

Whole Exome Sequencing Identifies Novel Recurrently Mutated Genes in Patients with Splenic Marginal Zone Lymphoma

Marina Parry¹✉, Matthew J. J. Rose-Zerilli¹✉, Jane Gibson²✉, Sarah Ennis², Renata Walewska³, Jade Forster¹, Helen Parker¹, Zadie Davis³, Anne Gardiner³, Andrew Collins², David G. Oscier^{1,3}, Jonathan C. Strefford^{1*}

1 Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom, **2** Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom, **3** Department of Pathology, Royal Bournemouth Hospital, Bournemouth, United Kingdom

Abstract

The pathogenesis of splenic marginal zone lymphoma (SMZL) remains largely unknown. Recent high-throughput sequencing studies have identified recurrent mutations in key pathways, most notably *NOTCH2* mutations in >25% of patients. These studies are based on small, heterogeneous discovery cohorts, and therefore only captured a fraction of the lesions present in the SMZL genome. To identify further novel pathogenic mutations within related biochemical pathways, we applied whole exome sequencing (WES) and copy number (CN) analysis to a biologically and clinically homogeneous cohort of seven SMZL patients with 7q abnormalities and *IGHV1-2*04* gene usage. We identified 173 somatic non-silent variants, affecting 160 distinct genes. In addition to providing independent validation of the presence of mutation in several previously reported genes (*NOTCH2*, *TNFAIP3*, *MAP3K14*, *MLL2* and *SPEN*), our study defined eight additional recurrently mutated genes in SMZL; these genes are *CREBBP*, *CBFA2T3*, *AMOTL1*, *FAT4*, *FBXO11*, *PLA2G4D*, *TRRAP* and *USH2A*. By integrating our WES and CN data we identified three mutated putative candidate genes targeted by 7q deletions (*CUL1*, *EZH2* and *FLNC*), with *FLNC* positioned within the well-characterized 7q minimally deleted region. Taken together, this work expands the reported directory of recurrently mutated cancer genes in this disease, thereby expanding our understanding of SMZL pathogenesis. Ultimately, this work will help to establish a stratified approach to care including the possibility of targeted therapy.

Citation: Parry M, Rose-Zerilli MJJ, Gibson J, Ennis S, Walewska R, et al. (2013) Whole Exome Sequencing Identifies Novel Recurrently Mutated Genes in Patients with Splenic Marginal Zone Lymphoma. PLoS ONE 8(12): e83244. doi:10.1371/journal.pone.0083244

Editor: Jose Angel Martinez Climent, University of Navarra, Spain

Received: July 9, 2013; **Accepted:** November 1, 2013; **Published:** December 13, 2013

Copyright: © 2013 Parry et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The manuscript was supported by: Leukaemia and Lymphoma Research, Cancer Research UK, Kay Kendall Leukaemia Fund, Wessex Medical Research, Bournemouth Leukaemia Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: JCS@soton.ac.uk

✉ These authors contributed equally to this work.

Introduction

Splenic Marginal Zone Lymphoma (SMZL) is a low grade chronic B cell lymphoproliferative disorder that predominantly affects elderly patients and involves the spleen, bone marrow, and peripheral blood [1]. Although the median survival is around 10 years, approximately 70% of SMZL patients require treatment, of whom 25% experience progressive disease, leading to early death [1].

Our understanding of the molecular pathogenesis of SMZL remains limited. Early cytogenetic studies identified recurrent deletions of 7q31-q32 and duplications of 3q in approx. 30% and 20% of cases, respectively [2], but subsequent molecular investigations have failed to identify causative genes within

these regions [3]. Candidate gene studies are limited to mutations in *TP53*, which is disrupted in 10-15% of cases [2], and to genes within the NF-KB pathway, which are mutated in a third of all cases [4,5]. The presence of a highly restricted immunoglobulin gene repertoire, in particular the selective usage of the immunoglobulin heavy chain variable (*IGHV*) 1-2*04 allele in 20-30% of patients, suggests that antigenic stimulation may be important in the pathogenesis of this disease [6].

The recent application of whole exome sequencing to frozen splenic tissue from 14 patients with SMZL followed by targeted resequencing of recurrent variants in larger cohorts has identified further biologically relevant genes [7,8]. Mutations in *NOTCH2*, which eliminate the C-terminal PEST domain and

result in compromised protein degradation, were identified in 20 - 25% of cases although there was no consensus as to the clinical significance of these mutations between studies [7,8]. Gene mutations in modulators or other members of the Notch signalling pathway and in other pathways, such as chromatin remodelling and transcriptional regulation were also implicated [8].

