*

Male and female human reference DNA

M-MLV reverse transcription enzyme ans 5X buffer

RNAsin

Rsa1

T4 DNA Ligase : *T4 DNA Ligase, T4 DNA ligase 10x buffer*

Oligo(dt) primer

QBiogene

21q subtelomere FISH probe

Qiagen

DNAse 1: *lyophilized DNAse 1, buffer RDD, nuclease free water*

DNeasy tissue kit: *DNeasy mini spin columns, AW1 buffer, AW2 buffer, proteinase K, AL buffer, AE buffer.*

QIAquick Gel extraction kit:

Qiagen midi kit: *RNase, buffer P1, buffer P2, buffer P3, Qiagen tips, QBT buffer, QC buffer, QT buffer.*

QIAprep spin miniprep kit: *QIAprep spin column, PB buffer, PE buffer, EB buffer*

QIAquick PCR purification kit: *QIAquick spin columns, PB buffer, PE buffer, EB buffer*

RNeasy mini kit: *RNeasy mini spin columns, RLT buffer, RPE buffer, RW1 buffer, RNAse free water.*

Roche

LightCycler 480 Instrument 96 well plate

Sanger Centre

BAC clones

Sigma

Acetonitrile

Chloroform

Chloramphenicol
Giemsa stain
Igepal
Kanamycin
methacryloxypropyltrimethoxysilane
Phosphate buffer tablets
Trypan blue
Tween 20

Spectral Genomics
Human BAC Array

Thermo electron corporation
EZ cytofunnel

Worthington
Trypsin

Vector laboratories
Vectashield mounting medium

Vysis
DAPI III
FISH probes; *TEL/AML1 extra signal, BCR/ABL dual colour dual fusion, p16(9p21(so)/CEP9(sg),MLL dual colour break-apart reaarangement, IGH dual colour break-apart, X centromere.*
Hybridisation buffer
MFISH probe
Nick translation kit: *dATP, dTTP, dGTP, dCTP, 10x nick translation buffer, nuclease free water, DNase and polymerase enzyme combination.*
Spectrum green dUTP
Spectrum red dUTP

Appendix 3

Home-grown clones used for FISH analysis. Unless otherwise stated, clones were obtained from The Wellcome Trust Sanger Institute. All were grown in the presence of chloramphenicol for colony selection, with the exception of RP1-80I19, which was grown in kanamycin.

Clone Name	Clone Type	Colour labelled	Chromosome band	Position (bp)	Gene Target
RP11-375H16	BAC	RED	2q23.1	148888403-149008484	*D5*
RP11-79A11	BAC	RED	2q23.3	151030186-151196407	*RND3*
RP11-484E7	BAC	GREEN	3p12.2	76248984-76475216	Control
RP11-499P15	BAC	RED	3q26.32	178316159-178404359	*TBL1XR1*
RP11-138C1	BAC	GREEN	5p13.1	39246335-39374627	Control
RP11-78J4	BAC	RED	5q33.3	158143025-158299537	*EBF1*
RP1-80I19	BAC	RED	6p22.1	29154434- 29299270	None
RP11-167P05	BAC	GREEN	6p12.3	49243156- 49420568	Control
RP11-246L12	BAC	RED	7q34	127471196-127495720	*TCRβ*
RP11-426J23	BAC	GREEN	7q34	142042930-142222513	*TCRβ*
RP11-313F12	BAC	GREEN	8p12	34467174-34617727	Control
RP11-108L23	BAC	RED	8q12.1	60223830-60381578	*TOX*
p16/9cen (WRGL)*	Unknown	RED & GREEN	9p21/9cen	Unknown	*CDKN2A*
RP11-469D03	BAC	RED	9p13	36928510- 37098070	*PAX5*
RP11-113O24	BAC	GREEN	9p13.1	38263089-38427296	Control
RP11-89K14	BAC	GREEN	9q22.1 - 9q22.2	90982108-91150061	*CKS2/SBP2*
RP11-477A7	BAC	GREEN	9q32	105154965-105331219	*TAL2*
RP11-18B3	BAC	RED	9q32	105552126-105552617	*TAL2*
RP11-169B16	BAC	RED	9q32	127471196-127495720	*TAL2*
RP11-167O6	BAC	GREEN	10p11.22	32116662-32303828	Control
CTD-2511A8 (Invitrogen)	BAC	RED	10q25.2	111769502- 111857085	*ADD3*
RP11-44H16	BAC	RED	11p15.5	Unknown	
RP11-646J21	BAC	GREEN	11p13	33693154- 33841461	*LMO2*
RP11-202M19	BAC	RED	11p13	33959194- 34004313	*LMO2*
RP11-734J23	BAC	RED	11p13	33999153- 34214460	*LMO2*
G248P81985C2/ WI2-1541F3	FOSMID	GREEN	11q13.1	65,380,496-65,419,297	*MUS81*

G248P89301D12/ WI2-3373H23	FOSMID	GREEN	11q13.1	65,422,680-65,463,891	*MUS82*
RP11-88H18	BAC	GREEN	11q22.2 - 11q22.3	102253679-102424014	*MMP13*
RP11-33F6	BAC	GREEN	11q22.3	102496919-102497333	*DYNC2H1*
RP11-180M15	BAC	RED	12p13.1	12614233-12772124	*CDKN1B*
G248P87878F1 /WI2-1689K2	FOSMID	GREEN	12p13.32	5,932,935-5,978,096	*VWF*
G248P84562E1 /WI2-1002I2	FOSMID	GREEN	12p13.32	5,977,783-6,016,668	*VWF*
RP11-91M21	BAC	RED	12q24.21	113305910-113477410	*TBX5*
RP11-526P6	BAC	RED	12q24.32	126676143-126835833	None
RP11-91B1	BAC	RED	12q24.32	127615070-127615435	None
RP11-81G12	BAC	RED	12q24.33	129104134-129104626	None
RP11-242H9	BAC	GREEN	14q11	19869665- 20012604	*TCRαδ*
RP11-447G18	BAC	RED	14q11	20965368- 20965676	*TCRαδ*
RP11-13J15 (21(A))	BAC	GREEN	21q21.2	24047487-24200441	None
RP11-30N6 (21(B))	BAC	RED	21q21.3	29766105-29836744	None
RP11-147H1 (21(C))	BAC	RED	21q22.1	32372362-32372918	None
RP11-396G11 (BA)	BAC	RED	21q22.12	35423157-35622771	*RUNX1* (Tel)
RP11-272A03 (BA)	BAC	GREEN	21q22.12	34768331- 34953503	*RUNX1* (Cen)
RP11-121782 (21(E))	BAC	RED	21q22.2	40015038-40157799	*IGSF5*
G248P87335G2/ WI2-1777N4	FOSMID	RED	21q22.3	42,862,482-42,898,892	*PDE9A*
G248P8283A3/ WI2-517B6	FOSMID	RED	21q22.3	42,892,637-42,933,023	*PDE9A*
G248P84236D6/ WI2-1402G11	FOSMID	RED/GREEN	21q22.3	42,936,865-42,979,542	*PDE9A*
G248P86904G6/ WI2-1902M11	FOSMID	GREEN	21q22.3	42,982,966-43,027,053	*PDE9A*
G248P83754D4/ WI2-1328G8	FOSMID	GREEN	21q22.3	43,025,731-43,069,905	*PDE9A*
RP11-88N2 (21(F))	BAC	GREEN	21q22.3	43556394-43769964	*SNF1LK*
21qtel	Unknown	RED	21q22.3	46144350-46311763	*COL6A1*
RP11-492H04	BAC	RED	Xq25	129238662-129413422	None
RP11-377B2	BAC	GREEN	Xp11	44271913-44399052	Control

*WRGL; Wessex Regional Genomics Laboratory

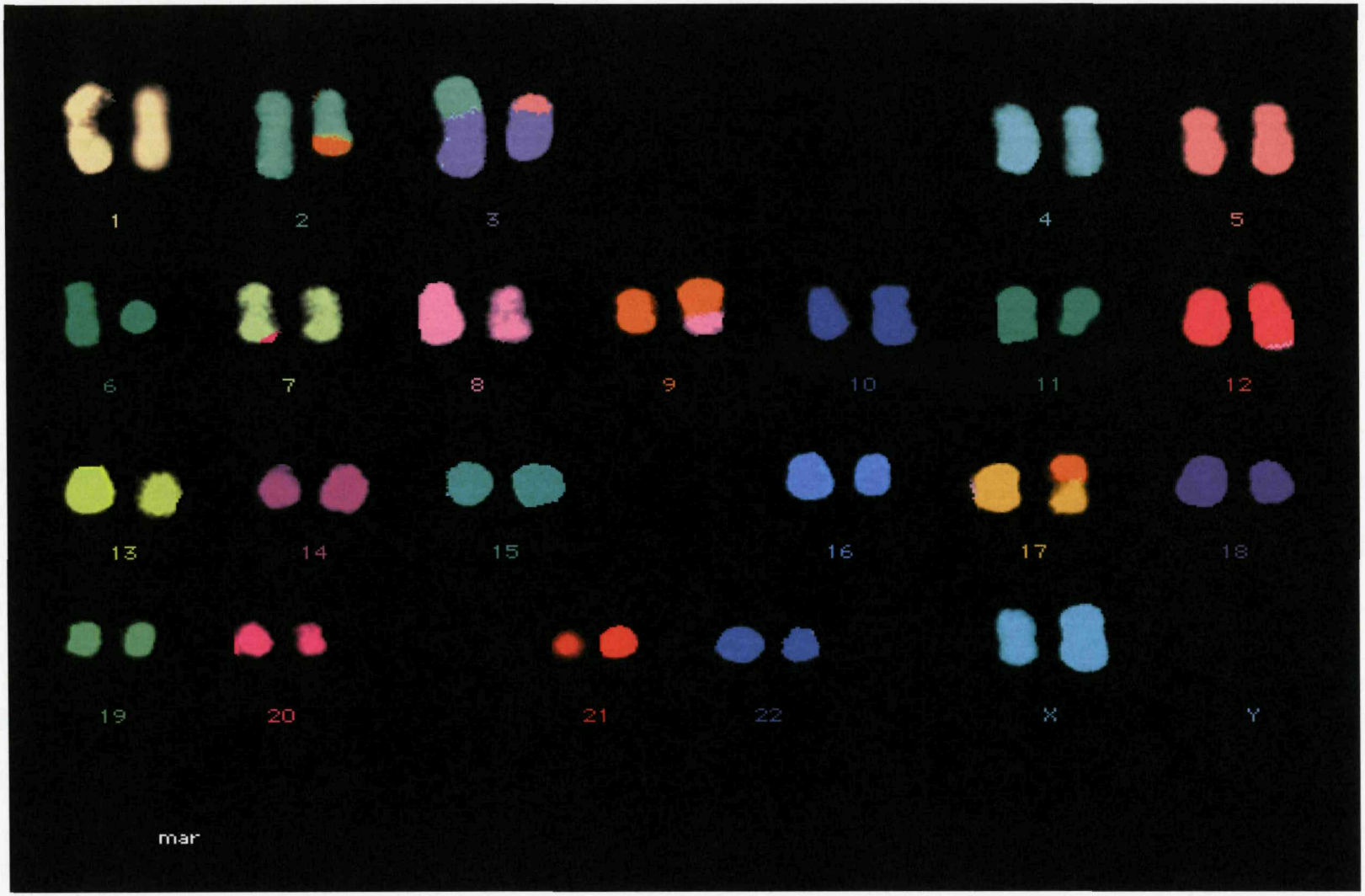
Commercially available FISH probes used in this study

Probe Name	Company	Chromosome band
SIL-TAL1 Sub-deletion signal	Dakocytomation	1p32
TEL/AML1 Extra Signal	Vysis	t(12;21)(p13;q22)
BCR/ABL Dual Colour Dual fusion	Vysis	t(9;22)(q34;q11.2)
p16(9p21(so)/CEP9(sg)	Vysis	9p21
MLL Dual colour break-apart rearrangement	Vysis	11q23
ETV6 break-apart	Dakocytomation	12p13
IGH Dual colour break-apart	Vysis	14q32.3
TCF3 Split signal	Dakocytomation	19p13
21q subtelomere	Qbiogene	21q22.12
X centromere	Vysis	

Appendix 4

G-banded karyotype for ARH77.

MFISH karyotype for ARH77



Appendix 5

Primer sequences for MCC analysis of *PDE9A* in iAMP21 patients

Marker	External Forward	Internal forward	Reverse
1	GTCTTGAGTCATGGCTGTG	GAAGACAGATGCGCCAGG	CTTCACAGAGACACTGAGG
2	GTAGATATTATACCTGAGG	GTCTGAGGAGCAAATGAAAG	CTCTGATTTTATCCATCTGG
3	GCTTTGGATATTGGGAGG	CTCCTTCACTTACCTCACG	GAAGGAAGGCTGCAACCG
4	CGTGTGAGTGATTTCAGG	GTCAAACTGTCAGACTTGG	GAGCAAAGCATTTAATCTGG
5	GAGTAGCTGGGATTACAGG	CAAACTCCTGACCTCACG	GAGGAGAGTGAGGCAGG
6	CTTCACTGTTACTTTAGAGG	CTAGAGCTTCTCAGGAGG	CTTACAGGAATCTTTATGG
7	CAGTTGCAGCAGATGAGG	AACTGGACCAGCGCAAGG	CCCATCCGAATGGAAGCC
8	AAGGAGAGGGAAAGGCAG	TACAAGCCAGGCTCTGCG	CCGGAGCCAAGAGAGATA
9	TGGCTCCAGCCTTGGTG	GGTGCCTCTCTGCACTG	CAATGGATGCATGCAGCTC

Restriction endonucleases for LDI-PDR digestion of *PDE9A* in iAMP21 patients

Patient	Proximal RE	Distal RE	Buffer
4279/8983/7619	BbvCI (CC↓TCAGC)	NcoI (C↓CATGG)	2
6996	NdeI (CA↓TATG)	HindIII (A↓AGCTT)	2
7024/7828	Bg1 II (A↓GATCT)	Bg1 II (A↓GATCT)	
2647/3745	BbvCI (CC↓TCAGC)	HindIII (A↓AGCTT)	2

Primer sequences for LDI-PCR analysis of *PDE9A* in iAMP21 patients

Patient	External Forward	Internal forward	Reverse
4279/8983/7619/2647/3745	TGGTGAGTCCACCATGTG	TGCCAGGCTTGGCTTCTG	GGCAGACGACTACCTGG
6996	TACAAGCCAGGCTCTGCG	TCGAGCTCAGGCTTCAGG	ATGGACAAATGTGTGTCTG
7024	AGCGTTAGGGAAAGCGTG	GACAGCCTCTTCAACTAGG	ACCCAATCCTCATCACAGG
7828	GTGAGAATGGCTCTAGAGG	TGAGAATGAAAGCGCCAGG	ACTGCTTGCCAAGGCAAAG

Primer sequences for RT-PCR analysis of *PDE9A*-fusion genes

patient	Forward	Reverse
4279	TCTACCTGGACATCGATGG (*PDE9A*)	GGTAATTCACACCCGCAGA (*MUS81*)

Primer sequences for HRM analysis of *TBL1XR1* in *ETV6-RUNX1* positive patients

Marker	Forward	Reverse
4	CTCATTCAGCATTTACCTTTGG	GATAGAAATCAATGCAGCGGG
5a	GGTACCTTGTTTGATGGTCG	CTGCTGCCTGTTGCTGTG
5b	AGGCAGCAGCTGCTGCAG	GTCTGAGCTCACTTGCTATAG
6	ACTGATATGATGGAAGTGGATG	AACAGGGTTCCAGGCACAG
7	GTGAGAACAGCACCAGTGG	GTGACATCCTTGTTGCTTGG

8	GGTACACTTCTAGCAACTGG	GATCAACAACAAAGTAAGAAGG
9	AACCTTGCTAGCACCTTAGG	TCTACTCCAGCACTTAGGATG
10	CAGACTACAATTATTTGGGACG	GCAAGCTCCAGCTTTCACG
11	GCATTGGATGTTGATTGGCAG	CTTACCGTATGTCCTTGGAATG
12	GAAGTAAATGCTATCAAATGGG	AACCCATTTCTCCCTACTAAG
13	GAAACAAGACAATTGTGTCCATG	ATTAGTCCCTGGTCCTGTTG
14	AGGTTATGGGATGTAGACCG	AACCACTTGCCAGATACCTG

Primer sequences for mutation sequencing of *TBL1XR1* in *ETV6-RUNX1* positive patients

Exon	Forward	Reverse
4	TTTACATGCACACACACGCG	TCCACAGCACAATGCTGTTG
5 + 6	GGTACCTTGTTTGATGGTCG	AACAGGGTTCCAGGCACAG
6 + 7	ACTGATATGATGGAAGTGGAT	GTGACATCCTTGTTGCTTGG
8	CCTGGTAAAGTAGGCAGTTG	GATCAACAACAAAGTAAGAAGG
10	TGAGTGTGGTGTCAAACAGG	GCAAGCTCCAGCTTTCACG
11 + 12	TATGCAGTGACAGAACCGAG	AACCCATTTCTCCCTACTAAG
13 + 14	GAAACAAGACAATTGTGTCCATG	AACCACTTGCCAGATACCTG

LDI-PCR Sequencing Data for patients with *PDE9A* gene rearrangements;

4279: *PDE9A* (intron 1)-*MUS81* (intron 12)
TGCCAGGCTTGGCTTCTGGAAGAGAAAGCTGGGGGTTTAGTGACTCCAGCCAACTGCTGTGAGCCCAAGGCATCAAACTAAAAATAGCGAG
GCCAAGAGGGAGTCTTTGTGTTCGGCTCAGTGGGCAATGGTGCCTCACCTCGCAGCATGACACAGGGCTGATGACACGGGGCACGTTGGT
CACAAAGTGCCAGTCTCAGCCTGGCACAGGGACGTGCCACCTGCAGGTGATGACACATCACAGTGGCAGCCGCTCCTGATGGTGCTGCTG
CATTGAGTAGGAGCCCAGATCCAGGGAGACAGTTGCAGCAGATGAGGCAGGAGGCCAGCGAGGGGTTCTACCTGCTTTCCGGGCTGGCCT
CTGCATGGACACAAAGCAGAAAGGGCATAAAGGTGACCCTGGCCACCAAAGCAGACCTGGAGCAAACTGGACCAGCGCAAGGCCCCGGTG
CCAGC↓↓TCTTCCCTCAGGGCCACACCCTACGCAGCCGCCCCTGGGGAACCCCTGGGAACCCTGAATCAGGGGCCATGACCTCTCCAAACC
CTCTCTGCTCACTCCTCACCTTCAGTGACTTCAACGCAGGAGCCATCAAGAATAAGGTACTGTCTCTGCCTAGCTTCTCAGACATGGCCTGG
CCCAGACCCCCACTGATCCAGCCCTTTCCCGAACCCAGGCCCAGTCGGTGCGAGAAGTGTTTGCCCGGCAGCTGATGCAGGTGCGCGGAG
TGAGTGGGGAGAAGGCAGCAGCCCTGGTGGATCGATACAGCACCCCTGCCAGGTAGGCCCTAAAGGGCCCTTAGGTGTCCTCAGC

6996: *PDE9A* (intron 1)-*VWF* (intron 3)
TCGAGCTCAGGCTTCAGGCCTCCGGAGCCAAGAGAGATAAACCTCTGTGGTTTCATGTCACCCAGTTTGTGGTCTTTGTTATAGCAGCCCCA
GGACACTCAGGGGGTTGCCTTGTCCACCCCCATCCTGGCCCTCTGAGGGACTGGTTGGCTCTTAGAAGACCGTTCAACCTGCCCTCTGGTC
CTCCTCCCCCACCCAGAGACCCCTGGCTACACCCTGCCCAAGGGAGCTCACC↓↓AGGCCTGACAGACAATAAAAGGTGGAGCTGAGGCCGG
GCACGGTGGCTCATGCCTGTAATCCCAGTACTTTGGGAGGCCAAGGTGGGTGGATCACCTGAGGTCAAGAGTTTGAGACCAGCCTGGCCAA
CATGGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGTGTGGTGGTGTGTGTCTATAATCCCAGCTACTTGGGAGGCTGAGGCA
GGAGAATCGCTTGAACCTGGTTGCAATAAGCTGAGATACACTCCAGCCTGGGCAACAGAGCGAGACTCCATCTCAAAAAAAAAAACAACCCA
CAAAAAACAAAAAAACTGGACTAAGCAGGCCAAGGACAGAGCCCAAGGCCAAGGCTTAATCTAGAAGAGGGCTCAGAAGTGCCCCACTCAA
GTTTGGTCAAGGAGGGAGTCTTTGGCAACACCTGGACACTTACCTGAGATCTGGGCTGTAGGGCTCCTGGGGTCATTGCTCCATCAGTCAG
CGGGGACTGACACAGGGTCCTCCATGTGCCCAGCACTGGGCTAGGCTCTGTCTAGCACCTGGCTATAGCTATGAGCTCCCCGCATTCCTGC
TTTCTCCCTGAGGCCATCTTTGCTACCTGCCCTGACTCTGCTGGGAAACTTATCACATATG

8983; *PDE9A* (intron 1)-*REXO1L1* (intron 2)
ACTGCTGTGAGCCCAAGGCATCAAACTAAAAATAGCGAGGCCAAGAGGGAGTCTTTGTGTTCGGCTCAGTGGGCAATGGTGCCTCACCTCG
CAGCATGACACAGGGCTGATGACACGGGGCACGTTGGTCACAAAGTGCCAGTCTCAGCCTGGCACAGGGACGTGCCACCTGCAGGTGATG
ACACATCACAGTGGCAGCCGCTCCTGATGGTGCTGCTGCATTGAGTAGGAGCCCAGATCCAGGGAGA↓↓AGCTCACCCTGCCCACCTCGCC
GCAAAGCGAAAGAAACGGAGCAGCCGGCGGCAGGAGAGGGCAAAAAGCCAAGAGTAACCCCAACCCCCCACTCCCGGTCCCCCGGAATCC
CTGCCGCGGCCCCTCGGGCCTGTCCACATCCCTCTGCCCCTCCCAGACCTCTGTCCTTCCACCAATCGCCTCCCGCAGCCCCGAGCCGCC
ACTCCCAGTCCCCCGAGTCCCTGCCGCGCGCCCTCGCGCCTGTTCACATCCCTCTGCCCATCCGAGACCTCTGTCCTTACACCACTAGCCA
CCCCACGTGGGACTTCCATGG

2647; *PDE9A* (intron 1)- Chromosome 21 (42,936,452 Mb) (5' to *PDE9A*)

TGCCAGGCTTGGCTTCTGGAAGAGAAAGCTGGGGGTTTAGTGACTCCAGCCAACTGCTGTGAGCCCAAGGCATCAAACTAAAAATAGCGAG
GCCAAGAGGGAGTCTTTGTGTTCGGCTCAGTGGGCAATGGTGCCTCACCTCGCAGCATGACACAGGGCTGATGACACGGGGCACGTTGGT
CACAAAGTGCCAGTCTCAGCCTGGCACAGGGACGTGCCACCTGCAGGTGATGACACATCACAGTGGCAGCCGCTCCTGATGGTGCTGCTG
CATTGAGTAGGAGCCCAGATCCAGGGAGACAGTTGCAGCAGATGAGGCAGGAGGCCAGCGAGGGGTTCTACCTGCTTTCCGGGCTGGCCT
CTGCATGGACACAAAGCAGAAAGGGCATAAAGGTGACCCTGGCCACCAAAGCAGACCTGGAGCAAACTGGACCAGCGCAAGGCCCCGGTG
CCAGCACTCAGAGACATTTCTTTTCTGGCAGGAAATTTGTGCTTCATTTTCTAATGAAATTCCAAGGCAATCAAAAGGCAAGGAAAGAGGAGG
GAAAGGAAGGGAGGGAGGGAGGGGGTTCCATTCCCATCCGAATGGAAGCCCCCAAGGCTTCTGTAGCTCCCCCCATCACCAACTTCACTCA
CCACTCTTATCTGATGCCTCCTATTTGGGTACACACGGCCGGGTCGGCCTTGCGAGAGACTCAGTTTCCTCTCGACTTCCGCCCGCCTGCAC
TGTCACCACGTCCTGCCATTGACCTCTGTTCATCCACTCATGGTCACTGCTGTGAGTTGAATTGTGTCTCCCAAAAAAGTGGATTTCTGTGCT
AACCTGTGAATGTCCCGTTAGTTGGAAATAGGGTATTTGCAGACATAATCACATCAAGATGAGGTCATACTGCATTGGGATGGGCCTCAAGCC
CATATGCCCGCCATCCTTCTAAGGAGAGGGAAAGGCAGACACACATTTGTCCATGGAAATGAGA↓↓AAGGAACAGAATTTTCCTACCTGTTAA
GAACTTGGTTTTCTTTTTTTTTTTTTTTAAACAGAGTCTTACCCTGTTGCCCAGGCTGGAGTGCAATGGTGCGATCTCAGGCCACTGCAGCCTCT
GCCTCCCAGGTTCAAGCAATTCTCCTGCCTCAGCGCTGTGTCACACTCCACACTGAGGCCCCACTGAAGCATGCCAGGCCCCATCCTATTTC
TTCCAGGTAGTCGTCTGCC

3745; *PDE9A* (intron 1)- Chromosome 21 (43,109,515 Mb) (3' to *PDE9A*)

TGCCAGGCTTGGCTTCTGGAAGAGAAAGCTGGGGGTTTAGTGACTCCAGCCAACTGCTGTGAGCCCAAGGCAT↓↓TTTAAAACACTAAAGGC
AAGTCCTGGCACACAAGTTTCTATACCTGCTTACTAAGTAACCATTTTCCTGCTCTGCCATCCTTTGTGAAGTACCTTATGGGCCAAAATGGTG
CTGAAGCACCAGCCATTATGACTGCATGACAGGAAGGAGGATAAGGCACAATATTTCCCCTCACACTTGGAAAAACAATTTTCTAAAATTAGG
ATTCCATCACCAAGGAGAAACATTCATTTTACTACACACTGCTGTGAATTTATCTCACTAAATAATGCTATTGTCTTTGTCCATTTTCTGTTGCTT
ATAACAGAATTCCTGAGGCCGGGTCATTTGTAAAGAAAGGGAATTTATTTCTTCCATTATGGAGGCTGAGAAATCTAAGGTTAAGGCACCACA
TCTGATGAGGGCCTTCCTGCTGGTGGGGACTCTCTGCAGCACAGGGTATCACGTGAGGGGCTGAGTGTGCTAGCCCAGGTGTCTCTTTCTC
CTCTTATAAAGCCATTAATGCCACCACCATGATAACCTGTTACTCATTTACCCATTAACCCATTAATGGACTAATGCATTCATGAGAGCAGAGC
CGTCAGAACACAACCACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTGGCACA
ATCTCAGCTCACTACAACCTCCTCCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCGCTGTGTCACACTCCACACTGAGGCCCCACTGAAGC
ATGCCAGGCCCCATCCTATTTCTTCCAGGTAGTCGTCTGCC

7024; *PDE9A* (intron 5)-*PDE9A* (intron 1)

GACAGCCTCTTCAACTAGGACAGTGCTGTAGGGGTCTGGTTCTGCTTGTTCTCCCCCACATTCCTGCCTGCCCTAGGACCCAGGACCTGTCT
CTAGTCACACACACCCTGTTCAGCCTCTGCATCTGGCCCCCGCGGAGACCCACACCACATTCTACTCCCCTGGCCAGAGACCGAGGGTGAC
TCGGAGGCTTCCAGTTCCCAGCAGCTTTATTGGCGTTGCTGTTTCCATAGAAGGACAATAAAAGGGGGCCATAAGGAGGCGCATCCGCAGA
GTGCGTGCTTGAGCTTCTGCAAGCCAGAGGAGGCCGGCGGCCGGCAGGGTGGGCAGGCTGGCTGGCTGTGCCCCACAACATGCATGTGG
AGAGGTTTTGGGGGCAGCCATGGAACCCCGGAGCTGCCCCCACATGAGAGTTCCCCCAGGTGGTGACCAAATGCATTTACAGTCCAGCCGC
AGTCCAGATGCAGATGCGCTGTAACTGTAGAACACACAGCAAATTCTGAAGGCTTAGTATGAAAAAGAATATAA↓↓TTTATTCCTCCTGACTAA
AGATTCTCGTCGGTGTAACCGGAGGTTCTGTACCTTCCGTTTATGATCACTGTTAGGGCAACTGTTAACACTCTTCCACTGCGTGACGGGGT
GTGATCTCTACTCAGCCTACGGGAGGTTGGGAGGTTACCCGCTCACGATTTGACCCCAGCCGGGTGTTGTACCTTGTGCCGAGTCGTTCCTT
TTCCACGTGTGTTGGTGAACCTACCTAGAGTTCCCTTACTACGGTCAACCTTTTTCGGTCAGAGTTCTCCAATGTATGACACATTAGGGTAAAT
ATATTATAAGAACTTTACTGCCTTAACTTCTCTATCTCTTGTCTAAGTACAAGTGGTCCTGCCCCACGACCTTTAGCTCAATCCGTACCCCACA
ACCTTTACCCACGTCCATCCCCTACACCAACCAATATTCTCCGTCGGAACGCCGCTACCTTACAAGATCTGACACTGACACCAACACCTACAT
TCTCAGATGTGTCCGTTCTTGTAGCACGTCTTCATTCACGTGTACGAGTTCACTCATGTTCACGTACTCCGACAAGAACGTAACGATATTTCTT
TATAGACTCTGACCCAGTAAATATTTCGGTTAAAGATTAACCGAGTGTCAAGACGTCCGACATATTCGTACCGTGGCTGTAGTGACCCGAAGA
CCCCTCCGGAGTCCCTCGAAAGTGAGTCCTACCTTCCACTTTGTCCTCGTCCGTATAGTGTACTACTTCCCTCCGCACGTGTGGGCACTCCC
TCTCTCTCCTCTACGGAAGATGGAAGTGTGTATTTGTTGGTCTAG

7828; *PDE9A* (intron 5)- Chromosome 21 (42,941,359 Mb) (5' to *PDE9A*)

CCACGGCCCAGGGAGGGGGTTGCTGCGGGTGGGCTTGGTCTGCTCATCAGACATCAGGTGCATGAGGAGCAGGCAGAGTCAGGGCAGGC
GGCCTCTCGGCTGTCGGCTCCCAATGGATGGTGTTGAAGTCTGGGACTTGGTGAGGCCAGCAGGGAGTGAGTGAGCACCAGCAGGAGGGG
CCAGGCCTGAGCCGGACGGAGACCGAACCCATCCGTGGGAGAAGGACGCTGGTGACAGGACACCATCGGGAGCAGAGGGGAAGCCATAA
AGCCCTGGAGGCAGGAAGAGGCAAAGGCTGGCCGGGTGTGGGTGAGGTGCTGCAGGGCTCTCGCAGGCCAGCGAAGACAGGCGAGATG
GCTGTGCTCAGAGTTAGAAGAGCATCGCTGGGAGCAGAGGAAAGAGGGATTTCTGTGTCACCTGGGGACCCATCCAGGACCAGTGGGAGA
TGGAGTTGGGCTCATTGAGGGAAAGAGGTGAATGGAAGGAAGTTGGAACCAGCTGTCCTAAGGCCGGAGCTGCTCCTCCCTGGAGGACTC
CGGCAAAGGCTGCTGGCCTTGGGCTGTGCGGGCACCCAGTGGCTGAGAGATCGTTGCCCTCTGCAGCCCTCCTGCAACTTGGATCCTAGG
AGCCACAAAGAGGCTTGAAAAATAATCCATCCTGAGA↓↓AAGGATATGGATATGGGTGTTCGTTCACACTGTTTACAACAACAACATATTGGGA
AAATCCAAACGCCCATCACACGAGGAGCTGGCCAGGAGTCGGAAGCACACAGGCATCAGGGGCCTGCTGGTCCATGGTGAGAAGGGGTGG
GTCTCCATTTTCTTTTTGTTTTTGTTTTTGTTTTTTTTTTTTGAGGTGGAATTTCACTCTTGTCGCCCAGGCTGGAGTGCAATGGCACGATCTCGG
CTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATTCTCCTACCTCAGCCTCCCGAGTAGCTGGGACTACAGGGACGCGCCACCACGCCCA
GCTAATTTTTGTATTTTTAGTAGAGATGTAGTTTCGCCATGTTGGCCAGGATGGTCTCAATCTCTTGACCTTGTGATCTGCCCTCCTCGGCCTC
CCAAAGTGCTGGGATTACAGGCATGAGCCACTGTGCCCGGCCTCCATTTTCTCTACTAGAAGAGGGACTTTGCCTTTTCTACCACATTGTTTG
AAATATTTACAACTAGTAATAATTTTGCAATTACCAAGACCATAAAAAGAAACAGGTGGAGGGTATATAAATCAAATACTACTACAACAGACAC
GGAAGTG

Appendix 6

An example of qRT-PCR data and analysis.

For each sample the average Ct values for *TBL1XR1* are subtracted from those for *GAPDH*, to generate the ΔCt. The ΔCt for each sample is subtracted from the ΔCt for BM to generate the ΔΔCt value, which is then used to calculate fold change.

Sample Id	Ct values *GAPDH*	Average Ct values *GAPDH*	Ct values *TBL1XR1*	Average Ct values *TBL1XR1*	Δ Ct	ΔΔCt	2^-ΔΔCt
BM	25.42	25.27	31.6	31.61	6.34	0	1
	25.18		31.56				
	25.22		31.66				
PB	23.26	23.21	30.38	30.39	7.18	0.86	0.55
	23.18		30.28				
	23.18		30.52				
7551	27.92	27.99	34.77	34.95	6.96	0.62	0.65
	28		35.25				
	28.04		34.82				
8898	28.66	28.62	33.54	33.1	4.48	-1.85	3.61
	28.59		32.66				
	28.61		33.11				
4281	24.01	23.9	28.69	28.57	4.67	-1.66	3.16
	23.97		28.53				
	23.71		28.49				
7361	28.75	28.63	35.32	35.8	7.17	0.84	0.56
	28.52		36.56				
	28.62		35.53				
4995	21.95	21.96	26.01	26.04	4.08	-2.24	4.72
	21.88		26.03				
	22.04		26.09				
10086	21.05	21	28.42	28.47	7.47	1.17	0.44
	20.94		28.48				
	21.02		28.51				
4323	24.18	24.25	27.8	27.95	3.7	-2.63	6.19
	24.27		28.02				
	24.3		28.02				
5062	21.34	21.32	25.87	25.81	4.49	-1.84	3.58
	21.33		25.78				
	21.29		25.77				

5848	21.47	21.49	25.39	25.55	4.06	-2.27	4.82
	21.5		25.64				
	21.5		25.63				
3700	23.54	23.27	27.26	27.18	3.91	-2.42	5.35
	23.37		27.15				
	22.9		27.13				
4932	22.67	22.72	26.6	26.62	3.9	-2.43	5.39
	22.87		26.64				
	22.61		26.62				
4598	25.2	24.93	28.96	28.81	3.88	-2.45	5.46
	24.89		28.75				
	24.69		28.71				

Appendix 7

Demographic, cytogenetic, FISH and aCGH data for *ETV6-RUNX1* positive patients. Underlining indicates cytogenetic aberrations that were confirmed or refined by aCGH. Patients highlighted in bold have relapsed.

