COLD SPRING HARBOR
Molecular Case Studies

RESEARCH REPORT

Molecular etiology of an indolent lymphoproliferative disorder determined by whole-genome sequencing

Jeremy D.K. Parker,[1] Yaoqing Shen,[1] Erin Pleasance,[1] Yvonne Li,[1] Jacqueline E. Schein,[1] Yongjun Zhao,[1] Richard Moore,[1] Joanna Wegrzyn-Woltosz,[1] Kerry J. Savage,[2] Andrew P. Weng,[3] Randy D. Gascoyne,[2] Steven Jones,[1] Marco Marra,[1] Janessa Laskin,[4] and Aly Karsan[5]

[1]Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada; [2]Centre for Lymphoid Cancer and Department of Pathology, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada; [3]Terry Fox Laboratory and Department of Pathology, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada; [4]Department of Medical Oncology, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 4E6, Canada; [5]Genome Sciences Centre and Department of Pathology, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada

Abstract In an attempt to assess potential treatment options, whole-genome and transcriptome sequencing were performed on a patient with an unclassifiable small lymphoproliferative disorder. Variants from genome sequencing were prioritized using a combination of comparative variant distributions in a spectrum of lymphomas, and meta-analyses of gene expression profiling. In this patient, the molecular variants that we believe to be most relevant to the disease presentation most strongly resemble a diffuse large B-cell lymphoma (DLBCL), whereas the gene expression data are most consistent with a low-grade chronic lymphocytic leukemia (CLL). The variant of greatest interest was a predicted NOTCH2-truncating mutation, which has been recently reported in various lymphomas.

Corresponding author: akarsan@bcgsc.ca

Ontology terms: increase in B cell number; lymphocytosis

Published by Cold Spring Harbor Laboratory Press

doi: 10.1101/mcs.a000679

[Supplemental material is available for this article.]

INTRODUCTION

B-cell neoplasms are clonal expansions of cells and are roughly classified according to the developmental B-cell stage from which the clone arose. The clonal expansion may first be apparent simply as increased white cell counts (lymphocytosis). B-cell development requires stereotypical expression of cell surface receptors, which in turn activate or repress specific signaling pathways and lead to changes in gene expression. The correspondence between the cellular morphology, immunophenotype, and the genotype (inferred by various genetic and cytogenetic assays) can typically be used to diagnose patients and indicate therapeutic options. "Mature" B-cell neoplasms include mantle cell lymphoma (MCL), follicular lymphoma (FL), Burkitt lymphoma (BL), Hodgkin lymphoma (HL), diffuse large B-cell lymphoma (DLBCL), marginal zone lymphomas (MZLs) and MALT (mucosa-associated lymphoid tissue) lymphomas, chronic lymphocytic leukemia (CLL), and others—each are proposed to

correspond to a different normal counterpart (Swerdlow et al. 2008), and typically each of them has specific treatment options. In practice, patients may present with a form of disease that defies easy categorization, and thus treatment options are limited.

The use of whole-genome sequencing as an assay to inform clinical decisions is an active area of research. We have established a pipeline capable of providing a clinician-readable report of the affected genes and pathways in individual cancer patients, in a relevant time frame. Ideally, this type of analysis will guide therapeutic options, potentially by repositioning agents already approved for other indications or by resolving the classification of a borderline case. Here, we highlight the utility of a combined analysis that takes into account gene expression data from RNA sequencing (RNA-seq) to assist with the prioritization of variants across the genome for a patient with a lymphocytosis of uncertain etiology.

RESULTS

Clinical Presentation

A female patient, age 65 yr and with no family history, was found to have a node-negative colorectal tumor. Mismatch repair deficiency (MLH1- and PMS2-low) was identified by immunohistochemistry (IHC), and microsatellite instability was observed. The tumor was resected, and because of the low likelihood of benefit of adjuvant therapy, none was provided. During follow-up 14 mo after the colorectal diagnosis, she was noted to have lymphocytosis of 8.3×10^9/L, comprising mainly small cells, in the absence of any lymphadenopathy or hepatosplenomegaly. Lymphocytosis (4.9×10^9/L) was first documented 18 mo prior to its diagnosis, meaning that its development predated the colorectal diagnosis. Peripheral blood flow cytometry demonstrated a κ-clonal B-cell population that was positive for CD5 and FMC-7-bright, but was negative for CD10, CD23, CD38, CD11c, CD25, and CD103 (Supplemental Fig. S1). The morphology most closely resembled CLL, although the quality of the chromatin was unusual and the cells lacked expression of CD23. Fluorescence in situ hybridization (FISH) for the *IGH-CCND1* t(11;14) rearrangement was negative. FISH for ATM (11q), CEP12 (trisomy 12), D13S319 (13q), and TP53 (17p) were normal. The remainder of the blood counts was normal, and there was no paraproteinemia. A bone marrow biopsy demonstrated involvement of the marrow with small-to-intermediate-sized cells and some plasmacytoid differentiation comprising 60%–70% of marrow cellularity (Fig. 1). No growth centers were identified. Additional immunostaining of the bone marrow biopsy for Cyclin D1 and SOX11 was negative. The patient was diagnosed with a low-grade $CD5^+$ B-cell lymphoproliferative disorder with the possibilities including atypical CLL, or a low-grade B-cell

