

Genome-wide DNA methylation analysis of transient neonatal diabetes type 1 patients with mutations in *ZFP57*

Mads Bak¹, Susanne E Boonen^{1,2}, Christina Dahl³, Johanne M D Hahnemann², Deborah J DG Mackay^{4,6}, Zeynep Tümer^{2,7}, Karen Grønskov², I Karen Temple^{4,5}, Per Guldborg³, Niels Tommerup¹

5

¹ Wilhelm Johannsen Center for Functional Genome Research, Institute of Cellular and Molecular Medicine, Panum Institute, University of Copenhagen, DK-2200 Copenhagen N, Denmark.

² Center for Applied Human Molecular Genetics, Kennedy Center, DK-2600 Glostrup, Denmark.

³ Institute of Cancer Biology, Danish Cancer Society, DK-2100 Copenhagen Ø, Denmark.

10 ⁴ Human Genetics and Genomic Medicine, Faculty of Medicine, University of Southampton, Southampton SO16 6YD, UK.

⁵ Wessex Clinical Genetics Service, Southampton University Hospitals Trust, Southampton SO16 5YA, UK.

15 ⁶ Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury NHS Foundation Trust, SP2 8BJ Salisbury, UK.

⁷ Institute of Cellular and Molecular Medicine, Panum Institute, University of Copenhagen, DK-2200N Copenhagen, Denmark.

Corresponding author: Mads Bak; Email: madsba@sund.ku.dk; Telephone: +45 35327844

20

Email addresses:

Susanne E Boonen: susanne@boonen.dk

Christina Dahl: chd@cancer.dk

Johanne M D Hahnemann: Johanne.Davidsen.Hahnemann@regionh.dk

25 Deborah J DG Mackay: D.J.G.Mackay@soton.ac.uk

Zeynep Tümer: asuman.zeynep.tuemer@regionh.dk

Karen Grønskov: Karen.Groenskov@regionh.dk

I Karen Temple: I.K.Temple@soton.ac.uk

Per Guldborg: perg@cancer.dk

30 Niels Tommerup: ntommerup@sund.ku.dk

Abstract

Background

Transient neonatal diabetes mellitus 1 (TNDM1) is a rare imprinting disorder characterized by
35 intrauterine growth retardation and diabetes mellitus usually presenting within the first six weeks of
life and resolves by the age of 18 months. However, patients have an increased risk of developing
diabetes mellitus type 2 later in life. Transient neonatal diabetes mellitus 1 is caused by
overexpression of the maternally imprinted genes *PLAGL1* and *HYMAI* on chromosome 6q24. One
of the mechanisms leading to overexpression of the locus is hypomethylation of the maternal allele
40 of *PLAGL1* and *HYMAI*. A subset of patients with maternal hypomethylation at *PLAGL1* have
hypomethylation at additional imprinted loci throughout the genome, including *GRB10*, *ZIM2*
(*PEG3*), *MEST* (*PEG1*), *KCNQ1OT1* and *NESPAS* (*GNAS-AS1*). About half of the TNDM1 patients
carry mutations in *ZFP57*, a transcription factor involved in establishment and maintenance of
methylation of imprinted loci. Our objective was to investigate whether additional regions are
45 aberrantly methylated in *ZFP57* mutation carriers.

Methods

Genome-wide DNA methylation analysis was performed on four individuals with homozygous or
compound heterozygous *ZFP57* mutations, three relatives with heterozygous *ZFP57* mutations and
five controls. Methylation status of selected regions showing aberrant methylation in the patients
50 was verified using bisulfite-sequencing.

Results

We found large variability among the patients concerning the number and identity of the
differentially methylated regions, but more than 60 regions were aberrantly methylated in two or
more patients and a novel region within *PPP1R13L* was found to be hypomethylated in all the
55 patients. The hypomethylated regions in common between the patients are enriched for the *ZFP57*
DNA binding motif.

Conclusions

We have expanded the epimutational spectrum of TNDM1 associated with *ZFP57* mutations and found one novel region within *PPP1R13L* which is hypomethylated in all TNDM1 patients
60 included in this study. Functional studies of the locus might provide further insight into the etiology of the disease.

Keywords

Next-generation sequencing, imprinting disorder, transient neonatal diabetes, DNA methylation

Background

Imprinting is an epigenetic mechanism in mammals by which one allele of certain genes is expressed in a parent-of-origin specific manner [1]. Imprinting is controlled by modifications of sequence elements termed "imprinting control regions" (ICRs). These modifications include DNA methylation as well as acetylation and methylation of histones [2]. Transient neonatal diabetes mellitus 1 (TNDM1) is an imprinting disorder caused by overexpression of the maternally imprinted genes *PLAGL1* and *HYMAI* on chromosome 6q24. There are three major mechanisms related to this disorder: paternal uniparental disomy of chromosome 6, paternal duplications of 6q24 and total loss of maternal methylation at the TNDM1 differentially methylated region (DMR), *PLAGL1*:alt-TSS-DMR, without any detectable DNA sequence changes [3]. Hypomethylation of additional imprinted loci is observed in half of the TNDM1 patients displaying loss of methylation at the TNDM1 locus. In almost half of the TNDM1 patients with multilocus methylation defects (TND-MLMD), homozygous and compound heterozygous mutations were identified in *ZFP57* [4]. Methylation analysis of imprinted loci in these patients have shown that in addition to loss of methylation at *PLAGL1*:alt-TSS-DMR, *PEG3*:TSS-DMR and *GRB10*:alt-TSS-DMR are hypomethylated to different extents, while the *MEST*:alt-TSS-DMR, *KCNQ1OT1*:TSS-DMR and *GNAS-AS1*:TSS-DMR are hypomethylated in some patients [4, 5]. The normal function of *ZFP57* is maintenance of imprinting during early embryonic development [6, 7]. Mouse and human *ZFP57* binds the methylated hexanucleotide TGCC^{me}GC [7, 8] and missense mutations in *ZFP57* can disrupt interaction between the protein and its target sequence [8].

Genome-wide methylation analyses have recently been carried out in patients with TNDM to identify additional aberrantly methylated regions [9–11]. These, analyses were performed using methylation microarrays and were thus restricted to analyzing regions covered by the probes on the arrays. In the present study, we employed a DNA sequencing based approach to detect aberrantly methylated regions in TNDM1 patients with *ZFP57* mutations.

Methods

Probands, heterozygotes, and controls

The methylation status at six specific known imprinted loci was previously investigated by
95 methylation specific PCR (MS-PCR) and verified by pyrosequencing [4] in four patients and three
heterozygotes selected for this study. The *ZFP57* mutations and methylation status of the patients
were:

Patient 1 (Proband II-1, family 1): *ZFP57* genotype: c.723C>A / c.723C>A (p.Cys241* /
p.Cys241*). Total loss of methylation at *PLAGL1*:alt-TSS-DMR, *GRB10*:alt-TSS-DMR and
100 *PEG3*:TSS-DMR. Partial loss of methylation at *MEST*:alt-TSS-DMR, *KCNQ1OT1*:TSS-DMR, and
GNAS-AS1:TSS-DMR.