Patient ID	Age (yrs)	Sex	Karyotype	FISH results (where available)	aCGH data
2196	2	M	46,X,del(Y)(q?),der(9)t(9;13)(p1?;q1?), der(11)t(Y;11)(q?;q2?),der(21)t(12;21)(p13;q22),inc [cp6]	TEL/AML1: 2R1G1F (85%)	enh(1)(68.0[p31.2]-68.5[p31.2]), dim(9)(20.4[p21.3]-21.0[p21.3]), dim(9)(21.0[p21.3]-23.2[p21.3]x2), dim(9)(23.2[p21.3]-38.7[p13.1]), dim(11)(0.18[p15.5]-6.7[p15.4]), dim(11)(45.7[p11.12]-50.6[p11.11]), dim(11)(85.6[q14.1]-134.4[q25]), dim(13)(18.9[p11]-24.7[q12.11]), dim(17)(27.4[q11]-29.7[q11.2]),dim(21)(34.2[q22.11]-34.5[q22.11], dim(Y)(6.5[p11.2]-10.5[p11.2]), dim(Y)(16.4[p11.2]-22.3[p11.21]).
2327	2	F	46,XX[20]	TEL/AML1: 2R1G1F (35%)	dim(12)(0.5[p13.33]-21.2[p12.2]).

3428	9	M	47,XY,?r(6)(::p?25q2?1::),-10, add(12)(p1?3),add(18) (q23),+21c,+r.ish del(12)(p13)(TEL), add(12)(p13) (TEL),der(21)t(12;21)(p13;q22), (AML1+,TEL+),+21c(AML1+).nuc ish 10cen(D10Z1x2),16cen (D16Z1x2)[89%]/ 10cen(D10Z1x1),16cen(D16Z1x2) [5%]/10cen(D10Z1x2),16cen(D16Z1x1) [4%]	TEL/AML1: 2R0G1F 6p22.1: 1R2G0F (96%) ETV6 DC Split: 1R1G0F (86%)/1R1G1F (3%)/0R0G2F (7%) RUNX1 BA: 1R1G2F (87%)/0R0G3F (4%)/0R0G2F (5%)	dim(1)(114.0[p13.3]-114.1[p13.3]), dim(1)(187.5[q25.3]-188.1[q25.3]), dim(4)(108.8[q25]-108.9[q25]), dim(6)(25.0[p22.1]-29.5[p21.33]), enh(6)(29.5[p21.33]-74.8[q14.1]), dim(6)(78.7[q14.2]-170.8[q27]), dim(9)(134.5[q34.3]-134.5[q34.3]), dim(10)(21.7[p12.33]-22.3[p12.33]), dim(10)(120.9[q26.12]-135.3[q26.13]), dim(12)(0.03[p13.33]-16.2[p13.1]), dim(12)(20.3[p12.3]-20.6[p12.3]), dim(16)(25.1[p12.1]-25.2[p12.1]), dim(17)(36.7[21.1]-36.7[q21.1]), dim(17)(71.4[q25.2]-71.9[q25.3]), dim(18)(68.2[q22.2]-76.1[q23]), enh(19)(15.6[p13.13]-15.6[p13.13]), dim(19)(38.0[q13.11]-38.3[q13.11]), dim(19)(39.5[q13.11]-39.6[q13.11]), dim(20)(10.3[p12.2]-10.4[p12.2]), enh(21).
3602	11	F	46,XX,del(11)(q23),der(12)t(12;21) (p13;q22)ins(12;?)(q2?2;?),der(21) t(12;21)(p13;q22),+21,inc	TEL/AML1: Undefined ADD3: 1R2G0F (85%)/2R2G0F (13%) ETV6 DC Split: 2R1G0F (55%)/0R0G2F (34%) RUNX1 BA: 1R2G1F (70%)/1R1G1F (9%)/0R0G2F (29%)	dim(2)(106.1[q12.2]-106.2[q12.2]), dim(2)128.7[q21.1]-128.8[q21.1]), dim(3)(28.9[p23]-29.2[p23]), dim(3)35.6[p22.1]-35.7[p22.1]), dim(3)(113.5[q13.11]-113.7[q13.11]), dim(3)(196.9[q29]-196.9[q29]), dim(6)(106.9[q21]-107.4[q21]), dim(6)(109.3[q21]-109.4[q21]), dim(6)(111.3[q22.1]-112.4[q22.1]), dim(6)(155.9[q25.2]-156.8[q25.2]), dim(7)(44.8-[p13]-44.9[p13]), dim(10)(111.7[q24.32]-111.8[q24.32]), dim(11)(85.4[q14.1]-128.1[q24.3]), dim(12)(11.9[p13.2]-21.3[p12.3]), dim(13)(18.9[p11.1]-19.5[p11.1]), dim(13)(21.0[q11]-21.9[q11]), dim(13)(23.3[q11]-23.5[q11]), dim(13)(27.3[q12.11]-27.4[q12.11]), dim(13)(47.8[q14.11]-47.9[q14.11]), dim(15)(23.5[q11.1]-23.60[q11.1]), dim(16)(86.6[q24.2]-87.7[q24.2]), enh(21)(14.4[p11.2]-35.2[q22.11]).

3700	4	M	46,XY[20]	TEL/AML1: 2R1G1F (69%) ETV6 DC Split: 1R1G0F (67%) RUNX1 BA: 1R1G1F (70%) Xq25: 2R2G0F (93%)	dim(5)(138.9[q31.3]-140.0[q31.3]), dim(8)(17.4[p22]-18.6[p22]), dim(12)(11.7[p13.2]-15.9[p12.3]).
3975	9	M	47,XY,t(1;15)(p3?;q1?5),+21[4]	TEL/AML1: 2F	dim(3)(176.4[q26.31]-179.5[q27.1]), dim(4)(86.8[q21.3]-86.9[q21.3]), dim(5)(80.8[q14.1]-81.0[q14.1]), dim(8)(60.1[q12.1]-60.4[q12.1]), dim(11)(23.7[p14.3]-24.1[p14.3]), enh(12)(0.3[p13.33]-11.9[p13.2]), dim(18)(51.4[q21.1]-51.8[q21.1]), enh(21)(14.4[p11.2]-35.2[q22.11]), dim(X)(44.4[p11.3]-44.6[p11.3]).
4013	4	F	44-46,XX,add(11)(q23),del(11)(q2?2), add(12)(p?),+mar [cp4]	TEL/AML1: 84% rearranged Xq25: 3R2G0F (46%)/ 2R3G0F (24%)	enh(4)(181.5[q34.3]-191.3[q35.2]), dim(6)(57.2[q11.2]-57.6[q11.2]), dim(6)(99.8[q16.3]-170.8[q27]), dim(7)(142.3[q35]-142.3[q35]), dim(11)(98.2[q22.1]-134.3[q25]), enh(15)(22.9[q11.1]-23.0[q11.1]), enh(16)(83.6[q24.1]-84.2[q24.1]), enh(X)(122.8[q25]-154.4[q28]).

4093	4	M	46,XY,?ins(1)(p13),del(6)(q1q2), add(9)(p13), add(12)(p1),?del(21)(q22)[9]	TEL/AML1: 2R1G1F (94%) ETV6 DC Split: 1R1G0F (94%)/0R0G2F (5%) RUNX1 BA: 1R1G1F (98%)	dim(1)(87.3[q22.2]-88.4[q22.2]), dim(1)(195.4[q31.2]-196.0[q31.2]), dim(3)(0.05[p26.3]-0.15[p26.3]), dim(3)(25.7[p24.1]-26.2[p24.1]), dim(5)(63.2[q12.3]-68.8[q13.1], dim(6)(3.0[p25.2]-3.7[p25.2]), dim(6)(86.2[q15]-126.2[q22.33]), dim(6)(127.6[q22.33]-128.9[q22.3]), dim(7)(41.6[p13]-52.8[p12.1]), dim(9)(34.0[p13.2]-34.7[p13.2]), dim(9)(35.8[p13.2]-36.1[p13.2]), dim(9)(128.3[q34.12]-133.7[q34.2]), dim(11)(36.5[p12]-36.5[p12]), dim(12)(11.1[p13.31]-15.9[p13.1]), enh(16)(85.1[q24.1]-88.0[q24.3]), dim(18)(51.4[q21.1]-51.9[q21.1]), dim(19)(11.3[p13.2]-11.4[p13.2]).
4250	3	M	46,XY[13]	TEL/AML1: Undefined TOX: 1R2G0F (90%) p27: 1R0G0F (78%) Xq25: 2R1G0F (89%)	dim(5)(158.3[q33.3]-158.4[q3.3]), dim(6)(91.0[q16.1]-170.8[q27]), dim(8)(60.1[q12.1]-60.4[q12.1]), dim(12)(0.5[p13.33-17.6[p12.3]), dim(21)(34.2[q22.11]-34.4[q22.11]), enh(X)(72.2[q13.1]-128.7[q25]), dim(X)(128.7[q26.1]-128.8[q26.1]), enh(X)(128.8[q26.1]-154.4[q28]).
4281	4	M	46,XY,add(6)(q1?)[14]	TEL/AML1: Undefined MTAP: 2R0G0F (89%) p16/CEP9: 0R2G0F (93%) Xq25: 2R1G0F (98%)	dim(3)(178.3[q26.33]-178.6[q26.33]), dim(4)(149.7[q31.21-150.3[q31.21), dim(6), dim(6)(81.6[q14.3]-170.8[q27]), dim(9)(18.3[p22.1]-19.4[p22.1]), dim(9)(19.4[p22.1]-22.1[p21.3]x2), dim(9)(22.1[p21.3]-22.4[p21.3]), dim(9)(112.3[q32]-112.5[q32]), dim(12)(0.59[p13.33]-11.9[p13.2.]), dim(14)(60.9[q22.3]-61.2[q22.3]), enh(14)(68.6[q23.2]-70.5[q23.2]), dim(15)(57.0[q21.1]-57.1[q21.1]), enh(21)(35.3[q22.11]-46.8[q22.3]), enh(X)(89.6[q21.31]-145.5[q27.3]).

4323	7	M	46,XY,?del(8)(p1?),del(12)(p1?)[2]	TEL/AML1: 0G (98%) p27: 1R0G0F (82%)	dim(2)(24.9[p24.1]-27.7[p23.3]), dim(3)(113.5[q13.2]-113.6[q13.2]), enh(3)(120.2[q13.33]-120.3[q13.33]), dim(4)(149.3[q31.21]-149.4[q31.21]), dim(6)(52.2[p12.3]-52.3[p12.3]), dim(6)(100.4[q16.3]-114.6[q22.1]), dim(6)(161.1[q25.3]-168.1[q27], dim(8)(18.4[p21.3]-19.2[p21.3]), dim(8)(23.2[p21.2]-34.8[p11.23]), dim(8)(48.0[q11.21]-49.0[q11.21]), dim(11)(36.5[p12]-36.6[p12]), dim(12)(0.03[p13.33]-21.0[p12.3]), dim(12)(90.7[q21.33]-91.0[q21.33]), dim(19)(0.2[p11.32]-19.7[p11.1]), dim(19)(36.0[q12.1]-40.3[q12.2]).
4598	3	M	46,XY[10]	TEL/AML1: Undefined	dim(1), dim(2), dim(3), dim(9), dim(11), dim(12)(11.9[p13.2]-132.3[q24.33]), dim(13), dim(15), dim(16), dim(21)(35.3[q22.11]-46.9[q22.3]). (Suspected hidden tetraploid clone with deletions of these chromosomes)
4675	11	F	80-82<4n>,XX,-X,-X,-2,-6,del(6)(q?),-11,-12,-13 [cp4]	TEL/AML1: Undefined	dim(1)(191.2[q31.1]-191.4[q31.1]), dim(2), dim(3)(187.3[q28]-187.7[q28]), dim(6), dim(6)(109.9[q21]-170.8[q27]), dim(8), enh(9)(0.19[p24.3]-8.9[p24.1]), dim(10)(27.9[p12.1]-27.9[p12.1]), dim(11), dim(12), dim(13), dim(14), dim(15), dim(X).
4932	6	F	Fail	TEL/AML1: 2R1G1F (82%)	dim(8)(0.17[p23.3]-38.6[p11.22]), dim(12)(0.3[p13.33]-28.3[p11.23]), dim(12)(90.7[q22]-91.0[q22]).

4969	5	F	46,XY,del(9)(p2?),+21[4]	TEL/AML1: 2R1F (73%)	dim(3)(113.5[q13.13]-113.6[q13.13]), enh(3)(128.1[q21.3]-128.2[q21.3]), dim(6)(90.4[q16.1]-92.2[q16.1]), dim(6)(95.9[q16.2]-170.8[q27]), dim(9)(0.15[p11.1]-20.7[p21.3]), dim(9)(20.7[p21.3]-22.5[p21.3]x2), dim(9)(22.5[p21.3]-26.0[p21.1]), enh(12)(0.03[p13.33]-11.9[p13.2]), dim(12)(11.9[p13.2]-14.4[p13.1]), enh(21)(13.5[p11.2]-35.3[q22.11]), enh(X)(87.6[q21.1]-154.4[q28]).
4995	8	F	44,X,-X,dic(8;12)(p1?;p1?),inv(9)(p11q13)c[9]	TEL/AML1: 2R1G1F (72%) EBF1: 1R?G0F (86%) 6p: 1R2G0F (39%) p27: 1R0G0F (88%)	dim(1)(59.7[p32.1]-59.7[p32.1]), dim(1)(191.2[q31.1]-191.4[p31.1]), dim(1)(238.4[q43]-238.5[q43]), dim(2)(106.1[q12.3]-106.2[q12.3]), dim(2)(145.1[q22.2]-145.6[q22.2]), enh(3)(101.8[q12.3]-101.9[q12.3]), dim(3)(178.3[q26.31]-178.9[q26.31]), dim(5)(157.9[q33.3]-158.4[q33.3]), dim(6)(29.2[p22.32]-29.2[p22.32]), dim(8)(0.16[p23.3]-28.8[p21.1]), dim(8)(29.5[p21.1]-35.7[p11.23]), dim(12)(0.05[p13.33]-25.2[p12.1]), dim(12)(31.7[p11.22]-31.9[p11.22]), dim(19)(6.4[p13.3]-6.4[p13.3]), dim(X).
5053	1	M	47,XY,del(6)q(1q2),+21,inc[cp3]	TEL/AML1: Undefined	enh(2)(182.2[q32.1]-182.2[q32.1]), dim(5)(21.7[p14.3]-23.8[p14.1]), dim(9)(36.9[p13.1]-37.0[p13.]), dim(12)(10.2[p13.31]-19.0[p12.3]), dim(16)(81.5[q23.3]-81.6[q23.3]), enh(21).

5062	2	F	46,XX[11]	TEL/AML1: 2F (28%) ETV6 DC Split: 1R1G1F (44%)/2R1G1F (34%)/ 0R0G2F (17%)/1R1R0F (3%) MTAP: 2R0G0F (88%) p16/CEP9: 0R2G0F (79%) RUNX1 BA: 1R1R1F (54%)/1R2G1F (30%)/ 0R0G2F (13%)	dim(8)(15.9[p22]-16.0[p22]), dim(9)(16.3[p22.1]-19.9[p21.3]), dim(9)(19.9[p21.3]-22.2[p21.3]x2), dim(9)(22.2[p21.3]-23.8[p21.2]), enh(10), enh(12)(0.05[p13.33]-11.9[p13.2]), enh(21)(13.6[p11.2]-35.3[q22.11]), enh(X)(7.6[p22.32]-8.0[p22.32]).
5073	3	M	45,XY,der(12)t(12;15)(p11.2;q1),-15[7]	TEL/AML1: Undefined (Loss of ETV6)	dim(12)(0.03[p13.33]-22.0[p12.2]), dim(12)(44.3[q12]-44.6[q12]), enh(22)(18.5[q11.21]-18.9[q11.21]).
5848	2	F	FAIL	TEL/AML1: Undefined PAX5: 1R2G0F (97%) Xq25: 3R2G0F (7%)	dim(1)(143.5[q21.2]-144.9[q21.2]), enh(5)(0.7[p15.33]-0.9[p15.33]), dim(6)(26.1[p22.1]-26.3[p22.1]), dim(7)(137.4[q35]-158.6[q36.3]), dim(9)(36.8[p13.1]-37.2[p13.1]), dim(12)(0.6[p13.33]-11.9[p13.2]), dim(19)(48.1[q13.2]-48.5[q13.2]), enh(21)(35.2[q22.11]-46.8[q22.3]), enh(X)(122.9[q25]-154.4[q28]).
5849	3	F	46,XX[5]	TEL/AML1: Undefined	dim(3)(196.9[q29]-196.9[q29]), dim(6)(26.2[p22.1]-26.3[p22.1]), enh(8)(35.3[p11.24]-39.5[p11.21]), dim(10)(111.7[q25.2]-111.7[q25.2]), dim(12)(0.05[p13.33]-19.0[p12.3]), dim(X)(113.7[q24]-115.8[q24]), enh(X)(115.8[q24]-154.4[q28).

6834	5	F	47,XX,dic(4;12)(p1;p1),add(7)(q22),add(10)(q2?),t(12;21)(p13;q22),+21,+mar, inc[cp8]	TEL/AML1: 3R0G1F (100%)	dim(4)(77.4[q21.1]-82.6[q21.23]), enh(7)(72.1[q11.23]-75.7[q11.23]), enh(7)(97.1[q22.1]-101.6[q22.1]), dim(10)(103.9[q24.31]-105.3[q24.32]), dim(12)(11.2[p13.31]-20.7[p12.3]), enh(19)(0.3[p13.3]-19.7[p13.11]), enh(21).
6956	5	M	46,XY,-6,add(7)(p22),add(10)(q2), t(12;21) (p13;q22),+r[7]/46,XY,-6,add(8)(p1),t(12;21),+r[5]	TEL/AML1: 2R1G1F (70%)/ 2R2G0F (27%) p16/9cen: 2R1G0F (51%)	dim(6)(0.11[p25.3]-2.9[p25.2]), enh(6)(2.9[p25.2]-87.3[q15]), dim(6)(87.3[q15]-170.8[q27]), dim(8)(0.06[p23.3]-41.4[p11.21]), enh(8)(67.5[q12.3]-146.2[q24.3]), dim(9)(19.9[p22.1]-23.0[p21.3]), dim(11)(13.4[p15.2]-13.5[p15.2]), dim(12)(10.8[p13.31]-14.7[p13.1]).
7361	5	M	46,XY,t(12;21)(p13;q22),add(16)(p13), inc[3]/47,idem,del(12)(p11p13),+21, inc[4]	TEL/AML1: 3R0G1F (35%) /2R0G1F (33%) /2R1G1F (22%) EBF1: 1R?G0F (92%) TOX: 1R2G0F (94%) Xq25: 2R1G0F (98%)	dim(3)(178.4[q26.33]178.5[q26.33]), dim(5)(157.9[q33.3]-158.4[q33.3]), dim(7)(26.9[p15.2]-27.0[p15.2]), dim(8)(60.1[q12.1]-60.2[q12.1]), dim(9)(109.7[q32]-109.8[q32]), dim(12)(0.03[p13.33]-11.4[p13.31]), dim(12)(11.4[p13.31]-27.4[p11.23]x2), dim(12)(44.4[q12]-44.5[q12]), dim(12)(90.7[q22]-91.0[q22]), dim(13)(31.9[q12.2]-114.0[q34]), enh(17)(9.6[p13.1]-9.7[p13.1]), enh(19)(56.9[q13.41]-57.3[q13.41]), enh(19)(61.0[q13.42]-61.1[q13.42]), enh(21)(25.8[q21.1]-46.9[q22.3]), dim(22)(20.8[q11.21]-20.9[q11.21]x2), enh(X)(2.7[p22.33]-88.9[q21.31]).
7551	4	M	46,XY,add(12)(p1),t(12;21)(p13;q22), inc[2]	TEL/AML1: 2R1G1F (38%)/2R0G1F (33%)/2R2G0F (30%) TBL1XR1: 1R2G0F (77%) TOX 1R2G0F (88%) Xq25: 3R2G0F (50%)	dim(3)(178.2[q26.33]-179.3[q26.33]), enh(3)(196.9[q29]-196.9[q29]), dim(8)(60.1[q12.1]-60.4[q12.1]), dim(12)(9.7[p13.31]-15.5[p13.1]), dim(12)(90.2[q21.33]-91.0[q22], enh(15)(83.6[q25.1]-83.8[q25.1]), enh(22)(30.0[q12.2]-49.5[q13.33], enh(X)(98.1[q21.33]-154.4[q28]).

7552	3	F	46,XX,add(12)(p1),t(12;21)(p13;q22),inc[2]	TEL/AML1: 2R1G1F (30%)/3R0G1F (30%)/2R0G1F (24%)/3R1G1F (11%)	dim(3)(4.0[p26.2]-4.1[p26.2]), dim(4)(1.0[p16.3]-1.1[p16.3]).
8898	4	F	46,XX,del(12)(p12p12)[9]		enh(2)(88.1[p11.2]-88.9[p11.2]), dim(3)(178.2[q26.33]-178.3[q26.33]), enh(4)(89.0[q22.1]-89.2[q22.1], dim(4)(149.1[q31.21]-150.2[q31.21]), dim(7)(26.9[p15.2]-26.9[p15.2]), dim(10)(111.7[q25.1]-111.8[q25.1], dim(11)56.0[q12.1]-65.1[q13.1]), dim(12)(4.7[p13.32]-16.1[p13.1]), dim(13)(43.7[q13.3]-43.9[q13.3]), dim(13)(76.7[q21.33]-76.9[q21.33]).
9150	3	M	46,XY,ider(21)(q?10)t(12;21)(p13;q22), inc	TEL/AML1: 2R1G2F (51%)	dim(8)(27.4[p21.1]-27.5[p21.1]), dim(9)(91.6[q22.31]-91.7[q22.31]), dim(9)(95.2[q22.32]-95.3[q22.32]), dim(22)(34.2[q12.3]-34.3[q12.3]).
10026	4	F	46,XX,del(11)(q2?2)[cp4]	TEL/AML1: 2R1G1F (59%) MLL: 0R0G1F (55%) T Xq25: 3R2G0F (50%)	enh(2)(0.02[p25.3]-15.9[p25.1]), dim(3)(137.8[q22.2]-137.9[q22.2]), enh(3)(163.9[q26.1]-164.1[q26.1]), dim(6)(29.2[p21.32]-29.2[p21.32]), enh(7)(118.3[q31.2]-118.6[q31.2]), dim(9)(21.9[p21.3]-21.9[p21.3]), dim(11)(77.7[q13.4]-134.4[q25]), dim(16)(6.9[p13.2]-7.0[p13.2]), enh(X)(122.8[q25]-154.4[q28]).
10073	2	F	FAIL	TEL/AML1: Fusion by RT-PCR TOX: 1R2G0F (83%) PAX5: 1R2G0F (24%)	enh(6)(131.0[q23.1]-131.3[q23.1]), dim(8)(60.1[q12.1]-60.7[q12.1]), dim(9)(36.9[p13.1]-37.2[p13.1]), enh(12)(0.8[p13.33]-1.5[p13.33]), dim(12)(1.5[p13.33]-2.0[p13.33]).

10086	4	M	49,XY,t(3;15)(q27;q1?5),+10,del(12)(p11 p13),+16,add(18)(q2),+21[19]	TEL/AML1: Fusion by RT-PCR p16: 3R3G0F (12%) TBL1XR1: 1R2G0F (84%)/ 2R4G0F (9%)/ 2R2G0F (5%) ADD3 3R2G0F (70%)/3R3G0F(18%)	dim(1)(227.7[q42.13]-237.2[q42.3]), dim(3)(177.8[q26.33]-180.6[q27.1]), dim(3)(181.9[q27.2]-182.2[q27.2]), dim(4)(162.2[q32.1]-162.3[q32.1]), dim(8)(7.0[p23.1]-8.1[p23.1]), enh(10), enh(12)(0.03[p13.33]-11.8[p12.3]), dim(15)(38.3[q13.3]-41.3[q13.3]), enh(16), dim(16)(66.1[q22.1-66.5[q22.1]), enh(21)(13.4[p11.2]-35.2[q22.11]), enh(Y)(13.6[p11.2]-13.9[p11.2]).
10628	5	F	46,XX,del(3)(p13p26),del(6)(q1q2),add (10)(p13).ish.t(12;21)(p13;q22)[5]/ 90,idemx2,-11,inc[6]	TEL/AML1: 2R1G2F/2R1G2F TBL1XR1: 2R1G1F (59%) BCR/ABL: 3R2G0F (50%) MLL: 0R0G1F (55%) Xq25: 5R5G0F (10%)/4R4G0F (5%)	dim(3)(175.5[q26.32]-179.5[q27.1]), dim(6)(80.6[q14.3]-170.9[q27]), dim(9)(19.6[p22.1]-24.8[p21.2]x2), dim(10)(46.3[q11.1]-47.1[q11.1]), dim(12)(0.07[p13.33]-11.2[p13.2]), dim(13)(43.5[q13.32]-43.9[q13.32]), enh(13)(70.8-[q21.31]-114.1[q33.1]), enh(21)(35.3[q22.11]-42.2[q22.3]).
10928	2	F	46,XX[11]	TEL/AML1: 2R1G1F (63%) 6p22.1: 1R2G0F (85%)	dim(3)(128.4[q21.2]-128.5[q21.2]), enh(4)(63.0[q13.1]-63.2[q13.1]), dim(4)(153.6[q31.23]-155.5[q31.3]), enh(4)(155.5[q31.3]-191.2[q35.2]), dim(6)(24.2[p22.1]-30.7[p21.32]), dim(9)(127.0[q34.12]-131.1[q34.2]), enh(10)(0.1[p15.3]-96.2[q23.33]), dim(10)(96.2[q23.33]-98.7[q24.1]), dim(11)(60.1[q12.3]-68.3[q13.2]), dim(15)(63.3[q21.3]-69.9[q22.31]).

11469	6	M	46,XY,t(3;9)(q29;p21)[11]	TEL/AML1: 3R1G1F (64%)/ 2R1G1F (20%) PAX5: 1R2G0F (30%)	dim(3)(34.9[p22.1]-35.6[p22.1]), dim(9)(36.9[p13.1]-37.4[p13.1]), dim(10)(18.6[p13]-19.0[p13]), dim(11)(102.5[q22.2]-115.4[q23.3]), dim(12)(11.9[p13.2]-11.9[p13.2]), enh(12)(129.2[q24.33]-132.3[q24.33]), dim(19)(38.3[q13.11]-38.6[q13.11]), dim(19)(39.0[q13.11]-39.8[q13.11]), dim(19)(40.6[q13.12]-43.6[q13.12]), dim(19)(46.6[q13.2]-48.6[q13.2]), dim(19)(51.0[q13.32]-53.3[q13.33]), enh(21), enh(22)(23.9[q11.21]-24.2[q11.22].

The normal clone has been omitted from cases with an abnormal karyotype

Appendix 8

The copy number variants observed in *ETV6-RUNX1*, iAMP21 and unclassified patients, using aCGH.

Chromosome	Start position (Mb)	Stop position (Mb)	Number of patients with CNA at these loci		
			ETV6-RUNX1 patients	iAMP21 patients	unclassified patients
1	16672053	16954894	1		2
1	16967194	17021370	2		2
1	25277399	25441068		1	
1	145820605	145951498	8		5
1	145951498	146473785	3		4
1	149358185	149432438			3
1	193470194	193543089		1	
1	195487682	196000000			1
1	244908381	245210850			1
2	88935100	91321535	30	15	24
2	242633011	242784096	1		
3	59707111	61197806	2		1
3	109190890	109262444	2		1
3	163957219	164112940	16	1	10
3	178756221	178927205	1		
3	196914968	197000000			1
4	69231557	69789502	9	3	11
4	70305861	70473970	2	1	3
4	173099471	174343913	2		
5	2347965	2478306	1		
5	68566952	70842896	1	4	1
6	97634	347075	1		3
6	8154909	26004621		1	
6	26371411	26571481			1
6	29962849	30062967			3
6	32558677	32777081	3	8	4
6	124167172	125188150			1
7	18284206	18655919			1
7	37732921	38149800	32	7	28
7	109897373	110795977	1		
7	112100408	112253417	1		
7	141494693	142019122	33	5	19
8	39356595	39505315		1	3
8	6867165	8221903		1	1
8	20032880	20092040	2		
8	39281207	39535714	22	1	8
8	71105403	71152232			1
8	100082337	100434742			1
8	113298265	114907920	1		
8	115260947	115261006	1		
8	115701268	115739178	1		
9	9783019	9793979		1	
9	24469961	24495824	1		

9	29908743	29996163	1		
9	38799072	69056983		7	2
10	46431590	47172593	3		4
11	54907484	55580831	10	1	3
13	19343261	19380767	1		
13	47883983	48036764		1	
14	18504575	19768421	7	1	1
14	21575909	22046156	26	7	24
14	105275857	106092884	30	16	29
15	18362555	21195262	13	6	12
15	21795947	22384475	1		
15	28441169	30648918			1
15	32414308	32676783	3	1	3
16	14956177	15023080	2		
16	21258019	21476732			1
16	32105404	33758141			3
16	34242095	34634745	1		1
16	54336421	54437612	3		2
16	68667006	68838427	1		1
16	73809039	73855302			1
17	21826262	21857788			1
17	26574318	26663366		1	
17	33794970	34031410			1
17	36742924	36936111	1		
17	40926154	41041718	1		
18	1715225	1828941			1
19	20463727	20523715			2
22	18084715	18127159			1
22	20586004	21597258	30	4	11
22	22583946	22905480	7	2	11
22	23913676	24344072	1		
22	37671764	37727445	1		3
X	140078611	140271252	2		
Total			325	94	263

Appendix 9

The results of the confirmatory FISH screen for CNA seen in *ETV6-RUNX1* positive patients by aCGH

Patient	6p22.1 deletion	*ADD3* deletion	*EBF1* deletion	*PAX5* deletion	*TBL1XR1* deletion	*TOX* deletion	Xq gain
3428	Confirmed						
3602		Unconfirmed (probe overlapped deletion by 44kb)					
4013							
4250			Unconfirmed (deletion downstream of probe)			Unconfirmed (probe upstream of deleted region)	Confirmed
4281							Confirmed
4995	Confirmed		Confirmed				
5848				Confirmed			Confirmed
7361			Confirmed			Unconfirmed (probe upstream of deleted region)	Confirmed
7551					Confirmed	Unconfirmed (probe upstream of deleted region)	Confirmed
8898		Unconfirmed (probe overlapped deletion by 38kb)			Confirmed		
10026	Failed						Confirmed
10073				Confirmed		Confirmed	
10086					Confirmed		
10628					Confirmed		Confirmed
10928	Confirmed						
11469				Confirmed			

Appendix 10

The *PDE9A* gene sequence, from genomic position 42.954831 to 42.968027 Mb (intron 1), and the position of the MCC markers 1 to 9. Markers 1-6 were designed approximately 2000bp apart, and markers 7-9 were designed 200 bp apart, between markers 2 and 3.