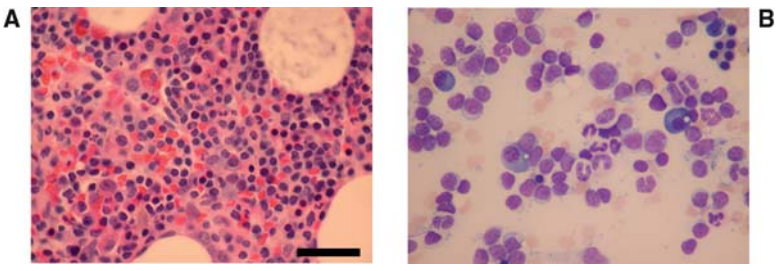

Figure 1. Photomicrographs demonstrating pathological features of the marrow and peripheral blood. (*A*) Bone marrow biopsy shows a clustered, atypical lymphoid infiltrate, with small-to-intermediate-sized cells and some plasmacytoid differentiation. (*B*) Peripheral blood smear shows small amounts of eccentric cytoplasm and confirms mild plasmacytoid differentiation. Scale bar, 50 μm.

lymphoproliferative disorder not otherwise specified, but a $CD5^+$ MZL with a leukemic presentation could not be excluded.

Genomic Analyses

In an attempt to refine the diagnosis and potentially inform the choice of therapeutics, the patient was enrolled in the Personal OncoGenomics (POG) initiative at the British Columbia Cancer Agency. POG aims to identify "actionable" targets of aberrant pathways in a clinically feasible timeline using deep sequencing (Laskin et al. 2015). After enrollment, a fresh bone marrow sample was flow-sorted for tumor cells using the viable $CD19^+/CD33^-/CD3^-$ B-cell fraction, with $CD19^-/CD3^+/CD33^-$ T cells used as the normal sample for germline comparison of variants (Supplemental Fig. S2). Genomic copy number in the B-cell population was inferred by reference to the T-cell population and, consistent with clinical FISH results, indicated a normal karyotype. There was also no evidence for either translocations or significant regions of copy-neutral loss of heterozygosity (LOH) from the genomic data, aside from LOH of the X Chromosome. RNA from the $CD19^+$ cells was also isolated and sequenced and expression levels were calculated using RSEM (RNA-Seq by Expectation Maximization; Mortazavi et al. 2008; Li and Dewey 2011).

Expression Analysis

The RNA-seq gene expression values from the $CD19^+$ B-cell population were compared with a normal B-cell library available at the Genome Sciences Centre. The most differentially expressed genes from this comparison were then compared with previously published microarray data consisting of multiple B-cell disorders (Supplemental Fig. S3; Fernàndez et al. 2010). This analysis suggested that the lymphocytosis resembled CLL more than MCL, but splenic marginal zone lymphoma (SMZL) could not be excluded—which is consistent with the final clinical diagnosis. CLL has recently been proposed to be divided into a low-risk, C1 subtype and a higher-risk, C2 subtype using RNA-seq expression analysis (Ferreira et al. 2014). Using the overall expression ranking (of the patient's RNA-seq compared with normal B cells) with Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005), we found that the gene expression in the patient resembles a C1-subtype CLL (Supplemental Fig. S4). This is consistent with the lack of symptomatic clinical progression or development of cytopenias, with a current observation of 2.5 yr, in this patient.

Variant Analysis

Variants identified from the whole-genome shotgun sequencing (WGS) by tumor/normal comparisons using paired calls were prioritized based on several filtering steps (Fig. 2; Table 1; details given in the Methods section). Of the expressed genes possessing variants, only three were present in the COSMIC database (Catalogue of Somatic Mutations in Cancer; Forbes et al. 2011) when filtered for lymphoid origin and population frequency of at least 1%: *APC*, *KLHL6*, and *NOTCH2*. We consider these to be the most relevant variants and discuss each in detail below.

A single-base insertion, expected to cause a frameshift, was detected in the region coding for the PEST domain of NOTCH2 (Supplemental Fig. S5A). Similar mutations have been reported in SMZL and other MZL, as well as at lower frequency in DLBCL. These mutations result in protein stabilization and constitutive activation of NOTCH2 (Trøen et al. 2008; Kiel et al. 2012; Rossi et al. 2012; Parry et al. 2015). The presentation of the disease in the case under consideration indicates a more primitive cell type of origin than seen in other B-cell malignancies with NOTCH2-activating mutations, indicating that other, cooperating mutations may be relevant in this case. Although mouse *Notch2* has only been shown to be essential in marginal zone B cells (Saito et al. 2003), the transcript is present in earlier B-cell

