doi:10.1038/nature10351

Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma

Ryan D. Morin[1]*, Maria Mendez-Lago[1]*, Andrew J. Mungall[1], Rodrigo Goya[1], Karen L. Mungall[1], Richard D. Corbett[1], Nathalie A. Johnson[2], Tesa M. Severson[1], Readman Chiu[1], Matthew Field[1], Shaun Jackman[1], Martin Krzywinski[1], David W. Scott[2], Diane L. Trinh[1], Jessica Tamura-Wells[1], Sa Li[1], Marlo R. Firme[1], Sanja Rogic[2], Malachi Griffith[1], Susanna Chan[1], Oleksandr Yakovenko[1], Irmtraud M. Meyer[3], Eric Y. Zhao[1], Duane Smailus[1], Michelle Moksa[1], Suganthi Chittaranjan[1], Lisa Rimsza[4], Angela Brooks-Wilson[1,5], John J. Spinelli[6,7], Susana Ben-Neriah[2], Barbara Meissner[2], Bruce Woolcock[2], Merrill Boyle[2], Helen McDonald[1], Angela Tam[1], Yongjun Zhao[1], Allen Delaney[1], Thomas Zeng[1], Kane Tse[1], Yaron Butterfield[1], Inanç Birol[1], Rob Holt[1], Jacqueline Schein[1], Douglas E. Horsman[2], Richard Moore[1], Steven J. M. Jones[1], Joseph M. Connors[2], Martin Hirst[1], Randy D. Gascoyne[2,8] & Marco A. Marra[1,9]

Follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL) are the two most common non-Hodgkin lymphomas (NHLs). Here we sequenced tumour and matched normal DNA from 13 DLBCL cases and one FL case to identify genes with mutations in B-cell NHL. We analysed RNA-seq data from these and another 113 NHLs to identify genes with candidate mutations, and then re-sequenced tumour and matched normal DNA from these cases to confirm 109 genes with multiple somatic mutations. Genes with roles in histone modification were frequent targets of somatic mutation. For example, 32% of DLBCL and 89% of FL cases had somatic mutations in *MLL2*, which encodes a histone methyltransferase, and 11.4% and 13.4% of DLBCL and FL cases, respectively, had mutations in *MEF2B*, a calcium-regulated gene that cooperates with CREBBP and EP300 in acetylating histones. Our analysis suggests a previously unappreciated disruption of chromatin biology in lymphomagenesis.

Non-Hodgkin lymphomas (NHLs) are cancers of B, T or natural killer lymphocytes. The two most common types of NHL, follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL), together comprise 60% of new B-cell NHL diagnoses each year in North America[1]. FL is an indolent and typically incurable disease characterized by clinical and genetic heterogeneity. DLBCL is aggressive and likewise heterogeneous, comprising at least two distinct subtypes that respond differently to standard treatments. Both FL and the germinal centre B-cell (GCB) cell of origin (COO) subtype of DLBCL derive from germinal centre B cells, whereas the activated B-cell (ABC) variety, which has a more aggressive clinical course, is thought to originate from B cells that have exited, or are poised to exit, the germinal centre[2]. Current knowledge of the specific genetic events leading to DLBCL and FL is limited to the presence of a few recurrent genetic abnormalities[2]. For example, 85–90% of FL and 30–40% of GCB DLBCL cases[3,4] harbour t(14;18)(q32;q21), which results in deregulated expression of the BCL2 oncoprotein. Other genetic abnormalities unique to GCB DLBCL include amplification of the *c-REL* gene and of the miR-17-92 microRNA cluster[5]. In contrast to GCB cases, 24% of ABC DLBCLs harbour structural alterations or inactivating mutations affecting *PRDM1*, which is involved in differentiation of GCB cells into antibody-secreting plasma cells[6]. ABC-specific mutations also affect genes regulating NF-κB signalling[7,8,9], with *TNFAIP3* (also known as *A20*) and *MYD88* (ref. 10) the most abundantly mutated in 24% and 39% of cases, respectively. To enhance our understanding of the genetic architecture of B-cell NHL, we undertook a study to (1) identify somatic mutations and (2) determine the prevalence, expression and focal recurrence of mutations in FL and DLBCL. Using strategies and techniques applied to cancer genome and transcriptome characterization by ourselves and others[11,12,13], we sequenced tumour DNA and/or RNA from 117 tumour samples and 10 cell lines (Supplementary Tables 1 and 2) and identified 651 genes (Supplementary Figure 1) with evidence of somatic mutation in B-cell NHL. After validation, we showed that 109 genes were somatically mutated in two or more NHL cases. We further characterized the frequency and nature of mutations within *MLL2* and *MEF2B*, which were among the most frequently mutated genes with no previously known role in lymphoma.