Marker 1 (M1) 42954961-42955232 Mb
Marker 2 (M2) 42957043-42957316 Mb
Marker 7 (M7) 42957596-42957890 Mb
Marker 8 (M8) 42958229-42958467 Mb
Marker 9 (M9) 42958734-42959003 Mb
Marker 3 (M3) 42959395-42959895 Mb
Marker 4 (M4) 42961638-42961869 Mb
Marker 5 (M5) 42964407-42964675 Mb
Marker 6 (M6) 42967283-42967567 Mb

CTGCCTGATATTTTTAAAGCCTCAGTCATTTTCATTCTTAGTTCTTCAGTTTGCAAAGG
GCCGGCCTGTTGTGATGGGTTTATTGTTGCATCTTCTGAAAAACTTTGTGAGTTCCC
GAAGGCCTCTCCAC***GTCTTGAGTCATGGCTGTG***CTGCTTTGCGAAGCTGTGTGCAT
GTCTGTGGTGTAGCCTGTGTACAGCCTGTGATGCGTCGCAGCGGAAGGGTAGGGG
GCTGCCCAGG***GAAGACAGATGCGCCAGG***TTCCGCTCCACGGAGCTCAGCTACCAG
CCCTGGTGGAGAGCGCAGATGGGGCCGATTTGTCATTTGTCTGCGCTTGAGCCCAG
CAGCTGTGTGGCCTCCCTGGGGCAAGTCATTCACCCTCTCTGTG***CCTCAGTGTCTC***
TGTGAAGTGGGGGGAAGTTTTTGGAAGATTAAGTGGGCTAATACGCGTCAAATGCT
AGAACACACTGGCACTTGGTAAATGCTATATAAGTGTTATCATTTTTGTTTGTTTATTA
TTTATTTATCCATGTCATGGATTTAAGGCAGTCTAATCTCAGTGCCTGGTGGAGACCC
GGGCGCCACCCCCCCAGGTTGTCAAGTGGTCCCTGTGCCTCCCACTATCGTCCCCA
GCACGGTGCCGTGTCCGGGCGTGCGTGAAGGGCCAGCTCACCTCACAGCTCTGCT
TCCCAGGTTAGCTCCCATCCACCCTCCCCCTACCTGTTTCCTCACCGAAGGAAGGG
GCAGAAAGATGCCCTCCATGCCCCTTCCGACTGTTGCAGAGGGATGGAGGACTATG
GTACATACACACATATACACGCACACACCCACCACACACACACACCATGCACATGCA
CATCACACACACGCACACACCATGCACACACACCACACACGTACACACCATGCACAT
CACACACAAGCACACCACACACACCAAGCACACGCACATCATACACATGCACACACA
CCACACACACCACACACCACACACACGCACATCCCACACCATGCACACACACACCA
CGCACACGCACTGCACACACATGCACACACCATGCACACGCACATCACACACTTGT
ACACACCAGGCACATCACACACAAGAACGCACACCACACACAAGGACACACACCAC
ACACACACACCAAGCACACACACACCACACACACGCACACACACGCACACACCCAC
CACACACCACACACACATGCACATCACACACGTACACACCACACACAGCACACACAAAA
ACACACCACACACACACCATGCACATGCACATCACACACATGCACACCACGCACACGC
ACATCACATGCACACGCCACGCGCCTCACACACAAGAACACACACCACACACACGC
ACACCACACACACGCACATATCACACACATGCACATCACATACATGCACACACCACA
CTTACACACATGCACATCACACACAAGAATGCACACCACATACACGCACATCACACA
CATGCACACACCACACACATGCACATCACACACCTGCATATACATCACACACACATCAC
ACACATCCACATGCACCACACCCACCACACACATGCACGCACCTACCACACATGCAT
ATCACACACATGCACACACACCACACACAAGCACATCACAAACACCACACAGGCACA
TCACACACATACATACACACCGCGCACACATCACACACATGCACACACACCACACAC
GCACAGTATGAAGATGACAGAGAAGTCAGCAGAGCCCCTGGGGGTTTCAGGAGTTG
AGTGGGGACGAGGCAGGTGGCCACACAGCTCGGAGGGAGGGCAGAAGGCACCCG
TCTTCGCAGAGTCACAGTGGGGCTTGCCTCCTATACGTGCCGGAAGGGTGAAGCCC
CTTATCTCATTGCCCGATGTGATGAGGTTTGAAGTGGTTGAGCTTGGAGGGTATGGG
CTGAGCCCACAGCAGACAGGCCTGGCAGTGCCCCCCGTCTTCCCACCCTGCGGTC
ATGCAGCCCGACTCTTGGGGGCGGGGTGCCACCGTGATGTGCCGTCACTTGCAGC
AGTATGTCCTCATGCCAGGCTGAGACCAGCACTGTGTGACCAGCCTGCCCCATGTC
CCTCAGCGCTGTGTCACACTCCACACTGAGGCCCCACTGAAGCATGCCAGGCCCCA
TCCTATTTCTTCCAGGTAGTCGTCTGCCCAGCACCTCCCCTCCCAGGCCTTGCCCTA
AAGCCCAGGCCTCT***CAGCATCCAGGACTCTGG***AATCTCCCAGGTGTTGATGGCTGA
GGTCATCAGCTGGATGCCAGGACACCCAGGATGCTCCAGGTCCTTGCTGGGGTCTG

GAGTTTGTGATTCAGCCTCAGTCCCTG*TGGTGAGTCCACCATGTG*TCCTCAGACAAT
GGATCAGAGAGCCCCCAGAGGACGGCCCGGGGCCGAGTCGGGTCCAGATGCCAA
AGGGGCTTCTCGAGGCATCTCCAGCTCCCACCTGCCCCTCGGAGTG*CCAGGCTTG*
*GCTTCTGGA*AGAGAAAGCTGGGGGTTTAGTGACTCCAGCCAACTGCTGTGAGCCCA
AGGCATCAAACTAAAAATAGCGAGGCCAAGAGGGAGTCTTTGTGTTCGGCTCAGTG
GGCAATGGTGCCTCACCTCGCAGCATGACACAGGGCTGATGACACGGGGCACGTT
GGTCACAAAGTGCCAGTCTCAGCCTGGCACAGGGACGTGCCACCTGCAGGTGATG
ACACATCACAGTGGCAGCCGCTCCTGATGGTGCTGCTGCATTGAGTAGGAGCCCAG
ATCCAGGGAGA*CAGTTGCAGCAGATGAGG*CAGGAGGCCAGCGAGGGGTTCTACCT
GCTTTCCGGGCTGGCCTCTGCATGGACACAAAGCAGAAAGGGCATAAAGGTGACCC
TGGCCACCAAAGCAGACCTGGAGCA*AACTGGACCAGCGCAAGG*CCCCGGTGCCA
GCACTCAGAGACATTTCTTTTCTGGCAGGAAATTTGTGCTTCATTTTCTAATGAAATT
CCAAGGCAATCAAAAGGCAAGGAAAGAGGAGGGAAAGGAAGGGAGGGAGGGAGG
GGGTTCCATT*CCCATCCGAATGGAAGCC*CCCAAGGCTTCTGTAGCTCCCCCCATCA
CCAACTTCACTCACCACTCTTATCTGATGCCTCCTATTTGGGTACACACGGCCGGGT
CGGCCTTGCGAGAGACTCAGTTTCCTCTCGACTTCCGCCCGCCTGCACTGTCACCA
CGTCCTGCCATTGACCTCTGTTCATCCACTCATGGTCACTGCTGTGAGTTGAATTGT
GTCTCCCAAAAAAGTGGATTTCTGTGCTAACCTGTGAATGTCCCGTTAGTTGGAAAT
AGGGTATTTGCAGACATAATCACATCAAGATGAGGTCATACTGCATTGGGATGGGCC
TCAAGCCCATATGCCCGCCATCCTTCT*AAGGAGAGGGAAAGGCAG*ACACACATTTG
TCCATGGAAATGAGAACGGAAGACGTAGAGACAGAATCAGGCCATGTAAAGACAGA
CTCAGAGACCAGAGTGATGCATG*TACAAGCCAGGCTCTGCG*GGCAGCCCCTAGAA
GCTGGAAGAGGCAAGCCTCCTCCCTGGAGTCCTGGGAAGGAGCCAGCCCCGCCCA
TGCCTCGAGCTCAGGCTTCAGGCCT*CCGGAGCCAAGAGAGATA*AACCTCTGTGGT
TTCATGTCACCCAGTTTGTGGTCTTTGTTATAGCAGCCCCAGGACACTCAGGGGGTT
GCCTTGTCCACCCCCATCCTGGCCCTCTGAGGGACTGGTTGGCTCTTAGAAGACCG
TTCAACCTGCCCTCTGGTCCTCCTCCCCCACCCAGAGACCCCTGGCTACACCCTGC
CCAAGGGAGCTCACCAAAGCCCACAGGGTCCCTGTGATCTTGTCACCTGGGCTCCC
ACAGGCCAGCAGTGTTGGAACCTGGACATC*TGGCTCCAGCCTTGGTG*GCCACCCC
CGACCTTGCTCTGCTCCTACAGCTCCCGGCCTTCCCTGCCCTGGCCGGCCCTTGCT
GAATGTCCCCGCTCCTAGGAAGCCGGATGTGTCGCGTGCA*GGTGCCTCTCTGCACT*
*G*TAATAACTCCTCACCAGGCACCCTTGGAGCCCAGGATCAAGTCACTCACCTTCCAA
CCCTGCCCCTGCCCAGAGACTGGCACGAGAACCAGTAACAGTTGTTGAATAATGAA
*CAATGGATGCATGCAGCTC*ACCCGTGAGAGGGCGTTGCTCGGCTGGGCCAGGGAT
GAAGCTTCTCCCGCGTGGCTCCACCGTCTATCCCTCAAGCTCCTGATGCTCCCCTG
GGCACCTCTGCGCCATGAGGCCCCGTTGAAGTCAGTTGGGGAGGCAGTGTGGACC
GTGAGAAGGGAGTCCTAAGCACAGGAAACGCTGCCTTCCTGGGGCCCTAAAGAACA
GGCTCACCTTCCCCAAGGGCCTTCAGGCCCTGACATCTAAAAAGATGTTCTTTATCA
AAGGGGAAAGCAGCCCTGGGGTGGGGGCGTGGCCACCAGTGCAGTGGGCGTGGC
CACAACCGAGAGGGCGTGGTCATTCGTGCAGTGGGCGTGGTCATCAAGAAAGGGC
GTGGCTGCAGCCGGAAGGGCGTG*GCTATCAGCTCCAGAAGG*CCCTTCTTGCTGTC
CTTGGACTTGGGTTCTCCTTGTCGACGAGGATCTCAGCTGGAGCTGCCAGGGGCAG
GGCT*GGCACCAAAGCACTTCTG*AAGTCTCCCCGCCTCCCAAGTGTGAGGCTGTGG
GGAAGGAGGGAGGGCAGCTGGCTTGTTTCTGGATGAGCC*CATCTGACTCAGCAGC*
*TG*GGACCCAGCAGGAGCGCAGAGCAGGTCCTGAAGGTGCCGAGGCCCCACCCGC
AGCTCCGTCTCAGCTTTGTGAGGGCTGCACTCCCCCGTGTGCATCCGGCGTCACAA
AGGACCACAGGCCCGGCGGCCTCCACAGCAGACACCAAGGCCCTCGCGGTCTGAA
GGCCGAATTCCCAGGTCAGGCTGCAGCAGGGCTGGTTCCTGAGGCCTCTCCTGGG
CTGGCACCTCCTCCCCGTGTCCATGCAGGGCCGAGCCTCTGCATGTCTGTGTCCTC
GGCGCCTCGTTTTACAAGGACATCTCATTTTAAAGGCCCGGTCTCCAAATAAGGTCA
TATCGTGAAGTGCTGGGGTCAGAACTACAGCATGAATTTTGACGGGACACAGTTCAG
TCCATAGCACCCGCTTTGGGAAGTGGGAATGACCTCTCCCTTCCCCATCTTCTGGCC
TTTGTAACCCTAATGTGGGAAAGCAGGAGGGATCCAGAACGACATTGGGTCACCAA
AGAGAGCATCGCAGCCCAAAAGGTTGAAAGGCGGCACCAAGGTCCCGCGCCTGGT
GACCGCCGACTCGCCCTCCCGGGCTCCAGGCTGTCCCCACTCCCCCCCGCCCCAA
CCCCAGCATGTTGAGGCGCCTGGTCCCAACTGCCCATTCACTCCCTGTACACAGAC
TGCTTCTTTGTCCTCCAGCCAGGGGTGAAGGGCCATGCTGGCGAAACAGCAAACAA
CCCCAACATGCTTTGCAGCCTGGCCAACCACATGACCAGTTTGAGGAGGTTGGCTG
GAGATCACAATGAGGAAGTTGGATGAGGGGGTCCAACTTCTAGCAGCCCATAAGCC

TATCCATTCAGACACCAAATCCATGGAAAACAGAGAGACAATATACTGATTTTCAAAA
GGGAGACATACCCCAGCAAGTACAGACCAGGGAACAGGGTGCTGACGTGGCCACA
GTTCTTAAACTACTCGACTTCATGGATGGGGCACAGTTTTTTTTAAAGGAGGATTTAG
AATTCTATTTGGCAGGTGGCATGAAAAAGAGTTTTGAATAGAAACACACCCCTCCCT
CCCAGTTGCCTGCCCAGTATGTTTGTTTTGTTCTCTTGAAACAAAGGAAGAGAGTGG
TGAGGAGGGAGGGAAGATGAACCAGGAAGAATGAAAGCACTTGGCCTGCTTCTCAG
ATGGTCTTGGACTGTCTCCGCTCACAGGAGTTGCAGGTTTTTGGGTTTTATGCAAGT
TCGGATCACAACAGGCCTCTTGCAAGAACCAGGCATTACCTTTATTCCAAAAGCCTT
GCTGGGCCTTCTTGAAAGGACTGGCAGCCACCACTTCCCTAGAAACCCGTGGCAAG
TTCCCTCCTCCCGACCCCCTGCTGTTTGGTCCAGGACCCCCTGCAGTCCTGACTGG
GACCCACTTGGCCCTTCTCAAGGGACCCAGCACAGGACGTGTGTTTCTGCTTTGAT
GAACCGCCCTTTCCCCCCTGCAGTTCTAGTGGAATCTCCCCTCCCTTATCCCATTCC
CACTTGGAAATGATGATGATGGAATGGGCGATGCCGGCTTACCAAGACATCCAAGC
TCTCTGGGTCGAAGTTGGAGAGAGGGGTGCACAAAAAGGATTCCACTCCTGTCTGC
CTTCCCCTGCCTCCCCGCTGGGGTGTCCCTTTCCCATGGTTTCCTCTAAATCTTTTC
ACCCATGTCCAGCCACACCATTCTCACATACTGCCTTGAAACTGGGCTCGAAGGGG
CTGCGTGCTGAGAGAAGGTCCTCTCCCCCAGTGATCCTCCAGAGGGGCTGCCGCC
TGGGTCCCCCGAGCACCTCCTACCCCACCCTCCCCATTCCTGCCATCCCCAGGGTC
CAGGGAGCCCAGATTCCAGGGAAGGGTTGCATTAGCTCCCACTCGGAGTCCTGATG
CAGCAGAGACAGACAGAGGCCCTGGGAGAAGTGAGCATGAATTATTAAGACAAGAC
AAGGGTGAGGCCCCAGAGAGGGGGTGGCGGAAGGGTCATGTTCATTGCAGAGCTC
CATCTCAGCACACTTCCTGTTCTTT***GCAGCGAGAGTTGCTTCG***AGCTTGAACCGCGT
ATCCAGGAGTCAAGCAGATTGCAACTGGCGAGAGGCCTTCAGAAATGCCCCGTGAG
A***GTCCTGTGTGGTTCGTCG***ATTTTTGCATTTTCAGTCCCCTGTGATCCATTATTTATA
ACAGTGGAGATTGGCCTCAGACACTAGCAGTGAGGAAAACAAAAGCGAAGCTACGC
AGAAAAATGA***CAAGAGTGATGAGCACAGC***AGTCATGACAAATGAGCCCTGTGCGGA
GGCCCGGGATCCGCGCAGATGCCGGCGCGGGGGAAATGGGCCCTGAAATCCCAC
CGTCAGGCCAGGCAGCTCTGAGCGTGACCTGGAGGGCTGTTCAGACGGTCTGGGT
AGCCGTGTCCTGCGCATGAACATCCTCCGTCGGGAGAGGAATTCCCCACGGATTAT
CAGAGCTGCTCCCTCCACCCCCCGCCACGTCCCACGCGGGCCACATCAACTCCCTC
TGCAGCCTCTGGCCAGCGGCTGAGCCCTCCGTGTCTCCCCTCGTTAATGCCTCCTT
CACCATCCCCTCCTGAAGTTTCCCCCATTGCATACACGCGCTGAGGCCCACCCGGT
ATCAAGGACTCCCATTGCTTGCGAAAAAGATTCCACCCCTCTTAGAACAGAGACCAG
GGCCGCTGTAGCAAATGGCCATAAATGCCACAGCTTAAAACAACAGAAACGGATTAT
CTCGCAGCTCTGGAGGATGGAGTCCAAAATCTGAATCGCTGGGCTGAAATCCAGGT
GTGGGCAGGGCCGCGCTCCCTCTAGAGGCTCCCCCGGAGATTCCCTTCCTTGCCT
CTTCCAGCTGCTGGTGGCTGCCAGCAGTTTGGGAATTGCGGCCGCATCACACCACC
TTTCTGTTTGTTGTTGACATCCCCGCCTCCCCTGCCTGCGGGGTCTTAGATGTCTCT
CTCCTTCCCACTGAGTTTCACTCCACATTTGAATTGGATTAACTCATGCCATGTTAGG
CAAACGTGCCCCTCAAATCCTTCCACTTAACAGACATTTATTGAAGGTTCCTGTGTGC
GGGGCCCAAGAGAAGGGACATTAAATTAGGAGATGAACCTTGGGAGGCCATGGCAT
GCTTCGGAGCCTCGGCGTCAAAATGACAACCCACACCCCACTTGGCGGTTCATCTG
GGGCCCTGGAATCAGCAGACTGAGGCTGGAACCCAGCAATGGCACATGCTGGCTG
TATGTTCTCGGGCCTCAGTTTCAACATCAGTGACATGGGGATAGAATCTGCCCAGCG
GAGTTAAGTCACCGCATCCCGGAGAGCCGAGTGTCTCATAGAGGCTCCTGGTGTGG
TCTCCATGCTGCTGCTCAGCACGGAATGAAGAGGTTCAGGAACTTGCCCACAGCCA
CGGATAATGAGGGACAGAGCTGGAATTGAGCCCGGGAACCTGCACACACAGCCCA
GTGTGACCGTGCCTTCTAGAGAGAGGCCAGCCCTGGCACCAGGTGGGCATGCCAG
CATCACCTTCAAGGTGTGCTTTAAGGATCCAAGATAATGTGGGTGCTTCAACAGATG
CATGGCAATAAAAGGAAAGCACCATGTTTCACCTTCACACACACACACTTACACATTC
ACACACTCACATACTTACACACATTCTCATACACTCACACATATGCTTACACACATTC
ACAGGCTCACACATGCTTACACACACATTCACACGCACACACATTCACACGCACACA
TACACTTACACACATTCACACACATACACTTACATTCACACACATACACTTACATTCAC
ACTCATACACATACATACATTCACACACATATTCACACATTCACACATACACATTCACA
CACATACACTTAGACACCACACTCACATTCACACACATTCACATACATTTACACACAC
ATCCACACACACATTCACACACATTCACATTTACATTCACAAACATTCACACATACATT
CACACGCACTTACAGTCACACATACACTTATACACAGTCACACAGACTTGCACATAG
ACTTACACACATACACTTACACACGCACTCACACCTTCACACACACAGGCAATCACA
CATTCACACTCACGTACACTTACCCACACATTCACACATACACACACCACATACACG

CATACACTTACACATGCACTCACACACTATCACATTCACACACACATACACTTACAAA
CTATCATTCACACACATACACACATACACATGCATTCACACACATGCTCACACACTCA
CACATACACTTACACTCACAGTCACATACACCCACATTCACACACATACACTCACATT
CACACGCACATACACTCTGACACACATATACATACATTCACTCACACATACATTCACA
CACACACGGATGCATTAGTTTTCAAGTCAGGATCCTAGCACAGCCCACACCCTGCAT
TTGCTTGGTGTCTCCTGAGGTCTCTAGTTATACTTCCCATCCCTTTTCTTGCCATGTA
TTTCCTGAACACATGTTATCTTTGATCCTTAAACCATGACTTAGAGATACAGATGTAT
GACAAGCTGAGAGATGATGTGAGACTGTTGTTAATTTTTTCTCCATGTACTAATTACA
TTGTGGTTATCTTAACCACAAAGGAAGCGTTCTACTCCAGAGAACCGCGTGGAAACA
GTGCAGATGGACTAGCACGATACCTGAGAGGAGCTTTAGGATAACTCCGGTGCTGG
GGAGGTTGGGGCCCAGTTCAAGCAGGAGGGGAGGCGTGGGGACAGGTGTTGAAG
CTGTGGGTGTGC***AGCGGAGGCTTCATGACG***CTTTCCTCCTGCTCTTGTGTACGTCT
GAGCATTTCCATAACAGGGAAGTGTTTTGTTTTGTTTTTTAAATAACATTTTAAATGCC
TCTG***GCACATGGGAATCCATGG***TAATTTCTGTCGTTGTTGTATATTCTTTTTCAAATG
CAAGGACCTGAAAAATATCTCCTAAACTGGAAAAGGAAATGAGGCTGAATACAGATT
TTATTTTAGCAAATAGAGAAGCAGTGATTGT***CTGGCGGTTAACTGTAATC***CATTCTGA
GTGTGACTTTACTGTTAAGTATTCTAAAAACAAGAGACCAAACAAAATGAACAATATG
TCCAAATGTGTATATATATAATTTGTAACTAGAGGTAAATTGCAGATGAATTGCAGAT
GTACACAGACAGATGGATGGACAGATGATTGATATAGATTGATTGACAGATGATTCA
TAGATAACTTGCAATGTTTCTTCAGTGAACGCATATTTATTTCTTGGGTAATAAGAAAA
AACAATAGAATGTGGAATTTGGGGCTTGGCCTCTTTTAGCTCATTTAAAAAAATATTA
AACACTTGGAATTATCGAGCGCTGAAAAAGGCAATAGAAAATTCACCTTTTAACTCCC
TCAAAGATAGAAATGTGTGAATCTGCTAGTGGGTTGCGTTTGGCTGGTCCTGCAGGC
TGATTTGGTCTGTTCTCTCTCCCAGAGAGATCCCCAAGCCCGGCTGAGATGTGCCT
GTTTGTCTTGAGGGCAGACAAAGCCTAGCAAAATCTTATGGAAGCTGATATAAAAGC
TTAATAGCTTATAAAAGCTCTAAAAATGATTCACAGTCTAGGATAGAGAAAATAAAAG
GAATTTGAATTGGTTTATTTTCCTTTAGAGGAGCCTAAATCAATATTTATTCTGAGTAT
GCAATCAGAAATATCAGTAATTTGAGAACAAGGCAAACAGAGCTGCACATCCCGTTA
TGGCCACGAGCCCACCTCGCCTTCGTTCAGACCACAGCTGAAGGTCTGACTTAGGT
GCAGCACGCACCCCTGGAGCTGAGGGCCTGGAGCAGGGCCAGGCCGGGCTCTTCT
CTCTCTTGTCTGTCCTCGTCTGTGGGGGAATGAAGATGACACAGCACTCAGGAGGT
CCTTTTCAGGACCCAGAGTCACTGGGCTTTTAGGGCGTTTGGCTCTGGTTCTAGCCC
GAGTGTGGTCTCAGTCCCCATTGGTGGTTGTGGTTGTGGCATGGGCATCATTAGTC
ACACAACATAAAACCACAGAACGCAGAATTTGCTCACCCTGGTTCACACAGGAGCAA
GCTTGGGCACCAGCACCCCCTGAACGCATGCCAAAGGGCAGGTCCAGCTCAGTGC
TGGGGTGTGGAGGAACCCACCTGGGGCCTGCCTGCAAAGACAGATGCCCTGCCGT
GTTTAGAAAAAGAGAAAACTGGCCTAAACCACAGTGAGGTCAAATGGATCATTTAAC
CCGTCGAGTTTGGTTTCTTGTTCATCCGACGCGGTTGATCTTATTAACGCATATGTGT
TTGTGTTGATGTGGCCTCCAGCTTCTGTCGCACTGTGGACACAGTTGGATGTGGCT
GCTAATCAATGCGGGGCTCCGTAGGTCAGTGAGAGCAATCAGAGCCTGCTGTGCGG
TGGCATCCCCGAGTTAAGACAGGCTTTTGTCCCATACTCTATTGAAACACCCAGAAA
TGCTGACATGTTGAAGTTAAGGCCATGTGCCTTGGTGCTACTTGCTTGCGAAACAGC
CTTTGAAGGAAAAAGGAGGAGGAGGATCCATGAGTTTCAAGGGTCCCCTCCTGTAA
TCACGGCTGCCACTCAACGGAGTAAGTAATGTGTAGCGGGCCCTCTATCTGCATGA
CGAATCCTCCTAACATTCCTGTGGGCGAGGTAGGATTTGTCTCGCTTTGCAGCCGAA
GAATCAGAGCCTCAGAAAAGTTAAGAAATCTTCCCTGGGTCAAGCTGGTAGGAGGA
GGGACTGGAACTGAACTCACATCTCCACCTCCTCTCTTCCTACTCCTCTGCCTTTCTT
GCACCTAGAGCAGTGTACAACATCAACAACTGACCAGGGCGATCCTAGAGGTAAGA
GACAGCCGAATGGCACGGACCCAGCCAAGCCCAGGCCTGCCCAGGTGGAACAGAG
CCAGCTAGGAGCATTCGGTAGCAGAGGTCAGTCCGCTGATTGACTGTTCTCACTTC
CAGGCCCACAGCCTCCTCCACCCTGATGGTCTATGAGTCTAGTGCGGCCCTTTGCA
CAATTTTAGCAACTGAACCCTTTTCCTCACAGAATTGTGATACCGTCCTAACATATGA
AGCATATATAAACAGTTGGGCAAATCATAATAGTTTCCATTATTTCATAAATTTGCATT
TGTGGCACCAAGGAGAGAATTAATGGGAACAAGAGGGTTCTTGTGTGAGTTCCGCA
TTTGGATTTGGGGGTTCGAGGTGGCAGCTATAGGCTCACATTAGTGTCTGCATCACT
TTGCTTCAGTTTGTCACGGTGGCCCCCCAAATCCAGAACCGCCCTCCTCGCTGTGT
GAGGAGGGTCCTGCCTGCACCGCTTTGCAGACATGGGCCAGCCTGGATGCCTGGG
CGACGCTGCCCTGCGACTCGCTGGCCTGCAGGACATCTGCAAAACATTCAGGTCTC
ACGGCAGTGGCCATGAACGGCCTCACAGACACTGAGTGGACCTGACCCCAAAGAG

AATGCGTCGTCCTTGAACGTGTCGATATTTAGGTTATTACACAATGAGTGTGTTGCTT
TGCAATTAGTGTCTGAAAGAGTTAGAGCAGATGTACATATAGAAACGAAGAAACGAA
ATCTGAATCCATGCTGACAGATTCCCCCTCCCCAAACCTTGCTCAAG*GTAACACAGG*
*CGACCAGG*AGCAAGTCCCGGCCGGCCCCCTGCTAGCGTGGTCTGAGTGGCCCAAG
CAGGACCCCTGCCCAGGCTCACTGGACTCTCCTGGAGCCCCCAGACACGACACCC
TTCCTCATCACTGAAAATCCACTTCCA*CAGTGTAGGAGATGCATGG*ATCTGAGGGC
CAGTAGGAAGTGCTGGAGCTTTAGGCTTTAGACATTCTTTTTTTTTTTTTTTTTGAGAC
AGGGTCTCCATTTGTGGCCCAGGCTGGGGTA*CAGTGGCGTGAGAAATTCTT*TAACA
CTCTGCCGAGTCCTGGCCCTTGGGGAGTGAGGCCTGGCTATGAGGCTTCTTGTCAG
AACTCCGTGTGACTCTCAGGATGGAAGAACAGGGTGGGGGCTGCTCCTCCCCTCCC
TCCCTCCAGCCCCTGGCCTTTGTAAAACGGCTCCGCCTACGCGCCTGTAGCCAGTG
CAGAACCAGGGAGGCCTCTCCCCTTCCCTTCCCAGGAGGGGCCCCGCCTTGTCCT
CCATGCCGGATTCTGTTCGTCCTCCCCCAGGTCCTCTCTGCTGCCTCCCACTGCCC
CATTTCCTCCAGGAACCTTCCTCTCTCCTACCACCAGCCTCTCTCCTACCCCCAGAC
ACACACTCTTTACATGCCGATTTCTGTCCTCTGGTTAACTGGCGTTTAATTTTCTCAG
CTGCTTGGCTTAATGCCAAAGAGTGTTCTTCCAAATAAAAATGGTAAAGACCTTTTTT
C

Appendix 11

Demographic, cytogenetic, FISH and aCGH data for iAMP21 patients. Underlining indicates cytogenetic aberrations that were confirmed or refined by aCGH. Patients highlighted in bold have relapsed. See Appendix 3 for further details on probes 21(A)-(F)

Patient ID	Age (yrs)	Sex	Karyotype	FISH results	aCGH data
4279	11	M	47,XY,-12,der(21)dup(21)(q?), +mar1,+mar2[3]	TEL/AML1; 4-9R2G0F 21(A); 2R3-4G0F 21(B); 4-5R2G0F 21(C); 4-7R2G0F 21(E); 3-5R2G0F 21(F); 2R1G0F 21qtel 0R1G0F	dim(1)(193.4[q31.1]-193.5[q31.1]), dim(9)(28.1[p21.1]-28.3[p21.1]), enh(21)(20.8[q11.2]-30.5[q21.2]), amp(21)(30.5[q21.2]-42.9[q22.3]), dim(21)(42.9[q22.3]-46.9[q22.3]),dim(X)(87.6[q21.31]-89.4[q21.31]), dim(X)(90.4[q21.31]-91.6[q21.31]),dim(Y)(2.9[p11.31]-5.6[p11.2]).
4414	7	M	45,XY,t(6;19)(p21;p13),der(7)t(7;15)(p1;q1),del(11) (p13),-15,del(16)(q2), ider(21)(q10)dup(21)(q?)[8]	TEL/AML1; Multiple copies of AML1	dim(7)(27.6[p15.2]-51.1[p12.2]), dim(11)(26.9[p14.2]-31.0[p13]), dim(11)(102.7[q22.2]-102.9[q22.3]), dim(15)(27.2[q11.2]-41.5[q13.3]), dim(15)(87.5[q25.3]-97.4[q26.2]), dim(16)(45.1[q11.2]-86.4[q24.1]), enh(21)(16.6[p11.2]-22.3[q11.2]), amp(21)(22.5[q11.2]-41.1[q22.2]), dim(21)(41.1[q22.2]-42.3[q22.3]).
5898	5	M	47,XY,+X,del(16)(q13),i(17)(q10),ider(21)(q10) dup(21)(q?)[3]/47,idem, add(7)(p1)[3]	TEL/AML1; 4-6R2G0F 21(A); 2R3G0F 21(B); 3-5R2G0F 21(C); 3-6R2G0F 21(E); 3-5R2G0F 21(F); 2R1G0F 21qtel 0R1G0F	dim(7)(36.5[p14.1]-51.0[p12.1]), dim(16)(52.6[q12.2]-88.6[q24.3]), dim(17)(0.01[p13.3]-21.4[p11.2]), enh(17)(21.4[p11.2.]-78.6[q25.3]), dim(21)(13.9[p11.2]-21.1[q11.2]), enh(21)(21.1[q11.2]-28.3[q21.1]), amp(21)(28.3[q21.1]-43.4[q22.3]), dim(21)(43.4[q22.3]-46.9[q22.3]), enh(X).

6111	4	F	FAIL	TEL/AML1; 4-11R2G0F 21qtel; 0R1G0F 21q sub-tel; 1R0G0F X cen; 0R3G0F (92%)	enh(9)(126.9[q34.11]-138.2[q34.3]), enh(16)(0.4[p13.3]-1.6[p13.3]), enh(17)(67.6[q25.1]-78.5[q25.3]), enh(18)(14.1[p11.21]-15.3[p11.21]), enh(20)(60.2[q13.33]-61.6[q.13.33]), enh(21)(17.5[p11.1]-18.1[p11.1.]), enh(21)(22.3[q11.2]-31.0[q21.2]), amp(21)(31.0[q21.2]-42.9[q22.3]), dim(21)(42.9[q22.3]-46.9[q22.3]).
6783	3	M	43-44,XY,del(5)(q11q13),der(15;21)(q10;q10)dup(15;21)(q10;q10),del(16)(q22),-20[cp5]/45,XY,der(15;21)(q10;q10)c[cp3]	TEL/AML1; 3-6R2G0F 21(A); 2R3G0F 21(B); 2R2G0F 21(C); 3-6R2G0F 21(E); 3R2G0F 21(F); 2R3-5G0F 21qtel 0R1-3G0F	dim(5)(60.3[q12.1]-80.8[q14.1]), dim(9)21.8[p21.3]-21.9[p21.3]), enh(15)(23.2[q11.1]-25.7[q11.2]), enh(15)(28.[q11.2]-35.2[q13.1]), dim(15)(36.3[q13.2]-36.5[q13.2]), dim(15)(36.8[q13.2]-57.7[q13.2]), amp(15)(57.7-[q21.1]-69.5[q22.31]), dim(15)(78.9[q24.2]-79.1[q24.2]), enh(15)(79.1[q24.2]-84.2[q25.2]), dim(15)(84.2[q25.3]-100.3[q26.3]), dim(16)(48.0[q12.1]-60.2[q21]), dim(20), dim(21)(9.8[p12]-13.9[q11.2]), enh(21)(14.4[p11.2]-15.9[p11.1]), dim(21)(24.2[q21.1]-25.6[q21.1]),enh(21)(25.7[q21.1]-28.1[q21.1]), enh(21)(31.1[q21.2]-32.3[q21.3]), amp(21)(32.8[q21.3]-39.4[q22.13]), enh(21)(39.4[q22.13]-41.5[q22.2]), amp(21)(41.5[q22.2]-46.9[q22.3]), dim(X)(87.7[q21.31]-91.8[q21.32]), dim(Y)(2.9[p11.31]-6.5[p11.2]).

6788	4	M	46,XY,add(13)(q?),ider(21)(q10)dup (21)(q?)[2]/47,idem,+14[1]/48,idem,+8, +14[5]	IGH; 0R0G3F (89%) TEL/AML1; 5R2G0F (95%) 21(A); 2R4-5G0F 21(B); 3-5R2G0F 21(C); 4-5R2G0F 21(E); 4-6R2G0F 21(F); 2R4-5G0F 21qtel 0R3-5G0F	dim(8)(111.6[p23.2]-146.2[p24.3]), dim(13)(107.6[q33.3]-114.1[q34]), enh(21)(19.9[q11.1]-23.9[q21.1]), amp(21)(23.9[q21.1]-46.9[q22.3]).
6996	12	F	FAIL	BCR/ABL; 4R4G0F p16/9cen; 4R4G0F MLL; 0R0G4F IGH; 0R0G5F TEL/AML1; 8-14R2G0F 21(A); 2R5-9G0F 21(C); 8-10R2G0F 21(E); 8-9R2G0F 21(F); 2R1G0F 21qtel 0R1G0F 21q sub-tel 1R0G0F	dim(9)(116.9[q33.1]-138.2[q34.3]), dim(11)(88.2[q14.2]-89.7[q14.3]), dim(16)(31.4[p11.2]-35.0[p11.1]), enh(16)(75.1[q23.1]-87.9[q24.3]), dim(18)(59.2[q21.32]-65.4[q22.1]), enh(20)(60.2[q13.33]-62.3[q13.33]), enh(21)(14.4.3[p11.2]-31.3[q21.2]), amp(21)(31.3[q21.2]-40.3[q22.2), dim(21)(40.3[q22.2]-46.9[q22.3]).
7045	10	F	47,XX,+X,del(9)(p?),-10,del(11)(q13),ider(21)(q2?) dup(21)(q?),+mar[cp11]/ 46,XX[3]	MLL; 0R0G1F TEL/AML1; 5-8R2G0F 21(A); 2R4-6G0F 21(B); 4-8R2G0F 21(C); 5-8R2G0F 21(E); 4-8R2G0F 21(F); 2R4-8G0F 21qtel 0R1G0F 21q sub-tel 1R0G0F (85%)	dim(8)(90.3[q21.3]-91.5[q21.3]), dim(9)(16.0[p22.2]-23.3[p21.2]), dim(9)(26.6[p21.1]-27.0[p21.1]), dim(9)(33.5[p13.3]-64.1[q21.1]), dim(9)(91.2[q22.2]-95.1[q22.32]), dim(9)(95.5[q22.32]-96.0[q22.32]), dim(10)(42.4[p11.1]-53.7[q11.22]), dim(10)(89.1[q23.21]-99.1[q23.33]), dim(11)(78.7[q13.4.]-132.1[q25]), amp(21)(18.9[q11.1]-46.6[q22.3]), dim(21)(46.6[q22.3]-46.9[q22.3]), enh(X).

7093	6	F	FAIL	TEL/AML1; 6-8R2G0F 21(A); 2R4G0F 21(B); 5-7R2G0F 21(C); 5-7R2G0F 21(E); 5-7R2G0F 21(F); 2R1G0F 21qtel 0R1G0F 21q sub-tel 0R0G1F	enh(1)(142.1[q21.1]-245.4[q44]), enh(6)(128.9[q22.33]-170.8[q27]), enh(21)(22.7[q11.1]-27.8[q21.1]), amp(21)(27.8[q21.1]-43.2[q22.3]), dim(21)(43.2[q22.3]-46.9[q22.3]).
7255	30	F	?45,XX,-21,+mar,inc[2]	TEL/AML1; 5-11R2G0F 21(A); 2R5-7G0F 21(B); 3-4R2G0F 21(C); 4-14R2G0F 21(E); 5-8R2G0F 21(F); 2R3G0F 21qtel 0R2G0F	enh(6)(8.15)[p24.3]-26.0[p22.1]), dim(7)(132.0[q32.3]-139.5[q34]), dim(12)(11.6[p13.31]-16.9[p13.1]), dim(20)(43.6[q13.12]-47.8[q13.13]), dim(21)(16.0[p11.1]-20.0[q.11.1]), enh(21)(20.0[q11.1]-23.1[q11.2]), amp(21)(23.1[q11.2]-36.3[q22.11]), enh(21)(36.3[q22.11]-37.0[q22.12]), amp(21)(37.0[q22.12]-40.5[q22.2]).
7583	14	M	47,XY,+X,der(21)r(21)(q?)dup(21)(q?)[4]	p16/9cen; 2R0G0F (23%)/ 2R1G0F (8%) TEL/AML1; 4-5R2G0F 21(A); 2R3G0F 21(B); 3R2G0F 21(C); 4-5R2G0F 21(E); 4R2G0F 21(F); 2R4-5G0F 21qtel 0R1G0F 21q sub-tel 0R1G0F	dim(3)(46.9[q21.32]-47.1[q21.32]), dim(3)(108.7[q13.13]-109.4[q13.13]), dim(9)(19.4[p21.3]-24.1[p21.2]), enh(10)(46.1[q11.1]-46.7[q11.1]), dim(16)(31.9[p11.2]-34.1[p11.1]), dim(18)(13.3[p11.21]-13.4[p11.21]), enh(21)(19.4[q11.1]-31.5[q21.2]), amp(21)(31.5[q21.2]-46.3[q.22.3]), dim(21)(46.3[q22.3]-46.9[q22.3]), enh(X).