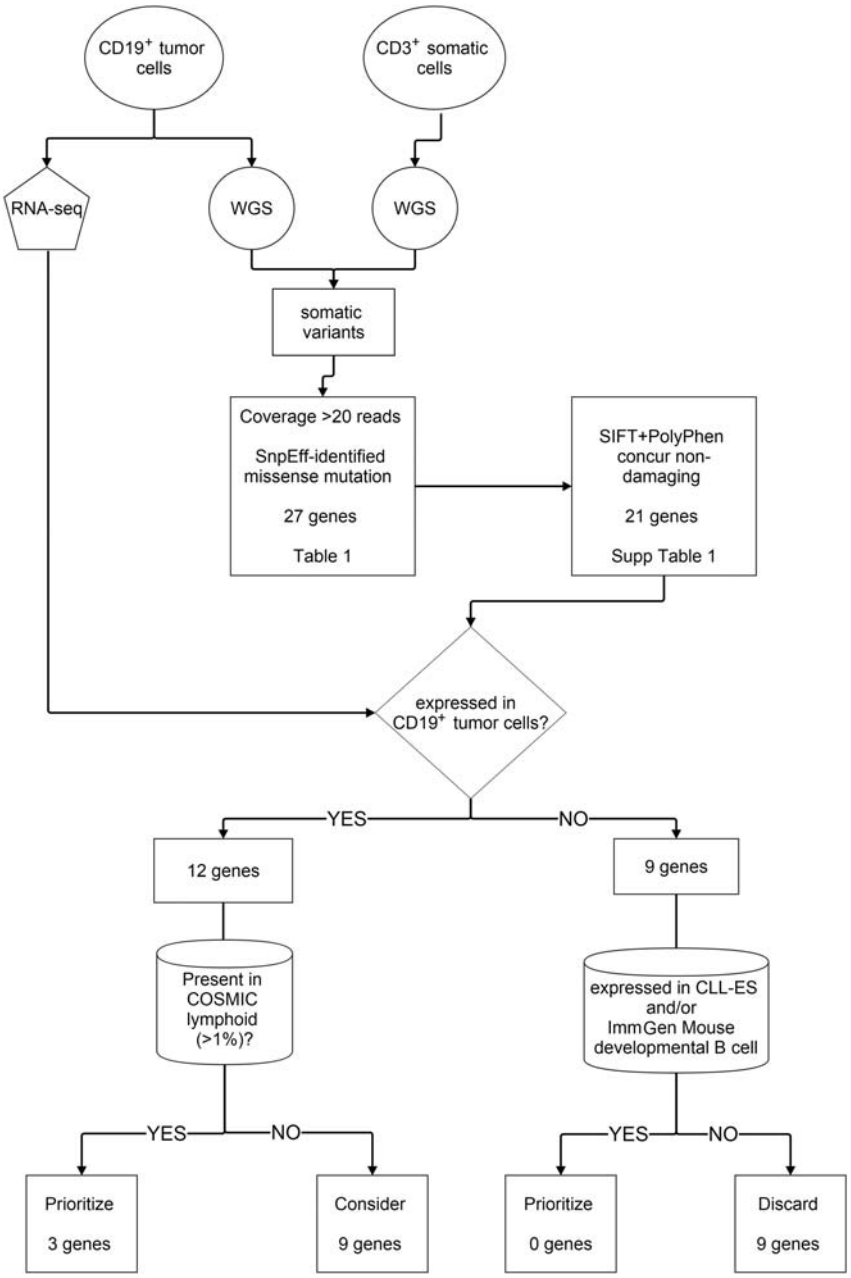

Figure 2. Overview of the filtering process used to prioritize variants. WGS, whole-genome shotgun sequencing; COSMIC, Catalogue of Somatic Mutations in Cancer; CLL-ES, chronic lymphocytic leukemia—Spain.

precursors (Saito et al. 2003; Heng and Painter 2008). An activating mutation could thus have a functional consequence in the precursors.

KLHL6 has previously been reported as a recurrently mutated gene in CLL (Puente et al. 2011, 2015; Ferreira et al. 2014), with an overall prevalence of ~2% coding mutations in CLL. The two *KLHL6* somatic missense mutations in this patient's B cells both occur in the amino-terminal POZ/BTB (Pox virus and zinc finger/bric-a-brac tramtrack broad complex) domain (Supplemental Fig. S5B) and occur in the same allele at approximately the same frequency

Table 1. List of genes found to have a somatic mutation in the patient with more than 20 reads supporting the variant and having a predicted moderate or high effect on coding sequence based on the SnpEff algorithm, as well as the predicted functional consequence on the protein as determined by a combination of both SIFT and PolyPhen