Identification of recurrently mutated genes

We sequenced the genomes or exomes of 14 NHL cases, all with matched constitutional DNA sequenced to comparable depths (Supplementary Tables 1 and 2). After screening for single nucleotide variants followed by subtraction of known polymorphisms and visual inspection of the sequence read alignments, we identified 717 non-synonymous variants (coding single nucleotide variants; cSNVs) affecting 651 genes (Supplementary Figure 1 and Supplementary Methods). We identified between 20 and 135 cSNVs in each of these genomes. Only 25 of the 651 genes with cSNVs were represented in the cancer gene census (December 2010 release)[14].

We performed RNA sequencing (RNA-seq) on these 14 NHL cases and an expanded set of 113 samples comprising 83 DLBCL, 12 FL and 8 B-cell NHL cases with other histologies and 10 DLBCL-derived cell lines (Supplementary Table 2). We analysed these data to identify

[1]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. [2]Centre for Lymphoid Cancer, BC Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. [3]Centre for High-throughput Biology, Department of Computer Science, Vancouver, British Columbia V6T 1Z4, Canada. [4]Department of Pathology, University of Arizona, Tucson, Arizona 85724, USA. [5]Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. [6]Cancer Control Research, BC Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. [7]School of Population and Public Health, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. [8]Department of Pathology, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. [9]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6H 3N1, Canada.
*These authors contributed equally to this work.

novel fusion transcripts (Supplementary Table 3) and cSNVs (Fig. 1). We identified 240 genes with at least one cSNV in a genome/exome or an RNA-seq 'mutation hot spot' (see later), and with cSNVs in at least three cases in total (Supplementary Table 4). We selected cSNVs from each of these 240 genes for re-sequencing to confirm their somatic status. We did not re-sequence genes with previously documented mutations in lymphoma (for example, *CD79B*, *BCL2*). We confirmed the somatic status of 543 cSNVs in 317 genes, with 109 genes having at least two confirmed somatic mutations (Supplementary Tables 4 and 5). Of the successfully re-sequenced cSNVs predicted from the genomes, 171 (94.5%) were confirmed somatic, 7 were false calls and 3 were present in the germ line. These 109 recurrently mutated genes were significantly enriched for genes implicated in lymphocyte activation ($P = 8.3 \times 10^{-4}$; for example, *STAT6*, *BCL10*), lymphocyte differentiation ($P = 3.5 \times 10^{-3}$; for example, *CARD11*), and regulation of apoptosis ($P = 1.9 \times 10^{-3}$; for example, *BTG1*, *BTG2*). Also significantly enriched were genes linked to transcriptional regulation ($P = 5.4 \times 10^{-4}$; for example, *TP53*) and genes involved in methylation ($P = 2.2 \times 10^{-4}$) and acetylation ($P = 1.2 \times 10^{-2}$), including histone methyltransferase (HMT) and acetyltransferase (HAT) enzymes known previously to be mutated in lymphoma (for example, *EZH2* (ref. 13) and *CREBBP* (ref. 15); Supplementary Methods).