In view of the relatively small number of patients investigated so far and the biological heterogeneity of SMZL, it is vital to perform additional gene discovery experiments to fully catalogue the molecular lesions that contribute to disease pathogenesis. To this aim, we performed whole exome sequencing and copy number analysis of tumour and germ-line DNA extracted from a clinically homogeneous cohort of SMZL patients. In doing so, we expand the reported directory of recurrently mutated cancer genes in this disease, thereby expanding our understanding of SMZL pathogenesis that will ultimately facilitate improvements in disease management and the promise of novel therapies.

Materials and Methods

Patients and biomarker analysis

Seven patients were included in this current study, all met established diagnostic criteria [1], and 5/7 underwent a splenectomy with histology typical of SMZL in each case and no evidence of transformation to a high-grade lymphoma. Each patient harboured chromosomal aberrations targeting 7q and *IGHV1-2*04* usage (Table S1), ensuring the exclusion of other types of splenic lymphoma from our analysis and maximizing the likelihood of identifying pathogenic mutations within related biochemical pathways. Informed patient consent was obtained according to the declaration of Helsinki, and the study was ethically approved by the local REC.

Chromosomal analysis was performed and described according to the International System for Human Cytogenetic Nomenclature [9]. Immunoglobulin variable region genes were sequenced from either cDNA or gDNA as previously described [6]. cDNA was synthesised by reverse transcription according to the manufacturers protocol (Promega). gDNA was extracted using the Qiagen Blood Mini Kit and amplified using the BIOMED 2 protocol [10]. PCR products were sequenced directly using an ABI 310 genetic analyser and sequences were aligned to the IMGT-V-Quest database.

High-throughput sequencing, variant calling and Sanger validation

Using targeted exome capture (SureSelect Human All Exon 51Mb V4, 50Mb V3, Agilent) we prepared sequencing libraries from high-molecular weight genomic DNA from CD19 positive-purified tumour cells (five cases extracted from the spleen and two from peripheral blood) and matched saliva cells (Oragene DNA kit, DNA Genotek) prior to high-throughput sequencing with the Illumina HiSeq system. The paired-end sequencing data were aligned against the human genome reference sequence (hg19/GRCh37) using the Novoalign software (novoalignMPI V2.08.02, Novocraft Technologies, Selangor, Malaysia). Duplicate reads, resulting from PCR clonality or

optical duplicates, and reads mapping to multiple locations were excluded from downstream analysis. Depth and breadth of sequence coverage was calculated with custom scripts and the BedTools package (v2.13.2) [11] and is included in table S2.

Germ-line-Tumour paired datasets were analysed to identify single nucleotide variations (SNVs) and small insertion and deletions using VarScan 2.3.3 [12] (<http://varscan.sourceforge.net>). The minimum variant allele frequency threshold was set to 10% with a minimum read depth of 4. Variants were filtered using the 'somaticFilter' command to remove clusters of false positives and SNV calls near indels with the same frequency and depth thresholds.

Variants were annotated with respect to genes and transcripts and filtered using the Annovar software tool (v2012Jun21) [13]. Variants were cross referenced with databases of known variation were downloaded from the Annovar website (June 2012); data from the 1000 Genomes Project (2012 April release)[14], dbSNP135 (and a version with SNPs flagged as rare <1% frequency or clinically associated by NCBI) and data from 4300 European American samples from The National Heart Lung and Blood Institute Exome Sequencing Project Exome Variant Server (<http://evs.gs.washington.edu/EVS/>), (ESP6500 release). Using conventional Sanger sequencing, we confirmed the presence of 38/45 somatic variants (84.4%) and those non-concordant cases were due to low exome read-depth in the tumour sample.