Gene name	Chr	DNA change	Change in representative protein model	Variant type	Alt	Ref	SIFT+ PolyPhen
AGBL4	1	50163019 A>G	NP_116174.3:p.Phe83Ser	SNV	54	63	Damaging
AGMO	7	15405188 C>T	NP_001004320.1:p.Arg405Glnln	SNV	43	61	Unknown
ALOX12	17	6903771 T>G	NP_000688.2:p.Asn308Lys	SNV	35	54	Benign
AMZ2	17	66247254 C>A	NP_001028741.1:p.Thr174Lys	SNV	27	58	Damaging
APC	5	112173449 A>G	NP_000029.2:p.Met720Val	SNV	45	50	Damaging
BAG1	9	33264387 C>G	NP_004314.5:p.Glu96Gln	SNV	38	59	Benign
CDH19	18	64211302 A>G	NP_066976.1:p.Tyr374His	SNV	44	44	Unknown
CDH2	18	25756984 C>T	NP_001783.2:p.Met1?	SL	52	48	Unknown
CNN2	19	1037681 G>A	NP_004359.1:p.Asp238Asn	SNV	45	51	Benign
CSMD3	8	113332130 C>T	NP_443132.3:p.Gly2312Ser	SNV	48	73	Damaging
DAZAP1	19	1434900 T>C	NP_061832.2:p.Tyr405His	SNV	47	54	Damaging
EYA1	8	72184078 C>T	NP_000494.2:p.Arg294Gln	SNV	39	67	Unknown
FCN1	9	137804876 C>T	NP_001994.2:p.Gly152Arg	SNV	21	59	Damaging
KLHL6	3	183273174 G>C	NP_569713.2:p.Phe83Leu	SNV	49	54	Damaging
KLHL6	3	183273195 A>G	NP_569713.2:p.Leu90Val	SNV	45	60	Damaging
NOTCH2	1	120458435 T>TG	NP_077719.2:p.Ile2304HisfsTer9	FS			
OR2B2	6	27879578 C>A	NP_149046.2:p.Val174Leu	SNV	44	42	Benign
PDGFRA	4	55138671 T>A	NP_006197.1:p.Cys450Ser	SNV	54	58	Damaging
RAPGEF1	9	134497376 T>C	NP_941372.1:p.Tyr572Cys	SNV	45	48	Damaging
RRBP1	20	17639820 C>T	XP_005260843.1:p.Ala445Thr	SNV	28	330	Benign
SAMD12	8	119593067 G>T	NP_997389.2:p.Gln27Lys	SNV	47	63	Damaging
SEC14L1	17	75186921 C>G	NP_002994.3:p.Pro34Ala	SNV	38	43	Damaging
SLC35A5	3	112289510 T>G	NP_060415.1:p.Leu118Arg	SNV	62	53	Damaging
SP4	7	21521702 C>T	NP_003103.2:p.Arg690Trp	SNV	42	67	Damaging
TAS2R16	7	122635631 T>C	NP_058641.1:p.Thr20Ala	SNV	65	57	Benign
TMED9	5	177019292 T>A	NP_059980.2:p.Leu26Gln	SNV	55	53	Unknown
UNC13A	19	17729273 G>T	NP_001073890.2:p.Pro1456Gln	SNV	29	45	Damaging
VMP1	17	57816223 TTC>T	NP_112200.2:p.Leu111PhefsTer40	FS			

Chr, Chromosome; Alt, coverage at the alternative allele; Ref, coverage of the reference allele; SNV, single-nucleotide variant; SL, translation start site lost; FS, frameshift.

(Supplemental Table S1). Mutations at these codons are observed at low frequencies (<0.1%) in the ExAC (Exome Aggregation Consortium) database, and are annotated in dbSNP (Database for Short Genetic Variations), indicating a low-level prevalence in the overall population. *KLHL6* was initially identified as a transcript up-regulated during germinal center B-cell differentiation (Gupta-Rossi et al. 2003). Whole-animal and B-cell-specific deletions of the mouse ortholog each show a block in B-cell differentiation between the immature (E) and mature (F) stages (Kroll et al. 2005). We hypothesize that the two BTB domain mutations in the B-cell compartment of the patient could contribute to a similar differentiation block that is relevant to the disease presentation (Fig. 3).

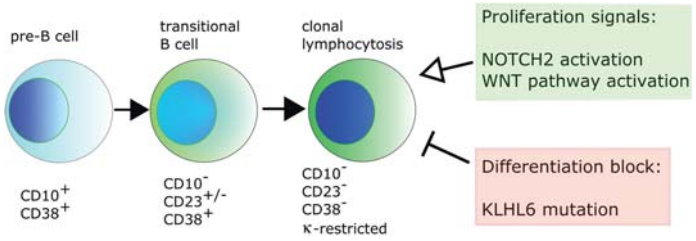

Figure 3. Model of the effects of mutated genes in this patient. We hypothesize that a partial block in normal B-cell development occurred in this patient because of the *KLHL6* mutation, and combined with activation of the cooperating pathways Notch and β-catenin, led to proliferation of a transitional cell type expressing a mix of classic B-cell surface markers.

The final gene of interest present in primary lymphoid samples in COSMIC is *APC*. The APC protein is a well-known tumor suppressor in colorectal cancer but is not believed to be relevant in lymphoid neoplasms (Sharma and Sen 2013). *APC* is expressed, albeit at low levels, in the patient's B cells, the CLL-ES cohort, and in all mouse B-cell populations examined (Supplemental Table S1). The predicted somatic missense mutation lies in the carboxy-terminal armadillo repeat (Supplemental Fig. S5C). This position is not found to be altered in ExAC, and the mutant allele is expressed in the tumor (Supplemental Table S1). One of the other genes having somatic mutations in this patient, but at a lower frequency in lymphoid samples in COSMIC, is *RAPGEF1*, which, like *APC*, encodes a negative regulator of β-catenin (Dayma et al. 2012). The mutation in *RAPGEF1*, Y572C, occurs at the same position as COSMIC mutation 1901064 (which is Y572H). β-catenin pathway–activating mutations have been previously reported in CLL (Wang et al. 2014), and thus we hypothesize that reduced activity of APC or RAPGEF1, or both, may contribute to the proliferation of the mutant clone (Fig. 3).