Patient 2 (Proband II-2, family 1): *ZFP57* genotype: c.723C>A / c.723C>A (p.Cys241* /
p.Cys241*). Total loss of methylation at *PLAGL1*:alt-TSS-DMR, *GRB10*:alt-TSS-DMR and
PEG3:TSS-DMR. Partial loss of methylation at *MEST*:alt-TSS-DMR and *KCNQ1OT1*:TSS-DMR.
105 Normal methylation at *GNAS-AS1*:TSS-DMR.

Patient 3 (Proband II-3, family 2): *ZFP57* genotype: c.257_258delAG / c.257_258delAG
(p.E86VfsX28 / p.E86VfsX28). Total loss of methylation at *PLAGL1*:alt-TSS-DMR and
GRB10:alt-TSS-DMR. Partial loss of methylation at *PEG3*:TSS-DMR. Normal methylation at
MEST:alt-TSS-DMR, *KCNQ1OT1*:TSS-DMR and *GNAS-AS1*:TSS-DMR.

110 **Patient 4** (Proband II-3, family 7): *ZFP57* genotype: c.683G>A / c.838_845delACCCAGGC
(p.R228H / p.279fsX1). Total loss of methylation at *PLAGL1*:alt-TSS-DMR and *GRB10*:alt-TSS-
DMR. Partial loss of methylation at *PEG3*:TSS-DMR, *MEST*:alt-TSS-DMR, and *GNAS-AS1*:TSS-
DMR. Normal methylation at *KCNQ1OT1*:TSS-DMR.

The three heterozygotes had normal methylation at these six loci.

115 Five unrelated individuals were used as normal controls.

MBD-seq

Genomic DNA (2 µg) was isolated from peripheral blood leukocytes (PBL) and randomly sheared by nebulization to fragments with an average size of 250 bp. Methylated DNA fragments were
120 isolated using His-tagged methyl binding protein MBD2b using MethylCollector (ActiveMotif) and
DNA adaptors were ligated to the ends of the isolated methylated DNA using the NEBNext DNA
library preparation kit (New England BioLabs). Ligation products were separated on a 2% agarose
gel and fragments of 200-400 bp in size were isolated from the gel. The isolated fragments were
amplified by 15 cycles of PCR using primers binding to the adapter sequences. The resulting library
125 was quantified with picogreen and 36 bases of the fragments were sequenced on a Genome
Analyzer IIx (Illumina) following the manufacturer's protocol.

Data analysis

Sequences were aligned to the human genome (GRCh37/hg19) using BWA version 0.6.1 [12].
Reads with a mapping quality < 37 and duplicate reads were removed. Methylated genomic regions
130 were identified using MACS version 1.4.2 ($p < 1e-8$) [13]. Methylated regions overlapping more
than 50% were merged. For each methylated region, the number of non-redundant reads was
counted for each patient and controls. Regions with less than 20 reads per million reads in all
samples were removed. For each patient and *ZFP57*-heterozygous individuals, methylation levels
were compared to controls. Aberrantly methylated regions were detected using edgeR [14] ($P \leq$
135 0.05, Benjamini-Hochberg adjusted). Overlap of aberrantly methylated regions with allele-specific
methylated regions and mapping relative to Ensembl transcripts were performed using python
scripts. The analysis was also performed using BALM [15] as peak caller yielding similar results.

Bisulfite sequencing

Genomic DNA (1 µg) from PBL was bisulfite converted (Zymo Research). Differentially
140 methylated regions were PCR amplified using 100 ng bisulfite converted DNA as template in a
reaction volume of 15 µl (10 µl 10x reaction buffer, 1 µl forward primer (10 µM), 1 µl reverse
primer (10 µM), 1 µl dNTPs (2.5 µM each), 0.45 µl MgCl₂ (50 mM), 0.1 µl Platinum TAQ

Polymerase (Invitrogen) and 1 µl DNA template, 8.95 µl H₂O). Primers for PCR amplification were designed using the Methyl Primer Express Software v1.0 (Applied Biosystems). Primer sequences

145 were: GDF7_F: AGTTGGGTATTGTTGTTAGGA, GDF7_R:
ACACRTAAACAACAACAACAAC, MAFF_F: TTGGGGTATTTAAAGGTGTTT, MAFF_R:
TACAACCTCCTCCTTCTAACACA, GAL3ST1_F: GATAGGGGAAAAAATTAAGGGT,
GAL3ST1_R: ACCAACTTAAACCCTACCTCAA, KCNAB3_F:
GGGGATATAGGAGTAAGGATTAGG, KCNAB3_R: CTAACAACCCTTAAATCCCAA,
150 DFNB31_F: TATGGAGTTATTGAAAAAGTTGTGTT, DFNB31_R:
ACCCCCTACAAACAACCC, ICAM5_F: AGATGGTTTTAGGTTTGAGGAGT, ICAM5_R:
AAACACCCCTACTCCCTACC, PPP1R13L_F: AGAAGGTTTGGGGTTTTTTT, PPP1R13L_R:
ATCCCTCAATACCCTAACCTC, ADRA2A_F: TGGTGTGTTGGTTTTTTTTTTT, ADRA2A_R:
AACTAAACACCCCTCRATACC. PCR conditions were: 96°C for 10 min., 5 cycles of [95°/15s;
155 65°/30s; 72°/30s], 5 cycles of [95°/15s; 60°/30s; 72°/30s], 30 cycles of [95°/15s; 55°/30s; 72°/30s]
and a final extension at 72°C for 5 min. For each patient, PCR products were combined and
libraries for sequencing on the Genome Analyzer were prepared. Briefly, the combined PCR
products were blunt ended and concatemerized by DNA ligation. Ligated DNA was randomly
sheared by sonication (Bioruptor, Diagenode), followed by blunt ending and ligation of indexed
160 sequencing adapters. Indexed adapters were designed in a way such that the first 5 bases of each
read identify the patient. Adapter oligo 1: 5'-
ACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXXXT, adapter oligo 2: 5'-phosphate-
xxxxxAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG, where XXXXX is the index and
xxxxx is the complementary bases of the index. The following indexes were used: ACCAAT,
165 AGAGAT, AGTCAT, CAGTCT, CCGGCT, CGATCT, CGTACT, GAGAGT and GCCGGT.
Indexed libraries were pooled and sequenced on a Genome Analyzer Iix (Illumina). Reads were
aligned to the amplicons using Bismark [16]. For each CpG site, a methylated-unmethylated ratio
was calculated. An overall methylated-unmethylated ratio was calculated for each amplicon as the

average of all CpG methylation ratios within each amplicon.