SNP6.0 array hybridization, data extraction and analysis

Tumour and germ-line DNA was purified, amplified, labelled and hybridized to the Affymetrix SNP6.0 platform (Affymetrix, Santa Clara, CA) as previously described [15]. For copy number analysis, two independent researchers visually inspected parallel copy number profiles (aligned to hg19/GRCh37) from tumour and germ-line samples using Partek Genomics Suite (Partek Inc, Missouri, USA), and lesions were classified as somatic if they were present and absent in the tumour and germ-line material, respectively. Copy number alterations (CNAs) were defined as a deviation of 50 consecutive array features (probes) from a normal value of 2 (± 0.3), within a consecutive genomic window of 50 Kilobases. The allele ratio was calculated for each sample using the HapMap Allele Reference baseline (Affymetrix) and copy number neutral loss of heterozygosity (CNNLOH) event were defined as somatic if they were present and absent in the tumour and germ-line material, respectively.

Results and Discussion

Exome-capture and high-throughput sequencing allowed us to align approx. 41.9 million reads per sample at a mean depth of 69x (range, 43-109x). In total, an average of 82.2% (range, 70-95%) of target sequences captured at 20x. Our analytical pipeline identified 176 somatic non-silent variants, affecting 165 distinct genes (Table S2). These variants were base-pair transitions (34%), transversions (28%), insertions (6%) and

Table 2. Summary of the pathways in which mutated genes in our SMZL cohort can be found and their predicted functional consequences.

DAVID Pathway	Genes	Accession numbers	Variant nomenclature		SIFT score prediction	Polyphen-2 score prediction	Case no ² .													
			Nucleotide change	Amino acid change			consequences	consequences	1	2	3	4	5	6	7					
MAP kinase	<i>CACNA1E</i>	NM_001205293	c.G1069C	p.E357Q	Damaging	Damaging													✓	
	<i>CACNA1H</i>	NM_021098	c.391delG	p.E131fs	Truncating	Truncating													✓	
	<i>CACNA2D2</i>	NM_001174051	c.2837delC	p.P946fs	Truncating	Truncating													✓	
	<i>FLNC</i>	NM_001458	c.C3179T	p.P1060L	Damaging	Probably damaging														✓
	<i>MAP3K14</i>	NM_003954	c.C200G	p. A67G	Truncating	Truncating														✓
	<i>MAPK8IP3</i>	NM_001040439	c.743delA	p.Q248fs	Truncating	Truncating														✓
	<i>RASA1</i>	NM_002890	c.C142A	p. P48T	Damaging	Truncating														✓
	<i>TAOK3</i>	NM_016281	c.438-7-T)	Splicing ¹	Truncating	Truncating													✓	
Notch	<i>NOTCH2</i>	NM_024408	c.C7081T	p.Q2361X	Truncating	Truncating													✓	
			c.6836delA	p.H2279fs	Truncating	Truncating													✓	
		<i>PIWIL3</i>	NM_001008496	c.2242delA	p.T748fs	Truncating	Truncating												✓	
		<i>NOTCH4</i>	NM_004557	c.C5877G	p.C1959W	Truncating	Damaging													✓
		<i>MAML3</i>	NM_018717	c.1513_1514del	p. 505_505del	Truncating	Truncating													✓
Cell cycle	<i>CUL1</i>	NM_003592	c.T469G	p.Y157D	Damaging	Probably damaging													✓	
	<i>CREBBP</i>	NM_001079846	c.A4349G	p.Y1450C	Damaging	Truncating													✓	
	<i>CDC27</i>	NM_001114091	c.A701C	p. Y234S	Tolerated	Benign													✓	
Cytokine-receptor interaction	<i>FLT1</i>	NM_002019	c.2594_splice	splicing	Truncating	Truncating													✓	
	<i>CRLF2</i>	NM_022148	c.G340C	p.V114L	Tolerated	Probably damaging													✓	

¹ The *TRRAP* mutation in case 5 occurred within a splice-site and is predicted to result in aberrant splicing

² Showed the presence (✓) and absence (white box) of each mutation in the patients in our series

doi: 10.1371/journal.pone.0083244.t002

somatic non-synonymous mutations in the majority of our cases (5/7, 71%). Whilst a biological role of these genes in SMZL required functional confirmation, our data does suggest that the MAPK signalling pathway is a major target for somatic mutations in this sub-group of SMZL.