The other genes implicated in this patient that are present below 1% in COSMIC are *AMZ2*, *DAZAP1*, *FCN1*, *SEC14L1*, *SLC35A5*, *SP4*, *TMED9*, and *VMP1*. The predicted nonsense mutation in the *VMP1* gene would result in a truncated protein very likely to be unstable, likely resulting in haplodeficiency of the gene product. VMP1 is an integral membrane protein that recruits the adaptor protein Beclin 1 in an early stage of autophagy (Ropolo et al. 2007; Molejon et al. 2013), suggesting that autophagy-dependent cell death may be reduced in the malignant clone.

Splicing factors, specifically SF3B1, have been previously implicated in CLL (Quesada et al. 2012; Jeromin et al. 2014), and the *DAZAP1* gene product has been implicated in splicing regulation (Choudhury et al. 2014). The missense mutation in DAZAP1 is in its carboxy-terminal domain, which has been shown to mediate interactions with, and antagonize the splicing-inhibitory functions of, hnRNP A1 proteins (Choudhury et al. 2014). This implies that the *DAZAP1* variant could have functional consequences via unknown downstream splicing effects.

Of the remaining candidate genes (*AMZ2*, *FCN1*, *SEC14L1*, *SLC35A5*, *SP4*, and *TMED9*), less is known, especially in cells of lymphoid origin, but all are expressed and predicted to have damaging missense mutations in this patient's B-cell clone. As with *DAZAP1*, any or all could have pleiotropic effects. AMZ2 is a protease (Díaz-Perales et al. 2005), although the observed mutation (T174K) does not fall within the predicted peptidase domain. *FCN1* is known to be highly polymorphic and encodes Ficolin 1, a soluble pattern-recognition molecule and activator of the complement system (Ferreira et al. 2014). *SEC14L1* is a putative biomarker for aggressive prostate cancers (Agell et al. 2012; Burdelski et al. 2015) and was independently identified as a binding partner and negative regulator of the RIG-I protein (encoded by *DDX58*), which stimulates IFN (interferon)-β after

binding single-stranded RNA viruses (Li et al. 2013). The observed mutation in SEC14L1 occurs in the amino terminal of the PRELI-MSF1 domain and results in a proline-to-alanine missense mutation that may disrupt folding of the domain by altering a loop between two β helices (Miliara et al. 2015). SLC35A5 is a solute carrier of unknown endogenous function, and the observed missense variant occurs in the nucleotide-sugar transporter domain. SP4 is a transcription factor best characterized in neurons and cardiac cells but which has paralogs implicated in cancers (Safe and Abdelrahim 2005), and TMED9 is a Golgi-associated protein, but the observed variant does not fall within annotated domains in either of these proteins.

Finally, as the patient's lymphoproliferative disorder was concurrent with a sporadic colorectal tumor, we verified that there were no germline nonsynonymous variants in genes relevant to hereditary nonpolyposis colorectal cancers (data not shown).

Integrative Analysis

RNA expression profiling can also be used to find evidence that supports the purely variant-based hypotheses above or to find new observations that may be worth further investigation. For example, consistent with the expectation of NOTCH2 activation due to its predicted truncating mutation, the known NOTCH2 target gene *DTX1* is overexpressed relative to normal B cells. The transcription factor KLF2 has emerged as another important player in SMZL, and loss-of-function mutations in this gene are proposed to contribute to the homing of B cells to the marginal zone (Clipson et al. 2015). Consistent with this model, in this patient, in whom the marginal zone is not involved, *KLF2* is one of the most up-regulated genes (Supplemental Table S2).

Inspection of the most overexpressed genes as compared with normal B cells (Supplemental Table S2) also led us to observe that one molecule that is up-regulated in the patient, perhaps as a consequence of activation of the β-catenin pathway (Lako et al. 2001), is the embryonic transcription factor Brachyury. This gene, which is not normally expressed during B-cell development (Heng and Painter 2008), is also expressed at very low levels in all samples in the CLL-ES RNA-seq data (Supplemental Fig. S6). The actual cause, or potential consequences, of Brachyury expression in B cells remains an open question for further research. Aberrant expression of Brachyury in some solid tumors contributes to epithelial-to-mesenchymal transformation, but its expression in lymphoid cells has not previously been reported.

DISCUSSION

The genomic analysis presented here highlights the utility of expression data as an adjunct to prioritizing variants of interest. Even in the absence of directly comparable samples, gene expression profiles from both the biopsy and literature provided useful information. By aggregating expression profiles, we performed a simple version of sample-based expression clustering, which indicates the patient's B cells resemble a low-grade CLL. This is at odds with both the mutation profile, which is overall most consistent with DLBCL, and the presence of the *NOTCH2* mutation, which is most strongly associated with SMZL. It is possible that this type of indolent lymphoproliferative disorder often goes undiagnosed, as it may have been in this case if it were not found secondary to the colorectal cancer. Another strong possibility is that similar cases with unclear diagnoses may be excluded from most current large-scale analyses.