Mutation hot spots can result from mutations at sites under strong selective pressure and we have previously identified such sites using RNA-seq data[13]. We searched our RNA-seq data for genes with mutation hot spots, and identified 10 genes that were not mutated in the 14 genomes (*PIM1*, *FOXO1*, *CCND3*, *TP53*, *IRF4*, *BTG2*, *CD79B*, *BCL7A*, *IKZF3* and *B2M*), of which five (*FOXO1*, *CCND3*, *BTG2*, *IKZF3* and *B2M*) were not previously known targets of point mutation in NHL (Supplementary Table 6 and Supplementary Methods). *FOXO1*, *BCL7A* and *B2M* had hot spots affecting their start codons. The effect of a *FOXO1* start codon mutation, which was observed in three cases, was further studied using a cell line in which the initiating ATG was mutated to TTG. Western blots probed with a FOXO1 antibody revealed a band with a reduced molecular weight, indicative of a FOXO1 amino-terminal truncation (Supplementary Figure 2), consistent with use of the next in-frame ATG for translation initiation. A second hot spot in *FOXO1* at T24 was mutated in two cases. T24 is reportedly phosphorylated by AKT subsequent to B-cell receptor (BCR) stimulation[16] inducing FOXO1 nuclear export.

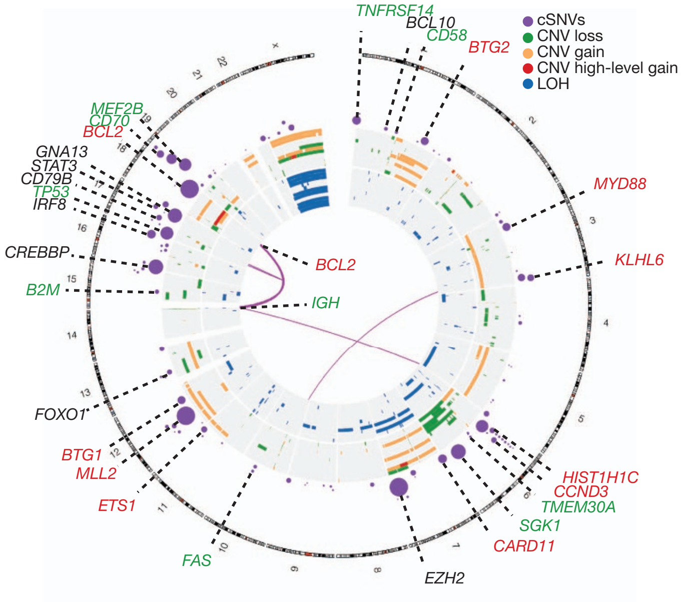

Figure 1 | Genome-wide visualization of somatic mutation targets in NHL. Overview of structural rearrangements and copy number variations (CNVs) in the 11 DLBCL genomes and cSNVs in the 109 recurrently mutated genes identified in our analysis. Inner arcs represent somatic fusion transcripts identified in at least one of the 11 genomes. The CNVs and LOH detected in each of the 11 DLBCL tumour/normal pairs are displayed on the concentric sets of rings. The inner 11 rings show regions of enhanced homozygosity plotted with blue (interpreted as LOH). The outer 11 rings show somatic CNVs. Purple circles indicate the position of genes with at least two confirmed somatic mutations with circle diameter proportional to the number of cases with cSNVs detected in that gene. Circles representing the genes with significant evidence for positive selection are labelled. Coincidence between recurrently mutated genes and regions of gain/loss are colour-coded in the labels (green, loss; red, gain). For example *B2M*, which encodes beta-2-microglobulin, is recurrently mutated and is deleted in two cases.

We analysed the RNA-seq data to determine whether any of the somatic mutations in the 109 recurrently mutated genes showed evidence for allelic imbalance with expression favouring one allele. Out of 380 expressed heterozygous mutant alleles, we observed preferential expression of the mutation for 16.8% (64/380) and preferential expression of the wild type for 27.8% (106/380; Supplementary Table 7). Seven genes showed evidence for significant preferential expression of the mutant allele in at least two cases: *BCL2*, *CARD11*, *CD79B*, *EZH2*, *IRF4*, *MEF2B* and *TP53*; Supplementary Methods. In 27 out of 43 cases with *BCL2* cSNVs, expression favoured the mutant allele, consistent with the previously-described hypothesis that the translocated (and hence, transcriptionally deregulated) allele of *BCL2* is targeted by somatic hypermutation[17]. Examples of mutations at known oncogenic hot spot sites such as F123I in *CARD11* (ref. 18) showed allelic imbalance favouring the mutant allele in some cases. Similarly, we noted expression favouring two novel hot spot mutations in *MEF2B* (Y69 and D83) and two sites in *EZH2* not previously reported as mutated in lymphoma (A682G and A692V).