Finally, we identified somatically acquired mutations in genes also targeted by 7q deletions in our patients. In doing so, we found *CUL1*, *FLNC* and *EZH2* mutations in individual cases (Table 2). Of these gene mutations, only *FLNC* was located within the published 7q MDR [3,20]. *FLNC* mutations have not been previously identified in a series eight del(7q) cases [20], suggesting that the prevalence of *FLNC* mutation is low in this sub-type of SMZL. However, further research will be required to establish if rare mutations represent only one mechanism of gene deregulation, as repression of *FLNC* transcription by promoter methylation is prevalent in several other human cancer types [21-24]. The somatic variant we identified in *EZH2* (p.K199N), which is located outside the SET protein domain, is not the activating mutation prevalent in follicular lymphoma and DLBCL [25] nor has it been previously reported in AML or MDS [26].

Herein, for the first time, we report the analysis of a homogeneous cohort of SMZL cases using whole exome sequencing and copy number analysis. In doing so, we validate the presence of recurrent mutations in several genes with established importance in SMZL. Furthermore, we expand the

reported directory of recurrently mutation cancer genes in this disease, with the most significant observation being the identification of recurrent mutations in *CREBBP* and *CBFA2T3*. The importance of *CREBBP* is further strengthened by the presence of a single SMZL case in the literature with a small deletion that juxtaposes 16 exons of *CREBBP* with the *ZNF434* gene, resulting in loss of the acetyltransferase domain of the *CREBBP* protein [8]. Furthermore, we show the majority of cases in our series carried mutations within MAPK signalling genes, suggesting that mutations in these genes are strongly associated with 7q-rearranged SMZL with *IGHV1-2*04* usage. Whilst our analysis identifies a series of novel genes mutated in SMZL, a larger study is required to determine the frequency of these events and any utility in the risk-adapted stratification of SMZL patients. To this aim, we are currently coordinating a pan-European study into the presence of somatic mutations in approx. 750 genes with a known or postulated role in SMZL pathophysiology in a cohort of more than 300 SMZL cases. This will ultimately establish the frequency and clinical importance of gene mutations in SMZL and help to establish a stratified approach to care including the possibility of targeted therapy.

Supporting Information

Table S1. Clinical characteristics of each patient included in the study.
(DOCX)

Table S2. Shows the 176 non-silent somatic mutations identified in our cases.