Results and discussion

We used methyl-CpG binding domain protein sequencing (MBD-seq) [17] to generate genome-wide methylation profiles of genomic DNA from peripheral blood leukocytes (PBL) from four TND-MLMD patients with homozygous or compound heterozygous *ZFP57* mutations (175 *ZFP57*mut/mut), three relatives with heterozygous *ZFP57* mutations (*ZFP57*norm/mut) and five unrelated controls. The four patients have previously been analyzed for changes in methylation levels at selected known imprinted loci [4, 5]. A total of 17,123 methylated regions with at least 20 reads per million reads (RPM) in at least one of the 12 individuals were detected by MBD-seq. More than 20% (3,712 / 17,123) of the methylated regions detected by this method are not covered 180 by probes on the Infinium HumanMethylation450 Beadchip (Illumina). As patients with imprinting defects can display substantial differences in the number and extent of affected loci [5, 9, 10] they were compared to the controls individually to allow detection of patient-specific methylation changes.

Validation of method

185 To validate the method we compared MBD-seq results with the previous studies of the patients [4, 5]. MBD-seq successfully detected all loci which have previously been shown to be completely hypomethylated (Figure 1). However, for most of the known partially hypomethylated regions the P-values were above 0.05. To further validate the method, we selected eight hypomethylated regions and verified their methylation status by bisulfite-sequencing (BS-seq) in three patients, 190 three heterozygotes and three controls. The selected regions included regions showing hypomethylation in four (*PPP1R13L*), three (*GDF7*, *DFNB31* and *KCNAB3*), two (*ADRA2A*, *PES1* and *ICAM5*) and one (*MAFF*) of the patients. BS-seq confirmed hypomethylation of these loci and there was good agreement between the changes in methylation levels detected by the two methods (Figure 2).

195

Aberrantly methylated loci

We found large variability among the patients regarding the number of affected loci, ranging from 12 regions (8 of these were hypomethylated) to 201 regions (123 of these were hypomethylated) regions in patient 4 and 2, respectively (Supplementary table 1-7 and Figure 3). In agreement with previous reports, we detected hypomethylation of several imprinted loci, including *PLAGL1*:alt-TSS-DMR [4, 5, 9, 10], *GRB10*:alt-TSS-DMR [4, 5, 9, 10], *PEG3*:TSS-DMR [4, 5, 9], *GNAS-AS1*:TSS-DMR [4, 5, 9, 10], *INPP5F*:Int2-DMR [10] and *NAP1L5*:TSS-DMR [10]. The following known imprinted loci were detected as hypomethylated in the patients: Patient 1: *PLAGL1*:alt-TSS-DMR, *GRB10*:alt-TSS-DMR, *PEG3*:TSS-DMR, *INPP5F*:Int2-DMR, *GNAS-AS1*:TSS-DMR, *NAP1L5*:TSS-DMR and *SLC22A18*; Patient 2: *PLAGL1*:alt-TSS-DMR, *GRB10*:alt-TSS-DMR, *PEG3*:TSS-DMR, *NAP1L5*:TSS-DMR, *SLC22A18*; Patient 3: *PLAGL1*:alt-TSS-DMR, *GRB10*:alt-TSS-DMR; Patient 4: *PLAGL1*:alt-TSS-DMR, *GRB10*:alt-TSS-DMR and *GNAS-AS1*:TSS-DMR. The hypomethylated region within *SLC22A18* is an internal CpG island encompassing exon 4 and 5 of the gene. Methylation changes were not detected at any other known imprinted loci. Docherty and colleagues identified three novel imprinted genes (*NHP2L1*, *WRB* and *PPIEL*) among the hypomethylated regions in TND patients without *ZFP57* mutations [9]. In agreement with the results of the present study, no hypomethylation was observed in patients with *ZFP57* mutations at any of these loci [9]. Three regions were aberrantly methylated - all hypomethylated - in all four patients and two of the them were the known *PLAGL1* and *GRB10* regions. The third region resides within the *PPP1R13L* gene (encoding Inhibitory member of the ASSP family (iASSP) or RelA-associated inhibitor (RAI)) on chromosome 19q13.32. This hypomethylated region is a CpG island encompassing exon 4-6 of *PPP1R13L* (Figure 4). *PPP1R13L* is an inhibitory member of the apoptosis stimulating protein of p53 family and overexpression of *PPP1R13L* has been detected in various cancers [18–21]. In normal tissue, it shows highest expression in the heart, placenta, skin, tongue and prostate (www.biogps.org) [22]. A splice-site mutation in *Ppp1r13l* was identified in Waved3 mice causing open eyelids at birth, wavy coat, and cardiomyopathy [23]. The normal function of gene body CpG islands like the hypomethylated CpG island within *PPP1R13L* is

currently not known. They might be involved in regulation of exonic inclusion/exclusion [24] or they could represent “orphan promoters” that are used in early development [25]. The nearest
225 known transcript is a short isoform of *PPP1R13L* (NM_006663) with transcription start site located
~3.5 kb from the hypomethylated CpG island.

In the *ZFP57* heterozygotes, all aberrantly methylated regions were private except for two regions within *ESPNP* and *MUC4* which were hypermethylated in two of the three heterozygotes and a region encompassing *ACTL10*, which was hypomethylated in all heterozygotes. In the patients,
230 *ESPNP* and *ACTL10* were both hypomethylated in three individuals, whereas no changes in *MUC4* methylation was detected.

Aberrantly methylated regions shared between two or more patients

To characterize the differentially methylated common regions, we focused on loci showing methylation changes in two or more patients. A total of 52 loci were hypomethylated and 9 were
235 hypermethylated in at least two patients (Supplementary table 8). Six shared aberrantly methylated regions are not covered by probes on the Infinium HumanMethylation450 Beadchip (Illumina) (Supplementary table 8). Six known imprinted loci were among the shared hypomethylated regions, including the maternally imprinted *PLAGL1:alt-TSS-DMR*, *GRB10:alt-TSS-DMR*, *PEG3:TSS-DMR*, *GNAS-AS1:TSS-DMR*, *NAP1L5:TSS-DMR* and the maternally expressed *SLC22A18*. The
240 majority of affected regions overlap a CpG island (50 out of 52 hypomethylated and 5 out of 9 hypermethylated). Measured from the center of the regions, 24 hypomethylated and 6 hypermethylated regions map more than 1 kb from the 5' end of an Ensembl transcript. Thus, more than half of the affected regions are not closely associated with known transcriptional start sites. They are, however, associated with gene bodies. Of those regions not closely associated with
245 transcription start, almost 90% localizes within Ensembl transcripts and many close to the 3' end. The normal function of gene body methylation is currently not known, but it is not associated with transcriptional silencing [26].