Together, the observations outlined in this paper lead us to a speculative model (Fig. 3) in which a combination of proliferative signals and differentiation blocks leads to the expansion of a postgerminal center B-cell type. Of the 12 genes that we identified as having somatic

variants that are expressed, only two of those genes, *NOTCH2* and *KLHL6*, have well-established roles in B cells. Several of the other genes of interest, including *RAPGEF1*, *APC*, and *VMP1*, have variants or known associations with B-cell neoplasms that make them interesting candidates for further study. However, the full implications for this patient and her treatment remain unclear. Should she require chemotherapy in the future, the mutations identified herein may be informative. For now, the findings that there are no mutations indicative of aggressive disease, and that the gene expression is consistent with a low-grade disease, are positive for the patient.

METHODS

FISH assays and IHC were performed by the clinical laboratories at the BC Cancer Agency according to established protocols. Two hundred nuclei were examined per FISH assay. Library construction was performed using standard protocols, and sequencing was performed using Illumina sequencers as indicated. One microgram of DNA from $CD19^+$ cells was used as input to a PCR (polymerase chain reaction)-free WGS protocol and sequenced on a HiSeq 2000 to 106× coverage; control ($CD3^+$ fraction) was similarly sequenced to 45× coverage. Additionally 4 µg of total RNA (RIN 8.0) was subjected to poly(A) selection followed by ssRNA-seq on a HiSeq 2000 instrument to a total of 370 million reads. Reads were aligned to the human genome (GRCh37-lite) using BWA (0.5.7) (Li et al. 2009). Reads from multiple lanes were merged and duplicate marked using Picard (v1.38, https://github.com/broadinstitute/picard). Variants were called using mpileup (SAMtools v0.1.17) (Li and Durbin 2009) and subsequently filtered with varFilter. The tumor sample was compared with the normal sample to identify somatic copy-number variants (CNAseq v0.0.6, http://www.bcgsc.ca/platform/bioinfo/software/cnaseq), LOH events (APOLLOH v0.1.1, using the "K18 params Illumina stromalRatio Hyper10k min10max200" parameter matrix, downloaded from http://compbio.bccrc.ca/software/apolloh/) (Ha et al. 2012), single-nucleotide variants (SAMtools v0.1.17, MutationSeq v1.0.2 [Ding et al. 2012], Strelka v0.4.6.2 [Saunders et al. 2012]), and small insertions and deletions (Strelka v0.4.6.2). RNA-seq reads were analyzed with JAGuaR (Junction Alignments to Genome for RNA-Seq Reads; Butterfield et al. 2014) to include alignments to a database of exon junction sequences and subsequent repositioning onto the genomic reference. RNA-seq data were processed using the Genome Sciences Centre (GSC)'s WTSS (whole-transcriptome shotgun sequencing) pipeline coverage analysis (v1.1) using the "stranded" option to determine gene and exon read counts and normalized expression level. Expressed variants were called with SNVMix2 (v0.12.1-rc1) (Goya et al. 2010) and SAMtools (v0.1.17). Gene expression in the tumor was compared with a compendium of normal tissues and to one or more normal libraries from the same tissue type to identify up- or down-regulated genes. Both genomic and RNA-seq tumor data were also assembled using Trans-ABySS (v1.4.3) (Robertson et al. 2010) to identify structural variants and fusion genes. Variants were annotated to genes using the Ensembl database (v59,69).

Expression profiling analysis was carried out with pairwise comparisons of the $CD19^+$ fraction of the patient's bone marrow relative to a normal B-cell library, against $CD19^+$ populations from 17 CLL patient samples, 15 conventional MCLs, seven indolent MCLs, and four SMZLs (NCBI [National Center for Biotechnology Information] GEO [Gene Expression Omnibus] data set GSE16455 [Fernàndez et al. 2010]). For these, we carried out pairwise comparisons between each group and the CLL samples by extracting lists of the 150 genes most significantly over- or underexpressed relative to CLLs. We then determined the extent to which genes in these lists were expressed in the patient's transcriptome relative to a control B-cell library.

Recurrent mutations were identified from COSMIC complete mutation data (CosmicCompleteExportIncFus_v68_210114) parsed by tissue to include only samples of lymphoid origin using the provided web tools. The CLL-ES RNA-seq data set of 109 samples was obtained from International Cancer Genome Consortium (ICGC) DCC version 16 (https://dcc.icgc.org/projects/CLLE-ES). In a publication describing the analysis of RNA-seq data from an overlapping cohort (Ferreira et al. 2014), 98 RNA-seq samples (41 IgVH [immunoglobulin heavy chain variable]-normal and 54 IgVH-mutated <98% identity) are described.

Twenty-seven genes, containing 28 coding somatic single-nucleotide variants (SNVs), were identified as having more than 20 reads representing the alternate allele and predicted to have altered amino acid sequence by SnpEff (Table 1; Cingolani et al. 2012). The SIFT (Forbes et al. 2011) and PolyPhen-2 (Ferreira et al. 2014) algorithms were then applied in tandem, and variants were filtered out if both algorithms provided a benign/nondamaging interpretation. The remaining 21 genes were then examined for expression in the $CD19^+$ tumor cells; genes not expressed were also checked for expression in alternative data sources (CLL RNA-seq data from the ICGC [Ferreira et al. 2014], and the ImmGen mouse developmental B-cell resource [Heng and Painter 2008]) to check whether those genes are expected to be expressed in B cells. None of the 21 genes was found to be expressed in these data sets.