7619	12	M	47,XY,+9,der(21)r(21)(q?)dup(21)(q?)	p16/9cen; 3R3G0F (33%) BCR/ABL; 3R3G0F TEL/AML1; 6-10R2G0F (64%) 21q sub-tel 0R1G0F	dim(3)(22.4[q24.3]-25.1[q24.1]), enh(9), dim(20)(43.1[q13.12]-47.2[q13.13]), enh(21)(13.3[p11.1]-19.9[q11.1]), dim(21)(19.9[q11.1]-20.9[q11.2]), amp(21)(20.9[q11.2]-42.5[q22.3]), dim(21)(42.5[q22.3]-46.9[q22.3]).
7732	11	F	FAIL	TEL/AML1; 4-7R2G0F 21(A); 2R3G0F 21(B); 3-5R2G0F 21(E); 5-7R2G0F 21(F); 2R4-6G0F 21qtel 0R1G0F	dim(3)(163.9[q26.1]-164.5[q26.1]), dim(7)(141.8[q34]-146.0[q35]), enh(8)(140.6[q24.3]-146.0[q24.3]), enh(11)(0.1[p15.5]-3.8[p15.5]), enh(11)(60.1[q12.3]-72.2[q13.3]), dim(11)(106.6[q22.3]-116.0[q23.3]), dim(12)(18.3[p12.3]-19.3[p12.3]), enh(16)(0.5[p13.3]-3.1[p13.3]), enh(16)(27.7[p11.2]-31.4[p11.2]), dim(16)(31.4[p11.2]-33.6[p11.1]), enh(21)(17.1[p11.1]-27.9[q21.1]), amp(21)(27.9[q21.1]-46.5[q22.3]), dim(21)(46.5[q22.3]-46.9[q22.3]), enh(X)(2.6[p22.33]-62.1[q11.1]).
7828	7	M	46,XY[20]	TEL/AML1; 3-9R2G0F 21(A); 2R4-6G0F 21(B); 5-8R2G0F 21(C); 7-9R2G0F 21(E); 5-8R2G0F 21(F); 2R1G0F 21qtel 0R1G0F 21q sub-tel 0R1G0F	dim(21)(14.5[p11.2]-17.4[p11.1]), enh(21)(18.8[p11.1]-28.7[q21.1]), amp(21)(28.7[q21.1]-42.8[q22.3]), enh(X).

8743	8	F	46,XX,t(7;9;17)(q22;p1?,p1?),del(11)(q23q2?5),dup(21)(?)[7]	MLL; 0R0G1F TEL/AML1; 3-5R2G0F 21(A); 2R3-5G0F 21(B); 4-6R2G0F 21(C); 3-5R2G0F 21(E); 4-5R2G0F 21(F); 2R4-5G0F 21qtel 0R1G0F	dim(7)(49.9[p12.2]-51.1[p12.1]), dim(11)(95.8[q21]-134.4[q25]), enh(21)(22.2[q11.1]-23.1[q11.1]), amp(21)(23.1[q11.1]-45.0[q22.3]), dim(21)(45.0[q22.3]-46.9[q22.3]).
8983	13	F	46,XX[20]	p16/9cen; 2R1G0F (62%) TEL/AML1; 5-9R2G0F 21(A); 0R1G0F 21(B); 4-6R2G0F 21(F); 1R0G0F 21q sub-tel 1R0G0F	dim(8)(0.1[p23.3]-43.0[p11.1]), enh(8)(43.0[p11.1]-146.2[q24.3]), dim(9)(0.1[p24.3]-39.1[p12]), dim(16)(31.8[p11.2]-33.7[p11.1]), enh(21)(9.8[p12]-18.9[p11.1]), amp(21)(18.9[p11.1]-22.4[q11.2]), dim(21)(22.4[q11.2]-22.9[q11.2]), amp(21)(22.9[q11.2]-26.6[q21.1]), dim(21)(26.6[q21.1]-32.7[q21.3]), amp(21)(32.8[q21.3]-37.9[q21.3]), dim(21)(37.9[q21.3]-46.9[q22.3]).
9028	10	F	46,X,add(X)(q26-q28),ider(21)(q10)dup(21)(q?)[8]	TEL/AML1; 5-12R2G0F 21q sub-tel 0R0G1F	dim(3)(106.7[q13.11]-109.6[q13.13]), dim(5)(68.8[q13.1]-70.6[q13.2]), dim(7)(109.1[q22.3]-112.4[q31.1]), dim(7)(113.1[q31.1]-158.3[q36.3]), dim(11)(102.8[q22.2]-114.4[q23.3]), dim(21)(14.2[p11.2]-14.4[p11.2]), enh(21)(14.6[p11.2]-25.8[q21.1]), amp(21)(25.8[q21.1]-27.1[q21.1]), enh(21)(27.1[q21.1]-28.3[q21.1]), enh(21)(28.3[q12.1]-46.3[q22.3]), dim(21)(46.3[q22.3]-46.9[q22.3]), enh(X)(2.6[p22.33]-55.0[p11.22]), enh(X)(94.4[q21.33]-154.4[q28]).

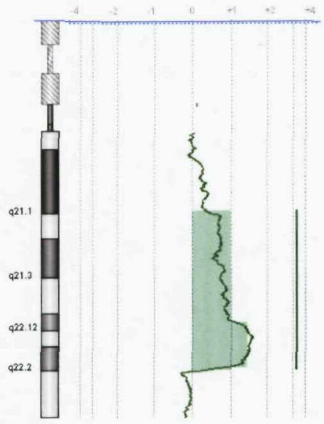
9864	10	M	47,XY,dup(21)(q?),+dup(21)(q?)[6]	TEL/AML1; 5-10R2G0F 21q sub-tel 0R3-6G1F	dim(7)(49.8[p12.2]-51.2[p12.1]), enh(21)(13.3[p11.2]-23.6[q21.1]), amp(21)(23.6[q21.1]-46.9[q22.3]).
11706	5	M	FAIL	TEL/AML1; 5-13R2G0F 21q sub-tel 1R0G0F X cen; 2R0G0F	dim(12)(11.3[p13.31]-20.8[p12.3]), dim(19)(4.0[p13.3]-4.0[p13.3]), enh(21)(14.3[p11.1]-18.5[p11.1]), enh (21)(20.7[q11.2]-21.1[q11.2]), enh(21)(21.7[q11.2]-24.3[q21.1]), amp(21)(24.3[q21.1]-27.8[q21.1]), enh(21)(27.8[q21.1]-30.8[q21.2]), amp(21)(30.8[q21.2]-44.7[q22.3]), dim(21)(44.7[q22.3]-46.9[q22.3]), enh(X).

The normal clone has been omitted from cases with an abnormal karyotype

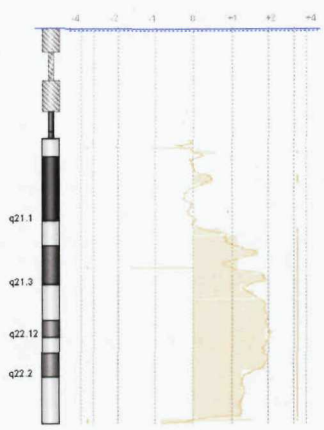
Appendix 12

The aCGH chromosome 21 profiles for all iAMP21 and non-iAMP21 patients. The shaded areas and the corresponding coloured sidebars indicate regions of copy number change. Deletions, gains and amplifications were classified as five or more consecutive probes that deviate from 0 by +/- 0.25, or +1, respectively.

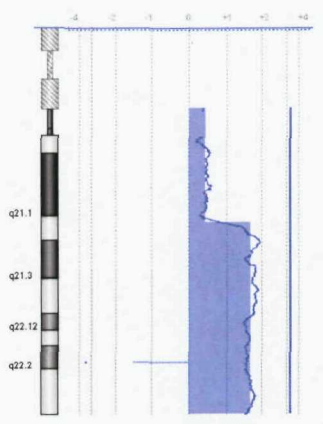
4414



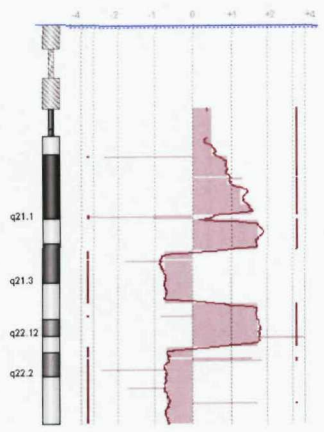
9028



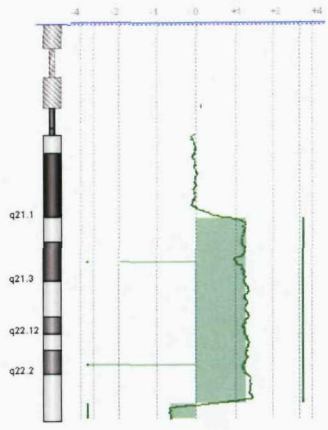
9864



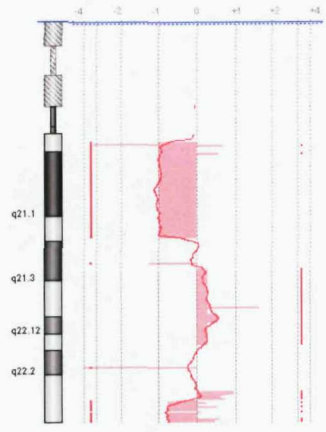
8983



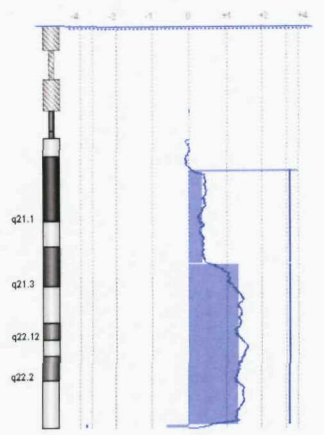
8743



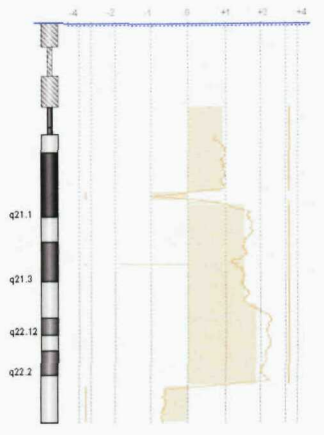
9010



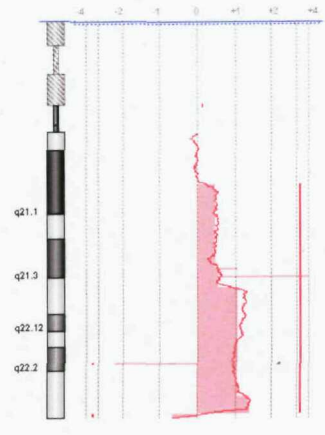
7732



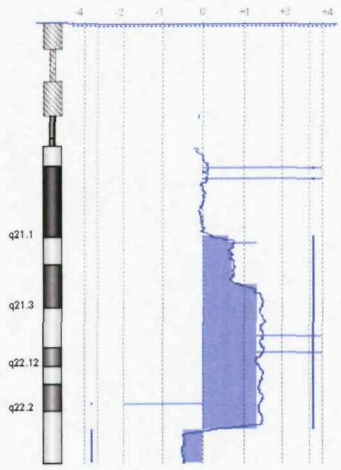
7619



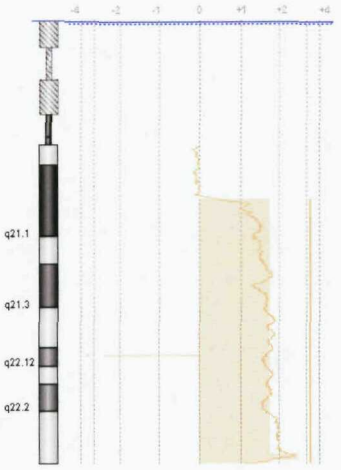
7583



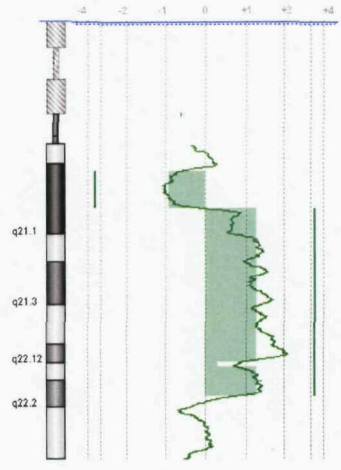
7093



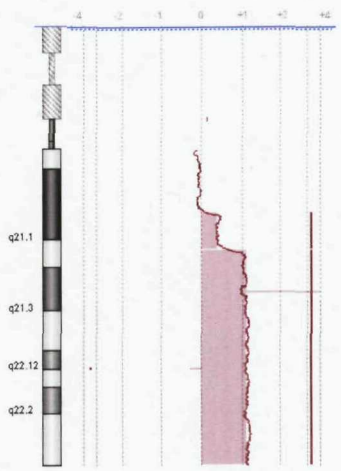
7045



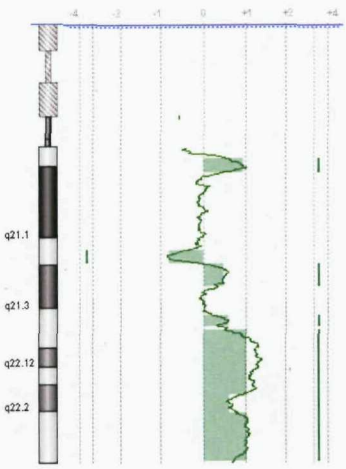
7255



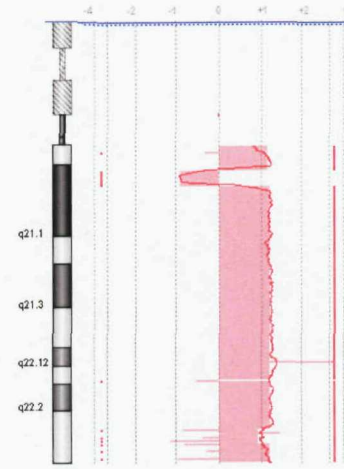
6788



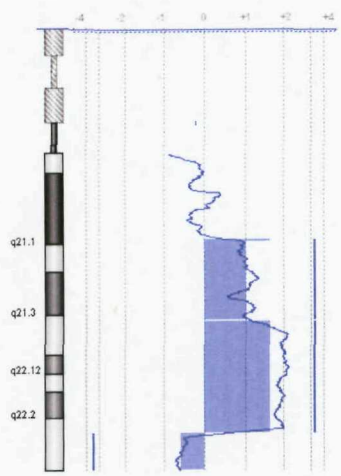
6783



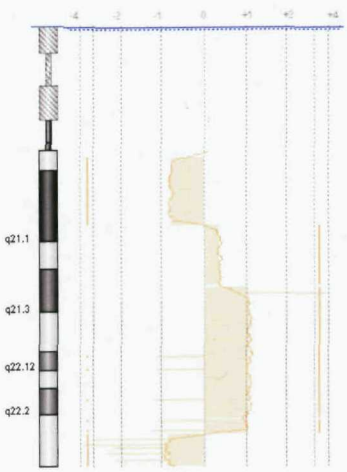
4746



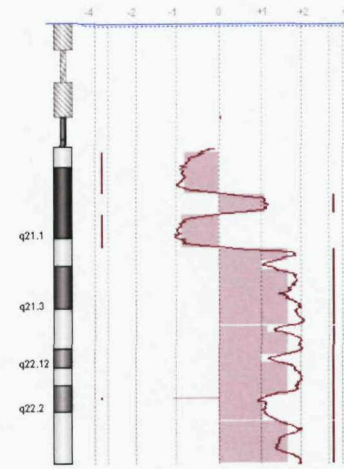
6111



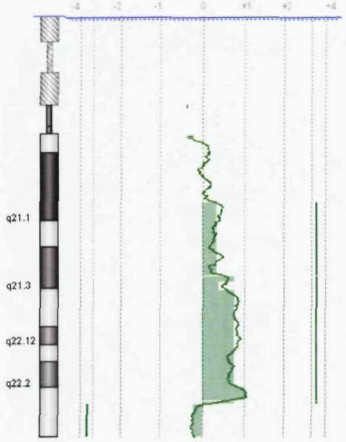
5898



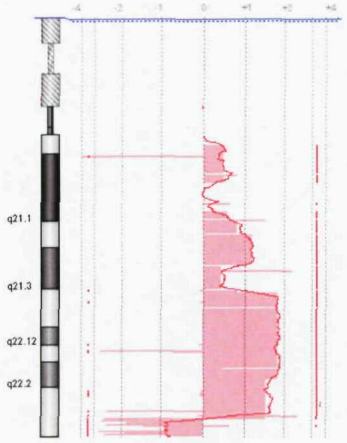
4676



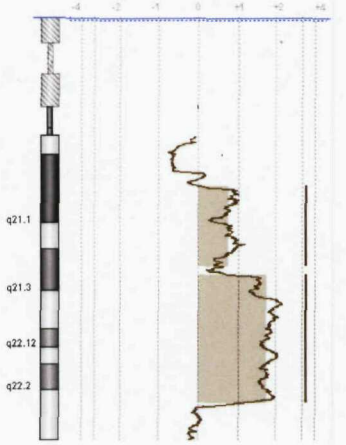
4279



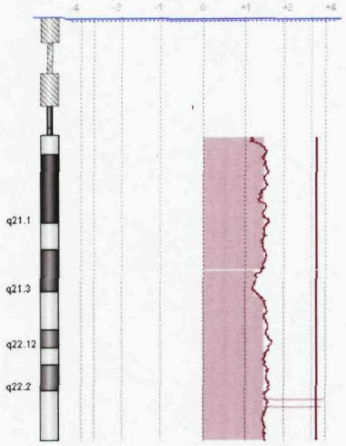
11706



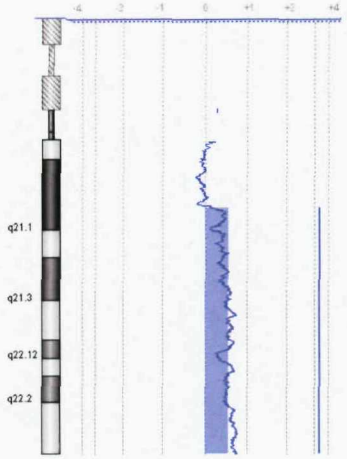
7828



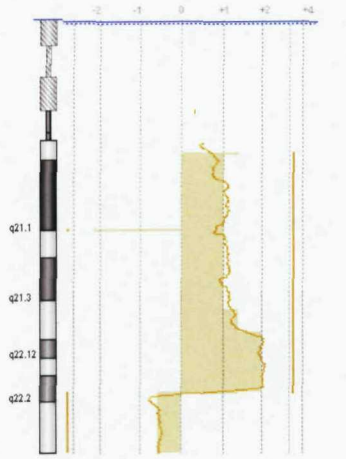
10442



4247



6996



Appendix 13

The *PDE9A* FISH screen results

Patient	FISH signal patterns	Overall *PDE9A* status	Array-CGH profile for *PDE9A*
2647	0R0G2F 14% 0R0GampF 66% ExtraR0G1-3F 20%	Amplification and telomeric deletion	
2776	0R0G2F 16% 0R0GampF 86%	Amplified	
3131	0R0G2F 5% 0R0GampF 95%	Amplified	
3368	0R0G2F 5% 0R0GampF 95%	Amplified	
3527	0R0G2F 99% 0R0G3F 1%	Normal	
3743	0R0G3F 87% 0R0G2F 13%	gained	
3745	extraR0GampF 70% 0R0GampF 30%	Amplification and telomeric deletion	
3956	0R0G2F 13% 0R0GampF 87%	Amplified	
3970	0R0G2F 28% 0R0G1F 72%	Deleted	
4135	0R0G2F 13% 0R0GampF 65% 0R0G1F 22%	Amplified	
4178	0R0G1F 90% 0R0G2F 10%	Deleted	
4237	0R0GampF 99% 0R0G2F 1%	Amplified	
4279	ampR0G1F 72% ampR0G0F 6% 0R0G2F 12%	Amplification and telomeric deletion	Amplification and telomeric deletion
4316	0R0G2F 64% 0R0G3F 20% 0R0G1F 6%	Normal	
4405	0R0G1F 100%	Deleted	
4444	0R0GampF 99% 0R0G2F 1%	Amplified	
4561	0R0G2F 3% 0R0G1F 97%	Deleted	
4623	0R0G2F 6% 0R0G1F 94%	Deleted	
4746	0R0G2F 1% 0R0GampF 99%	Amplified	
4780	0R0G2F 100%	Normal	
5607	0R0G2F 36%	Amplified	

	0R0GampF 58% 0R0G3F 6%		
5655	0R0G2F 4% 0R0G1F 96%	Deleted	
5674	0R0G1F 85% 0R0G2F 15%	Deleted	
5754	0R0G2F 54% 0R0GampF 43% 0R0G1F 3%	Amplified	
5759	0R0G2F 3% 0R0GampF 97%	Amplified	
5809	0R0G1F 97% 0R0G2F 3%	Deleted	
5858	0R0GampF 80% 0R0G2F 20%	Amplified	
5898	0R0GampF 38% 0R0G2F 16% 0R0G3-4F 46%	Amplified	Amplified
6008	0R0G2F 3% 0R0G1F 97%	Deleted	
6020	0R0G2F 11% 0R0G1F 89%	Deleted	
6092	0R0G2F 96% 0R0GF 2% 1R0G2F 2%	Normal	
6111	ampR0G1F 82% 0R0G2F 10%	Amplification and Telomeric deletion	Amplification and Telomeric deletion
6783	0R0G3-4F 76% 0R0GampF 10% 0R0G2F 14%	Amplified	Amplified
6788	0R0G4F 10% 0R0GampF 64% 0R0G2F 11% 1R0GampF 15%	Amplified	Amplified
6868	0R0GampF 17% 0R0G2F 83%	Normal	
6937	0R0G2F 7% 0R0GampF 93%	Amplified	
6957	0R0GampF 10% 0R0G2F 1%	Amplified	
7024	extraR0G1F 76% extraR0G0F 22% 0R0G2F 2%	Amplification and Telomeric deletion	
7045	0R0GampF 69% extraR0GampF 26% 0R0G2F 5%	Amplification and Telomeric deletion	Amplified
7100	0R0G2F 77% 0R0G1F 11%	Normal	
	0R0GampF 12%		

7219	0R0G1F 88% 0R0G2F 12%	Deleted	
7255	0R0G3F 80% 0R0G2F 20%	gain	Normal
7583	0R0G2F 5% 0R0G3F 20% 0R0G4F 42% 0R0GampF 29% 1R0G3F 4%	Amplified	Amplified
7619	0R0G1F 39% 1R0G1F 31% 1R0G0F 6% 0R0G2F 19%	Deleted/Telomeric deletion	Deleted
7650	0R0GampF 100%	Amplified	
7732	0R0GampF 75% 0R0G2F 25%	Amplified	Amplified
7828	0R0G2F 16% 0R0GampF 8% ampR0G1F 69%	Amplification and Telomeric deletion	Normal
7829	0R0G2F 94% 0R0G1F 6%	Normal	
7980	0R0G2F 89% 0R0GampF 11%	Normal	
8743	0R0GampF 81% extraR0GampF 5% 0R0G2F 3%	Amplified	Amplified
8767	0R0G2F 4% 0R0G1F 96%	Deleted	
8983	0R0G1F 55% 1R0G1F 22% 2R0G0F 1% 0R0G2F 19%	Deleted/Telomeric deletion	Deleted
8987	0R0GampF 65% 0R0G2F 30% 0R0G1F 5%	Amplified	
9010	0R0G2F 88% 0R0G1F 10% 0R0G3F 2%	Normal	
9028	0R0G1F 65% 1R0G1F 25% 0R0G2F 10%	Deleted	Gained
9864	0R0GampF 92% 1R1GampF 3% 0R0G2F 5%	Amplified	Amplified
10442	0R0G2F 6% 0R0GampF 94%	Amplified	
10448	0R0G2F 5% 0R0GampF 95%	Amplified	
10468	0R0GampF 95%	Amplified	

	0R0G2F	5%		
10542	0R0G2F 0R0G1F	6% 94%	Deleted	
10697	0R0G2F 0R0GampF	10% 90%	Amplified	
11005	0R0G2F 0R0G1F	12% 88%	Deleted	
11056	0R0G2F 0R0GampF 0R0G1F	28% 68% 4%	Amplified	
11061	0R0G1F 0R0G2F	89% 11%	Deleted	
11158	0R0G2F 0R0GampF 0R0G3F	13% 65% 22%	Amplified	
11551	0R0G1F	100%	Deleted	
11706	0R0GampF 0R0G2F	93% 7%	Amplified	Amplified
11961	0R0G2F 0R0GampF 0R0G1F	58% 40% 2%	Normal/Amplified	
12122	0R0G2F	100%	Normal	
12209	0R0G2F 0R0G3F 0R0G4F	15% 29% 56%	Amplified	

Appendix 14

Demographic, cytogenetic, FISH and aCGH data for the unclassified patients. Underlining indicates cytogenetic aberrations that were confirmed or refined by aCGH. Patients higlighted in bold have relapsed. Patients higlighted in green were included in the iAMP21 study for comparison purposes.

Patient ID	Age (yrs)	Sex	Karyotype	FISH results	aCGH data
186	8	M	48,XY,add(7)(q?),add(9)(p),+10,add(14)(p),+16[6]/48,idem,del(6)(q?15q21)[5]/47,idem,del(6)(q?15q21),-16[5]		dim(9)(20.48[p21.3]-23.56[p21.3]), enh(10), dim(10)(111.76[q25.1]-111.86[q25.1]), dim(15)(38.04[q13.3]-39.51[q13.3]), enh(16), enh(17)(37.53[q21.1]-78.65[q25.3]), dim(18)(27.32[q11.2]-76.11[q23], enh(21)(22.05[q11.2]-46.91[q22.3]), enh(X)(130.30[q26.1]-154.34[q26]).
772	6	M	44,XY,?del(1)(p),?del(2)(p),add(3)(q),-4,add(5)(q),-8,?del(9)(p),-10,-19,+mar,+mar[1]		enh(7)(74.91[q11.23]-74.98[q11.23]), enh(7)(129.23[q32.3]-129.27[q32.3]), dim(9)(9.81[p22.33]-10.01[p22.33]), enh(15)(88.97[q25.3]-89.04[q25.3]).
2189	5	M	46,XY,inc[31]		dim(6)(26.24[p22.1]-26.36[p22.1]), dim(7)(104.28[q22.1]-158.59[q36.3]), dim(10)(111.76[q25.1]-111.86[q25.1]), dim(12)(0.06[p13.33]-23.80[p12.1]), enh(12)(23.80[p12.1]-36.16[p11.1]), dim(13)(53.97[q14.13]-54.37[q14.13]), enh(X)(122.97[q25]-154.49[q28]).

2320	1	M	46,XY,der(4,12,13)t(4;12)(q23;q21)t(12;13)(p12;q22)[18]/46,idem,del(6)(q?)[9].ish der(4,12,13)t(4;12)(wcp4+,wcp12+;wcp4+,wcp12+)t(12;13)(wcp12+;wcp12+)		dim(5)(88.16[q14.3]-88.22[q14.3]), dim(6)(52.74[p12.3]-52.78[p12.3]), dim(8)(26.21[p21.1]-26.28[p21.1]), dim(10)(69.73[q22.1]-96.84[q23.33]).
2326	13	M	46,XY[30]		dim(1)(40.57[p34.1]-40.64[p34.1]), dim(1)(169.72[q24.1]-169.91[q24.1]), dim(2)(191.57[q32.3]-191.63[q32.3]), dim(4)(53.91[q11]-54.14[q11]), dim(4)(86.86[q21.3]-86.99[q21.3]), dim(6)(26.25[p22.1]-26.38[p22.1]), dim(7)(39.77[p14.1]-39.86[p14.1]), dim(7)(50.13[p12.1]-50.26[p12.1]), dim(10)(111.76[q25.1]-111.86[q25.1]), dim(12)(11.75[p13.2]-11.97[p13.2]), dim(12)(90.81[q22]-91.04[q22]), enh(14)(99.13[q32.31]-99.20[q32.31]), enh(17)(76.53[q25.3]-76.59[q25.3]), dim(X)(70.25[q13.1]-70.28[q13.1]).
2741	12	F	46,XX,t(4;18)(q21;q22),del(7)(q22q32),add(9)(p22),add(19)(q?13.3)[19]		dim(3)(178.40[q26.33]-17.88[q26.33]), dim(4)(83.84[q21.23]-86.31[q21.23]), dim(7)(107.46[q22.2]-116.39[q31.2]), dim(7)(118.49[q31.31]-124.12[q31.33]), dim(7)(126.39[q32.12]-133.60[q33]), dim(7)(135.02[q33]-138.58[q34]), enh(8)(76.85[q13.3]-146.25[q24.3]), dim(9)(0.15[p24.3]-21.90[p21.3]), dim(9)(21.90[p21.3]-22.00[p21.3])x2, dim(9)(22.00[p21.3]-36.98[p13.3]), enh(18)(17.26[p11.1]-55.07[q21.23]), enh(18)(64.64[q22.1]-76.11[q23.33]), enh(19)(47.52[q13.2]-48.71[q13.31]), dim(19)(56.87[q13.33]-63.78[q13.43]).

3043	13	M	47,XY,+21[15]		enh(21),dim(21)(38.69[q22.13]-38.79[q22.13]).
3068	12	M	46,XY,del(11)(p?11p?15),del(12)(p11p13)[1]/46,idem,del(6)(q?13q?23)[9].ish del(12)(p13p13)(TEL-),21q22(AML1x2)		dim(3)(109.30[q13.13]-109.44[q13.13]), dim(8)(96.98[q22.1]-99.01[q22.13]), dim(11)(22.24[p14.3]-50.64[p11.11]), dim(12)(6.65[p13.33]-26.44[p12.1]), dim(22)(39.88[q13.2]-43.96[q13.31]).
3141	10	M	46,XY[20]		dim(7)(50.14[p12.1]-50.27[p12.1]).
3435	14	M	46,XY,del(11)(q23)[5]		dim(3)(105.80[q13.11]-108.83[q13.12]), dim(11)(99.32[q22.1]-115.42[q22.3]).
4017	14	M	46,XY[31]		dim(5)(0.74[p15.32]-0.84[p15.31]), dim(17)(27.19[q11.1]-28.81[q11.1]).
4051	1	F	46,XX,del(9)(p?)[3]		
4073	2	F	48,XX,+X,+21c[10]		dim(6)(26.25[p22.1]-26.36[p22.1])x2, dim(10)(27.92[p12.2]-27.97[p12.2]), enh(21), enh(X).
4247	4	F	46,XX,add(21)(q22)[23]	TEL/AML1; 3R2G0F (70%) 21(A); 2R3G0F 21(B); 3R2G0F 21(C); 3R2G0F 21(E); 3R2G0F 21(F); 2R3G0F 21q sub-tel 0R3G0F (70%)	dim(2)(241.4[q27.2]-241.7[q37.2]), enh(8)(142.3[q24.3]-145.8[q24.3]), enh(11)(63.0[q13.1]-67.6[q13.1]), enh(21)(20.4[q11.2]-46.9[q22.3]).

4676	7	M	47,XY,del(9)?(p11p13),i(21)(q10), +dup(21)(q?21q22)[15]	TEL/AML1; 4-5R2G0F (88%) 21(A); 2R4-5G0F 21(B); 5R2G0F 21(C); 5R2G0F 21(E); 5R2G0F 21(F); 2R5G0F	dim(9)(19.9[p21.3]-39.2[p12]), dim(12)(110.3[q24.13]-110.3[q24.13]), amp(21)(13.3[p11.2]-15.9[p11.1]), dim(21)(15.9[p11.1]-17.6[p11.1]), amp(21)(17.6[p11.1]-46.9[q22.3]).
4746	9	M	50,XY,+X,+14,der(20)t(20;21)(q13;q?), rea(21),+rea(21)x2[16]	TEL/AML1; 7R2G0F (95%)	enh(14), enh(17)(45.2[q21.31]-78.6[q25.3]), dim(20)(31.2[q11.21]-33.5[q11.22]), dim(20)(34.7[q11.23]-60.2[q13.33]), dim(21)(13.7[p11.2]-18.1[p11.1]), enh(21)(18.1[p11.1]-20.2[q11.2]), dim(21)(20.2[q11.2]-23.9[q21.1]), amp(21)(23.9[q21.1]-46.9[q22.3]), enh(X).
5052	14	M	46,XY,del(9)(p1?3)[9]	p16/9cen; 1R2G0F (45%)	dim(9)(0.15[p24.3]-21.97[p21.3]), dim(9)(21.97[p21.3]-22.00[p21.3])x2, dim(9)(22.00[p21.3]-36.82[p13.1]).
5846	13	F	46,XX[10]		dim(1)(49.63[p33]-49.70[p33]), dim(5)(157.90[q33.3]-158.45[q33.3]), dim(6)(108.99[q21]-109.44[q21]), dim(9)(36.87[p13.1]-37.01[p13.1]), dim(12)(63.79[q14.3]-63.89[q14.3]), dim(14)(92.88[q32.12]-92.96[q32.12]), dim(16)(65.09[q22.1]-65.18[q22.1]).
5944	2	F	46,XX,der(7)del(7)(q?q?)t(7;9)(p?;p?),del(9)(p?),der(9)t(7;9)(p?;p?)t(7;12)(p?;p13)ins(9;7)(q?;??),del(12)(p1?),der(12)t(7;12)(p?:p13),inc[cp3]	TEL/AML1; 2R3G0F (35%) TCRβ; 1R1G1F (95.5%)	dim(4)(70.3[q13.3]-70.3[q13.3]), dim(9)(19.9[p21.3]-27.7[p21.1]).