To analyse whether the identified hypomethylated regions are differentially methylated regions

(DMRs) of novel imprinted genes or imprinting control regions (ICRs) regulating parental-of-origin
250 expression of nearby genes, we analysed them for enrichment of regions reported to show allele-
specific methylation, enrichment for ZFP57 binding motifs and experimentally investigated a subset
of them for allele-specific expression.

We compared the 52 hypomethylated regions with allele-specific methylated regions (AMRs) from
MethBase (<http://smithlabresearch.org/software/methbase/>) [27]. We extracted 92 data sets with
255 AMRs from 22 studies and counted number of datasets showing AMR in the hypomethylated
regions. As expected, the known human imprinted DMRs [28, 29] are highly enriched for AMRs.
More than 77% of human imprinted DMRs are detected as AMRs in more than five of the data sets
and all germline DMRs (gDMRs) are detected as AMR in more than 15 data sets. Half (51%) of the
hypomethylated regions identified in this study are AMRs in more than 5 data sets whereas only
260 11% of all detected methylated regions (1,933 / 17,123) was reported as AMRs in more than 5 data
sets. However, the five known maternally imprinted loci identified as hypomethylated in this study
(*PLAGL1*:alt-TSS-DMR, *GRB10*:alt-TSS-DMR, *PEG3*:TSS-DMR, *GNAS-AS1*:TSS-DMR,
NAP1L5:TSS-DMR) top the list with more than 36 datasets detecting them as AMRs. Thus, the
hypomethylated regions are enriched for AMR. However, allele-specific methylation is not
265 necessarily indicative of imprinting as it is well established that the genotype can influence upon the
DNA methylation in *cis* [30–33].

Most of the regions show intermediate methylation in blood cells or other tissues
[24] (Supplementary table 8). This was also evident from the BS-seq validation. In the controls the
average methylation levels ranged from 25% to 88%. The chromatin of the unmethylated allele of
270 many imprinted DMRs are marked by H3K4me2/3 [34]. We downloaded H3K4me2/3 data from
blood cells from the ENCODE website (<https://www.encodeproject.org/>). Of the BS-seq validated
regions *ICAM5*, *DFNB31*, *KCNAB3* and *MAFF* showed H3K4me2/3 in most or all cell types
whereas *PPP1R13L* and *PES1* did not show the modifications in any of the cell types
(Supplementary table 8).

275 Mouse and human Zfp57 is involved in maintenance of genomic imprints during early embryogenesis [6]. The protein recognizes the methylated DNA motif TGCCGC (TGCC^{me}GC) which is present at all known mouse and some human imprinting control regions [7]. In mice imprinting control regions, the motif is present with an average of two copies [7]. Sequence analysis of the 52 hypomethylated regions in *ZPF57*mut/mut individuals revealed enrichment of the
280 TGCCGC motif compared to all analyzed regions (17,123) (Figure 5). Approximately 54% (28/52) of the hypomethylated regions contain two or more ZFP57 binding motif while 21% (3,560 / 17,123) of all methylated regions has two or more copies of this motif. The five hypomethylated regions at the maternally imprinted *PLAGL1*:alt-TSS-DMR, *NAP1L5*:TSS-DMR, *GRB10*:alt-TSS-DMR, *GNAS-AS1*:TSS-DMR and *PEG3*:TSS-DMR have the top five most TGCCGC motifs with 6,
285 7, 8, 10 and 11 motifs, respectively. Very similar to the hypomethylated regions, 58% (18/31) of the known human gDMRs [29] contain 2 or more copies of the binding motif.

To further investigate if the hypomethylated regions are imprinted DMRs we genotyped nine trios in the genes adjacent to the eight BS-seq validated hypomethylated regions and sequenced cDNA from peripheral blood samples to investigate for allele-specific expression. We identified
290 informative SNPs in *DFNB31*, *GAL3ST1*, *ICAM5* and *MAFF*, which also were expressed in blood. All showed biallelic expression, suggesting that they are not imprinted. However, we cannot rule out the possibility that they exhibit tissue-specific imprinting.

Thus, the hypomethylated regions in the patients are enriched for allele specific methylated regions and for the ZFP57 binding motif, but not to an extent found at the affected regions that are known
295 imprinted genes. Additionally, none of the regions have been identified as imprinted or candidate imprinted in genome-wide or chromosome-wide screens [35–40], suggesting that they are not associated with imprinted genes. This is in agreement with recent genome-wide studies of patients with multilocus methylation defects other than TND also reporting aberrant methylation at apparently non-imprinted loci [41, 42]. Experiments in mice have shown that ZFP57 in complex

300 with KAP1 and SETDB1 also binds to methylated non-imprint control regions which loses methylation in *Zfp57* knockout cells [7]. It might be speculated that ZFP57 is also involved in maintenance of methylation at non-imprinted genes in humans.

Conclusions

305 In summary, by genome-wide methylation analysis we have expanded the epimutational spectrum of TNDM1 associated with mutations in *ZFP57*. We detected large differences in the number and extent of affected regions in individual patients, but 61 regions were aberrantly methylated in two or more patients. We found one novel region within *PPP1R13L* which was hypomethylated in all TNDM1 patients included in this study. Expansion of the patient cohort will reveal if

310 hypomethylation of this locus is a common feature of TNDM1 and functional studies of the aberrantly methylated regions identified in this study might provide further insight into the etiology of the disease.

Declarations

315

Acknowledgements

Patients and their families are thanked for their participation. Hanne Mølgaard is gratefully acknowledged for technical assistance. Wilhelm Johannsen Centre for Functional Genome Research was established by the Danish National Research Foundation. This work was supported by Diabetes
320 UK, the Danish Agency for Science, Technology and Innovation and the University of Copenhagen. Furthermore, financial support was granted by the “Director Jacob Madsen and wife Olga Madsens Foundation” and “King Christian the 10th Foundation”.

Ethics

325 Patients were recruited as part of a Danish Imprinting and Methylation study and as part of the UK clinical local research network study: “Imprinting Disorders finding out why”. The Danish study was approved by The National Committee on Health Research Ethics (H-D-2008-079) and Danish Data Protection Agency (2008-41-2565). The UK study was approved by Southampton and South West Hampshire Research Ethics committee (07/H0502/85). Written consent was obtained from
330 patients and guardians for participation in research.

Consent to publish

Written informed consent was obtained from families for presentation of clinical data and for publication of results. Written consent was given by patients and guardians for the publication of
335 patient and family numbers in anonymised form.

Competing interests

The authors declare no conflict of interests.

340 **Author's contributions**

MB, SEB, ZT and NT conceived and designed the experiment. MB, SEB and CD performed the experiments. MB and SEB analyzed the data. MB drafted the manuscript. SEB, DJGM and IKT collected the biological samples and performed the clinical investigations of the patients. MB, SEB, CD, JMDH, DJGM, ZT, KG, IKT, PG and NT participated in discussions related to the analysis and
345 interpretation of the data. All authors have read and approved the final version of the manuscript.