ADDITIONAL INFORMATION

Ethics Statement

This work, including data deposition, was approved by the Research Ethics Board at the British Columbia Cancer Agency, protocol H12-00137. Written consent was obtained from the patient after discussion with her oncologists.

Database Deposition and Access

Whole-genome sequencing and RNA-seq data (.bam files) are available from EGA (www.ebi.ac.uk/ega/home) under accession number EGAD00001001656. Variants have been submitted to ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) under accession numbers SCV000258944–SCV000258956.

Acknowledgments

We thank the patient for agreeing to participate in this study; Philip Beers and David Ko for flow cytometry; Alexandra Fok and Robyn Roscoe for project management; and the British Columbia Cancer Foundation for funding.

Author Contributions

S.J., M.M., J.L., and A.K. contributed to conception/design of study. J.E.S., J.D.K.P., and J.W.-W. provided the study material or patients. Pathological study was contributed by K.J.S., A.P.W., R.D.G., and A.K. Y.Z., R.M., Y.L., Y.S., and E.P. contributed to collection and/or assembly of data. J.D.K.P., Y.L., Y.S., and E.P. contributed to data analysis and interpretation. J.D.K.P. and A.K. contributed to manuscript writing. All authors approved the final manuscript.

Funding

This work was supported by the British Columbia Cancer Foundation funding held by J.L. and M.M., Terry Fox Foundation Program Project Grant funding held by A.K. (TFF-122869), and the Michael Smith Foundation for Health Research held by J.W.-W.

Competing Interest Statement

The authors have declared no competing interest.

Referees

Megan A. Cooper
Anonymous

Received August 21, 2015; accepted in revised form December 30, 2015.

REFERENCES

Agell L, Hernández S, Nonell L, Lorenzo M, Puigdecanet E, de Muga S, Juanpere N, Bermudo R, Fernández PL, Lorente JA, et al. 2012. A 12-gene expression signature is associated with aggressive histological in prostate cancer: *SEC14L1* and *TCEB1* genes are potential markers of progression. *Am J Pathol* **181:** 1585–1594.

Burdelski C, Barreau Y, Simon R, Hube-Magg C, Minner S, Koop C, Graefen M, Heinzer H, Sauter G, Wittmer C, et al. 2015. *Saccharomyces cerevisiae*–like 1 overexpression is frequent in prostate cancer and has markedly different effects in Ets-related gene fusion-positive and fusion-negative cancers. *Hum Pathol* **46:** 514–523.

Butterfield YS, Kreitzman M, Thiessen N, Corbett RD, Li Y, Pang J, Ma YP, Jones SJ, Birol I. 2014. JAGuaR: Junction Alignments to Genome for RNA-seq Reads. *PLoS One* **9:** e102398.

Choudhury R, Roy SG, Tsai YS, Tripathy A, Graves LM, Wang Z. 2014. The splicing activator DAZAP1 integrates splicing control into MEK/Erk-regulated cell proliferation and migration. *Nat Commun* **5:** 3078.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6:** 80–92.

Clipson A, Wang M, de Leval L, Ashton-Key M, Wotherspoon A, Vassiliou G, Bolli N, Grove C, Moody S, Escudero-Ibarz L, et al. 2015. KLF2 mutation is the most frequent somatic change in splenic marginal zone lymphoma and identifies a subset with distinct genotype. *Leukemia* **29:** 1177–1185.

Dayma K, Ramadhas A, Sasikumar K, Radha V. 2012. Reciprocal negative regulation between the guanine nucleotide exchange factor C3G and β-catenin. *Genes Cancer* **3:** 564–577.

Díaz-Perales A, Quesada V, Peinado JR, Ugalde AP, Álvarez J, Suárez MF, Gomis-Rüth FX, López-Otín C. 2005. Identification and characterization of human archaemetzincin-1 and -2, two novel members of a family of metalloproteases widely distributed in Archaea. *J Biol Chem* **280:** 30367–30375.

Ding J, Bashashati A, Roth A, Oloumi A, Tse K, Zeng T, Haffari G, Hirst M, Marra MA, Condon A, et al. 2012. Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* **28:** 167–175.

Fernàndez V, Salamero O, Espinet B, Solé F, Royo C, Navarro A, Camacho F, Beà S, Hartmann E, Amador V, et al. 2010. Genomic and gene expression profiling defines indolent forms of mantle cell lymphoma. *Cancer Res* **70:** 1408–1418.

Ferreira PG, Jares P, Rico D, Gómez-López G, Martínez-Trillos A, Villamor N, Ecker S, González-Pérez A, Knowles DG, Monlong J, et al. 2014. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res* **24:** 212–226.

Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, et al. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39(**Database issue**):** D945–D950.

Goya R, Sun MG, Morin RD, Leung G, Ha G, Wiegand KC, Senz J, Crisan A, Marra MA, Hirst M, et al. 2010. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* **26:** 730–736.

Gupta-Rossi N, Storck S, Griebel PJ, Reynaud C-A, Weill J-C, Dahan A. 2003. Specific over-expression of deltex and a new Kelch-like protein in human germinal center B cells. *Mol Immunol* **39:** 791–799.

Ha G, Roth A, Lai D, Bashashati A, Ding J, Goya R, Giuliany R, Rosner J, Oloumi A, Shumansky K, et al. 2012. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res* **22:** 1995–2007.