6037	3	M	46,XY,del(2)(p2?),del(3)(q2?),add(9)(p2?),add(9)(p2?),add(14)(q3?2)[10]		dim(2)(18.08[p24.3]-43.98[p22.1]), dim(2)(179.53[q31.2]-179.86[q31.2]), dim(3)(130.36[q21.3]-131.63[q21.3]), dim(8)(7.26[p23.1]-8.14[p23.1]), dim(9)(21.52[p21.3]-21.91[p21.3]), dim(9)(21.91[p21.3]-21.98[p21.3])x2, dim(9)(21.98[p21.3]-22.79[p21.3]), dim(9)(24.80[p21.2]-26.64[p21.1]), dim(14)(77.40[q24.2]-78.12[q24.2]), dim(14)(79.38[q24.3]-80.41[q24.3]), dim(16)(71.47[q22.3]-71.73[q22.3]).
6779		M	46,XY[5]		
6985	15	M	46,XY[2]		
7262	10	F	46,XX,dic(7;12)(p1;p1),+21,add(21)(p1?)[10]	TEL/AML1; 3R1G0F (88%)	dim(4)(69.06[q13.3]-69.17[q13.3]), dim(7)(3.12[p22.3]-56.06[p11.2]), dim(12)(0.03[p13.33]-23.53[p12.1]), enh(21).
7280	5	M	46,XY,del(12)(p11p12)[18]	TEL/AML1; 2R1G0F (59%)	dim(4)(69.06[q13.3]-69.17[q13.3]), dim(6)(164.58[q26]-166.13[q27]), dim(8)(12.08[p23.1]-12.51[p23.1]), enh(9)(135.99[q34.3]-136.37[q34.3]), dim(12)(11.68[p13.2]-14.69[p13.2]), dim(14)(100.15[q32.32]-100.26[q32.32]), enh(17)(19.21[p11.2]-19.23[p11.2]),enh(22)(46.29[q13.31]-46.4[q13.31]).
7322	4	M	46,XY,add(7)(p1?3),dup(21)(q22q22)[15]	TEL/AML1; 3R2G0F (92%)	
8443	1	M	Fail		dim(4)(97.78[q22.3]-109.30[q22.3]), dim(9)(0.19[p24.3]-21.08[p21.3]), dim(9)(21.08[p21.3]-22.35[p21.3])x2, dim(9)(22.35[p21.3]-36.13[p13.2]), dim(18)(35.63[q12.1]-37.73[q12.2]), dim(20)(30.50[q11.21]-62.38[q13.33]).

8451	7	M	42-45,XY,-9,-13,+mar[cp6]	p19/9cen; 2R0G0F (53%) BCR/ABL1; 1R2G0F (32%)	enh(1)(141.52[q21.1]-143.57[q21.1]), enh(1)(144.32[q21.1]-147.57[q21.2]), enh(1)(153.26[q22]-154.01[q22]), dim(9)(0.15[p24.3]-21.17[p21.3]), dim(9)(21.17[p21.3]-22.46[p21.3]), dim(9)(22.46[p21.3]-127.55[q34.11]), dim(13)(29.95[q12.13]-114.12[q33.1]).
8878	8	F	42,X,-X,-3,add(4)(p1),add(8)(q2), del(12)(p1),-14,-16,-19,+mar[1]		dim(12)(6.67[p13.32]-15.09[p13.1]), dim(16)(0.03[p13.3]-4.74[p13.3]), dim(X)(75.92[q13.2]-76.92[q13.2]).

9010	18	M	46,XY,t(5;9)(p13;p21),der(9)t(9;21)(p1;q?)dup(21)(q?),del(21)(q?),add(22)(q13)[4]/45,idem,-del(21)(q?)[10]	p16/CEP9; 2R0G0F (83%) /2R1G0F (4%) TEL/AML1; 3-4R2G0F (43%)	dim(1)(66.5[p31.1]-66.6[p31.1]), dim(1)(155.8[q23.1]-155.8[q23.1]), dim(3)(37.1[p22.1]-37.2[p22.1]), dim(3)(57.9[p21.1]-58.0[p21.1]), dim(3)(196.7[q29]-196.9[q29]), dim(4)(48.1[p12]-48.3[p12]), dim(4)(153.3[q31.23]-153.6[q31.23]), dim(5)(59.5[q12.1]-59.8[q12.1), dim(5)(157.5[q33.3]-158.0[q33.3]), dim(6)(13.0[p24.1]-13.3[p24.1]), dim(6)(151.0[q25.1]-151.[q25.1]), dim(6)(156.7[q25.2]-156.9[q25.2]), dim(7)(50.1[p12.1]-50.2[p12.1]), dim(8)(60.2[q12.1]-60.5[12.1]), dim(8)(96.0[q22.1]-96.0[q22.1]), dim(8)(132.9[q24.21]-133.1[q24.21]), dim(9)(0.03[p24.3]-38.5[p13.1]), dim(10)(68.0[q21.3]-68.2[q213.]), dim(10)(111.7[q25.1]-111.8[q25.1]), dim(12)(90.8[q22]-91.0[q22]), dim(12)(109.9[q24.13]-110.7[q24.13]), enh(13)(27.6[q12.12]-27.7[q12.12]), dim(13)(47.8[q14.11]-48.0[q14.11]), dim(14)(21.3[p11.2]-21.4[p11.2]), dim(14)(74.2[q24.1]-74.3[q24.1]), dim(16)(30.8[p11.2]-31.3[p11.2]), dim(19)(4.8[p13.3]-4.9[p13.3]), dim(20)(42.5[q13.11]-42.7[q13.11]), dim(21)(14.2[p11.2]-25.5[q21.1]), enh(21)(29.0[q21.1]-29.2[q21.2]), enh(21)(30.8[q21.2]-38.0[q22.12]), dim(21)(44.4[q22.3]-46.9[q22.3]).
9051	15	M	46,XY[40]		dim(4)(190.41[q35.2]-191.31[q35.2]).

9880	2	F	47,XX,+5,del(9)(p13p24)[14].ish del(9)(LSI9p21-)[5]	p16/9cen; 1R2G0F (82%)	dim(3)(60.02[p21.1]-60.55[p21.1]), enh(5), dim(8)(60.2[q12.1]-60.83[q12.1), dim(9)(17.4[p22.1]-21.92[p21.3), dim(9)(21.92[p21.3]-21.99[p21.3])x2, dim(9)(21.99[p21.3]-33.45[p13.3]), dim(14)(21.43[p11.2]-22.05[p11.1])x2, dim(15)(36.2[q13.2]-36.33[q13.2])x2, dim(20)(4.95[p13]-14.65[p12.1]).
10283	6	M	46,XY,del(6)(q1q2)[4]	p16/9cen; 1R2G0F (22.5%)	dim(1)(47.42[p33]-47.49[p33]), dim(3)(36.10[p22.1]-36.31[p22.1]), dim(6)(63.41[q11.2]-95.96[q16.2]), dim(9)(21.81[p21.3]-22.37[p21.3]), dim(11)(7.76[p15.4]-7.80[p15.4]), dim(Y)(2.97[p11.31]-6.83[p11.2]).
10284	7	F	46,XX,del(9)(p2?),del(9)(p2?)[8]	p16/9cen; 0R2G0F (44%)	enh(2)(25.22[p24.1]-25.27[p24.1]), dim(9)(0.15[p24.3]-20.78[p21.3]), dim(9)(20.78[p21.3]-22.69[p21.3])x2, dim(9)(22.69[p21.3]-41.90[p11.2]), enh(9)(91.77[q22.31]-91.81[q22.31]), dim(10)(111.80[q25.1]-111.86[q25.1]), dim(19)(23.68[p13.11]-23.72[p13.11]), dim(20)(28.08[q11.1]-62.38[q13.33]), dim(21)(38.69[q22.13]-38.74[q22.13]), enh(21)(43.94[q22.3]-43.97[q22.3]).
10297	15	F	46,XX,invdup(1)(q3q21),t(9;12)(p13;p1?3)[8]	ETV6 DC split; 1R0G2F (67%)	enh(5)(161.9[q34]-163.07[q34]).

10442	12	M	FAIL	TEL/AML1; 6R1G0F/ 6R2G0F 21q sub-tel 0R4-6G1F	dim(4)(69.0[q13.2]-69.7[q13.2]), enh(5), dim(7)(49.9[p12.2]-50.2[p12.2]), enh(9)(3.6[p24.2]-37.2[p13.1]), dim(12)(0.03[p13.33]-19.7[p12.3]), dim(12)(109.9[q24.12]-110.7[q24.13]), dim(16)(45.0[q11.2]-88.6[q24.3]), enh(21)(13.3[p11.2]-46.9[q22.3]).
10591	11	F	46,XX[20]		enh(4)(14.18[p15.33]-14.27[p15.33]), dim(7)(50.19[p12.1]-50.25[p12.1]).
11536	14	M	46,XY,der(3)del(3)(q1q2)?t(3;7;8)(q2;p2;q1),der(7)?t(3;7;8)(q2;p2;q1),der(8)?t(3;7;8)(q2;p2;q1)[9].ish der(8)?t(3;7;8)(wcp8+,wcp7+)[2]		dim(3)(99.35[q12.3]-125.22[q21.2]), dim(3)(143.98[q23]-152.46[q25.1]), dim(6)(74.04[q14.1]-76.09[q14.1), dim(7)(14.64[p21.3]-50.63[p12.1]), dim(8)(52.63[q11.22]-62.94[q12.2]), dim(9)(21.85[p21.3]-22.41[p21.3])x2.
11835	14	F	46,XX,t(12;17)(p13;q11)[10]		dim(9)(19.5[p21.3]-22.2[p21.3])x2, dim(17)(26.46[p11.1]-26.63[p11.1]).
11894	29	M	46,XY[20]		dim(1)(0.79[36.23]-3.54[p34.3]), dim(2)(1.49[p25.3]-1.49[p25.3]), dim(2)(10.43[p25.1]-10.45[p25.1]), dim(5)(0.66[p15.33]-1.79[p15.33]), dim(7)(0.14[p22.3]-25.04[p15.2]), dim(7)(71.71[q11.22]-75.79[q11.23]), dim(8)(142.21[q24.3]-146.03[q24.3]), dim(11)(66.80[q13.1]-66.83[q13.1]), dim(11)(75.53[q13.3]-75.70[q13.3]), dim(11)(118.27[q23.3]-118.35[q23.4]), dim(12)(6.15[p13.32]-6.25[p13.32]), dim(16)(28.54[p11.2]-30.50[p11.2]), enh(X)(2.70[p22.33]-47.80[p11.23]).

12377	11	F	46,XX,-20,+mar[7]		dim(9)(21.89[p21.3]-22.04[p21.3]), dim(13)(47.88[q14.11]-48.06[q14.11]), enh(13)(89.37[q31.1]-92.03[q31.3]), dim(X)(4.4[p22.33]-4.99[p22.33]).
12462	7	M	46,XY,add(7)(q11),?del(9)(p12p12)[10]	p16/9; 1R2G0F (69%)	dim(4)(186.18[q35.2]-188.19[q35.2]), dim(9)(9.77[p23]-11.23[p23]), dim(9)(21.65[p21.3]-21.82[p21.3]), dim(9)(21.82[p21.3]-21.99[p21.3])x2, dim(9)(21.99[p21.3]-22.58[p21.3]), dim(13)(43.74[q13.3]-43.92[q13.3]), dim(16)(52.54[q12.2]-52.66[q12.2]), dim(19)(10.61[p13.2]-11.02[p13.2]).

Appendix 15.

The 1200 genes selected for printing onto the CodeLink 16-assay gene expression arrays.

Gene symbol	Gene name
TBXAS1	Human TBXAS1 gene for thromboxane synthase
CD44	Human cell surface glycoprotein CD44
KIAA0056	Human mRNA for KIAA0056 gene
DHPS	deoxyhypusine synthase
KIAA1536	KIAA1536 protein
HAGH	hydroxyacyl glutathione hydrolase
JMJD3	KIAA0346 protein
STCH	stress 70 protein chaperone microsome-associated 60kD
TAF12	TAF12 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 20kDa (TAF12)
DDX3Y	DEAD/H Asp-Glu-Ala-Asp/His box polypeptide Y chromosome
RPS4Y1	ribosomal protein S4 Y-linked
NULL	Homo sapiens cDNA FLJ22448 fis clone HRC09541
NR2C1	nuclear receptor subfamily 2 group C member 1
IMPA1	inositol myo 1 or 4 monophosphatase 1
KIAA0033	KIAA0033 protein
PIP3E	phosphoinositide-binding protein PIP3-E (PIP3-E)
CD3Z	CD3Z antigen zeta polypeptide TiT3 complex
CRY1	cryptochrome 1 photolyase-like
RNF111	ring finger protein 11
MAN2B2	KIAA0935 protein
IL13RA1	interleukin 13 receptor, alpha 1 (IL13RA1)
STK17B	Human death-associated protein kinase
L3MBTL	Human l(3)mbt protein homolog mRNA
RTCD1	RNA terminal phosphate cyclase domain 1 (RTCD1)
MACF1	Homo sapiens mRNA for KIAA0754 protein
MAP3K11	mitogen-activated protein kinase kinase kinase 11 (MAP3K11)
YWHAQ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide (YWHAQ)
FLJ10803	hypothetical protein FLJ10803
MSH3	mutS E. coli homolog 3
NP25	neuronal protein
38241	septin 11 (SEPT11)
RSU1	Ras suppressor protein 1 (RSU1), transcript variant 1
LOC113251	c-Mpl binding protein (LOC113251), transcript variant 1
RBM16	RNA binding motif protein 16 (RBM16)
UCKL1	uridine kinase-like 1 (URKL1)
SERPINH1	serine (or cysteine) proteinase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1) (SERPINH1)
HIST1H4F	H4 histone family member G
NAB1	NGFI-A binding protein 1 ERG1 binding protein 1
POLR2C	polymerase RNA II DNA directed polypeptide C33kD
KIAA0117	RBP1-like protein
PON2	Paraoxonase 2
CD99	CD99 antigen (CD99)

CCND2	Cyclin D2
GYPC	glycophorin C (Gerbich blood group)
SEMA6A	Semaphorin 6A
SLC2A5	Solute carrier family 2 (facilitated glucose/fructose transporter), member 5
PRX	Periaxin
CTDSPL	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like (CTDSPL)
ECM1	Extracellular matrix protein 1
GPR110	GPR110 G protein-coupled receptor 110
DKFZP564A2416	DKFZP564A2416 unknown protein with a histone H5 signature.
FXYD5	FXYD domain-containing ion transport regulator 5
TUBA1	Tubulin, alpha 1 (testis specific)
CNN3	Calponin 3, acidic
KIAA0233	KIAA0233 gene product
FLJ21897	FLJ21897
DCTN4	Dynactin 4 (p62)
LIMD1	LIM domains containing 1
CASP10	Caspase 10, apoptosis related cysteine protease
FLJ21140	FLJ21140 (Ser/Thr protein kinase)
C5orf13	P311 protein. Similar to gastrin/cholecystokinin type B receptor.
CDC16	CDC16 cell division cycle 16 homolog
ITGAE	Integrin, alpha E (antigen CD103, human mucosal lymphocyte antigen 1)
C1orf24	Chromosome 1 ORF 24
ITGA6	Integrin, alpha 6
TNIP1	Nef-associated factor 1
PSTPIP2	proline-serine-threonine phosphatase interacting protein 2
IAN4L1	Immune associated nucleotide 4 like 1
FBXW7	F-box and WD-40 domain protein 7 (archipelago homolog, Drosophila)
RAB11A	RAB11A, member RAS oncogene family
PRSS25	Protease, serine, 25
INPP4B	Inositol polyphosphate-4-phosphatase, type II
KIAA1115	KIAA1115 protein
FLJ13197	Hypothetical protein FLJ13197
IL18BP	Interleukin 18 binding protein
MMP28	Matrix metalloproteinase 28
IFITM3	Interferon induced transmembrane protein 3 (I-8U)
MUC4	Mucin 4
TRAM2	TRAM-like protein
IGJ	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides
FLJ23058	hypothetical protein FLJ23058
INPP1	Inositol polyphosphate-1-phosphatase
ZC3HDC7	HSPC055 protein
LCP2	Lymphocyte cytosolic protein 2
TM4SF2	Transmembrane 4 superfamily member 2
MGC4677	Hypothetical protein MGC4677
TNFRSF1B	Tumor necrosis factor receptor superfamily, member 1B
LAIR1	Leukocyte-associated Ig-like receptor 1
ADCY3	Adenylate cyclase 3.

IDI1	Isopentenyl-diphosphate delta isomerase
CDS2	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 2
CHN2	Chimerin 2
CLN2	Ceroid-lipofuscinosis, neuronal 2, late infantile (Jansky-Bielschowsky disease). A pepstatin-insensitive lysosomal peptidase.
SIAT9	Sialyltransferase 9
NR3C2	Nuclear receptor subfamily 3, group C, member 2
TBX21	T-box 21
LTBR	Lymphotoxin beta receptor (TNFR superfamily, member 3)
PAWR	PRKC, apoptosis, WT1, regulator
KIAA0643	KIAA0643 protein
STX3A	Syntaxin 3A
FAT	FAT tumor suppressor homolog 1 (Drosophila)
LRMP	lymphoid-restricted membrane protein
PBX1	pre-B-cell leukemia transcription factor 1
DKFZp586C1019	DKFZp586C1019
RDGBB	retinal degeneration B beta
MGC10485	hypothetical protein MGC10485
EB1	E2a-Pbx1-associated protein
HELO1	homolog of yeast long chain polyunsaturated fatty acid elongation enzyme 2
KIAA0922	KIAA0922 protein
ADPRT	ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase)
WNT16	wingless-type MMTV integration site family, member 16
NULL	FLJ14167
FLJ20489	EST
ICB-1	basement membrane-induced gene
CKLFSF3	chemokine-like factor super family 3
NID2	nidogen 2 (osteonidogen)
FLJ10324	hypothetical protein FLJ10324
ACAD8	acyl-Coenzyme A dehydrogenase family, member 8
SS18L2	synovial sarcoma translocation gene on chromosome 18-like 2
SNAP23	synaptosomal-associated protein, 23kDa
HTRA3	serine protease HTRA3
SYNPO	synaptopodin
PLEKHF2	phafin 2
FAM3C	family with sequence similarity 3, member C
GOLGA3	golgi autoantigen, golgin subfamily a, 3
ARL7	ADP-ribosylation factor-like 7
KIAA0802	KIAA0802 protein
ARHGAP8	Rho GTPase activating protein 8
SLC27A2	solute carrier family 27 (fatty acid transporter), member 2
FLI1	Friend leukemia virus integration 1
KIAA0247	KIAA0247 gene product
TBCE	tubulin-specific chaperone e
SYT1	synaptotagmin I
KIAA0143	KIAA0143 protein
HA-1	minor histocompatibility antigen HA-1
SESN1	p53 regulated PA26 nuclear protein
FKBP5	FK506 binding protein 5

CIRBP	cold inducible RNA binding protein
GALNT1	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 1 (GalNAc-T1)
NULL	cDNA FLJ12010 fis
IGSF3	immunoglobulin superfamily, member 3
RNASET2	ribonuclease 6 precursor
SORBS1	sorbin and SH3 domain containing 1
PTBP2	polypyrimidine tract binding protein 2
FLJ31978	hypothetical protein FLJ31978
APBB2	amyloid beta (A4) precursor protein-binding, family B, member 2 (Fe65-like)
MGC4170	MGC4170 protein
KIAA0040	KIAA0040 gene product
BUB3	BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast)
H2AFZ	H2A histone family, member Z (H2AFZ)
SEMA4C	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4C
TOP2A	topoisomerase (DNA) II alpha 170kDa
CLTC	clathrin, heavy polypeptide (Hc)
UAP1	UDP-N-acteylglucosamine pyrophosphorylase 1
MSN	Moesin (membrane-organizing extensio spike protein)
KCNJ16	potassium inwardly-rectifying channel, subfamily J, member 16 (KCNJ16), transcript variant 1
SLAMF1	signaling lymphocytic activation molecule family member 1 (SLAMF1)
PGK1	Phosphoglycerate kinase 1
PDHA1	Pyruvate dehydrogenase (lipoamide) alpha 1
PLP2	Proteolipid protein 2 (colonic epithelium-enriched)
VPS26	Vacuolar protein sorting 26 (yeast)
CUL4B	Cullin 4B
MIR16	membrane interacting protein of RGS16
PRKAR2B	Protein kinase, cAMP-dependent, regulatory, type II, beta
BACH1	BTB and CNC homology 1, basic leucine zipper transcription factor 1
FTSJ1	FtsJ homolog 1 (E. coli)
UREB1	Upstream regulatory element binding protein 1
ATRX	Alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)
TNRC11	trinucleotide repeat containing 11 (THR-associated protein, 230 kDa subunit)
PSMD10	proteasome (prosome, macropain) 26S subunit, non-ATPase, 10
CALM1	Calmodulin 1 (phosphorylase kinase, delta)
RBBP7	Retinoblastoma binding protein 7
USP9X	Ubiquitin specific protease 9
TPD52	Tumor protein D52
UBE2A	Ubiquitin-conjugating enzyme E2A (RAD6 homolog)
ATP5J	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6
SYBL1	Synaptobrevin-like 1
HPRT1	Hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome)
HDAC6	Histone deacetylase 6

SH3BP2	SH3-domain binding protein 2
ZNF183	zinc finger protein 183 (RING finger, C3HC4 type)
KIAA0179	KIAA0179 protein.
MCTS1	MCT-1 protein
USP16	Ubiquitin specific protease 16; de-ubiquitinates histone H2A; ubiquitous expression.
HPS4	Hermansky-Pudlak syndrome 4
UXT	Ubiquitously-expressed transcript
MST4	Mst3 and SOK1-related kinase/STE20-like kinase; contains a Ser/Thr protein kinase domain
UPF3B	Similar to yeast Upf3, variant B
ZCWCC2	Hypothetical protein FLJ11565
CKLFSF2	Chemokine-like factor super family 2.
HNRPH2	Heterogeneous nuclear ribonucleoprotein H2 (H')
SH3BGRL	SH3 domain binding glutamic acid-rich protein like
PRDX4	Peroxiredoxin 4
FLJ21174	Hypothetical protein FLJ21174
IMPA2	Inositol(myo)-1(or 4)-monophosphatase 2
PSMC1	proteasome (prosome, macropain) 26S subunit, ATPase, 1
SPTBN1	Spectrin, beta, non-erythrocytic 1
LOC91316	Homo sapiens cDNA FLJ32313 fis, clone PROST2003232, weakly similar to BETA-GLUCURONIDASE PRECURSOR (EC 3.2.1.31)
TMSNB	Thymosin, beta
MGAT4A	Mannosyl (alpha-1,3)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase
MGC15737	hypothetical protein MGC15737
SOD1	Superoxide dismutase 1, soluble
HSPA1A	Heat shock 70kD protein 1A
HMGN1	High-mobility group (nonhistone chromosomal) protein 14; member of the HMG 14/17 family
VBP1	Von Hippel-Lindau binding protein 1
HTATSF1	HIV TAT specific factor 1
LAMP2	Lysosomal-associated membrane protein 2
MGAT2	Mannosyl (alpha-1,6-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
HMGB3	High-mobility group (nonhistone chromosomal) protein 4
CAPN2	Calpain 2, (m/II) large subunit; calcium-dependent Cys protease.
PRPS1	Phosphoribosyl pyrophosphate synthetase 1; purine biosynthesis.
GPR56	G protein-coupled receptor 56
TREX2	Three prime repair exonuclease 2
RAG2	Recombination activating gene 2; V(D)J recombinase.
ARMCX1	ALEX1 protein
SH3KBP1	SH3-domain kinase binding protein 1
MGC23937	Hypothetical protein MGC23937
DHX16	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 16
ZHX2	KIAA0854 protein
HCCS	Holocytochrome c synthase (cytochrome c heme-lyase)
SLC9A6	Solute carrier family 9 (sodium/hydrogen exchanger), isoform 6
RP2	Retinitis pigmentosa 2 (X-linked recessive)

SLK	Ste20-related serine/threonine kinase
TTC3	Tetratricopeptide repeat domain 3
CDC5L	CDC5 cell division cycle 5-like (S. pombe)
ADPRTL2	ADP-ribosyltransferase (NAD+; poly(ADP-ribose) polymerase)-like 2
UBQLN2	Ubiquilin 2
FUT7	fucosyltransferase 7 (alpha (1,3) fucosyltransferase)
DLG3	discs, large (Drosophila) homolog 3 (neuroendocrine-dlg)
SH3BP5	SH3-domain binding protein 5 (BTK-associated)
SLC16A2	solute carrier family 16 (monocarboxylic acid transporters), member 2 (putative transporter)
NPY	neuropeptide Y
ITSN1	intersectin 1 (SH3 domain protein)
PHF3	PHD finger protein 3
ADAM10	a disintegrin and metalloproteinase domain 10
SETBP1	SET binding protein 1 (SETBP1)
IGFBP4	insulin-like growth factor binding protein 4 (IGFBP4)
C20orf103	chromosome 20 open reading frame 103
NAV1	neuron navigator 1
MEIS1	Meis1, myeloid ecotropic viral integration site 1 homolog
MYO5C	myosin 5C
ALOX5	arachidonate 5-lipoxygenase
FHIT	fragile histidine triad gene
PTPRC	protein tyrosine phosphatase, receptor type, C
SCP2	sterol carrier protein 2
PRKCH	Protein Kinase C eta isoform.
IMP-2	IGF-II mRNA-binding protein 2
CORO1C	coronin, actin binding protein, 1C
MBNL1	muscleblind-like (Drosophila)
DNTT	deoxynucleotidyltransferase, terminal
TPP2	tripeptidyl peptidase II
ELK3	ELK3, ETS-domain protein (SRF accessory protein 2)
IGFBP7	insulin-like growth factor binding protein 7
MAP3K5	mitogen-activated protein kinase kinase kinase 5
KIAA1576	KIAA1576 protein
SLC35E3	bladder cancer over expressed protein
CAMK2D	EST; by BLAT calcium/calmodulin-dependent Protine Kinase type II Delta chain (CAMK GROUP I)
RHOBTB3	Rho-related BTB domain containing 3
LGALS13	lectin, galactoside-binding, soluble, 1 (galectin 1)
ATP2B4	calcium transporting ATPase plasma membrane protein.
FLJ30525	hypothetical protein FLJ30525
WDSAM1	hypothetical protein FLJ36175
HIF1A	hypoxia-inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
SORD	sorbitol dehydrogenase
TCF4	transcription factor 4
LY75	lymphocyte antigen 75
PLXNC1	plexin C1
PDE4B	phosphodiesterase 4B, cAMP-specific
D2S448	Melanoma associated gene
SRPRB	APMCF1 protein

FLJ10597	hypothetical protein FLJ10597
EGLN1	egl nine homolog 1 (C. elegans)
BITE	p10-binding protein
TM9SF2	transmembrane 9 superfamily member 2
BRE	brain and reproductive organ-expressed (TNFRSF1A modulator)
C11orf24	chromosome 11 open reading frame 24
NULL	KIAA0239 protein
ARPC2	actin related protein 2/3 complex, subunit 2, 34kDa
BASP1	brain abundant, membrane attached signal protein 1
M96	likely ortholog of mouse metal response element binding transcription factor 2
FEZ2	fasciculation and elongation protein zeta 2 (zygin II)
FAIM	Fas apoptotic inhibitory molecule
SCHIP1	schwannomin interacting protein 1
WBSCR20A	Williams Beuren syndrome chromosome region 20A
DDEF1	development and differentiation enhancing factor 1
POLE	polymerase (DNA directed), epsilon
C22orf9	KIAA0930 protein
MPZL1	hypothetical protein FLJ21047
MOSPD3	motile sperm domain containing 3 (MOSPD3)
TSAP6	dudulin 2 (TSAP6)
C9orf10	chromosome 9 open reading frame 10 (C9orf10)
PDCD6IP	programmed cell death 6 interacting protein (PDCD6IP)
IRF5	interferon regulatory factor 5 (IRF5), transcript variant 2
CENTA1	centaurin, alpha 1 (CENTA1)
PIP5K2C	phosphatidylinositol-4-phosphate 5-kinase, type II, gamma (PIP5K2C)
HLA-DPB1	major histocompatibility complex, class II, DP beta 1
SNX2	sorting nexin 2
NUCB2	nucleobindin 2
HLA-DRB3	major histocompatibility complex, class II, DR beta 3
CD79B	CD79B antigen (immunoglobulin-associated beta)
CD3E	CD3E antigen, epsilon polypeptide (TiT3 complex)
CD19	CD19 antigen
HLA-DRA	major histocompatibility complex, class II, DR alpha
CD74	CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated)
SH2D1A	SH2 domain protein 1A, Duncan's disease (lymphoproliferative syndrome)
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1
CD3D	CD3D antigen, delta polypeptide (TiT3 complex)
CD7	CD7 antigen (p41)
TRIM	T-cell receptor interacting molecule
HLA-DMA	MHC, class IIa, HLA-DMA
PAX5	paired box gene 5 (B-cell lineage specific activator protein)
EBF	cDNA FLJ39389 fis
USP20	ubiquitin specific protease 20
TCF7	transcription factor 7 (T-cell specific, HMG-box)
BLNK	B-cell linker
CD24	CD24 antigen (small cell lung carcinoma cluster 4 antigen)
BIN2	bridging integrator 2

C14orf139	hypothetical protein FLJ21276
KIAA0748	KIAA0748 gene product
HMGCS1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)
TCL1A	T-cell leukemia/lymphoma 1A
RAB32	RAB32, member RAS oncogene family
MAL	mal, T-cell differentiation protein
CD79A	CD79A antigen (immunoglobulin-associated alpha)
TRD@	T cell receptor delta locus
UBASH3A	ubiquitin associated and SH3 domain containing, A
DOK3	hypothetical protein FLJ22570
HRD1	HRD1 protein
NULL	cDNA FLJ37485 fis
M17S2	membrane component, chromosome 17, surface marker 2 (ovarian carcinoma antigen CA125)
HMGCR	3-hydroxy-3-methylglutaryl-Coenzyme A reductase
HLA-DMB	major histocompatibility complex, class II, DM beta
PLCG2	phospholipase C, gamma 2 (phosphatidylinositol-specific)
POU2AF1	POU domain, class 2, associating factor 1
PRKCQ	protein kinase C, theta
CSRP2	cysteine and glycine-rich protein 2
VPREB3	pre-B lymphocyte gene 3
NULL	cDNA FLJ31057 fis
SEPW1	selenoprotein W, 1
SLC9A3R1	solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 1
TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)
PTPN7	protein tyrosine phosphatase, non-receptor type 7
FXYD2	FXYD domain containing ion transport regulator 2
KIAA1033	KIAA1033 protein
IGHM	immunoglobulin heavy constant mu
DC-TM4F2	tetraspanin similar to TM4SF9
MGC45416	hypothetical protein MGC45416
INPP5D	inositol polyphosphate-5-phosphatase, 145kDa
LAT	Linker for activation of T cells
SLIT1	slit homolog 1 (Drosophila)
OIP106	KIAA1042 protein
KIAA1323	KIAA1323
GRAP2	GRB2-related adaptor protein 2 (GRAP2)
CD3G	CD3G antigen, gamma polypeptide (TiT3 complex) (CD3G)
FLJ12722	FLJ12722
	EST
FBN2	Fibrillin 2 (congenital contractural arachnodactyly)
ARHGEF4	Rho guanine nucleotide exchange factor (GEF) 4
PCLO	Piccolo (presynaptic cytomatrix protein)
BIRC7	baculoviral IAP repeat-containing 7 (livin)
FLJ20154	Hypothetical protein FLJ20154
DSC3	Desmocollin 3
N33	Putative prostate cancer tumor suppressor gene
PARG1	PTPL1-associated RhoGAP 1
TCFL5	Transcription factor-like 5 (basic helix-loop-helix)

KCNN1	Potassium intermediate/small conductance calcium-activated channel, subfamily N, member 1
CBFA2T3	Core-binding factor, runt domain, alpha subunit 2; translocated to, 3
HAP1	Huntingtin-associated protein 1 (neuroan 1)
RGM	hypothetical protein from EUROIMAGE 363668 RGM: likely ortholog of chicken repulsive guidance molecule
FYB	FYN-binding protein
EPOR	Eerythropoietin receptor
EPN2	Epsin 2
CLIC5	chloride intracellular channel 5
DSC2	Desmocollin 2
FLJ11259	Hypothetical protein FLJ11259
TNS	Tensin
FKBP1A	FK506 binding protein 1A (12kD)
TIF1	Transcriptional intermediary factor 1
PRKCB1	Protein kinase C, beta 1
SLC12A2	solute carrier family 12 (sodium/potassium/chloride transporters), member 2
FLJ32658	Homo sapiens cDNA FLJ32658
PIK3C3	Phosphoinositide-3-kinase, class 3
RAG1	Recombination activating gene 1
IQGAP2	IQ motif containing GTPase activating protein 2
MAPK13	Mitogen-activated protein kinase 13
HMG20A	High-mobility group 20A
FLJ21415	hypothetical protein FLJ21415 (FLJ21415)
NOVA1	Neuro-oncological ventral antigen 1
TXNRD1	Thioredoxin reductase 1
TERF2	Telomeric repeat binding factor 2
ABHD3	abhydrolase domain containing 3
MGC23911	hypothetical protein MGC23911
MDK	Midkine (neurite growth-promoting factor 2).
PIM1	Pim-1 oncogene
NRN1	Neuritin 1
FLJ23749	Hypothetical protein FLJ23749
ENPP4	Ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative function)
B4GALT6	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 6
SH3GLB2	SH3-domain GRB2-like endophilin B2
GNPDA1	Glucosamine-6-phosphate isomerase
FUCA1	Fucosidase, alpha-L- 1, tissue
KIAA1223	Hypothetical protein KIAA1223
DSG2	Desmoglein 2 desmosomal cadherin
STMN1	Stathmin 1/oncoprotein 18 leukemia-associated phosphoprotein
ARHGEF12	Rho guanine nucleotide exchange factor (GEF) 12
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2
RY1	Putative nucleic acid binding protein RY-1
HLA-DOB	major histocompatibility complex, class II, DO beta
CTGF	connective tissue growth factor
PTP4A3	protein tyrosine phosphatase type IVA, member 3
KHDRBS3	KH domain containing, RNA binding, signal transduction

	associated 3
FADS3	fatty acid desaturase 3
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4 (SMARCA4)
EFNB2	ephrin-B2 (EFNB2)
AKAP12	A kinase (PRKA) anchor protein (gravin) 12 (AKAP12), transcript variant 2
FARP1	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived) (FARP1), transcript variant 1, mRNA
SCN3A	sodium channel, voltage-gated, type III, alpha (SCN3A)
FOXO1A	forkhead box O1A (rhabdomyosarcoma) (FOXO1A)
CBFA2T1	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related
FCHSD2	FCH and double SH3 domains 2 (FCHSD2)
GBA3	glucosidase, beta, acid 3 (cytosolic) (GBA3)
POU4F1	POU domain, class 4, transcription factor 1
TRH	thyrotropin-releasing hormone
CAV1	caveolin 1, caveolae protein, 22kDa
CACNA2D2	calcium channel, voltage-dependent, alpha 2/delta subunit 2
CD58	CD58 antigen, (lymphocyte function-associated antigen 3)
NCALD	neurocalcin delta
ROBO1	roundabout, axon guidance receptor, homolog 1 (Drosophila)
PTGIR	prostaglandin I2 (prostacyclin) receptor (IP)
SLC25A1	solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1
FLJ12428	hypothetical protein FLJ12428
MAN1A1	mannosidase, alpha, class 1A, member 1
C11orf9	chromosome 11 open reading frame 9
HSA250839	gene for serine/threonine protein kinase
HYAL2	hyaluronoglucosaminidase 2
IL5RA	interleukin 5 receptor, alpha
PSD3	ADP-ribosylation factor guanine nucleotide factor 6
ISG20	interferon stimulated gene 20kDa
RCBTB1	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 1
ITGB4	integrin, beta 4
ADRM1	adhesion regulating molecule 1
PRAME	preferentially expressed antigen in melanoma
PYGO2	pygopus 2
HIST1H2BD	histone 1, H2bd
EGFL7	EGF-like-domain, multiple 7
MGC3222	hypothetical protein MGC3222
DISC1	disrupted in schizophrenia 1
PALM	paralemmin
WFDC1	WAP four-disulfide core domain 1
TCN1	transcobalamin I (vitamin B12 binding protein, R binder family)
SLCO4A1	solute carrier family 21 (organic anion transporter), member 12
POLR2J2	DNA directed RNA polymerase II polypeptide J-related gene
TWSG1	twisted gastrulation homolog 1 (Drosophila)
DKFZP434H132	DKFZP434H132 protein
BAIAP3	BAI1-associated protein 3
SCML2	sex comb on midleg-like 2 (Drosophila)