Availability of data and materials

Data from this study that do not pertain to individual patients are freely available and can be obtained by contacting the corresponding author. We do not have the consent of patients to publish
350 sequencing data in a format required by SRA (<http://www.ncbi.nlm.nih.gov/sra>), ENA (<http://www.ebi.ac.uk/ena>) or EGA (<https://www.ebi.ac.uk/ega/>). Sequencing data will be shared as aligned reads without sequence information. Files with alignments to hg18, hg19 and hg38 are available.

355 **References**

1. Biliya S, Bulla LA: **Genomic imprinting: the influence of differential methylation in the two sexes.** *Exp Biol Med (Maywood)* 2010; **235**:139–47.
2. Delaval K, Feil R: **Epigenetic regulation of mammalian genomic imprinting.** *Curr Opin Genet Dev* 2004; **14**:188–95.
- 360 3. Mackay DJG, Temple IK: **Transient neonatal diabetes mellitus type 1.** *Am J Med Genet C Semin Med Genet* 2010; **154C**:335–42.
4. Mackay DJG, Callaway JLA, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth H V, Goodship JA, Haemers AP, Hahnemann JMD, Kordonouri O, Masoud AF, Oestergaard E, Storr J, Ellard S, Hattersley AT, Robinson DO, Temple IK: **Hypomethylation of multiple**
- 365 **imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57.** *Nat Genet* 2008; **40**:949–51.
5. Boonen SE, Mackay DJG, Hahnemann JMD, Docherty L, Grønskov K, Lehmann A, Larsen LG, Haemers AP, Kockaerts Y, Dooms L, Vu DC, Ngoc CTB, Nguyen PB, Kordonouri O, Sundberg F, Dayanikli P, Puthi V, Acerini C, Massoud AF, Tümer Z, Temple IK: **Transient neonatal diabetes,**
- 370 **ZFP57, and hypomethylation of multiple imprinted loci: a detailed follow-up.** *Diabetes Care* 2013; **36**:505–12.
6. Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC: **A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints.** *Dev Cell* 2008; **15**:547–57.
7. Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone P
- 375 V., Grimaldi G, Riccio A, Trono D: **In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions.** *Mol Cell* 2011; **44**:361–72.
8. Baglivo I, Esposito S, De Cesare L, Sparago A, Anvar Z, Riso V, Cammisa M, Fattorusso R, Grimaldi G, Riccio A, Pedone P V: **Genetic and epigenetic mutations affect the DNA binding**
- 380 **capability of human ZFP57 in transient neonatal diabetes type 1.** *FEBS Lett* 2013; **587**:1474–81.
9. Docherty LE, Rezwan FI, Poole RL, Jagoe H, Lake H, Lockett G a, Arshad H, Wilson DI, Holloway JW, Temple IK, Mackay DJG: **Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel**
- 385 **candidate imprinted genes.** *J Med Genet* 2014; **51**:229–38.
10. Court F, Martin-Trujillo A, Romanelli V, Garin I, Iglesias-Platas I, Salafsky I, Guitart M, Perez de Nanclares G, Lapunzina P, Monk D: **Genome-wide allelic methylation analysis reveals disease-specific susceptibility to multiple methylation defects in imprinting syndromes.** *Hum Mutat* 2013; **34**:595–602.
- 390 11. Martin-Subero JI, Bibikova M, Mackay D, Wickham-Garcia E, Sellami N, Richter J, Santer R, Caliebe A, Fan J-B, Temple IK, Siebert R: **Microarray-based DNA methylation analysis of imprinted loci in a patient with transient neonatal diabetes mellitus.** *Am J Med Genet A* 2008; **146A**:3227–9.

- 395 12. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009; **25**:1754–60.
13. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008; **9**:R137.
- 400 14. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: A Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2009; **26**:139–140.
15. Lan X, Adams C, Landers M, Dudas M, Krissinger D, Marnellos G, Bonneville R, Xu M, Wang J, Huang THM, Meredith G, Jin VX: **High resolution detection and analysis of CpG dinucleotides methylation using MBD-seq technology.** *PLoS One* 2011; **6**.
- 405 16. Krueger F, Andrews SR: **Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications.** *Bioinformatics* 2011; **27**:1571–1572.
17. Serre D, Lee BH, Ting AH: **MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome.** *Nucleic Acids Res* 2010; **38**:391–9.
- 410 18. Bergamaschi D, Samuels Y, O’Neil NJ, Trigiante G, Crook T, Hsieh J-K, O’Connor DJ, Zhong S, Campargue I, Tomlinson ML, Kuwabara PE, Lu X: **iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human.** *Nat Genet* 2003; **33**:162–167.
19. Zhang X, Wang M, Zhou C, Chen S, Wang J: **The expression of iASPP in acute leukemias.** *Leuk Res* 2005; **29**:179–183.
- 415 20. Jiang L, Siu MKY, Wong OGW, Tam KF, Lu X, Lam EWF, Ngan HYS, Le XF, Wong ESY, Monteiro LJ, Chan HY, Cheung ANY: **iASPP and chemoresistance in ovarian cancers: Effects on paclitaxel-mediated mitotic catastrophe.** *Clin Cancer Res* 2011; **17**:6924–6933.
21. Liu Z, Zhang X, Huang D, Liu Y, Zhang X, Liu L, Li G, Dai Y, Tan H, Xiao J, Tian Y: **Elevated expression of iASPP in head and neck squamous cell carcinoma and its clinical significance.** *Med Oncol* 2012; **29**:3381–3388.
- 420 22. Petryszak R, Burdett T, Fiorelli B, Fonseca N a., Gonzalez-Porta M, Hastings E, Huber W, Jupp S, Keays M, Kryvych N, McMurry J, Marioni JC, Malone J, Megy K, Rustici G, Tang AY, Taubert J, Williams E, Mannion O, Parkinson HE, Brazma A: **Expression Atlas update - A database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments.** *Nucleic Acids Res* 2014; **42**(December 2013):926–932.
- 425 23. Herron BJ, Rao C, Liu S, Laprade L, Richardson J a., Oliveri E, Semsarian C, Millar SE, Stubbs L, Beier DR: **A mutation in NFκB interacting protein 1 results in cardiomyopathy and abnormal skin development in wa3 mice.** *Hum Mol Genet* 2005; **14**:667–677.
- 430 24. Elliott G, Hong C, Xing X, Zhou X, Li D, Coarfa C, Bell RJ a., Maire CL, Ligon KL, Sigaroudinia M, Gascard P, Tlsty TD, Harris RA, Schalkwyk LC, Bilenky M, Mill J, Farnham PJ, Kellis M, Marra M a., Milosavljevic A, Hirst M, Stormo GD, Wang T, Costello JF: **Intermediate DNA methylation is a conserved signature of genome regulation.** *Nat Commun* 2015; **6**:6363.

25. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP: **Orphan CpG islands identify numerous conserved promoters in the mammalian genome.** *PLoS Genet* 2010; **6**:e1001134.
- 435 26. Jones P a: **Functions of DNA methylation: islands, start sites, gene bodies and beyond.** *Nat Rev Genet* 2012; **13**:484–92.
27. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, Garvin T, Kessler M, Zhou J, Smith AD: **A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics.** *PLoS One* 2013; **8**.
- 440 28. Court F, Tayama C, Romanelli V, Martin Trujillo A, Iglesias-Platas I, Okamura K, Sugahara N, Simón C, Moore H, Harness J V, Keirstead H, Vicente Sanchez-Mut J, Kaneki E, Lapunzina P, Soejima H, Wake N, Esteller M, Ogata T, Hata K, Nakabayashi K, Monk D: **Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of the human imprintome and suggests a germline methylation independent establishment of imprinting.** *Genome Res*
445 2014:554–569.
29. Okae H, Chiba H, Hiura H, Hamada H, Sato A: **Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development.** *PLoS Genet* 2014; **10**:1–12.
30. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A: **Widespread monoallelic expression on human autosomes.** *Science* 2007; **318**:1136–40.
- 450 31. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty V V, Schupf N, Vilain E, Morris M, Haghghi F, Tycko B: **Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation.** *Nat Genet* 2008; **40**:904–8.
32. Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, Plomin R, Mill J: **Allelic skewing of DNA methylation is widespread across the genome.** *Am J Hum Genet* 2010;
455 **86**:196–212.
33. Zhang Y, Rohde C, Reinhardt R, Voelcker-Rehage C, Jeltsch A: **Non-imprinted allele-specific DNA methylation on human autosomes.** *Genome Biol* 2009; **10**:R138.
34. Kacem S, Feil R: **Chromatin mechanisms in genomic imprinting.** *Mamm Genome* 2009; **20**:544–556.
- 460 35. Choufani S, Shapiro JS, Susiarjo M, Butcher DT, Grafodatskaya D, Lou Y, Ferreira JC, Pinto D, Scherer SW, Shaffer LG, Coullin P, Caniggia I, Beyene J, Slim R, Bartolomei MS, Weksberg R: **A novel approach identifies new differentially methylated regions (DMRs) associated with imprinted genes.** *Genome Res* 2011; **21**:465–76.
- 465 36. Metsalu T, Viltrop T, Tiirats A, Rajashekar B, Reimann E, Kõks S, Rull K, Milani L, Acharya G, Basnet P, Vilo J, Mägi R, Metspalu A, Peters M, Haller-Kikkatalo K, Salumets A: **Using RNA sequencing for identifying gene imprinting and random monoallelic expression in human placenta.** *Epigenetics* 2014; **9**:1397–409.
37. Stelzer Y, Ronen D, Bock C, Boyle P, Meissner A, Benvenisty N: **Identification of novel imprinted differentially methylated regions by global analysis of human-parthenogenetic-induced pluripotent stem cells.** *Stem cell reports* 2013; **1**:79–89.
- 470

38. Barbaux S, Gascoin-Lachambre G, Buffat C, Monnier P, Mondon F, Tonanny MB, Pinard A, Auer J, Bessières B, Barlier A, Jacques S, Simeoni U, Dandolo L, Letourneur F, Jammes H, Vaiman D: **A genome-wide approach reveals novel imprinted genes expressed in the human placenta.** *Epigenetics* 2012; 7(February 2015):1079–1090.
- 475 39. Yuen RK, Jiang R, Peñaherrera MS, McFadden DE, Robinson WP: **Genome-wide mapping of imprinted differentially methylated regions by DNA methylation profiling of human placentas from triploidies.** *Epigenetics Chromatin* 2011; 4:10.
40. Baran Y, Subramaniam M, Biton A, Tukiainen T, Tsang EK, Rivas MA, Pirinen M, Gutierrez-Arcelus M, Smith KS, Kukurba KR, Zhang R, Eng C, Torgerson DG, Urbanek C, Li JB, Rodriguez-Santana JR, Burchard EG, Seibold MA, MacArthur DG, Montgomery SB, Zaitlen NA, Lappalainen T: **The landscape of genomic imprinting across diverse adult human tissues.** *Genome Res* 2015:gr.192278.115–.
- 480 41. Beygo J, Ammerpohl O, Gritzan D, Heitmann M, Rademacher K, Richter J, Caliebe A, Siebert R, Horsthemke B, Buiting K: **Deep bisulfite sequencing of aberrantly methylated Loci in a patient with multiple methylation defects.** *PLoS One* 2013; 8:e76953.
- 485 42. Caliebe A, Richter J, Ammerpohl O, Kanber D, Beygo J, Bens S, Haake A, Jüttner E, Korn B, Mackay DJG, Martin-Subero JI, Nagel I, Sebire NJ, Seidmann L, Vater I, von Kaisenberg CS, Temple IK, Horsthemke B, Buiting K, Siebert R: **A familial disorder of altered DNA-methylation.** *J Med Genet* 2014; 51:407–12.

Legends to figures

Figure 1. MBD-seq results of previously identified hypomethylated DMRs. Bars show methylation levels relative to controls. Error bars show 1X standard deviation. ^a: Letters T, P and N below bars indicate methylation results from a previous studies [4, 5]. T: Total loss of methylation; P: Partial loss of methylation; N: Normal methylation. Stars indicate level of significance of MBD-seq results: *:P(adjusted) < 0.05; **: P(adjusted) < 0.005; ***: P(adjusted) < 0.0005

Figure 2. Verification of MBD-seq results by bisulfite sequencing. The plot shows log₂ ratios of methylation levels in patients and *ZFP57* heterozygotes versus controls obtained by MBD-seq and BS-seq.

Figure 3. Venn diagram of overlapping aberrantly methylated regions in the patients. Up-arrow: hypermethylated regions. Down-arrow: hypomethylated regions. The three hypomethylated regions in common of the four patients are *PLAGL1*:alt-TSS-DMR, *GRB10*:alt-TSS-DMR and *PPP1R13L*.

Figure 4. Hypomethylation of *PPP1R13L* in the *TNDM1* patients. UCSC Genome Browser screenshot with bedgraph tracks showing MBD-seq reads from the patients (*TNDM1*) and controls. All patients show hypomethylation of the CpG island indicated by red dots, whereas no changes in methylation is detected elsewhere in the gene. Genomic positions of *ZFP57* binding motifs (TGCCGC) are indicated below the CpG islands.

Figure 5. Enrichment of the *ZFP57* binding sequence in the hypomethylated regions. The number of occurrences of the TGCCGC sequence were counted in the 52 hypomethylated regions shared between at least two patients (red bars) and in the 17,123 methylated regions detected in any of the samples (blue bars).