Heng TS, Painter MW. 2008. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* **9:** 1091–1094.

Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, Alpermann T, Roller A, Kohlmann A, Haferlach T, et al. 2014. *SF3B1* mutations correlated to cytogenetics and mutations in *NOTCH1*, *FBXW7*, *MYD88*, *XPO1* and *TP53* in 1160 untreated CLL patients. *Leukemia* **28:** 108–117.

Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigelin HG, Chiang MY, Huebner-Chan DR, Bailey NG, Yang DT, Bhagat G, et al. 2012. Whole-genome sequencing identifies recurrent somatic *NOTCH2* mutations in splenic marginal zone lymphoma. *J Exp Med* **209:** 1553–1565.

Kroll J, Shi X, Caprioli A, Liu H-H, Waskow C, Lin K-M, Miyazaki T, Rodewald H-R, Sato TN. 2005. The BTB-kelch protein KLHL6 is involved in B-lymphocyte antigen receptor signaling and germinal center formation. *Mol Cell Biol* **25:** 8531–8540.

Lako M, Lindsay S, Lincoln J, Cairns PM, Armstrong L, Hole N. 2001. Characterisation of *Wnt* gene expression during the differentiation of murine embryonic stem cells in vitro: role of Wnt3 in enhancing haematopoietic differentiation. *Mech Dev* **103:** 49–59.

Laskin J, Jones S, Aparicio S, Chia S, Ch'ng C, Deyell R, Eirew P, Fok A, Gelmon K, Ho C, et al. 2015. Lessons learned from the application of whole-genome analysis to the treatment of patients with advanced cancers. *Mol Case Stud* **1:** a000570.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Li M-T, Di W, Xu H, Yang Y-K, Chen H-W, Zhang F-X, Zhai Z-H, Chen D-Y. 2013. Negative regulation of RIG-I-mediated innate antiviral signaling by SEC14L1. *J Virol* **87:** 10037–10046.

Miliara X, Garnett JA, Tatsuta T, Abid Ali F, Baldie H, Pérez-Dorado I, Simpson P, Yague E, Langer T, Matthews S. 2015. Structural insight into the TRIAP1/PRELI-like domain family of mitochondrial phospholipid transfer complexes. *EMBO Rep* **16:** 824–835.

Molejon MI, Ropolo A, Re AL, Boggio V, Vaccaro MI. 2013. The VMP1–Beclin 1 interaction regulates autophagy induction. *Sci Rep* **3:** 1055.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.

Parry M, Rose-Zerilli MJ, Ljungström V, Gibson J, Wang J, Walewska R, Parker H, Parker AE, Davis Z, Gardiner A, et al. 2015. Genetics and prognostication in splenic marginal zone lymphoma: revelations from deep sequencing. *Clin Cancer Res* **21:** 4174–4183.

Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, Escaramis G, Jares P, Beà S, Gonzàlez-Díaz M, et al. 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475:** 101–105.

Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martin-Subero JI, Munar M, Rubio-Pérez C, Jares P, Aymerich M, et al. 2015. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526:** 519–524.

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, et al. 2012. Exome sequencing identifies recurrent mutations of the splicing factor *SF3B1* gene in chronic lymphocytic leukemia. *Nat Genet* **44:** 47–52.

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, et al. 2010. De novo assembly and analysis of RNA-seq data. *Nat Methods* **7:** 909–912.

Ropolo A, Grasso D, Pardo R, Sacchetti ML, Archange C, Re AL, Seux M, Nowak J, Gonzalez CD, Iovanna JL, et al. 2007. The pancreatitis-induced vacuole membrane protein 1 triggers autophagy in mammalian cells. *J Biol Chem* **282:** 37124–37133.

Rossi D, Trifonov V, Fangazio M, Bruscaggin A, Rasi S, Spina V, Monti S, Vaisitti T, Arruga F, Fama R, 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.

Safe S, Abdelrahim M. 2005. Sp transcription factor family and its role in cancer. *Eur J Cancer* **41:** 2438–2448.

Saito T, Chiba S, Ichikawa M, Kunisato A, Asai T, Shimizu K, Yamaguchi T, Yamamoto G, Seo S, Kumano K, et al. 2003. Notch2 is preferentially expressed in mature B cells and indispensable for marginal zone B lineage development. *Immunity* **18:** 675–685.

Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. 2012. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28:** 1811–1817.

Sharma A, Sen JM. 2013. Molecular basis for the tissue specificity of β-catenin oncogenesis. *Oncogene* **32:** 1901–1909.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102:** 15545–15550.

Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW, eds. 2008. *WHO classification of tumours of haematopoietic and lymphoid tissues. IARC WHO Classification of Tumours*. World Health Organization, Geneva.

Trøen G, Wlodarska I, Warsame A, Hernández Llodrà S, De Wolf-Peeters C, Delabie J. 2008. NOTCH2 mutations in marginal zone lymphoma. *Haematologica* **93:** 1107–1109.

Wang L, Shalek AK, Lawrence M, Ding R, Gaublomme JT, Pochet N, Stojanov P, Sougnez C, Shukla SA, Stevenson KE, et al. 2014. Somatic mutation as a mechanism of Wnt/β-catenin pathway activation in CLL. *Blood* **124:** 1089–1098.