PGDS	prostaglandin D2 synthase, hematopoietic
KIAA0830	KIAA0830 protein
IL17RB	interleukin 17 receptor B
BEX1	brain expressed, X-linked 1
TACSTD1	tumor-associated calcium signal transducer 1
LAPTM5	Lysosomal-associated multispanning membrane protein-5
BAALC	brain and acute leukemia, cytoplasmic
CLECSF5	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 5
USP10	ubiquitin specific protease 10
NARFL	protein related to Narf
MKNK2	MAP kinase-interacting serine/threonine kinase 2
RGS10	regulator of G-protein signalling 10
RPS6KA1	ribosomal protein S6 kinase, 90kDa, polypeptide 1
NIT2	Nit protein 2
VAMP5	vesicle-associated membrane protein 5 (myobrevin)
HIST1H2BH	histone 1, H2bh
CD244	natural killer cell receptor 2B4
IER2	immediate early protein
ARHGAP1	Rho GTPase activating protein 1
PTPN12	protein tyrosine phosphatase, non-receptor type 12
HIST1H2BG	histone 1, H2bf
LCP1	lymphocyte cytosolic protein 1 (L-plastin)
C6orf4	chromosome 6 open reading frame 4
WBSCR5	Williams-Beuren syndrome chromosome region 5
DKFZP564K0822	hypothetical protein DKFZp564K0822
ARHGEF6	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6
HIST1H2BI	histone 1, H2bi
AP1B1	adaptor-related protein complex 1, beta 1 subunit
MDS006	x 006 protein
MHC2TA	MHC class II transactivator
TIEG2	TGFB inducible early growth response 2
ITPR1	inositol 1,4,5-triphosphate receptor, type 1
SAH	SA hypertension-associated homolog (rat)
VEGF	vascular endothelial growth factor (VEGF)
PDGFC	platelet derived growth factor C (PDGFC)
PRSS21	protease, serine, 21 (testisin)
PTX3	pentaxin-related gene, rapidly induced by IL-1 beta (PTX3)
MAN2A2	mannosidase, alpha, class 2A, member 2 (MAN2A2)
HCA112	hepatocellular carcinoma-associated antigen 112 (HCA112)
HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
PKD1like	polycystic kidney disease 1-like (PKD1-like), transcript variant 1
SIX3	sine oculis homeobox homolog 3 (Drosophila) (SIX3)
STAB1	stabilin 1
HT011	uncharacterized hypothalamus protein HT011
PTGDS	prostaglandin D2 synthase 21kDa (brain)
FAD104	FAD104
PGBD5	piggyBac transposable element derived 5
MRC2	mannose receptor, C type 2
AGRN	agrin
AQP3	aquaporin 3

MMP2	matrix metalloproteinase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)
CST7	cystatin F (leukocystatin)
RCN1	reticulocalbin 1, EF-hand calcium binding domain
ANXA8	annexin A8
CALR	calreticulin
MST1	macrophage stimulating 1 (hepatocyte growth factor-like)
FGF13	fibroblast growth factor 13
	Homo sapiens transcribed sequence with strong similarity to protein sp:P05023 (H.sapiens) A1A1_HUMAN Sodium/potassium-transporting ATPase alpha-1 chain precursor (Sodium pump 1) (Na+/K+ ATPase 1)
COL2A1	collagen, type II, alpha 1 (primary osteoarthritis, spondyloepiphyseal dysplasia, congenital)
STXBP1	syntaxin binding protein 1
NULL	macrophage stimulating, pseudogene 9
IGFBP2	insulin-like growth factor binding protein 2, 36kDa
MGC3121	hypothetical protein MGC3121
CTSW	cathepsin W (lymphopain)
LASS4	hypothetical protein FLJ12089
FLJ20605	hypothetical protein FLJ20605
NISCH	nischarin
ARHGAP4	Rho GTPase activating protein 4
AUTS2	autism susceptibility candidate 2
NDST2	N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 2
CPA3	carboxypeptidase A3 (mast cell)
NAALADL1	N-acetylated alpha-linked acidic dipeptidase-like
MAP1A	microtubule-associated protein 1A
MPO	myeloperoxidase
ALCAM	activated leukocyte cell adhesion molecule
C18orf1	chromosome 18 open reading frame 1
ARMET	arginine-rich, mutated in early stage tumors
LRFN4	hypothetical protein MGC3103
DKFZP564G2022	DKFZP564G2022 protein
JAG1	jagged 1 (Alagille syndrome)
SIAT10	alpha2,3-sialyltransferase
TLE1	transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)
PRODH	proline dehydrogenase (oxidase) 1
CSGlcA-T	KIAA1402 protein
SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4
SERPING1	serine (or cysteine) proteinase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)
KRT18	keratin 18
ASNS	asparagine synthetase
P4HB	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)
KIAA0652	KIAA0652 gene product
RASL12	Ras family member Ris
PDE3B	phosphodiesterase 3B, cGMP-inhibited
MEG3	maternally expressed 3

EGFL3	EGF-like-domain, multiple 3
SEC31L1	yeast Sec31p homolog
SMARCD3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3
GALNS	galactosamine (N-acetyl)-6-sulfate sulfatase (Morquio syndrome, mucopolysaccharidosis type IVA)
S100B	S100 calcium binding protein, beta (neural)
FLJ20551	hypothetical protein FLJ20551
KIAA0195	KIAA0195 gene product
FMR2	fragile X mental retardation 2
DF	D component of complement (adipsin)
BIG1	brefeldin A-inhibited guanine nucleotide-exchange protein 1
NULL	Homo sapiens mRNA for CMP-N-acetylneuraminic acid hydroxylase, complete cds.
ANKFY1	ankyrin repeat and FYVE domain containing 1
MAP4	microtubule-associated protein 4
HSPA5	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)
CDKN1C	cyclin-dependent kinase inhibitor 1C (p57, Kip2)
CLMN	calmin (calponin-like, transmembrane)
ITPR2	inositol 1,4,5-triphosphate receptor, type 2
HYOU1	hypoxia up-regulated 1
GCNT1	glucosaminyl (N-acetyl) transferase 1, core 2 (beta-1,6-N-acetylglucosaminyltransferase)
GRB10	growth factor receptor-bound protein 10
DJ971N18.2	hypothetical protein DJ971N18.2
MYH11	myosin, heavy polypeptide 11, smooth muscle
MN1	meningioma (disrupted in balanced translocation) 1
SPARC	secreted protein, acidic, cysteine-rich (osteonectin)
NRP1	neuropilin 1
CBFB	core-binding factor, beta subunit
HIMAP4	immunity associated protein 4
CDW52	CDW52 antigen (CAMPATH-1 antigen)
IL10RA	interleukin 10 receptor, alpha
CCR2	chemokine (C-C motif) receptor 2
ST18	suppression of tumorigenicity 18 (breast carcinoma) (zinc finger protein)
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
AK5	adenylate kinase 5
SLC7A7	solute carrier family 7 (cationic amino acid transporter, y+ system), member 7
FLJ22662	hypothetical protein FLJ22662
SLC38A1	solute carrier family 38, member 1
CTSS	cathepsin S
QPCT	glutaminyl-peptide cyclotransferase (glutaminyl cyclase)
PTPRM	protein tyrosine phosphatase, receptor type, M
HPCAL1	hippocalcin-like 1
EDG2	endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2
NAGA	N-acetylgalactosaminidase, alpha-
KIF17	kinesin family member 17
CD1C	CD1C antigen, c polypeptide
S100A8	S100 calcium binding protein A8 (calgranulin A)

LGALS2	lectin, galactoside-binding, soluble, 2 (galectin 2)
PLXNB2	plexin B2
CYBB	cytochrome b-245, beta polypeptide (chronic granulomatous disease)
MEF2A	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
CD1E	CD1E antigen, e polypeptide
CRA	cisplatin resistance associated
TNFAIP2	tumor necrosis factor, alpha-induced protein 2
KCTD12	hypothetical protein BC013764
CLECSF14	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 13 (macrophage-derived)
CLECSF12	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 12
DUSP6	dual specificity phosphatase 6
ANXA11	annexin A11
PGM1	phosphoglucomutase 1
TCTEL1	t-complex-associated-testis-expressed 1-like 1
PFC	properdin P factor, complement
BTBD14A	hypothetical protein MGC23427
HIC	I-mfa domain-containing protein
TGFBI	transforming growth factor, beta-induced, 68kDa
CLECSF6	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 6
GRHPR	glyoxylate reductase/hydroxypyruvate reductase
EGFL5	EGF-like-domain, multiple 5
TES	testis derived transcript (3 LIM domains)
TNFSF13	tumor necrosis factor (ligand) superfamily, member 13
DACH1	dachshund homolog (Drosophila)
NICAL	NEDD9 interacting protein with calponin homology and LIM domains
PBX3	pre-B-cell leukemia transcription factor 3
ALDH3A2	aldehyde dehydrogenase 3 family, member A2
AK2	adenylate kinase 2
CES1	carboxylesterase 1 (monocyte/macrophage serine esterase 1)
PRG1	proteoglycan 1, secretory granule
CAPG	capping protein (actin filament), gelsolin-like
CCL5	chemokine (C-C motif) ligand 5
SOCS2	suppressor of cytokine signaling 2
RNASE2	ribonuclease, RNase A family, 2 (liver, eosinophil-derived neurotoxin)
RNASE3	ribonuclease, RNase A family, 3 (eosinophil cationic protein)
FCGR1A	Fc fragment of IgG, high affinity Ia, receptor for (CD64)
SCPEP1	serine carboxypeptidase 1 (SCPEP1)
MEF2C	MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)
NKG7	natural killer cell group 7 sequence
RNPEP	arginyl aminopeptidase (aminopeptidase B)
ANXA5	annexin A5
TRPM4	transient receptor potential cation channel, subfamily M, member 4
UBE2V1	ubiquitin-conjugating enzyme E2 variant 1

HEXB	hexosaminidase B (beta polypeptide)
PLD3	phospholipase D3
FBP1	fructose-1,6-bisphosphatase 1
AKR7A2	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)
RPS6KA5	ribosomal protein S6 kinase, 90kDa, polypeptide 5
ITGA7	integrin, alpha 7
PREB	prolactin regulatory element binding
RNH	ribonuclease/angiogenin inhibitor
ADCY9	adenylate cyclase 9
PHKA2	phosphorylase kinase, alpha 2 (liver)
LOC51760	B/K protein
FLJ14153	hypothetical protein FLJ14153
CCL23	chemokine (C-C motif) ligand 23
MLL	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
C2	complement component 2
CPM	carboxypeptidase M
GAS7	growth arrest-specific 7
MRPL33	mitochondrial ribosomal protein L33
APOC2	apolipoprotein C-II
MAGEF1	melanoma antigen, family F, 1
SNX10	sorting nexin 10
DEXI	dexamethasone-induced transcript
TDRD7	tudor repeat associator with PCTAIRE 2
COL9A2	collagen, type IX, alpha 2
MRPL34	mitochondrial ribosomal protein L34
COX8A	cytochrome c oxidase subunit VIII
PSFL	anterior pharynx defective 1B-like
GAGED2	G antigen, family D, 2
GGA2	golgi associated, gamma adaptin ear containing, ARF binding protein 2
CTSL	cathepsin L
FEZ1	fasciculation and elongation protein zeta 1 (zygin I)
LGALS1	lectin, galactoside-binding, soluble, 1 (galectin 1)
ITGAX	integrin, alpha X (antigen CD11C (p150), alpha polypeptide)
KCNE1L	potassium voltage-gated channel, Isk-related family, member 1-like
BLOC1S1	GCN5 general control of amino-acid synthesis 5-like 1 (yeast)
PYCARD	apoptosis-associated speck-like protein containing a CARD
HK3	hexokinase 3 (white cell)
GUSB	glucuronidase, beta
LY86	lymphocyte antigen 86
LOC339005	Homo sapiens hypothetical protein LOC339005, mRNA (cDNA clone IMAGE:4837016), partial cds
GBA	glucosidase, beta; acid (includes glucosylceramidase)
PGAM1	phosphoglycerate mutase 1 (brain)
SAGE1	sarcoma antigen
GLG1	golgi apparatus protein 1
RPL22	ribosomal protein L22
ALDH3B1	aldehyde dehydrogenase 3 family, member B1
SPG21	acid cluster protein 33

PTPNS1	protein tyrosine phosphatase, non-receptor type substrate 1
CUL1	cullin 1
FLJ22318	hypothetical protein FLJ22318
STXBP2	syntaxin binding protein 2
ALOX5AP	arachidonate 5-lipoxygenase-activating protein
ITGA2B	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41B)
TIMP3	tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory)
HIG1	likely ortholog of mouse hypoxia induced gene 1
GP1BB	glycoprotein Ib (platelet), beta polypeptide
TAL1	T-cell acute lymphocytic leukemia 1
GATA1	GATA binding protein 1 (globin transcription factor 1)
SERPINI1	serine (or cysteine) proteinase inhibitor, clade I (neuroserpin), member 1
UROD	uroporphyrinogen decarboxylase
RHAG	Rhesus blood group-associated glycoprotein
LAPTM4B	lysosomal associated protein transmembrane 4 beta
MYH10	myosin, heavy polypeptide 10, non-muscle
MYL4	myosin, light polypeptide 4, alkali; atrial, embryonic
LIM	LIM protein (similar to rat protein kinase C-binding enigma)
CTNNBL1	catenin, beta like 1
ANK1	ankyrin 1, erythrocytic
PCCB	propionyl Coenzyme A carboxylase, beta polypeptide
SDPR	serum deprivation response (phosphatidylserine binding protein)
BMP2K	BMP2 inducible kinase
MINPP1	multiple inositol polyphosphate histidine phosphatase, 1
PROS1	protein S (alpha)
TNIK	Traf2 and NCK interacting kinase
TFR2	transferrin receptor 2
PTGS1	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)
KLF1	Kruppel-like factor 1 (erythroid)
NET1	neuroepithelial cell transforming gene 1
ICAM4	intercellular adhesion molecule 4, Landsteiner-Wiener blood group
p44S10	proteasome regulatory particle subunit p44S10
TPM1	tropomyosin 1 (alpha)
HTCD37	TcD37 homolog
NGFRAP1	nerve growth factor receptor (TNFRSF16) associated protein 1
CMAS	cytidine monophosphate N-acetylneuraminic acid synthetase
KCNH2	potassium voltage-gated channel, subfamily H (eag-related), member 2
MRPS12	mitochondrial ribosomal protein S12
GP1BA	glycoprotein Ib (platelet), alpha polypeptide
FADS2	fatty acid desaturase 2
ABCC4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4
KEL	Kell blood group
NEO1	neogenin homolog 1 (chicken)
PDCD10	programmed cell death 10
RYR3	ryanodine receptor 3
SQLE	squalene epoxidase

SLC39A4	solute carrier family 39 (zinc transporter), member 4
C14orf2	chromosome 14 open reading frame 2
FHL2	four and a half LIM domains 2
GJA4	gap junction protein, alpha 4, 37kDa (connexin 37)
GATA2	GATA binding protein 2
RDX	radixin
RASA1	RAS p21 protein activator (GTPase activating protein) 1
STXBP6	syntaxin binding protein 6 (amisyn)
PLOD2	procollagen-lysine, 2-oxoglutarate 5-dioxygenase (lysine hydroxylase) 2
DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)
PNMT	phenylethanolamine N-methyltransferase
PRKCBP1	protein kinase C binding protein 1
ARMC8	HSPC056 protein
SMARCA3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3
C6orf82	chromosome 6 open reading frame 82
DNAJC9	DnaJ (Hsp40) homolog, subfamily C, member 9
FLJ12750	hypothetical protein FLJ12750
CD164	CD164 antigen, sialomucin
NDUFB6	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 6, 17kDa
DUSP5	dual specificity phosphatase 5 (DUSP5)
BTG3	BTG family, member 3 (BTG3)
ESRRG	estrogen-related receptor gamma (ESRRG), transcript variant 3
CALD1	caldesmon 1 (CALD1), transcript variant 2
PRDX2	peroxiredoxin 2 (PRDX2), nuclear gene encoding mitochondrial protein, transcript variant 1
TMOD1	tropomodulin 1 (TMOD1)
FLJ23221	chromosome 1 open reading frame 54
BHLHB2	basic helix-loop-helix domain containing, class B, 2
ARPC5L	actin related protein 2/3 complex, subunit 5-like
TBCD	tubulin-specific chaperone d
LSM7	LSM7 homolog, U6 small nuclear RNA associated (S. cerevisiae)
GADD45B	growth arrest and DNA-damage-inducible, beta
C20orf111	chromosome 20 open reading frame 111
C21orf66	chromosome 21 open reading frame 66
ATP5O	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (oligomycin sensitivity conferring protein)
NIPBL	Nipped-B homolog (Drosophila)
BAT2	HLA-B associated transcript 2
CDYL	chromodomain protein, Y-like
C6orf114	glucose-fructose oxidoreductase domain containing 1
KIAA0265	KIAA0265 protein
CAMSAP1	calmodulin regulated spectrin-associated protein 1
PELI2	pellino homolog 2 (Drosophila)
KIAA0100	KIAA0100 gene product
SS18	synovial sarcoma translocation, chromosome 18
MBD1	methyl-CpG binding domain protein 1
TARBP1	TAR (HIV) RNA binding protein 1 (TARBP1)
PPARA	peroxisome proliferative activated receptor, alpha

LGMN	legumain (LGMN)
ERG	v-ets erythroblastosis virus E26 oncogene like (avian)
MAWBP	MAWD binding protein (MAWBP)
TNFRSF7	Homo sapiens T cell activation antigen (CD27) mRNA
LRRC2	leucine rich repeat containing 2 (LRRC2)
FLJ21106	hypothetical protein FLJ21106 (FLJ21106)
SEC22L1	SEC22 vesicle trafficking protein-like 1 (S cerevisiae) (SEC22L1)
WBP4	WW domain binding protein 4 (formin binding protein 21) (WBP4)
RXRB	retinoid X receptor, beta (RXRB)
CHD9	chromodomain helicase DNA binding protein 9 (CHD9), mRNA
DUSP21	ubiquitously transcribed tetratricopeptide repeat, X chromosome (UTX)
BLCAP	bladder cancer associated protein (BLCAP)
PLXND1	plexin D1 (PLXND1)
RNASE6	Human ribonuclease k6 precursor gene
DFNA5	deafness, autosomal dominant 5 (DFNA5)
ATF5	activating transcription factor 5 (ATF5)
PCP4	Purkinje cell protein 4 (PCP4)
APOC1	Human apolipoprotein C-I (VLDL) gene
LYZ	Human lysozyme mRNA
PYGL	Human liver glycogen phosphorylase mRNA
S100A9	Human cystic fibrosis antigen mRNA
CD72	CD72 antigen (CD72)
CTSH	Human mRNA for cathepsin H
APOE	Human apolipoprotein E mRNA
ITGB2	Human leukocyte adhesion protein (LFA-1/Mac-1/p150,95 family) beta subunit mRNA
COL1A2	H.sapiens mRNA for prepro-alpha2(I) collagen
CD163	CD163 antigen
DPYSL2	Homo sapiens N2A3 mRNA
WNT10B	Human Wnt10B mRNA
KIAA0146	Human NF-IL6-beta protein mRNA
S100A12	Human DNA for CAAF1 (calcium-binding protein in amniotic fluid 1)
A2M	Human alpha-2-macroglobulin mRNA
PLA2G7	Human LDL-phospholipase A2 mRNA
CDK8	cyclin-dependent kinase 8 (CDK8)
PF4V1	Human platelet factor 4 varation 1 (PF4var1) gene
HNMT	Human histamine N-methyltransferase (HNMT) gene, exon 6
HHEX	H.sapiens mRNA for proline rich homeobox (Prh) protein
AIF1	Human allograft-inflammatory factor-1 mRNA
MSX1	Homo sapiens (region 7) homeobox protein (HOX7) mRNA
GLDC	Human glycine decarboxylase mRNA
VCAM1	Human vascular cell adhesion molecule 1 mRNA
MRC1	Human macrophage mannose receptor (MRC1) gene, exon 30
LCN2	HNL=neutrophil lipocalin [human, ovarian cancer cell line OC6
GP5	H.sapiens GPV gene encoding platelet glycoprotein V precursor
PF4	Human platelet factor 4 (PF4) mRNA
ALDH1A1	Homo sapiens aldehyde dehydrogenase (ALDH1) gene, exon 13

LIPA	H.sapiens (HepG2) LAL mRNA for lysosomal acid lipase
CDKN1A	Human melanoma differentiation associated (mda-6) mRNA
EIF3S9	Homo sapiens eukaryotic translation initiation factor (eIF3) mRNA
GMPS	guanine monphosphate synthetase (GMPS)
LAMC1	Human laminin B2 chain gene, exon 28.
PTPRK	Human protein tyrosine phosphatase mRNA
HYAL3	Homo sapiens cosmid clone LUCA14 from 3p21.3
GCN5L2	Human GCN5 (hGCN5) gene
SSA2	Human SS-A/Ro ribonucleoprotein autoantigen 60 kd subunit mRNA
COX6B	Human DNA from overlapping chromosome 19 cosmids R31396, F25451,and R31076 containing COX6B and UPKA, genomic sequence
CDC34	Human ubiquitin conjugating enzyme mRNA
EWSR1	H.sapiens EWS mRNA
RBM10	RNA binding motif protein 10 (RBM10), transcript variant 1
CBS	cystathionine-beta-synthase (CBS)
ILF3	Human nuclear factor NF90 mRNA
USF2	Homo sapiens DNA from chromosome 19-cosmid R30879 containing USF2
C14orf116	Human checkpoint suppressor 1 mRNA
PFKM	Human phosphofructokinase (PFKM) mRNA
KIAA0020	Homo sapiens KIAA0020 mRNA
PHB	prohibitin [human, mRNA, 1043 nt].
PPP2R4	H.sapiens hPTPA mRNA
POPDC2	Homo sapiens COX17 mRNA
USP4	Human ubiquitin protease (Unph) proto-oncogene mRNA
SMPD1	Human acid sphingomyelinase (ASM) mRNA
MLLT3	Homo sapiens mRNA for LTG9/MLLT3 protein
DNASE1L1	H.sapiens mRNA for DNase X gene
FARSLA	phenylalanine-tRNA synthetase-like, alpha subunit (FARSLA)
FARSLB	phenylalanine-tRNA synthetase-like, beta subunit (FARSLB)
GART	Human mRNA for GARS-AIRS-GART
EVI2A	Human EV12 protein gene, exon 1
NEB	Human nebulin mRNA
CARS	Human cysteinyl-tRNA synthetase mRNA
SNRPN	Human SNRPN mRNA, 3' UTR, partial sequence
EXTL2	Homo sapiens EXTL2 (EXTL2) mRNA
SNX1	Human sorting nexin 1 (SNX1) mRNA
CAPN1	Human mRNA for calcium activated neutral protease large subunit (muCANP, calpain, EC 3.4.22.17
NAP1L4	Human nucleosome assembly protein 2 mRNA
QSCN6	Homo sapiens bone-derived growth factor (BPGF-1) mRNA
MAPKAPK3	Homo sapiens MAPKAP kinase (3pK) mRNA
HIST1H2BE	histone 1, H2be (HIST1H2BE), mRNA
APC	Human APC gene mRNA
HIST2H2BE	histone 2, H2be (HIST2H2BE)
MVP	Homo sapiens lrp mRNA
CTSO	H.sapiens mRNA for cathepsin-O
ZNF76	Human zinc-finger protein (ZNF76) gene,
EDG5	H.sapiens mRNA for DNA (cytosin-5)-methyltransferase

HIST1H2AE	histone 1, H2ae (HIST1H2AE)
PURA	H.sapiens Pur (pur-alpha) mRNA
ECH1	Homo sapiens peroxisomal enoyl-CoA hydratase-like protein (HPXEL) mRNA
SSRP1	Human high mobility group box (SSRP1) mRNA
MTA1	Human metastasis-associated mta1 mRNA
ICT1	Homo sapiens ICT1 (alias DS-1) mRNA
ARID5A	AT rich interactive domain 5A (MRF1-like) (ARID5A), transcript variant 2
ZNF136	Human zinc finger protein ZNF136
BCL7B	B-cell CLL/lymphoma 7B (BCL7B), transcript variant 1
ATF3	Human activating transcription factor 3 (ATF3) mRNA
MOX2	Human MOX2 gene for OX-2 membrane glycoprotein, exon 1 and joined CDS
TSC1	tuberous sclerosis 1 (TSC1)
MASP1	mannan-binding lectin serine protease 1 (C4/C2 activating component of Ra-reactive factor) (MASP1), transcript variant 1
CCND3	Cyclin D3
CCND1	Cyclin D1
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian) (MAFB)
CKS1B	CDC28 protein kinase regulatory subunit 1B (CKS1B)
RB1	retinoblastoma 1 (including osteosarcoma) (RB1)
CX3CR1	Chemokine (C-X3-C motif) receptor 1
WHSC1	Wolf-Hirschhorn syndrome candidate 1
ASPM	asp (abnormal spindle)-like, microcephaly associated (Drosophila) (ASPM)
CDC20	CDC20 cell division cycle 20 homolog (S cerevisiae) (CDC20)
MAD2L2	MAD2 mitotic arrest deficient-like 2 (yeast) (MAD2L2)
MUM1	melanoma associated antigen (mutated) 1 (MUM1)
IRF4	interferon regulatory factor 4 (IRF4)
MAD1L1	MAD1 mitotic arrest deficient-like 1 (yeast) (MAD1L1)
DKK1	dickkopf homolog 1 (Xenopus laevis) (DKK1)
IL6R	interleukin 6 receptor (IL6R), transcript variant 1
PDZK1	PDZ domain containing 1 (PDZK1)
MUC1	mucin 1, transmembrane (MUC1)
IRTA1	immunoglobulin superfamily receptor translocation associated 1 (IRTA1)
IRTA2	immunoglobulin superfamily receptor translocation associated 2 (IRTA2)
MCL1	myeloid cell leukemia sequence 1 (BCL2-related) (MCL1), transcript variant 1
COAS2	cyclophilin-LC (COAS2)
NULL	COAS3 (COAS3) mRNA
CDH1	cadherin 1, type 1, E-cadherin (epithelial) (CDH1)
LMO1	LIM domain only 1 (rhombotin 1) (LMO1)
LMO2	LIM domain only 2 (rhombotin-like 1) (LMO2)
TLX1	T-cell leukemia, homeobox 1 (TLX1)
TLX3	T-cell leukemia, homeobox 3 (TLX3), mRNA
LYL1	lymphoblastic leukemia derived sequence 1 (LYL1)
TAL2	T-cell acute lymphocytic leukemia 2 (TAL2)
STIL	TAL1 (SCL) interrupting locus (STIL)
MYC	v-myc myelocytomatosis viral oncogene homolog (avian) (MYC)

LOC51198	Homo sapiens cyclin-dependent kinase inhibitor 2A
CDKN2B	Homo sapiens cyclin-dependent kinase inhibitor 2B
MTAP	methylthioadenosine phosphorylase
HOXA1	Homeobox A1
HOXA2	Homeobox A2
HOXA3	Homeobox A3
HOXA4	Homeobox A4
HOXA5	Homeobox A5
HOXA6	Homeobox A6 /A7
HOXA9	Homeobox A9
HOXA10	Homeobox A10
HOXA11	Homeobox A11
HOXA13	Homeobox A13
HOXC6	Homeobox C6
NOTCH1	Notch homolog 1, translocation-associated (Drosophila)
GATA3	GATA binding protein 3
NFIL3	Nuclear factor, interleukin 3 regulated
SMAD1	SMAD, mothers against DPP homolog 1 (Drosophila)
CD34	CD34 antigen
LSP1	lymphocyte-specific protein 1
IL7R	Interleukin 7 receptor
MME	Membrane metallo-endopeptidase
CD6	CD6 antigen
ID3	Inhibitor of DNA binding 3, dominant negative helix-loop-helix protein
TCRG	T cell receptor gamma
TCRD	T cell receptor delta
TCRA	T-cell antigen receptor alpha (TCRA)
TCRB	T cell receptor beta
CDK2	Cyclin-dependant kinase 2
PCNA	Proliferating cell nuclear antigen
POLA	Polymerase (DNA directed), alpha
DHFR	Dihydrofolate reductase
TYMS	Thymidylate synthetase
CCNA1	Cyclin A1
NUP98	Nucleoporin 98kDa
LRIG1	Leucine-rich repeats and immunoglobulin-like domains 1
BCL2	B-cell CLL/lymphoma 2
BCL11B	B-cell CLL/lymphoma 11B
S100A10	S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))
RARA	Retinoic acid receptor, alpha MMP11, FOS
CDKN1B	Cyclin-dependent kinase inhibitor 1B
CDC25B	Cell division cycle 25B
JUN	v-jun sarcoma virus homolog
MAPK12	Mitogen-activated protein kinase 12
RUNX2	Runt-related transcription factor 2
NFKBIA	Inhibitor of nuclear factor of light chain gene enhancer in B cells
IGFBP3	Insulin-like growth factor binding protein 3
IGFBP5	Insulin-like growth factor binding protein 5

TNFRSF1A	Tumor necrosis factor receptor superfamily 1A
BCL6	B-cell CLL/Lymphoma 6
CD36	Thrombospondin receptor
CXCL12	SDF-1
HMOX1	Heme Oxygenase 1
DCN	Decorin
BTRC	Beta-transducin repeat containing
APP	Amyloid beta (A4) precursor protein, appican
SERPINA1	Serine proteinase inhibitor (-anti-trypsin)
TNFRSF6	Tumor necrosis factor receptor superfamily, 6
TP53	Tumor protein p53
TCF3	E2A Ig enhancer binding factor (E12/E47)
XRCC4	X-ray repair factor 4
XRCC5	Ku 80
TXN	Thioredoxin
IL6ST	IL6 signal transducer
TGFB1	Transforming growth factor B1
CASP3	Caspase 3
CD47	CD47 antigen
CDK4	Cyclin-dependent kinase 4
CDK5	Cyclin-dependent kinase 5
CDK9	Cyclin-dependent kinase 9
CDC37	Cell division cycle 37 homolog
CCNB1	Cyclin B1
PSEN1	Presenilin 1
PSEN2	Presenilin 2
UCHL1	Ubiquitin carboxyl-terminal esterase L1
PTEN	Phosphatase and tensin homolog
ABL1	V-abl Abelson murine leukemia viral oncogene homolog 1
ABL2	V-abl Abelson murine leukemia viral oncogene homolog 2 (arg, Abelson-related gene)
ACK1	activated Cdc42-associated kinase 1 (ACK1)
ACK2/TNK1	tyrosine kinase, non-receptor, 1 (TNK1)
ALK	Anaplastic lymphoma kinase (Ki-1)
AXL	AXL receptor tyrosine kinase (AXL)
BLK	B lymphoid tyrosine kinase (BLK)
BMX	BMX non-receptor tyrosine kinase
BRAF	V-raf murine sarcoma viral oncogene homolog B1
BTK	Bruton agammaglobulinemia tyrosine kinase
CSF1R	colony stimulating factor 1 receptor, formerly McDonough feline sarcoma viral (v-fms) oncogene homolog (CSF1R)
CSK	C-src tyrosine kinase
DDR1	Discoidin domain receptor family, member 1
DDR2	discoidin domain receptor family, member 2 (DDR2)
EGFR	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)
EPHA1	EphA1 (EPHA1)
EPHA2	EphA2 (EPHA2)
EPHA3	EphA3 (EPHA3)
EPHA4	EphA4 (EPHA4)
EPHA5	EphA5 (EPHA5)

EPHA6	EphA6 (EPHA6)
EPHA7	EphA7 (EPHA7)
EPHA8	EphA8 (EPHA8)
EPHB1	EphB1 (EPHB1)
EPHB2	EphB2 (EPHB2)
EPHB3	EphB3 (EPHB3)
EPHB4	EphB4 (EPHB4)
EPHB6	EphB6 (EPHB6)
ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian) (ERBB3)
ERBB4	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian) (ERBB4)
FAK (PTK2)	PTK2 protein tyrosine kinase 2 (PTK2)
FER	fer (fps/fes related) tyrosine kinase (phosphoprotein NCP94) (FER)
FES	Feline sarcoma oncogene
FGFR1	Fibroblast growth factor receptor 1
FGFR2	Fibroblast growth factor receptor 2
FGFR3	Fibroblast growth factor receptor 3
FGFR4	Fibroblast growth factor receptor 4
FGR	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog (FGR)
FLT1	Fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)
FLT3	fms-related tyrosine kinase 3 (FLT3)
FLT4	Fms-related tyrosine kinase 4
FRK	Fyn-related kinase
FYN	FYN oncogene related to SRC, FGR, YES (FYN), transcript variant 2
HCK	Hemopoietic cell kinase
HER2/ERBB2	V-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)
IGF1R	insulin-like growth factor 1 receptor (IGF1R)
INSR	insulin receptor precursor
INSRR	insulin receptor-related receptor (INSRR)
ITK	IL2-inducible T-cell kinase
JAK1	Janus kinase 1 (a protein tyrosine kinase) (JAK1)
JAK2	Janus kinase 2
JAK3	Janus kinase 3
KDR	kinase insert domain receptor (a type III receptor tyrosine kinase) (KDR)
KIT	V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
LCK	Lymphocyte-specific protein tyrosine kinase
LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog (LYN)
MATK	Megakaryocyte-associated tyrosine kinase
MERTK	c-mer proto-oncogene tyrosine kinase (MERTK)
MET	Met proto-oncogene (hepatocyte growth factor receptor)
MUSK	muscle, skeletal, receptor tyrosine kinase (MUSK)
NTRK1	neurotrophic tyrosine kinase, receptor, type 1 (NTRK1)
NTRK2	neurotrophic tyrosine kinase, receptor, type 2 (NTRK2)
NTRK3	neurotrophic tyrosine kinase, receptor, type 3 (NTRK3)

PDGFRA	Platelet-derived growth factor receptor, alpha polypeptide
PDGFRB	Platelet-derived growth factor receptor, beta polypeptide
PTK2B	PTK2B protein tyrosine kinase 2 beta
PTK6	PTK6 protein tyrosine kinase 6
PTK7	PTK7 protein tyrosine kinase 7 (PTK7), transcript variant PTK7-1
RET	ret proto-oncogene (multiple endocrine neoplasia and medullary thyroid carcinoma 1, Hirschsprung disease) (RET), transcript variant 2
RON/MST1R	Macrophage stimulating 1 receptor (c-met-related tyrosine kinase)
ROR1	receptor tyrosine kinase-like orphan receptor 1 (ROR1)
ROR2	receptor tyrosine kinase-like orphan receptor 2 (ROR2)
ROS	V-ros UR2 sarcoma virus oncogene homolog 1 (avian)
RYK	RYK receptor-like tyrosine kinase (RYK)
SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian) (SRC), transcript variant 1
SRMS	src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristylation sites (SRMS)
STYK1	Serine/threonine/tyrosine kinase 1
SYK	Homo sapiens spleen tyrosine kinase
TEC	Tec protein tyrosine kinase
TEK	TEK tyrosine kinase, endothelial (venous malformations, multiple cutaneous and mucosal)
TIE	Tyrosine kinase with immunoglobulin-like and EGF-like domains 1
TNNI3K	TNNI3 interacting kinase (TNNI3K)
TXK	TXK tyrosine kinase (TXK)
TYK2	Tyrosine kinase 2
TYRO3	TYRO3 protein tyrosine kinase (TYRO3)
ZAP70	Zeta-chain (TCR) associated protein kinase 70kDa
LTK	leukocyte tyrosine kinase

Appendix 16

Demographic, cytogenetic, FISH and molecular data for all patients processed using the CodeLink gene expression arrays.