We sought to distinguish new cancer-related mutations from passenger mutations using the approach proposed previously[19]. We reasoned that this would reveal genes with strong selection signatures, and mutations in such genes would be good candidate cancer drivers. We identified 26 genes with significant evidence for positive selection (false discovery rate = 0.03, Supplementary Methods), with either selective pressure for acquiring non-synonymous point mutations or truncating/nonsense mutations (Supplementary Methods; Table 1 and Supplementary Table 8). Included were known lymphoma oncogenes (*BCL2*, *CD79B* (ref. 9), *CARD11* (ref. 18), *MYD88* (ref. 10) and *EZH2* (ref. 13)), all of which showed signatures indicative of selection for non-synonymous variants.

Evidence for selection of inactivating changes

We expected tumour suppressor genes to show strong selection for the acquisition of nonsense mutations. In our analysis, the eight most significant genes included seven with strong selective pressure for nonsense mutations, including the known tumour suppressor genes *TP53* and *TNFRSF14* (ref. 20 ; Table 1). *CREBBP*, recently reported as commonly inactivated in DLBCL[15], also showed some evidence for acquisition of nonsense mutations and cSNVs (Supplementary Figure 3 and Supplementary Table 9). We also observed enrichment for nonsense mutations in *BCL10*, a positive regulator of NF-κB, in which oncogenic truncated products have been described in lymphomas[21]. The remaining strongly significant genes (*BTG1*, *GNA13*, *SGK1* and *MLL2*) had no reported role in lymphoma. *GNA13* was affected by mutations in 22 cases including multiple nonsense mutations. *GNA13* encodes the alpha subunit of a heterotrimeric G-protein coupled receptor responsible for modulating RhoA activity[22]. Some of the mutated residues negatively affect its function[23,24], including a T203A mutation, which also showed allelic imbalance favouring the mutant allele (Supplementary Table 7). GNA13 protein was reduced or absent on western blots in cell lines harbouring either a nonsense mutation, a stop codon deletion, a frame shifting deletion, or changes affecting splice sites (Supplementary Methods and Supplementary Figure 4).

Table 1 | Overview of cSNVs and confirmed somatic mutations in most frequently mutated genes