Supplementary tables

515 **Supplementary_tables_1-7.xls:** Hypo- and hypermethylated regions detected in patients and heterozygotes.

Supplementary_table_8.xls: Hypo- and hypermethylated regions detected in at least 2 patients.

Additionally, the following characteristics of the regions are included in the table: known imprinting status, CpG island, ZFP57 motifs, microarray probes, bisulphite-sequencing results, allele-specific

520 methylation, intermediate methylation and H3K4me2/me3.

Figure 1

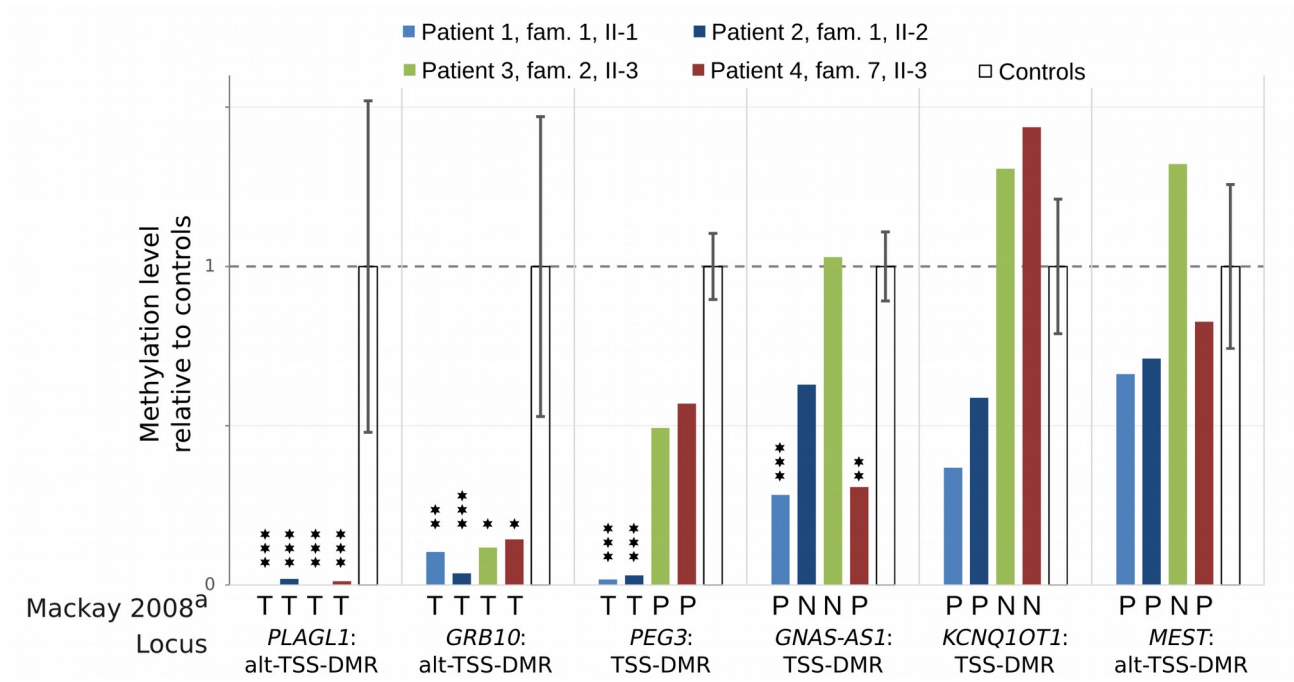


Figure 2

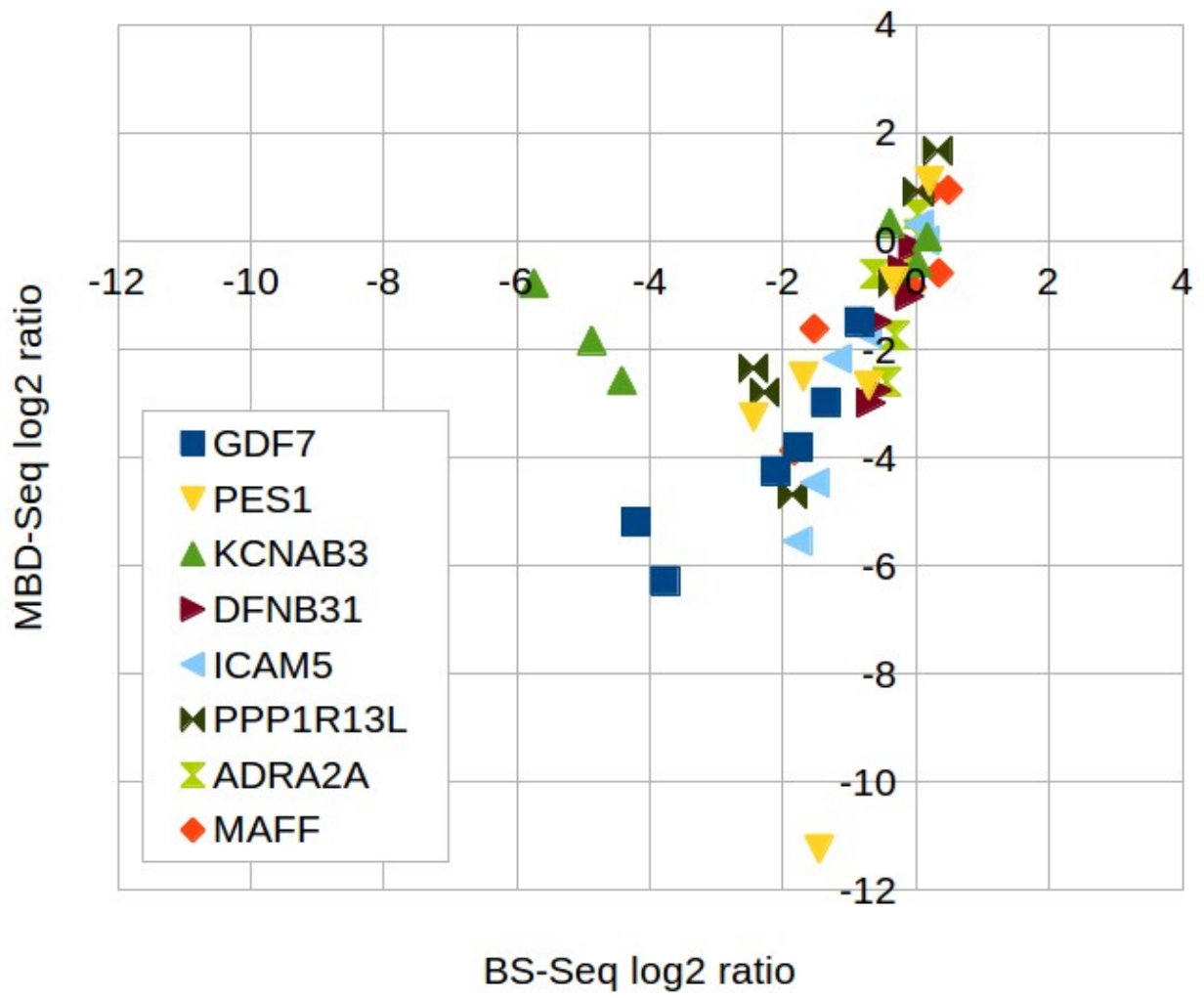


Figure 3

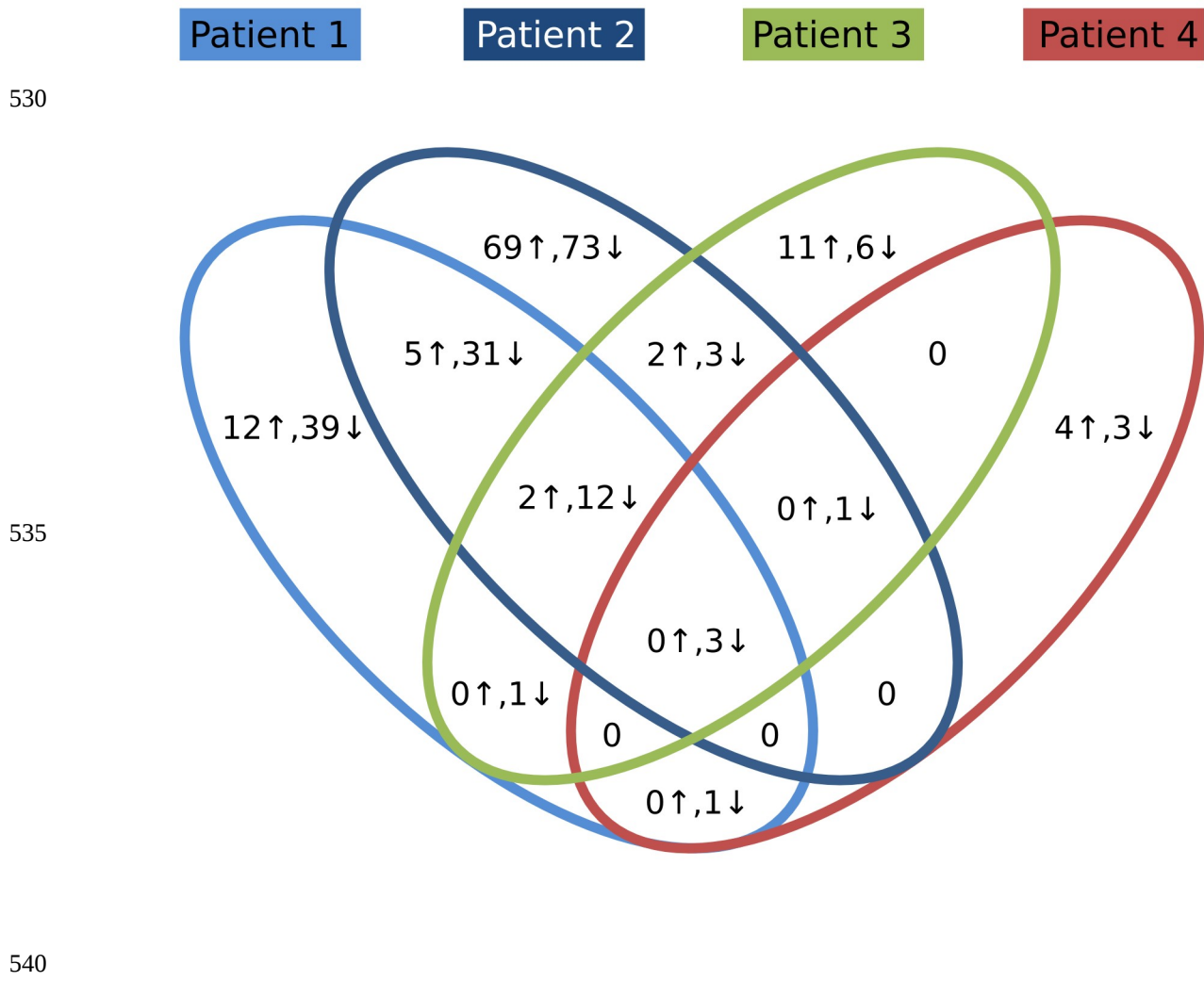


Figure 4

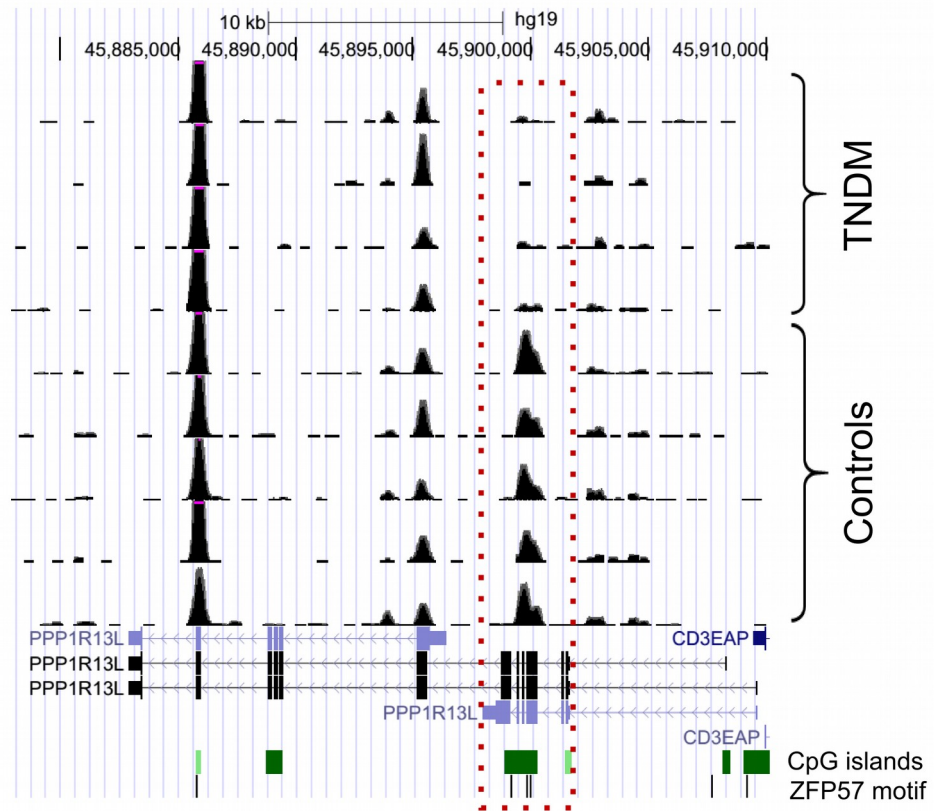


Figure 5