Patient id	Age	Sex	Disease subtype	Karyotype	FISH	RT-PCR
2179	4	F	*ETV6-RUNX1*	45,X,-X,add(3)(p13)[15]		TEL/AML positive
5972	11	F	*ETV6-RUNX1*	45,X,-X,t(1;17)(q24;q25)[8]	TEL/AML fusion; 75%	
5991	2	M	*ETV6-RUNX1*	46,XY,del(12)(p11),inc[2]	TEL/AML fusion; 87%	
6845	4	F	*ETV6-RUNX1*	46,XX,add(9)(p1?),del(11)(q?21q?23),add(12)(p11)[20]	TEL/AML fusion; 83%	
7328	3	F	*ETV6-RUNX1*	49,XX,+6,+16,+19[4]	TEL/AML fusion; 58%	
8161	13	M	*ETV6-RUNX1*	46,XY,del(8)(q13q22),add(12)(p13),add(15)(q?26)[20]	TEL/AML fusion; 94%	
10073	2	F	*ETV6-RUNX1*	Fail		TEL/AML positive
10086	4	M	*ETV6-RUNX1*	49,XY,t(3;15)(q27;q1?5),+10,del(12)(p11p13),+16,add(18)(q2),+21[19]		TEL/AML positive
10952	3	M	*ETV6-RUNX1*	46,XY,del(6)(q1?4q2?3)[4]	TEL/AML fusion; 93%	TEL/AML positive
10545	41	M	*BCR-ABL1*	46,XY,t(9;22)(q34;q11)[20]		BCR/ABL1 positive
10655	13	F	*BCR-ABL1*	46,XX,der(9)add(9)(p1)t(9;22)(q34;q11),der(22)t(9;22)[6]		BCR/ABL1 positive
7152	2	F	*TCF3-PBX1*	46,XX,del(6)(q1q2),i(9)(q10),der(19)t(1;19)(q23;p13)[11]	TCF3 split; 75%	
7204	4	M	*TCF3-PBX1*	47,XY,t(1;19)(q23;p13),+8[20]	TCF3 split; 78%	TCF3-PBX1 positive
11050	4	F	*TCF3-PBX1*	46,XX,der(15)t(1;15)(q11;p11),der(19)t(1;19)(q23;p13)[15]/ 49,XX,+5,+8,+18,der(19)t(1;19)(q23;p13),+21[2]	TCF3 split; 81%	
10430	0	M	*MLL*-rearranged	46,XY,t(4;11)(q21;q23)[15]		MLL/AF4 positive
1333	4	F	HeH	56,XX,+X,+X,+4,+6,+9,?add(9)(p),+10,+14,+17,+18,+21[7]		
2058	5	F	HeH	52,XX,+X,+9,add(9)(p13),+14,+18,+21,+21[25]		

2064	5	F	HeH	61,XX,+der(2)t(1;2)(q21;q31),+4,+5,+6,+8,+10,+11,+12,+14, +16,+17,+18,+21x2,+22[19]		
2328	4	F	HeH	58,XX,+X,+X,dup(1)(q23q42),+4,+6,+10,+11,+14,+16,+17,+18,+21 ,+21[20]		
2586	4	M	HeH	55,XY,+X,+4,+6,+9,+10,add(10)(q26),+14,?16,+?18,+21,+21 [14]		
5957	40	F	HeH	~80,XX,+X,+1,+1,+21,idic(21)(p1?),inc[15]		
6118	3	M	HeH	54,XY,+X,+4,+6,+17,+18,+21,+21[10]		
6948	28	M	HeH	58,XY,+X,+Y,+2,+4,+6,+10,+11,+14,+18,+19,+21,+21[14]		
6971	53	F	HeH	62-72,XX,+X,+1,+2,+4,+4,+5,+6,+6,+8,+8,add(8)(q24),+10,+10, +11,+11,+12,+12,+13,+14,+14,+18,+19,+19,+21,+21,+22,+22, +mar		
7072	19	M	HeH	53,XY,+X,+4,+9,t(10;14)(p1?3;q1?1),+14,+21,+21,+mar[3]		
8756	4	M	HeH	54,XY,+X,+8,+8,+10,+10,+18,+21,+21[20]		
9497	9	F	HeH	55,XX,+X,?invdup(1)(q10q32),+4,+6,+10,+14,+17,+18,+21,+21 [16]/ 55,idem,del(16)(q13)[4]		
10189	4	M	HeH	57,XY,+X,+4,+5,+6,+8,+10,+14,+17,+18,+21,+21[6]		
10440	4	F	HeH	53-56,XX,+X,+X,+4,+10,+14,+17,+18,+21[cp5]		
10940	19	F	HeH	59-61,XX,+X,+4,+5,+6,add(6)(q15),+?8,+10,+14,+14,+16,+17, +18,+?18,+20,+21,+21,~50dmin[cp13]		
11234	8	F	HeH	Fail		
3034	11	M	T-ALL	47,XY,der(1)t(1;17)(p3;q?),del(9)(p22),del(9)(p?22),+10,del(10) (q22q24),der(21)t(9;21)(p;q22)[8]/ 47,idem,del(6)(q13q21)[5]	SIL/TAL1 sub deletion; 69%	
6867	10	M	T-ALL	46,XY,add(9)(p1)[13]	SIL/TAL1 sub deletion; 95%	SIL/TAL1 positive
8988	15	M	T-ALL	46,XY[20].ish t(7;9)(q34;q32)(5'TCRB+,3'TCRB- ,wcp9+;wcp9+,3'TCRB+)[7].nuc ish 9q32(TAL2x2)(TAL2-UsepTAL2-Dx1)[89/205]	TAL2 rearrangement; 43% TCRβ rearrangement; 50%	
10028	14	F	T-ALL	46,XX,add(5)(q2?2),del(9)(p1?2),add(10)(q2?6), t(11;14)(p13;q11.2),add(20)(q1?3)[18]/47,idem,+add(5)(q2?2), add(10)(q2?6),+10[3]	LMO2 rearrangement; 73% TCRα/δ rearrangement; 83%	

10283	6	M	T-ALL	46,XY,del(6)(q1q2)[4]/ 46,XY[26]		
11216	48	M	T-ALL	Fail		
11208		M	Non-classified			
11210		M	Non-classified			
11213		M	Non-classified			
11225		M	Non-classified			
11226		M	Non-classified			
11227		M	Non-classified			
11229		F	Non-classified			
11231		F	Non-classified			
12883		M	Non-classified			
12884		F	Non-classified			
12885		M	Non-classified			
12886		M	Non-classified			
3043	12	M	Other	47,XY,+21[15]		
3047	7	M	Other	46,XY,t(9;14)(q22;q11)[29]		
6074	0	M	Other	46,XY,t(7;15)(q22;q21)[28]		
7046	7	F	Other	46,XX,del(6)(q1q2),?t(7;10)(q34;q24)[20]		
7344	11	F	Other	46,XX[20]		
8431	5	F	Other	46,XX,t(3;9)(p?21;p?22)[2]/46,idem,-16,+r[18]		
10186	9	F	Other	47,XX,+?8[3]		
10817	3	F	Other	47,XX,+21[13]		
11739	14	F	Other	47,XX,t(5;14)(q1?5;q32.?3),add(9)(p?13),der(20)t(5;20)(q1?5:q13),+mar[19]	IGH rearrangement; 85%	

The normal clone has been omitted from cases with an abnormal karyotype

Appendix 17.

The top 100 class discriminating genes for each of the eight ALL subgroups, ranked according to the degree of association with that subgroup (T-value). Genes that were previously identified as discriminatory for a subgroup are referred to as 'true'.

ETV6-RUNX1

Rank	Gene Name	Accession number	Fold Change	Below/ Above mean	T-value	Previously published
1	HAP1	NM_003949.2	32.44	Above	10.35	TRUE
2	BIRC7	NM_139317.1	7.22	Above	9.04	TRUE
3	NR3C2	NM_000901.1	6.97	Above	8.21	FALSE
4	RGMA	NM_020211.1	13.45	Above	8.12	TRUE
5	CBFA2T3	NM_005187.4	3.05	Above	8.02	TRUE
6	IMP-1	NM_006546.3	5.75	Above	7.65	FALSE
7	TCFL5	NM_006602.2	7.19	Above	7.03	TRUE
8	FHIT	NM_002012.1	3.54	Above	6.65	FALSE
9	DSC3	NM_024423.1	11.06	Above	6.40	TRUE
10	TERF2	NM_005652.2	5.52	Above	6.35	TRUE
11	SEMA6A	NM_020796.2	6.14	Above	5.90	FALSE
12	NRN1	NM_016588.2	16.82	Above	5.75	TRUE
13	CD44	M59040.1	0.28	Below	5.74	FALSE
14	TAGLN3	NM_013259.2	2.43	Above	5.72	FALSE
15	FBN2	NM_001999.2	7.22	Above	5.47	TRUE
16	FLJ23749	NM_152271.2	3.16	Above	5.35	TRUE
17	MGC23911	NM_144676.1	15.08	Above	5.34	TRUE
18	ARHGEF4	NM_032995.1	5.37	Above	5.26	TRUE
19	CSF1R	NM_005211.2	0.14	Below	5.16	FALSE
20	CD164	NM_006016.3	0.35	Below	5.02	FALSE
21	CTGF	NM_001901.1	10.56	Above	4.74	FALSE
22	PON2	NM_000305.1	0.19	Below	4.49	FALSE
23	HPS4	NM_022081.4	3.16	Above	4.47	FALSE
24	MYC	NM_002467.2	0.21	Below	4.42	FALSE
25	PRKCB1	NM_002738.5	2.54	Above	4.36	TRUE
26	IMPA2	NM_014214.1	0.38	Below	4.35	FALSE
27	INSR	X02160.1	3.60	Above	4.34	FALSE
28	MAPK13	NM_002754.3	2.33	Above	4.31	TRUE
29	SMAD1	NM_005900.2	2.64	Above	4.25	FALSE
30		BF513468.1	4.98	Above	4.22	FALSE
31	LSP1	NM_002339.1	0.33	Below	4.22	FALSE
32	LAIR1	NM_021706.1	2.04	Above	4.19	FALSE
33	ITPR1	NM_002222.1	2.63	Above	4.19	FALSE
34	AKAP12	NM_144497.1	10.14	Above	4.16	FALSE
35	MDK	NM_002391.2	4.83	Above	3.94	TRUE
36	MME	NM_007289.1	6.40	Above	3.94	FALSE
37	HSPCA	NM_005348.2	0.42	Below	3.94	FALSE
38	NKCC1	U30246.1	1.98	Above	3.79	TRUE
39	MYO5C	NM_018728.1	3.97	Above	3.77	FALSE
40	ERG	NM_182918.2	2.27	Above	3.76	FALSE
41	SLC35E3	NM_018656.2	3.01	Above	3.74	FALSE
42	FARP1	NM_005766.2	4.25	Above	3.73	FALSE
43	RY1	NM_006857.1	2.27	Above	3.73	TRUE

44	PTPRK	NM_002844.2	3.29	Above	3.73	TRUE
45	FYB	U93049.1	2.18	Above	3.72	TRUE
46	TUSC3	NM_006765.2	7.88	Above	3.64	TRUE
47	KHDRBS3	NM_006558.1	4.54	Above	3.63	FALSE
48	AUTS2	NM_015570.1	4.04	Above	3.62	FALSE
49	C11orf24	NM_022338.2	0.42	Below	3.60	FALSE
50	BLCAP	NM_006698.2	1.68	Above	3.60	FALSE
51	ISG20	NM_002201.4	3.03	Above	3.59	FALSE
52	IDI1	NM_004508.2	2.15	Above	3.56	TRUE
53	HSPCA	NM_005348.2	0.47	Below	3.54	FALSE
54	DRAM	BC013773.1	2.36	Above	3.51	TRUE
55	NOTCH1	NM_017617.2	2.37	Above	3.50	FALSE
56	RAG2	NM_000536.1	4.88	Above	3.42	FALSE
57	PLP2	NM_002668.1	0.51	Below	3.39	FALSE
58	MIB1	NM_020774.1	2.48	Above	3.33	FALSE
59	HMG20A	NM_018200.2	2.00	Above	3.32	TRUE
60	CDW52	NM_001803.1	2.21	Above	3.30	FALSE
61	TNK1	NM_003985.1	1.96	Above	3.29	FALSE
62	VPREB3	NM_013378.1	6.55	Above	3.27	FALSE
63	ABCG2	NM_004827.1	1.81	Above	3.22	TRUE
64	TM4SF14	NM_030927.1	1.77	Above	3.20	FALSE
65	C1orf38	NM_004848.1	1.85	Above	3.17	FALSE
66	CSRP2	NM_001321.1	3.61	Above	3.10	FALSE
67	S100A10	NM_002966.1	0.51	Below	3.09	FALSE
68	APP	NM_000484.2	2.49	Above	3.08	FALSE
69	TIF1	NM_003852.3	2.37	Above	3.04	TRUE
70	CLECSF6	NM_194447.1	0.49	Below	3.00	FALSE
71	FLJ21415	NM_024738.1	1.75	Above	2.93	TRUE
72		N27245.1	0.49	Below	2.92	FALSE
73	KCNJ16	NM_018658.1	3.95	Above	2.92	FALSE
74	SNRPN	NM_022805.2	3.16	Above	2.87	FALSE
75	ALDOAP2	M21191.1	0.49	Below	2.86	FALSE
76	PRG1	NM_002727.2	0.41	Below	2.84	FALSE
77	GIMAP4	NM_018326.2	2.77	Above	2.82	FALSE
78	HIST2H2BE	NM_003528.2	1.89	Above	2.80	FALSE
79	EBF1	AK096708.1	3.78	Above	2.77	FALSE
80	LTA4H	NM_000895.1	0.50	Below	2.76	FALSE
81	FGFR4	NM_002011.3	0.62	Below	2.74	FALSE
82		BI869593.1	3.85	Above	2.73	FALSE
83	C21orf66	NM_013329.2	0.72	Below	2.70	FALSE
84	MEF2C	NM_002397.2	2.41	Above	2.69	FALSE
85	HOXA5	NM_019102.2	0.35	Below	2.69	FALSE
86	DF	NM_001928.2	0.29	Below	2.66	FALSE
87	SESN1	NM_014454.1	1.62	Above	2.65	FALSE
88	TNS	NM_022648.2	2.48	Above	2.64	TRUE
89	MSX1	NM_002448.1	1.90	Above	2.64	FALSE
90	HLA-DPA1	NM_033554.2	2.44	Above	2.61	FALSE
91	TLE1	NM_005077.3	2.28	Above	2.61	FALSE
92	TCL1A	NM_021966.1	2.98	Above	2.59	FALSE
93	CDYL	NM_004824.2	1.80	Above	2.58	FALSE
94	ACK1	NM_005781.3	0.68	Below	2.58	FALSE
95	MRPL3	NM_007208.2	0.59	Below	2.57	FALSE
96	SERPINI1	NM_005025.2	2.32	Above	2.56	FALSE

97	FKBP1A	NM_000801.2	1.60	Above	2.54	FALSE
98	PCCB	NM_000532.2	0.64	Below	2.50	FALSE
99	NOVA1	NM_006489.2	1.92	Above	2.50	TRUE
100	NISCH	NM_007184.1	1.28	Above	2.50	FALSE

BCR-ABL

Rank	Gene Name	Accession number	Fold Change	Below/ Above mean	T-Value	Previously published
1	DNAPTP6	NM_015535.1	22.65	Above	5.69	TRUE
2	NAV1	NM_020443.2	5.69	Above	3.87	FALSE
3	GPR110	NM_153840.1	8.80	Above	3.69	TRUE
4	TNFRSF7	NM_001242.3	18.86	Above	3.61	TRUE
5	PON2	NM_000305.1	12.56	Above	3.49	TRUE
6	VAMP5	NM_006634.2	2.93	Above	3.41	FALSE
7	MAP1A	NM_002373.4	0.19	Below	3.40	FALSE
8	FAM38A	NM_014745.1	4.30	Above	3.39	TRUE
9	ECM1	NM_022664.1	13.72	Above	3.16	TRUE
10	C15orf39	AK128205.1	0.53	Below	3.09	FALSE
11	ERBB2	NM_004448.2	6.74	Above	3.05	FALSE
12	BTG3	NM_006806.3	0.12	Below	3.04	FALSE
13	KIAA0652	NM_014741.2	0.46	Below	3.00	FALSE
14	SEMA6A	NM_020796.2	6.96	Above	2.99	TRUE
15	DUSP6	NM_001946.2	3.56	Above	2.99	FALSE
16	PRX	NM_020956.1	4.36	Above	2.88	TRUE
17	CES1	NM_001266.3	0.33	Below	2.81	FALSE
18	PTPNS1	NM_080792.1	0.35	Below	2.79	FALSE
19	TNFRSF6	NM_000043.3	0.47	Below	2.78	FALSE
20	SLC2A5	NM_003039.1	2.68	Above	2.77	TRUE
21	AKR7A2	NM_003689.2	2.71	Above	2.72	FALSE
22	C21orf66	NM_013329.2	0.54	Below	2.72	FALSE
23	CTDSPL	NM_005808.2	3.87	Above	2.68	FALSE
24	TMSNB	NM_021992.1	0.08	Below	2.64	FALSE
25	MMP28	NM_024302.2	4.53	Above	2.62	TRUE
26	ALDH1A1	NM_000689.3	9.48	Above	2.53	FALSE
27	TOP1	NM_003286.2	0.56	Below	2.50	FALSE
28		W85860.1	4.87	Above	2.50	FALSE
29	CPM	J04970.2	3.41	Above	2.48	FALSE
30	GADD45B	NM_015675.1	3.67	Above	2.48	FALSE
31	FLJ21415	NM_024738.1	0.40	Below	2.47	FALSE
32	ITGA6	NM_000210.1	3.38	Above	2.47	TRUE
33	stch	U04735.1	0.40	Below	2.46	FALSE
34	CD19	NM_001770.3	2.86	Above	2.45	FALSE
35	FGFR1	NM_023111.1	0.46	Below	2.44	FALSE
36	HIST2H2BE	NM_003528.2	2.91	Above	2.44	FALSE
37	PLXNB2	AB002313.1	3.46	Above	2.44	FALSE
38	MUC1	NM_182741.1	0.53	Below	2.44	FALSE
39	SIX3	NM_005413.1	0.48	Below	2.41	FALSE
40	BAALC	NM_024812.1	3.54	Above	2.40	FALSE
41	TOP1	NM_003286.2	0.58	Below	2.38	FALSE
42	HIST2H2BE	NM_003528.2	2.89	Above	2.36	FALSE
43	HLA-DOB	NM_002120.2	0.51	Below	2.36	FALSE

44	CTSL	NM_001912.2	0.33	Below	2.34	FALSE
45	CDK9	NM_001261.2	2.13	Above	2.33	FALSE
46	HIST1H2BK	NM_080593.1	3.43	Above	2.32	FALSE
47	CD79B	NM_021602.1	3.67	Above	2.32	FALSE
48	SLC35E3	NM_018656.2	3.88	Above	2.32	FALSE
49		INCYTE UNIQUE	0.44	Below	2.32	FALSE
50	HLA-DMB	NM_002118.3	0.57	Below	2.31	FALSE
51	HTRA3	NM_053044.2	5.43	Above	2.29	FALSE
52	PSTPIP2	NM_024430.2	7.57	Above	2.28	TRUE
53	IL18BP	NM_005699.2	2.13	Above	2.25	TRUE
54	ARHGAP8	NM_181334.1	0.48	Below	2.24	FALSE
55	ABL1	NM_005157.2	3.37	Above	2.23	FALSE
56	CBFA2T3	NM_005187.4	0.48	Below	2.23	FALSE
57	CHN2	NM_004067.1	2.36	Above	2.21	TRUE
58	BIN2	NM_016293.2	0.45	Below	2.21	FALSE
59	POU4F1	NM_006237.2	0.49	Below	2.20	FALSE
60	NFKBIA	NM_020529.1	2.27	Above	2.20	FALSE
61	NTRK3	NM_002530.1	0.56	Below	2.17	FALSE
62	IGF1R	NM_000875.2	0.45	Below	2.16	FALSE
63	FUT7	NM_004479.2	0.47	Below	2.16	FALSE
64	PTGS1	M59979.1	4.12	Above	2.15	FALSE
65	RGMA	NM_020211.1	0.20	Below	2.14	FALSE
66	MHC2TA	NM_000246.1	0.50	Below	2.14	FALSE
67	EPHB3	NM_004443.3	0.48	Below	2.14	FALSE
68	IQGAP2	NM_006633.1	0.62	Below	2.13	FALSE
69	ANK1	NM_020477.1	0.37	Below	2.13	FALSE
70	GUSB	NM_000181.1	4.28	Above	2.12	FALSE
71	CD34	NM_001773.1	4.19	Above	2.12	FALSE
72	FLJ20605	NM_017898.3	0.29	Below	2.11	FALSE
73	PSD3	NM_015310.2	5.07	Above	2.11	FALSE
74	FLJ23221	NM_024579.1	3.20	Above	2.10	FALSE
75	PDCD10	NM_007217.3	0.51	Below	2.10	FALSE
76	LAIR1	NM_021706.1	2.05	Above	2.09	TRUE
77	TCTEL1	NM_006519.1	0.42	Below	2.09	FALSE
78	DKK1	NM_012242.1	5.11	Above	2.07	FALSE
79	COL1A1	Z74615.1	0.57	Below	2.07	FALSE
80	KIT	NM_000222.1	0.43	Below	2.07	FALSE
81	DISC1	NM_018662.1	0.43	Below	2.06	FALSE
82	SCN3A	AJ251507.1	0.44	Below	2.06	FALSE
83	HIF1A	NM_001530.2	2.69	Above	2.06	FALSE
84	CSK	NM_004383.1	0.24	Below	2.06	FALSE
85	MYO5C	NM_018728.1	4.47	Above	2.05	TRUE
86	WHDC1L1	BC048987.1	5.00	Above	2.05	FALSE
87	DHPS	NM_001930.2	0.43	Below	2.03	FALSE
88	HHEX	NM_002729.2	3.42	Above	2.03	FALSE
89	FLJ12750	NM_024667.1	0.40	Below	2.03	FALSE
90	UBQLN2	NM_013444.2	0.63	Below	2.02	FALSE
91	JUN	NM_002228.3	6.75	Above	2.01	FALSE
92	JAG1	NM_000214.1	0.22	Below	2.01	FALSE
93	SLC16A2	NM_006517.1	0.56	Below	2.01	FALSE
94	HOXA3	NM_030661.3	0.56	Below	2.00	FALSE
95	FARP1	NM_005766.2	4.69	Above	2.00	FALSE
96	INSRR	NM_014215.1	0.36	Below	1.99	FALSE

97	HLA-DRB5	NM_002125.3	4.03	Above	1.99	FALSE
98	ERBB4	NM_005235.1	0.47	Below	1.98	FALSE
99	PLSCR3	NM_020360.2	0.48	Below	1.98	FALSE
100	HOXC6	NM_004503.2	0.58	Below	1.97	FALSE

TCF3-PBX1

Rank	Gene Name	Accession number	Fold Change	Below/ Above mean	T-value	Previously published
1	NID2	NM_007361.2	30.03	Above	6.43	TRUE
2		AK055619.1	0.17	Below	6.17	FALSE
3	ROR1	NM_005012.1	7.56	Above	5.36	FALSE
4	LRMP	NM_006152.2	7.09	Above	5.30	TRUE
5	SLAMF1	NM_003037.1	12.72	Above	5.11	TRUE
6	SORBS1	NM_015385.1	4.61	Above	5.07	TRUE
7	PBX1	NM_002585.1	19.02	Above	5.02	TRUE
8	KIAA0922	NM_015196.2	3.66	Above	4.79	TRUE
9	TCF4	NM_003199.1	12.42	Above	4.64	FALSE
10	LAIR1	NM_021706.1	0.31	Below	4.57	FALSE
11	FAT	NM_005245.1	6.97	Above	4.38	TRUE
12	FAM38A	NM_014745.1	0.23	Below	4.32	FALSE
13	BLK	NM_001715.2	12.54	Above	4.18	FALSE
14	SYNPO	AB028952.1	5.37	Above	4.09	TRUE
15	FAM3C	NM_014888.1	4.41	Above	4.05	TRUE
16	HPCAL1	NM_002149.2	0.49	Below	3.93	TRUE
17	LMO2	NM_005574.2	0.30	Below	3.81	FALSE
18	CTNNBL1	NM_030877.3	2.02	Above	3.56	FALSE
19	IRF4	NM_002460.1	12.21	Above	3.52	FALSE
20		INCYTE UNIQUE	2.64	Above	3.50	FALSE
21	ACK1	NM_005781.3	2.21	Above	3.50	FALSE
22	IGHM	X58529.1	4.44	Above	3.46	FALSE
23	MAPKAPK3	NM_004635.3	0.31	Below	3.41	FALSE
24	EBF1	AK096708.1	11.76	Above	3.39	FALSE
25	LIPA	NM_000235.2	0.37	Below	3.38	FALSE
26	C14orf139	NM_024633.2	4.30	Above	3.32	FALSE
27	ARHGEF6	NM_004840.2	0.31	Below	3.31	FALSE
28	AKAP12	NM_144497.1	17.73	Above	3.26	FALSE
29	SPG21	NM_016630.3	0.55	Below	3.25	FALSE
30	PNAS-4	NM_016076.2	2.45	Above	3.21	TRUE
31	FLJ14153	NM_022736.1	0.30	Below	3.14	FALSE
32	TM4SF2	NM_004615.2	0.15	Below	3.05	FALSE
33	TCL1A	NM_021966.1	6.91	Above	3.01	FALSE
34	C18orf1	NM_181481.2	3.06	Above	3.01	FALSE
35	PRKCB1	NM_002738.5	0.36	Below	2.99	FALSE
36	LYL1	NM_005583.3	0.35	Below	2.94	FALSE
37	CD34	NM_001773.1	0.21	Below	2.93	FALSE
38	AUTS2	NM_015570.1	5.89	Above	2.93	FALSE
39	BLCAP	NM_006698.2	0.52	Below	2.93	FALSE
40	ASPM	NM_018136.2	2.95	Above	2.92	FALSE
41	ROR1	NM_005012.1	23.00	Above	2.91	FALSE
42	ARID5A	NM_006673.2	2.17	Above	2.90	FALSE
43	FYB	U93049.1	0.39	Below	2.88	FALSE

44	RAG2	NM_000536.1	7.97	Above	2.87	FALSE
45	RYK	NM_002958.2	0.30	Below	2.82	FALSE
46	EPHA3	NM_005233.3	8.22	Above	2.79	FALSE
47	FGFR4	NM_002011.3	2.13	Above	2.78	FALSE
48	SESN1	NM_014454.1	0.46	Below	2.78	TRUE
49	LRIG1	NM_015541.2	6.20	Above	2.75	FALSE
50	ILK	NM_004517.1	0.49	Below	2.73	FALSE
51	CKLFSF3	NM_144601.2	0.58	Below	2.72	TRUE
52	HTRA3	NM_053044.2	4.97	Above	2.72	TRUE
53	NR2C1	NM_003297.1	2.01	Above	2.72	FALSE
54	ARHGAP8	NM_181334.1	2.05	Above	2.72	TRUE
55	CD34	NM_001773.1	0.20	Below	2.69	FALSE
56	FLJ31978	NM_144669.1	3.70	Above	2.68	TRUE
57	KIAA0040	NM_014656.1	2.09	Above	2.68	TRUE
58	RGMA	NM_020211.1	4.99	Above	2.67	FALSE
59	CDS2	NM_003818.2	1.96	Above	2.66	FALSE
60	PDE4B	NM_002600.2	0.33	Below	2.65	FALSE
61	ZHX2	NM_014943.2	4.80	Above	2.65	FALSE
62	MGC10485	NM_052875.2	2.48	Above	2.61	TRUE
63	VPREB3	NM_013378.1	10.39	Above	2.60	FALSE
64	ITPR1	NM_002222.1	0.38	Below	2.60	FALSE
65	GOLGA3	NM_005895.2	2.21	Above	2.59	TRUE
66	SEMA4C	NM_017789.3	3.45	Above	2.59	FALSE
67	FLJ30525	NM_144584.1	0.46	Below	2.58	FALSE
68	TMSNB	NM_021992.1	7.40	Above	2.58	FALSE
69	GLDC	NM_000170.1	5.43	Above	2.57	FALSE
70	MEF2C	NM_002397.2	3.65	Above	2.57	FALSE
71	SEPW1	NM_003009.2	0.46	Below	2.56	FALSE
72	CKLFSF2	NM_144673.2	0.21	Below	2.56	FALSE
73	PURA	NM_005859.2	0.61	Below	2.52	FALSE
74	CD19	NM_001770.3	2.38	Above	2.52	FALSE
75	SLC2A5	NM_003039.1	0.49	Below	2.52	FALSE
76	C11orf9	NM_013279.1	1.85	Above	2.51	FALSE
77	FURIN	NM_002569.2	0.39	Below	2.49	FALSE
78	BLNK	NM_013314.2	2.19	Above	2.46	FALSE
79	GFOD1	NM_018988.1	0.49	Below	2.46	FALSE
80	TRAK1	AB028965.1	1.77	Above	2.44	FALSE
81	EPHB4	NM_004444.4	0.45	Below	2.42	FALSE
82	CAMK2D	NM_001221.2	2.38	Above	2.42	FALSE
83	RNASET2	NM_003730.3	0.37	Below	2.40	TRUE
84	PDHA1	NM_000284.1	0.60	Below	2.37	FALSE
85	CD24	NM_013230.1	3.88	Above	2.34	FALSE
86	KHDRBS3	NM_006558.1	4.68	Above	2.33	FALSE
87	CD99	NM_002414.3	0.29	Below	2.32	FALSE
88	IFITM3	NM_021034.1	0.42	Below	2.31	FALSE
89	AIF1	NM_032955.1	0.42	Below	2.31	FALSE
90	NPY	NM_000905.2	6.28	Above	2.31	FALSE
91	JAG1	NM_000214.1	3.96	Above	2.30	FALSE
92	MERTK	NM_006343.1	7.13	Above	2.27	FALSE
93	SLC9A3R1	NM_004252.1	0.53	Below	2.25	FALSE
94	UCKL1	NM_017859.1	1.67	Above	2.24	FALSE
95	RPS6KA5	NM_004755.2	0.45	Below	2.19	FALSE
96	KIAA1576	NM_020927.1	7.38	Above	2.17	FALSE

97	ALOX5AP	NM_001629.2	0.46	Below	2.15	FALSE
98	ROBO1	NM_133631.1	2.04	Above	2.14	FALSE
99	ALOX5	NM_000698.1	0.46	Below	2.13	FALSE
100	FLJ14167	AK024229.1	6.55	Above	2.13	TRUE

***MLL* rearranged**

Rank	Gene Name	Accession number	Fold Change	Below/ Above mean	T-Value	Previously published
1	MEIS1	NM_002398.2	96.97	Above	4.64	TRUE
2	SESN1	NM_014454.1	5.78	Above	3.80	FALSE
3	KIAA0056	NM_015261.1	0.08	Below	3.63	FALSE
4	HOXA5	NM_019102.2	35.48	Above	3.56	FALSE
5	C20orf103	NM_012261.2	26.25	Above	3.45	TRUE
6	RB1	NM_000321.1	0.07	Below	3.39	FALSE
7	LMO2	NM_005574.2	5.89	Above	3.28	FALSE
8	PSMB2	NM_002794.3	2.98	Above	3.17	FALSE
9	NAV1	NM_020443.2	0.14	Below	3.13	TRUE
10	PSFL	NM_031301.2	3.50	Above	3.11	FALSE
11	LCK	NM_005356.2	0.06	Below	2.94	FALSE
12	S100A10	NM_002966.1	5.12	Above	2.87	FALSE
13	BAALC	NM_024812.1	7.54	Above	2.84	FALSE
14	PIK4CA	NM_058004.1	0.13	Below	2.81	FALSE
15	HOXB5	NM_002147.2	37.55	Above	2.78	FALSE
16	PSMB2	NM_002794.3	2.49	Above	2.73	FALSE
17	MME	NM_007289.1	0.03	Below	2.68	TRUE
18	SCHIP1	NM_014575.1	0.04	Below	2.63	TRUE
19	CD44	M59040.1	5.27	Above	2.62	TRUE
20	MDK	NM_002391.2	0.06	Below	2.55	FALSE
21	DDEF1	NM_018482.2	0.33	Below	2.54	TRUE
22	HIF1A	NM_001530.2	0.19	Below	2.54	TRUE
23	SLAMF1	NM_003037.1	10.06	Above	2.53	FALSE
24	DF	NM_001928.2	21.05	Above	2.51	FALSE
25	ALCAM	NM_001627.1	0.10	Below	2.49	FALSE
26	CCNA1	NM_003914.2	7.23	Above	2.43	FALSE
27	KIAA1576	NM_020927.1	41.53	Above	2.39	TRUE
28	KIAA0754	AB018297.2	0.26	Below	2.37	FALSE
29	TCF4	NM_003199.1	0.10	Below	2.37	TRUE
30	H2AFX	NM_002105.2	0.32	Below	2.35	FALSE
31	MSN	NM_002444.2	0.20	Below	2.34	FALSE
32	C18orf1	NM_181481.2	0.23	Below	2.31	FALSE
33	KHDRBS3	NM_006558.1	0.07	Below	2.31	FALSE
34	RYK	NM_002958.2	0.18	Below	2.30	FALSE
35	SMAD1	NM_005900.2	0.24	Below	2.27	FALSE
36	FKBP5	NM_004117.2	4.46	Above	2.18	FALSE
37	FLJ30525	NM_144584.1	3.08	Above	2.17	TRUE
38	ARPC2	NM_005731.2	1.94	Above	2.12	TRUE
39	IL5RA	NM_000564.2	13.16	Above	2.11	FALSE
40	ITPK1	NM_014216.3	1.91	Above	2.09	FALSE
41	WHDC1L1	BC048987.1	0.11	Below	2.09	FALSE
42	CDC16	NM_003903.2	0.16	Below	2.09	FALSE
43		R59027.1	1.86	Above	2.09	FALSE
44	MYO5C	NM_018728.1	0.13	Below	2.09	TRUE