Gene	Cases			Total			Somatic cSNVs (RNA-seq cohort)*	*P* (raw)	*q*	NS SP	T SP	Skew (M, WT, both)†
	NS	S	T	NS	S	T						
MLL2‡	16	8	17	17	8	18	10	$\mathbf{6.85 \times 10^{-8}}$	8.50×10^{-7}	0.834	14.4	WT
TNFRSF14 ^{G}‡	7	1	7	8	1	7	11	$\mathbf{6.85 \times 10^{-8}}$	8.50×10^{-7}	7.52	118	Both
SGK1 ^{G}‡	18	6	6	37	10	6	9	$\mathbf{6.85 \times 10^{-8}}$	8.50×10^{-7}	19.5	61.7	–
BCL10‡	2	0	4	3	0	4	4	$\mathbf{6.85 \times 10^{-8}}$	8.50×10^{-7}	3.62	112	WT
GNA13 ^{G}‡	21	1	2	33	1	2	5	$\mathbf{6.85 \times 10^{-8}}$	8.50×10^{-7}	24.1	25.7	Both
TP53 ^{G}‡	20	2	1	23	3	1	22	$\mathbf{6.85 \times 10^{-8}}$	8.50×10^{-7}	15.6	14.1	Both
EZH2 ^{G}‡	33	0	0	33	0	0	33	$\mathbf{6.85 \times 10^{-8}}$	8.50×10^{-7}	11.4	0.00	Both
BTG2‡	12	6	1	14	6	1	2	6.85×10^{-8}	8.50×10^{-7}	23.9	35.1	–
BCL2 ^{G}‡	42	45	0	96	105	0	43	$\mathbf{9.35 \times 10^{-8}}$	8.50×10^{-7}	3.78	0.00	M
BCL6‡§	11	2	0	12	2	0	2	$\mathbf{9.35 \times 10^{-8}}$	8.50×10^{-7}	0.175	0.00	M
CIITA‡§	5	3	0	6	3	0	2	$\mathbf{9.35 \times 10^{-8}}$	8.50×10^{-7}	0.086	0.00	
FAS‡	2	0	4	3	0	4	2	1.52×10^{-7}	1.17×10^{-6}	2.54	66.5	WT
BTG1‡	11	6	2	11	7	2	10	1.52×10^{-7}	1.17×10^{-6}	17.5	52.5	Both
MEF2B ^{G}‡	20	2	0	20	2	0	10	2.05×10^{-7}	1.47×10^{-6}	14.2	0.00	M
IRF8‡	11	5	3	14	5	3	3	4.55×10^{-7}	3.03×10^{-6}	8.82	28.2	WT
TMEM30A‡	1	0	4	1	0	4	4	6.06×10^{-7}	3.79×10^{-6}	0.785	65.0	WT
CD58‡	2	0	3	2	0	3	2	2.42×10^{-6}	1.43×10^{-5}	2.29	69.2	–
KLHL6‡	10	2	2	12	2	2	4	1.00×10^{-5}	5.26×10^{-5}	5.42	16.4	–
MYD88 ^{A}‡	13	2	0	14	2	0	9	1.00×10^{-5}	5.26×10^{-5}	12.4	0.00	WT
CD70‡	5	0	1	5	0	2	3	1.70×10^{-5}	8.48×10^{-5}	7.08	44.0	–
CD79B ^{A}‡	7	2	1	9	2	1	5	2.00×10^{-5}	9.52×10^{-5}	10.9	18.3	M
CCND3‡	7	1	2	7	1	2	6	2.80×10^{-5}	1.27×10^{-4}	6.55	36.3	WT
CREBBP‡	20	7	4	24	7	4	9	1.00×10^{-4}	4.35×10^{-4}	2.72	6.04	Both
HIST1H1C‡	9	0	0	10	0	0	6	1.80×10^{-4}	7.50×10^{-4}	11.9	0.00	Both
B2M‡	7	0	0	7	0	0	4	3.90×10^{-4}	1.56×10^{-3}	16.6	0.00	WT
ETS1‡	10	1	0	10	1	0	4	4.10×10^{-4}	1.58×10^{-3}	5.76	0.00	WT
CARD11‡	14	3	0	14	3	0	3	1.90×10^{-3}	7.04×10^{-3}	3.37	0.00	Both
FAT2‡§	2	1	0	2	1	0	2	6.30×10^{-3}	2.25×10^{-2}	0.128	0.00	–
IRF4‡§	9	4	0	26	5	0	5	7.00×10^{-3}	2.41×10^{-2}	0.569	0.00	Both
FOXO1‡	8	4	0	10	4	0	4	7.60×10^{3}	2.53×10^{-2}	4.02	0.00	–
STAT3	9	0	0	9	0	0	4	2.19×10^{-2}	6.08×10^{-2}	–	–	Both
RAPGEF1	8	3	0	10	3	0	3	2.98×10^{-2}	7.45×10^{-2}	–	–	WT
ABCA7	12	3	0	15	3	0	2	7.76×10^{-2}	1.67×10^{-1}	–	–	WT
RNF213	10	8	0	10	8	0	2	7.87×10^{-2}	1.67×10^{-1}	–	–	–
MUC16	17	12	0	39	25	0	2	8.32×10^{-2}	1.73×10^{-1}	–	–	–
HDAC7	8	4	0	8	4	0	2	8.94×10^{-2}	1.82×10^{-1}	–	–	WT
PRKDC	7	3	0	7	4	0	2	1.06×10^{-1}	2.05×10^{-1}	–	–	–
SAMD9	9	2	0	9	2	0	2	1.79×10^{-1}	3.01×10^{-1}	–	–	–
TAF1	10	0	0	10	0	0	2	3.03×10^{-1}	4.74×10^{-1}	–	–	–
PIM1	20	19	0	33	34	0	11	3.40×10^{-1}	5.23×10^{-1}	–	–	WT
COL4A2	8	2	0	8	2	0	2	7.64×10^{-1}	8.99×10^{-1}	–	–	–
EP300	8	7	1	8	7	1	3	9.54×10^{-1}	1.00	–	–	WT