References

- Matutes E, Oscier D, Montalban C, Berger F, Callet-Bauchu E et al. (2008) Splenic marginal zone lymphoma proposals for a revision of diagnostic, staging and therapeutic criteria. *Leukemia* 22: 487-495. doi: 10.1038/sj.leu.2405068. PubMed: 18094718.
- Salido M, Baró C, Oscier D, Stamatopoulos K, Dierlamm J et al. (2010) Cytogenetic aberrations and their prognostic value in a series of 330 splenic marginal zone B-cell lymphomas: a multicenter study of the Splenic B-Cell Lymphoma Group. *Blood* 116: 1479-1488. doi:10.1182/blood-2010-02-267476. PubMed: 20479288.
- Watkins AJ, Hamoudi RA, Zeng N, Yan Q, Huang Y et al. (2012) An integrated genomic and expression analysis of 7q deletion in splenic marginal zone lymphoma. *PLOS ONE* 7: e44997. doi:10.1371/journal.pone.0044997. PubMed: 23028731.
- Rossi D, Deaglio S, Dominguez-Sola D, Rasi S, Vaisitti T et al. (2011) Alteration of BIRC3 and multiple other NF- κ B pathway genes in splenic marginal zone lymphoma. *Blood* 118: 4930-4934. doi:10.1182/blood-2011-06-359166. PubMed: 21881048.
- Novak U, Rinaldi A, Kwee I, Nandula SV, Rancoita PM et al. (2009) The NF- κ B negative regulator TNFAIP3 (A20) is inactivated by somatic mutations and genomic deletions in marginal zone lymphomas. *Blood* 113: 4918-4921. doi:10.1182/blood-2008-08-174110. PubMed: 19258598.
- Bikos V, Darzentas N, Hadzidimitriou A, Davis Z, Hockley S et al. (2012) Over 30% of patients with splenic marginal zone lymphoma express the same immunoglobulin heavy variable gene: ontogenetic implications. *Leukemia* 26: 1638-1646. doi:10.1038/leu.2012.3. PubMed: 22222599.
- Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigelin HG et al. (2012) Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. *J Exp Med* 209: 1553-1565. doi:10.1084/jem.20120910. PubMed: 22891276.
- Rossi D, Trifonov V, Fangazio M, Brusca G, Rasi S et al. (2012) The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *J Exp Med* 209: 1537-1551. doi:10.1084/jem.20120904. PubMed: 22891273.
- ISCN (2009) An International System for Human Cytogenetic Nomenclature: Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature Shaffer LG, Slovak ML, Campbell LJ, editors. Basel: S.Karger.
- Matthews C, Catherwood M, Morris TC, Alexander HD (2004) Routine analysis of IgVH mutational status in CLL patients using BIOMED-2 standardized primers and protocols. *Leuk Lymphoma* 45: 1899-1904. doi:10.1080/10428190410001710812. PubMed: 15223652.
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842. doi:10.1093/bioinformatics/btq033. PubMed: 20110278.
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD et al. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22: 568-576. doi: 10.1101/gr.129684.111. PubMed: 22300766.
- Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164. doi:10.1093/nar/gkq603. PubMed: 20601685.
- Consortium. GP, Durbin RM, Abecasis GR, Altshuler DL, Auton A et al. (2010) A map of human genome variation from population-scale sequencing. *Nature* 467: 1061-1073. doi:10.1038/nature09534. PubMed: 20981092. Available online at: doi:10.1038/nature09534 Available online at: PubMed: 20981092
- Parker H, Rose-Zerilli MJ, Parker A, Chaplin T, Chen X et al. (2011) 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia* 25: 489-497. doi:10.1038/leu.2010.288. PubMed: 21151023.
- Duan S, Cermak L, Pagan JK, Rossi M, Martinengo C et al. (2012) FBXO11 targets BCL6 for degradation and is inactivated in diffuse large B-cell lymphomas. *Nature* 481: 90-93. PubMed: 22113614.
- Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A et al. (2011) Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* 471: 189-195. doi:10.1038/nature09730. PubMed: 21390126.
- Salaverria I, Akasaka T, Gesk S, Szczepanowski M, Burkhardt B et al. (2012) The CBA2T3/ACSF3 locus is recurrently involved in IGH chromosomal translocation t(14;16)(q32;q24) in pediatric B-cell lymphoma with germinal center phenotype. *Genes Chromosomes Cancer* 51: 338-343. doi:10.1002/gcc.21919. PubMed: 22420028.
- Gruber TA, Larson Gedman A, Zhang J, Koss CS, Marada S, et al. (2012) An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell* 22: 683-697
- Fresquet V, Robles EF, Parker A, Martinez-Useros J, Mena M et al. (2012) High-throughput sequencing analysis of the chromosome 7q32 deletion reveals IRF5 as a potential tumour suppressor in splenic marginal-zone lymphoma. *Br J Haematol* 158: 712-726. doi:10.1111/j.1365-2141.2012.09226.x. PubMed: 22816737.
- Yi JM, Dhir M, Van Neste L, Downing SR, Jeschke J et al. (2011) Genomic and epigenomic integration identifies a prognostic signature in colon cancer. *Clin Cancer Res* 17: 1535-1545. doi: 10.1158/1078-0432.CCR-10-2509. PubMed: 21278247.
- Vanaja DK, Ehrlich M, Van den Boom D, Chevillet JC, Karnes RJ et al. (2009) Hypermethylation of genes for diagnosis and risk stratification of prostate cancer. *Cancer Invest* 27: 549-560. doi: 10.1080/07357900802620794. PubMed: 19229700.
- Kim JH, Jung EJ, Lee HS, Kim MA, Kim WH (2009) Comparative analysis of DNA methylation between primary and metastatic gastric carcinoma. *Oncol Rep* 21: 1251-1259. PubMed: 19360301.
- Imura M, Yamashita S, Cai LY, Furuta J, Wakabayashi M et al. (2006) Methylation and expression analysis of 15 genes and three normally-methylated genes in 13 Ovarian cancer cell lines. *Cancer Lett* 241: 213-220. doi:10.1016/j.canlet.2005.10.010. PubMed: 16303245.
- Morin RD, Johnson NA, Severson TM, Mungall AJ, An J et al. (2010) Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* 42: 181-185. doi:10.1038/ng.518. PubMed: 20081860.
- Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C et al. (2010) Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* 42: 722-726. doi:10.1038/ng.621. PubMed: 20601953.

(XLSX)

Author Contributions

Conceived and designed the experiments: JS DO. Performed the experiments: MP MRZ JF HP ZD AG. Analyzed the data: JG SE AC. Contributed reagents/materials/analysis tools: JG SE AC DO RW. Wrote the manuscript: MP DO JS.