45	VBP1	NM_003372.3	0.19	Below	2.09	FALSE
46	TNRC11	NM_005120.1	0.32	Below	2.08	FALSE
47	JAK1	BX648044.1	2.55	Above	2.08	FALSE
48	SCML2	NM_006089.1	0.23	Below	2.07	FALSE
49	HNRPL	NM_001533.2	0.44	Below	2.06	FALSE
50	PTGDS	NM_000954.5	9.66	Above	2.05	FALSE
51	SMARCA3	NM_003071.2	0.25	Below	2.05	FALSE
52	LGALS2	NM_006498.2	11.83	Above	2.03	FALSE
53	GFOD1	NM_018988.1	0.36	Below	2.01	FALSE
54	SEC22L1	NM_004892.3	0.28	Below	2.01	FALSE
55	PTK7	NM_002821.3	0.25	Below	2.00	FALSE
56	PYGL	NM_002863.2	4.15	Above	2.00	FALSE
57	MGAT4A	NM_012214.1	2.90	Above	1.99	FALSE
58	FBP1	NM_000507.2	3.09	Above	1.99	FALSE
59	CUL1	NM_003592.2	0.16	Below	1.98	FALSE
60	HOXA9	NM_002142.3	2.57	Above	1.97	TRUE
61	ITPK1	NM_014216.3	1.96	Above	1.96	FALSE
62	FLI1	NM_002017.2	0.28	Below	1.94	FALSE
63	MAPK13	NM_002754.3	2.83	Above	1.91	FALSE
64	FHIT	NM_002012.1	0.34	Below	1.90	TRUE
65	TM4SF2	NM_004615.2	0.13	Below	1.88	FALSE
66	ZHX2	NM_014943.2	0.15	Below	1.87	FALSE
67	EIF4G2	NM_001418.1	0.17	Below	1.87	FALSE
68	C5orf13	NM_004772.1	0.15	Below	1.84	FALSE
69	H2AFZ	NM_002106.3	0.37	Below	1.84	FALSE
70	EPHA7	NM_004440.2	1.82	Above	1.83	FALSE
71	DLG3	NM_021120.1	0.36	Below	1.82	FALSE
72	CHN2	NM_004067.1	0.38	Below	1.81	FALSE
73	GAGED2	NM_020411.1	2.31	Above	1.81	FALSE
74	SS18	NM_005637.2	0.42	Below	1.81	FALSE
75	C1orf38	NM_004848.1	2.56	Above	1.81	FALSE
76	APP	NM_000484.2	0.24	Below	1.80	FALSE
77	IMPA2	NM_014214.1	3.01	Above	1.80	FALSE
78	CNN3	NM_001839.2	0.16	Below	1.80	FALSE
79	FLJ23749	NM_152271.2	3.04	Above	1.79	FALSE
80	TARBP1	NM_005646.2	2.28	Above	1.79	FALSE
81	MRPL33	NM_004891.2	2.74	Above	1.78	FALSE
82	LCP2	NM_005565.3	0.26	Below	1.78	FALSE
83	C11orf24	NM_022338.2	3.18	Above	1.78	TRUE
84	CSK	NM_004383.1	0.21	Below	1.77	FALSE
85	MUC1	NM_182741.1	1.90	Above	1.77	FALSE
86	UREB1	NM_031407.2	0.51	Below	1.75	FALSE
87	STMN1	NM_203401.1	0.38	Below	1.74	FALSE
88	MINPP1	NM_004897.2	0.24	Below	1.74	FALSE
89	XRCC5	NM_021141.2	0.53	Below	1.74	FALSE
90	DRAM	BC013773.1	3.17	Above	1.74	FALSE
91	HMGB3	NM_005342.1	0.49	Below	1.73	FALSE
92	PCCB	NM_000532.2	2.27	Above	1.72	FALSE
93	SLC35E3	NM_018656.2	0.25	Below	1.71	FALSE
94	CTGF	NM_001901.1	10.97	Above	1.71	FALSE
95	QARS	NM_005051.1	2.03	Above	1.70	FALSE
96	DDR1	NM_013994.1	0.30	Below	1.70	FALSE
97	NKCC1	U30246.1	2.30	Above	1.69	FALSE

98	HTRA3	NM_053044.2	5.60	Above	1.69	FALSE
99	BCL6	NM_001706.2	0.30	Below	1.68	FALSE
100	HNRPL	NM_001533.2	0.51	Below	1.67	FALSE

High Hyperdiploidy

Rank	Gene Name	Accession number	Fold Change	Below/ Above mean	T-Value	Previously published
1	SH3KBP1	NM_031892.1	1.86	Above	5.45	TRUE
2	FLT3	NM_004119.1	3.25	Above	4.92	FALSE
3	ALDH3B1	NM_000694.1	1.99	Above	4.49	FALSE
4	SH3BP5	NM_004844.1	2.05	Above	4.42	FALSE
5	CDKN1A	NM_000389.2	2.47	Above	4.30	FALSE
6	SEPW1	NM_003009.2	0.54	Below	4.22	FALSE
7	HMGB3	NM_005342.1	1.63	Above	4.21	FALSE
8	OCIAD2	NM_152398.1	0.42	Below	4.20	FALSE
9	ROR1	NM_005012.1	0.45	Below	3.90	FALSE
10	SERPINA1	NM_000295.3	0.27	Below	3.89	FALSE
11	ERG	NM_182918.2	1.93	Above	3.86	FALSE
12	PSFL	NM_031301.2	0.63	Below	3.82	FALSE
13	NPY	NM_000905.2	4.43	Above	3.80	FALSE
14	CD1C	NM_001765.1	1.51	Above	3.77	FALSE
15	MAPK13	NM_002754.3	0.56	Below	3.74	FALSE
16	CSGlcA-T	NM_019015.1	0.69	Below	3.71	FALSE
17	ITK	NM_005546.3	0.53	Below	3.63	FALSE
18	LMO2	NM_005574.2	1.78	Above	3.58	FALSE
19	SCML2	NM_006089.1	2.11	Above	3.56	FALSE
20	PGM1	NM_002633.2	0.67	Below	3.47	FALSE
21	RAG2	NM_000536.1	0.28	Below	3.46	TRUE
22	S100A8	NM_002964.3	0.31	Below	3.41	FALSE
23	TCF7	NM_201632.1	0.43	Below	3.41	FALSE
24	WBSCR5	NM_032464.1	2.06	Above	3.36	FALSE
25	IFITM3	NM_021034.1	1.86	Above	3.31	FALSE
26	KIAA0922	NM_015196.2	0.62	Below	3.29	FALSE
27	PLSCR3	NM_020360.2	0.62	Below	3.28	FALSE
28	S100A12	NM_005621.1	0.37	Below	3.28	FALSE
29	MGAT4A	NM_012214.1	1.67	Above	3.27	TRUE
30	MAN2A2	NM_006122.2	0.66	Below	3.24	FALSE
31	EEF2	NM_001961.2	0.56	Below	3.23	FALSE
32	HOXC6	NM_004503.2	1.43	Above	3.22	FALSE
33	stch	U04735.1	1.61	Above	3.18	TRUE
34	NRN1	NM_016588.2	0.27	Below	3.15	FALSE
35	USP9X	NM_021906.1	1.40	Above	3.11	TRUE
36	NOTCH1	NM_017617.2	0.55	Below	3.07	FALSE
37	CD3D	NM_000732.3	0.46	Below	3.07	FALSE
38	COAS2	NM_178230.1	0.57	Below	3.03	FALSE
39	ELK3	NM_005230.2	1.37	Above	3.02	FALSE
40	SIX3	NM_005413.1	1.44	Above	3.01	FALSE
41	PTPNS1	NM_080792.1	1.59	Above	3.01	FALSE
42	CDC37	NM_007065.3	0.74	Below	3.00	FALSE
43	MST4	NM_016542.2	1.56	Above	3.00	TRUE
44	FYB	U93049.1	0.61	Below	2.98	FALSE
45	MAPKAPK3	NM_004635.3	0.59	Below	2.97	FALSE

46	WBSCR20A	NM_018044.2	0.65	Below	2.96	FALSE
47	BASP1	NM_006317.3	0.51	Below	2.94	FALSE
48	UCKL1	NM_017859.1	0.71	Below	2.94	FALSE
49	PDHA1	NM_000284.1	1.38	Above	2.92	TRUE
50	FARP1	NM_005766.2	0.41	Below	2.89	FALSE
51	CTNNBL1	NM_030877.3	0.74	Below	2.89	FALSE
52	FGR	NM_005248.1	0.51	Below	2.89	FALSE
53	RPL4	NM_000968.2	0.60	Below	2.86	FALSE
54	HPS4	NM_022081.4	0.55	Below	2.86	TRUE
55	QARS	NM_005051.1	0.67	Below	2.85	FALSE
56	SYK	NM_003177.3	0.63	Below	2.85	FALSE
57	FLJ23749	NM_152271.2	0.59	Below	2.84	FALSE
58	SCHIP1	NM_014575.1	2.79	Above	2.84	FALSE
59	PRDX4	NM_006406.1	1.56	Above	2.82	TRUE
60	ARMCX1	NM_016608.1	1.95	Above	2.80	TRUE
61	ATRX	NM_138271.1	1.81	Above	2.78	TRUE
62	HARSL	NM_012208.2	0.71	Below	2.78	FALSE
63	SYBL1	NM_005638.3	1.41	Above	2.78	TRUE
64	SPARC	NM_003118.2	1.61	Above	2.74	FALSE
65	AKAP12	NM_144497.1	0.29	Below	2.74	FALSE
66	PIP3-E	NM_015553.1	1.83	Above	2.72	FALSE
67	SESN1	NM_014454.1	0.68	Below	2.71	FALSE
68	MGC23937	NM_145052.1	1.58	Above	2.70	TRUE
69	MAP3K11	NM_002419.3	0.65	Below	2.70	FALSE
70	H2AFX	NM_002105.2	1.47	Above	2.69	FALSE
71	TIF1	NM_003852.3	0.55	Below	2.67	FALSE
72	CKLFSF2	NM_144673.2	2.29	Above	2.67	TRUE
73	INPP1	NM_002194.2	0.48	Below	2.66	FALSE
74		INCYTE UNIQUE	0.63	Below	2.66	FALSE
75	TMEM41B	BC035034.1	1.38	Above	2.66	FALSE
76	FBN2	NM_001999.2	0.44	Below	2.65	FALSE
77	IMP-1	NM_006546.3	0.57	Below	2.64	FALSE
78	CBFA2T3	NM_005187.4	0.71	Below	2.62	FALSE
79	MDS006	NM_020233.3	0.73	Below	2.61	FALSE
80	S100A9	NM_002965.2	0.48	Below	2.61	FALSE
81	SYNPO	AB028952.1	0.57	Below	2.60	FALSE
82	UBQLN2	NM_013444.2	1.28	Above	2.60	TRUE
83	DUSP5	NM_004419.2	2.15	Above	2.58	FALSE
84	FBP1	NM_000507.2	0.65	Below	2.58	FALSE
85	RTCD1	NM_003729.1	0.62	Below	2.57	FALSE
86	DCN	NM_133504.2	2.47	Above	2.57	TRUE
87	RPL4	NM_000968.2	0.60	Below	2.56	FALSE
88	USP20	NM_001008563.1	0.74	Below	2.55	FALSE
89	UXT	NM_004182.2	1.29	Above	2.55	TRUE
90	ELOVL5	NM_021814.3	0.58	Below	2.54	FALSE
91	MAL	NM_022440.1	0.44	Below	2.53	FALSE
92	H2AFX	NM_002105.2	1.44	Above	2.49	FALSE
93	COX6B1	NM_001863.3	0.78	Below	2.49	FALSE
94	STK17B	NM_004226.2	1.70	Above	2.47	FALSE
95	ERBB4	NM_005235.1	1.47	Above	2.47	FALSE
96	RBBP7	NM_002893.2	1.37	Above	2.46	TRUE
97		INCYTE	0.66	Below	2.45	FALSE

		UNIQUE				
98	TCEAL4	NM_024863.3	1.65	Above	0.44	TRUE
99	S100A10	NM_002966.1	0.66	Below	2.43	FALSE
100	POU4F1	NM_006237.2	1.38	Above	0.42	FALSE

T-ALL

Rank	Gene Name	Accession number	Fold Change	Below/ Above mean	T-Value	Previously published
1	MAL	NM_022440.1	35.56	Above	10.37	TRUE
2	CD3D	NM_000732.3	14.60	Above	9.60	TRUE
3	CD3G	NM_000073.1	9.79	Above	8.14	FALSE
4	CD79B	NM_021602.1	0.12	Below	7.81	TRUE
5	BTK	NM_000061.1	0.14	Below	7.39	FALSE
6	TFDP2	NM_006286.1	4.45	Above	7.34	TRUE
7	SEPW1	NM_003009.2	3.67	Above	6.95	TRUE
8	CD3E	NM_000733.2	6.66	Above	6.80	TRUE
9	HLA-DMA	NM_006120.2	0.16	Below	6.55	TRUE
10	TCL1A	NM_021966.1	0.07	Below	6.40	TRUE
11	MEF2C	NM_002397.2	0.13	Below	6.32	TRUE
12	VPREB3	NM_013378.1	0.03	Below	6.18	TRUE
13	SH2D1A	NM_002351.1	4.92	Above	6.14	TRUE
14	CD7	NM_006137.5	3.76	Above	6.10	TRUE
15	TCF7	NM_201632.1	6.99	Above	6.01	TRUE
16	PTPN12	NM_002835.2	0.33	Below	5.76	FALSE
17	WBSCR5	NM_032464.1	0.20	Below	5.70	FALSE
18	AUTS2	NM_015570.1	0.11	Below	5.33	FALSE
19	LAT	NM_014387.2	5.91	Above	5.25	TRUE
20	ITK	NM_005546.3	3.41	Above	5.11	FALSE
21	IGHM	X58529.1	0.22	Below	5.06	TRUE
22	ZAP70	NM_207519.1	4.24	Above	5.06	FALSE
23	SLC35E3	NM_018656.2	0.20	Below	5.03	FALSE
24	HHEX	NM_002729.2	0.20	Below	4.97	FALSE
25	NPY	NM_000905.2	0.07	Below	4.96	FALSE
26	OCIAD2	NM_152398.1	4.07	Above	4.92	TRUE
27	HLA-DPA1	NM_033554.2	0.17	Below	4.92	TRUE
28	HLA-DPB1	NM_002121.4	0.18	Below	4.90	TRUE
29	MAP1A	NM_002373.4	3.73	Above	4.87	FALSE
30	PLCG2	NM_002661.1	0.28	Below	4.80	TRUE
31	BEX1	NM_018476.2	6.69	Above	4.77	FALSE
32	MYO5C	NM_018728.1	0.15	Below	4.75	FALSE
33	STAB1	NM_015136.1	0.25	Below	4.75	FALSE
34	NAV1	NM_020443.2	0.30	Below	4.71	FALSE
35	ASPM	NM_018136.2	3.29	Above	4.67	FALSE
36	LCK	NM_005356.2	6.26	Above	4.66	FALSE
37	EBF1	AK096708.1	0.09	Below	4.64	TRUE
38	CD1E	NM_030893.1	7.42	Above	4.57	FALSE
39	AQP3	NM_004925.3	2.51	Above	4.37	FALSE
40	CSRP2	NM_001321.1	0.15	Below	4.31	TRUE
41	HLA-DRB5	NM_002125.3	0.19	Below	4.30	FALSE
42	SLIT1	NM_003061.1	13.34	Above	4.19	TRUE
43	FLI1	NM_002017.2	0.33	Below	4.19	FALSE
44	EPHB4	NM_004444.4	0.39	Below	4.16	FALSE

45	SERPINI1	NM_005025.2	0.24	Below	4.05	FALSE
46	SERPING1	NM_000062.1	0.20	Below	4.03	FALSE
47	USP20	NM_001008563.1	1.92	Above	4.00	TRUE
48	CPM	J04970.2	0.33	Below	3.94	FALSE
49	CDKN1A	NM_000389.2	0.30	Below	3.92	FALSE
50	INSR	X02160.1	0.27	Below	3.86	FALSE
51		NM_014796.1	2.93	Above	3.82	TRUE
52	CD3Z	NM_198053.1	2.33	Above	3.80	FALSE
53	BTG3	NM_006806.3	4.53	Above	3.79	FALSE
54	BRE	NM_199194.1	0.43	Below	3.78	FALSE
55	CD24	NM_013230.1	0.22	Below	3.78	TRUE
56	FHL2	NM_201557.1	4.47	Above	3.77	FALSE
57	TRAK1	AB028965.1	0.54	Below	3.77	TRUE
58	MDK	NM_002391.2	0.19	Below	3.74	FALSE
59	ITGAE	NM_002208.3	2.26	Above	3.73	FALSE
60	DNAPTP6	NM_015535.1	0.28	Below	3.73	FALSE
61	PLXNB2	AB002313.1	0.34	Below	3.72	FALSE
62	CTGF	NM_001901.1	0.12	Below	3.68	FALSE
63	PCCB	NM_000532.2	2.05	Above	3.67	FALSE
64	LY86	NM_004271.3	0.37	Below	3.65	FALSE
65	HMGCR	NM_000859.1	2.79	Above	3.63	TRUE
66	DUSP6	NM_001946.2	0.42	Below	3.60	FALSE
67		W74645.1	0.47	Below	3.59	FALSE
68	SLC2A5	NM_003039.1	0.48	Below	3.57	FALSE
69	FLT3	NM_004119.1	0.28	Below	3.55	FALSE
70	CMAS	NM_018686.3	1.72	Above	3.55	FALSE
71	ZC3HDC7	NM_014153.2	2.81	Above	3.54	FALSE
72	MME	NM_007289.1	0.15	Below	3.54	FALSE
73	IGJ	NM_144646.2	0.25	Below	3.53	FALSE
74	S100A10	NM_002966.1	2.33	Above	3.53	FALSE
75	TNFRSF6	NM_000043.3	1.73	Above	3.49	FALSE
76	GATA3	NM_001002295.1	5.60	Above	3.47	FALSE
77	MLLT3	NM_004529.1	3.30	Above	3.46	FALSE
78	TERF2	NM_005652.2	0.32	Below	3.39	FALSE
79	CD74	NM_004355.1	0.17	Below	3.39	TRUE
80	TMSNB	NM_021992.1	6.45	Above	3.38	FALSE
81	LAPTM5	NM_006762.1	0.38	Below	3.37	FALSE
82	SYK	NM_003177.3	0.47	Below	3.34	FALSE
83	FAM3C	NM_014888.1	0.41	Below	3.32	FALSE
84	HIST1H2BK	NM_080593.1	0.36	Below	3.32	FALSE
85	CDW52	NM_001803.1	0.41	Below	3.31	FALSE
86	EPHB6	NM_004445.2	3.71	Above	3.28	FALSE
87		AK055619.1	2.14	Above	3.27	TRUE
88	PARP1	NM_001618.2	0.46	Below	3.27	FALSE
89	DDEF1	NM_018482.2	0.55	Below	3.27	FALSE
90	CRA	NM_006697.1	0.40	Below	3.23	FALSE
91	MRPL34	NM_023937.2	1.61	Above	3.23	FALSE
92	PNMT	NM_002686.2	0.49	Below	3.19	FALSE
93	CD19	NM_001770.3	0.46	Below	3.19	TRUE
94	STIL	NM_003035.1	2.48	Above	3.16	FALSE
95	CTSL	NM_001912.2	2.36	Above	3.16	FALSE
96	ARMC8	NM_014154.2	2.46	Above	3.14	FALSE

97	FLJ20551	NM_017875.1	1.77	Above	3.14	FALSE
98	Cep70	NM_024491.2	3.22	Above	3.13	FALSE
99	TYMS	NM_001071.1	2.31	Above	3.12	FALSE
100	C1orf38	NM_004848.1	0.51	Below	3.12	FALSE

Other

Rank	Gene Name	Accession number	Fold Change	Below/ Above mean	T-Value	Previously published
1	TERF2	NM_005652.2	0.31	Below	3.91	FALSE
2	HMGB3	NM_005342.1	0.58	Below	3.56	FALSE
3	EEF2	NM_001961.2	2.42	Above	3.51	FALSE
4	QARS	NM_005051.1	1.87	Above	3.47	FALSE
5	EEF2	NM_001961.2	2.23	Above	3.47	FALSE
6	SEMA6A	NM_020796.2	0.31	Below	3.45	FALSE
7	UCKL1	NM_017859.1	1.67	Above	3.42	FALSE
8	CTSL	NM_001912.2	0.43	Below	3.39	FALSE
9	CCNA1	NM_003914.2	2.86	Above	3.35	FALSE
10	CBFB	NM_001755.2	1.74	Above	3.29	FALSE
11		INCYTE UNIQUE	2.07	Above	3.29	FALSE
12	HIST1H2BD	NM_021063.2	0.34	Below	3.27	FALSE
13	SH3BP5	NM_004844.1	0.49	Below	3.22	FALSE
14	EPHA7	NM_004440.2	1.47	Above	3.09	FALSE
15	SH3KBP1	NM_031892.1	0.61	Below	3.07	FALSE
16	RPL39	NM_001000.2	1.74	Above	3.05	FALSE
17		INCYTE UNIQUE	1.96	Above	3.01	FALSE
18	RPL38	NM_000999.2	1.86	Above	2.92	FALSE
19	ELK3	NM_005230.2	0.67	Below	2.89	FALSE
20	ALDH3B1	NM_000694.1	0.55	Below	2.85	FALSE
21	ERBB2	NM_004448.2	2.53	Above	2.74	FALSE
22	H2AFX	NM_002105.2	0.60	Below	2.71	FALSE
23	C6orf82	NM_015921.1	1.42	Above	2.66	FALSE
24	RPL39	NM_001000.2	1.59	Above	2.65	FALSE
25	EPHB4	NM_004444.4	1.76	Above	2.61	FALSE
26	TPP2	NM_003291.1	1.47	Above	2.61	FALSE
27	RPL38	NM_000999.2	1.77	Above	2.61	FALSE
28	GATA3	NM_001002295.1	3.36	Above	2.60	FALSE
29		BF091792.1	1.43	Above	2.58	FALSE
30	PLSCR3	NM_020360.2	1.65	Above	2.56	FALSE
31	CD244	NM_016382.2	0.63	Below	2.55	FALSE
32	TNK1	NM_003985.1	0.58	Below	2.54	FALSE
33	PRX	NM_020956.1	0.50	Below	2.54	FALSE
34	TM4SF2	NM_004615.2	0.34	Below	2.53	FALSE
35	TCFL5	NM_006602.2	0.43	Below	2.52	FALSE
36	CDYL	NM_004824.2	0.55	Below	2.51	FALSE
37	APOC1	NM_001645.3	0.29	Below	2.50	FALSE
38	TRAM2	NM_012288.3	0.57	Below	2.50	FALSE
39	RPL4	NM_000968.2	1.92	Above	2.49	FALSE
40	TP53	NM_000546.2	1.49	Above	2.49	FALSE
41	C11orf9	NM_013279.1	0.67	Below	2.49	FALSE

42	CKLFSF2	NM_144673.2	0.36	Below	2.47	FALSE
43	PIP3-E	NM_015553.1	0.49	Below	2.47	FALSE
44	H2AFX	NM_002105.2	0.62	Below	2.46	FALSE
45	RPL4	NM_000968.2	1.77	Above	2.46	FALSE
46	ARHGDIA	NM_004309.3	1.43	Above	2.44	FALSE
47	PNMT	NM_002686.2	1.66	Above	2.43	FALSE
48	ERG	NM_182918.2	0.57	Below	2.42	FALSE
49	SLC35E3	NM_018656.2	0.47	Below	2.42	FALSE
50	HOXC6	NM_004503.2	0.70	Below	2.41	FALSE
51	SYNPO	AB028952.1	1.99	Above	2.41	FALSE
52	QARS	NM_005051.1	1.47	Above	2.40	FALSE
53		BM510582.1	1.62	Above	2.39	FALSE
54	MGC23911	NM_144676.1	0.26	Below	2.38	FALSE
55	TNS	NM_022648.2	0.43	Below	2.38	FALSE
56	EPHA3	NM_005233.3	3.35	Above	2.37	FALSE
57	MDS006	NM_020233.3	1.45	Above	2.36	FALSE
58	GFOD1	NM_018988.1	1.59	Above	2.36	FALSE
59	FHL2	NM_201557.1	0.41	Below	2.33	FALSE
60	C9orf10	NM_014612.2	1.79	Above	2.31	FALSE
61	TYRO3	NM_006293.2	0.63	Below	2.31	FALSE
62	SORBS1	NM_015385.1	0.61	Below	2.30	FALSE
63	ARHGDIA	NM_004309.3	1.37	Above	2.29	FALSE
64	S100A10	NM_002966.1	1.67	Above	2.26	FALSE
65	LMO1	NM_002315.1	0.70	Below	2.26	FALSE
66	CAPG	NM_001747.1	0.64	Below	2.25	FALSE
67	GRHPR	NM_012203.1	1.31	Above	2.23	FALSE
68	BHLHB2	NM_003670.1	1.43	Above	2.22	FALSE
69	SCHIP1	NM_014575.1	0.35	Below	2.20	FALSE
70	CD1C	NM_001765.1	0.73	Below	2.15	FALSE
71	RAG2	NM_000536.1	2.85	Above	2.14	FALSE
72	FLJ40504	NM_173624.1	0.56	Below	2.13	FALSE
73	B4GALT6	AF038664.1	0.65	Below	2.13	FALSE
74	RADIL	AK001186.1	0.67	Below	2.11	FALSE
75	EIF2B5	NM_003907.1	1.35	Above	2.11	FALSE
76	IMPA2	NM_014214.1	0.60	Below	2.11	FALSE
77	COAS2	NM_178230.1	1.69	Above	2.10	FALSE
78	INSRR	NM_014215.1	0.51	Below	2.10	FALSE
79	NR3C2	NM_000901.1	0.53	Below	2.09	FALSE
80	PTK7	NM_002821.3	1.77	Above	2.09	FALSE
81	CMAS	NM_018686.3	0.74	Below	2.08	FALSE
82	BCL6	NM_001706.2	1.78	Above	2.08	FALSE
83	ALDOAP2	M21191.1	1.71	Above	2.08	FALSE
84	COX6B1	NM_001863.3	1.32	Above	2.07	FALSE
85	DSG2	NM_001943.1	0.66	Below	2.06	FALSE
86	HMGCL	NM_000191.1	1.39	Above	2.06	FALSE
87	USP20	NM_001008563.1	1.38	Above	2.05	FALSE
88	DDX3Y	NM_004660.2	0.30	Below	2.05	FALSE
89	VAMP5	NM_006634.2	0.70	Below	2.05	FALSE
90	AQP3	NM_004925.3	0.66	Below	2.05	FALSE
91	PHF15	BC021169.1	0.63	Below	2.04	FALSE
92	MRPL3	NM_007208.2	1.42	Above	2.02	FALSE
93	EIF2B5	NM_003907.1	1.38	Above	2.02	FALSE
94	FEZ2	NM_005102.1	1.40	Above	2.01	FALSE

95	ST18	NM_014682.1	0.74	Below	2.00	FALSE
96	QSCN6	NM_002826.4	0.58	Below	2.00	FALSE
97	RB1	NM_000321.1	1.90	Above	1.98	FALSE
98	CYC1	NM_001916.2	1.31	Above	1.97	FALSE
99	TAGLN2	NM_003564.1	1.96	Above	1.97	FALSE
100	MPZL1	BC007881.2	2.01	Above	1.96	FALSE

Non- classified

Rank	Gene Name	Accession number	Fold Change	Below/ Above mean	T-Value	Previously published
1	CST7	NM_003650.2	4.65	Above	6.04	FALSE
2	PYGL	NM_002863.2	3.53	Above	4.91	FALSE
3	ILK	NM_004517.1	2.15	Above	4.64	FALSE
4	DNTT	NM_004088.2	0.13	Below	4.61	FALSE
5	TTN	AK127826.1	3.51	Above	4.41	FALSE
6	ERG	NM_182918.2	0.39	Below	4.35	FALSE
7	SYK	NM_003177.3	2.38	Above	4.33	FALSE
8	stch	U04735.1	0.45	Below	4.24	FALSE
9	SERPINA1	NM_000295.3	5.79	Above	4.02	FALSE
10	CCL5	NM_002985.2	2.26	Above	4.00	FALSE
11	MGAT4A	NM_012214.1	0.45	Below	3.96	FALSE
12	FHIT	NM_002012.1	0.44	Below	3.86	FALSE
13	CPA3	NM_001870.1	4.69	Above	3.76	FALSE
14	DJ971N18.2	NM_021156.2	1.65	Above	3.72	FALSE
15	NPY	NM_000905.2	0.15	Below	3.70	FALSE
16	JAK2	NM_004972.2	1.71	Above	3.67	FALSE
17	PRG1	NM_002727.2	3.12	Above	3.64	FALSE
18	PARP2	NM_005484.2	0.57	Below	3.59	FALSE
19	NKG7	NM_005601.3	2.10	Above	3.59	FALSE
20	LYL1	NM_005583.3	2.33	Above	3.58	FALSE
21	LTK	NM_002344.3	4.98	Above	3.46	FALSE
22	MPO	NM_000250.1	3.87	Above	3.41	FALSE
23	SMARCD3	NM_003078.3	3.83	Above	3.38	FALSE
24	SPG21	NM_016630.3	1.51	Above	3.34	FALSE
25	IMPA2	NM_014214.1	2.16	Above	3.33	FALSE
26	LYZ	NM_000239.1	4.15	Above	3.32	FALSE
27	DF	NM_001928.2	4.66	Above	3.30	FALSE
28	PHF3	NM_015153.1	0.55	Below	3.29	FALSE
29	SPTBN1	NM_178313.1	0.42	Below	3.28	FALSE
30	COL9A2	NM_001852.3	1.98	Above	3.24	FALSE
31	RAG2	NM_000536.1	0.21	Below	3.24	FALSE
32	PXDN	AF200348.1	0.23	Below	3.23	FALSE
33	THBS4	NM_003248.3	3.78	Above	3.16	FALSE
34	ARMCX1	NM_016608.1	0.38	Below	3.15	FALSE
35	FARSLA	NM_004461.1	1.96	Above	3.14	FALSE
36	ZFP36L1	NM_004926.2	0.48	Below	3.10	FALSE
37	SEC22L1	NM_004892.3	2.14	Above	3.10	FALSE
38	KLF1	NM_006563.2	3.80	Above	3.10	FALSE
39	APOC2	NM_000483.3	5.24	Above	3.08	FALSE
40	LYN	NM_002350.1	2.73	Above	3.07	FALSE
41	S100A9	NM_002965.2	3.02	Above	3.07	FALSE
42	ILK	NM_004517.1	2.19	Above	3.06	FALSE

43	BRE	NM_199194.1	1.88	Above	3.06	FALSE
44	THBS4	NM_003248.3	3.74	Above	3.06	FALSE
45	MGC23911	NM_144676.1	0.18	Below	3.04	FALSE
46	C14orf139	NM_024633.2	0.41	Below	3.02	FALSE
47	FCGR1A	NM_000566.2	4.48	Above	2.99	FALSE
48	HIST2H2BE	NM_003528.2	0.50	Below	2.97	FALSE
49	IRF5	NM_032643.3	1.93	Above	2.97	FALSE
50	CTSW	NM_001335.2	3.80	Above	2.95	FALSE
51	EPHA7	NM_004440.2	0.70	Below	2.89	FALSE
52	LTA4H	NM_000895.1	2.09	Above	2.88	FALSE
53	TGFB1	NM_000660.2	1.76	Above	2.87	FALSE
54	ZFP36L1	NM_004926.2	0.51	Below	2.86	FALSE
55	CD36	NM_001001547.1	3.32	Above	2.84	FALSE
56	PSFL	NM_031301.2	1.57	Above	2.83	FALSE
57	SNAP23	NM_003825.2	2.17	Above	2.82	FALSE
58	IGFBP7	NM_001553.1	3.19	Above	2.81	FALSE
59		AK022072.1	0.55	Below	2.80	FALSE
60	ACK1	NM_005781.3	1.53	Above	2.79	FALSE
61	PTGDS	NM_000954.5	3.27	Above	2.78	FALSE
62	CALR	NM_004343.2	2.14	Above	2.78	FALSE
63	BTK	NM_000061.1	2.21	Above	2.76	FALSE
64	TAGLN3	NM_013259.2	0.62	Below	2.75	FALSE
65	HIST2H2BE	NM_003528.2	0.52	Below	2.74	FALSE
66	CD164	NM_006016.3	1.88	Above	2.74	FALSE
67	CX3CR1	NM_001337.2	2.72	Above	2.74	FALSE
68	RAB32	NM_006834.2	3.07	Above	2.73	FALSE
69	MAPKAPK3	NM_004635.3	1.86	Above	2.72	FALSE
70	NAP1L4	NM_005969.3	1.96	Above	2.72	FALSE
71	C20orf111	NM_016470.6	0.65	Below	2.71	FALSE
72		BF513468.1	0.33	Below	2.71	FALSE
73	RPS4Y1	NM_001008.2	8.07	Above	2.70	FALSE
74	MINPP1	NM_004897.2	2.33	Above	2.69	FALSE
75	TEC	NM_003215.1	2.45	Above	2.67	FALSE
76	CCL23	NM_005064.3	2.07	Above	2.66	FALSE
77	MYC	NM_002467.2	2.70	Above	2.65	FALSE
78	HCK	NM_002110.2	2.40	Above	2.64	FALSE
79	C18orf1	NM_181481.2	0.52	Below	-2.63	FALSE
80	VAMP5	NM_006634.2	0.64	Below	2.62	FALSE
81	LCP2	NM_005565.3	2.11	Above	2.59	FALSE
82	GP1BB	NM_000407.3	2.35	Above	2.57	FALSE
83	CDKN1A	NM_000389.2	0.49	Below	2.52	FALSE
84	DCTN4	NM_016221.2	0.67	Below	2.51	FALSE
85	BLK	NM_001715.2	0.35	Below	2.50	FALSE
86	RPS6KA5	NM_004755.2	1.82	Above	2.50	FALSE
87	CSRP2	NM_001321.1	0.35	Below	2.48	FALSE
88	M17S2	NM_031862.1	2.68	Above	2.48	FALSE
89	CDKN1C	NM_000076.1	2.02	Above	2.47	FALSE
90	FAM38A	NM_014745.1	1.78	Above	2.46	FALSE
91	MRPL3	NM_007208.2	1.68	Above	2.46	FALSE
92	C11orf9	NM_013279.1	0.67	Below	2.44	FALSE
93	FURIN	NM_002569.2	1.84	Above	2.42	FALSE
94	EBF1	AK096708.1	0.30	Below	2.42	FALSE
95	RNPEP	NM_020216.3	1.74	Above	2.42	FALSE

96	PGAM1	NM_002629.2	2.07	Above	2.41	FALSE
97	NFIL3	NM_005384.2	2.22	Above	2.39	FALSE
98	TCEAL4	NM_024863.3	0.53	Below	2.39	FALSE
99	EIF4A1	NM_001416.1	2.31	Above	2.39	FALSE
100	SRC	NM_198291.1	0.67	Below	2.37	FALSE