Individual cases with non-synonymous (NS), synonymous (S) and truncating (T) mutations and the total number of mutations of each class are shown separately because some genes contained multiple mutations in the same case. The *P* values indicated in bold are the upper limit on the *P* value for that gene determined with the approach described in ref. 19 (see Supplementary Methods), *q* is the Benjamini-corrected *q* value, and NS SP and T SP refer to selective pressure estimates from this model for the acquisition of non-synonymous or truncating mutations, respectively. Genes with a superscript of either A or G were found to have mutations significantly enriched in ABC or GCB cases, respectively ($P < 0.05$, Fisher's exact test).

* Additional somatic mutations identified in larger cohorts and insertion/deletion mutations are not included in this total.

† 'Both' indicates that we observed separate cases in which skewed expression was seen but where this skew was not consistent for the mutant or wild-type allele.

‡ Genes significant at a false discovery rate of 0.03. SNVs in BCL2 and previously confirmed hot spot mutations in *EZH2* and *CD79B* are probably somatic in these samples based on published observations of others.

§ Selective pressure estimates are both < 1 indicating purifying selection rather than positive selection acting on this gene.

SGK1 encodes a phosphatidylinositol-3-OH kinase (PI(3)K)-regulated kinase with functions including regulation of FOXO transcription factors[25], regulation of NF-κB by phosphorylating IκB kinase[26], and negative regulation of NOTCH signalling[27]. *SGK1* also resides within a region of chromosome 6 commonly deleted in DLBCL (Fig. 1)[5]. The mechanism by which *SGK1* and *GNA13* inactivation may contribute to lymphoma is unclear, but the strong degree of apparent selection towards their inactivation and their overall high mutation frequency (each mutated in 18 of 106 DLBCL cases) suggests that their loss contributes to B-cell NHL. Certain genes are known to be mutated more commonly in GCB DLBCLs (for example, *TP53* (ref. 28) and *EZH2* (ref. 13)). Here, both *SGK1* and *GNA13* mutations were found only in GCB cases ($P = 1.93 \times 10^{-3}$ and 2.28×10^{-4}, Fisher's exact test; $n = 15$ and 18, respectively) (Fig. 2). Two additional genes (*MEF2B* and *TNFRSF14*) with no previously described role in DLBCL showed a similar restriction to GCB cases (Fig. 2).

Inactivating *MLL2* mutations

MLL2 showed the most significant evidence for selection and the largest number of nonsense SNVs. Our RNA-seq analysis indicated that 26.0% (33/127) of cases carried at least one *MLL2* cSNV. To address the possibility that variable RNA-seq coverage of *MLL2* failed to capture some mutations, we PCR-amplified the entire *MLL2* locus (~36 kilobases) in 89 cases (35 primary FLs, 17 DLBCL cell lines, and 37 DLBCLs). Of these cases 58 were among the RNA-seq cohort. Illumina amplicon re-sequencing (Supplementary Methods) revealed 78 mutations, confirming the RNA-seq mutations in the overlapping cases and identifying 33 additional mutations. We confirmed the somatic status of 46 variants using Sanger sequencing (Supplementary Table 10), and showed that 20 of the 33 additional mutations were insertions or deletions (indels). Three SNVs at splice sites were also detected, as were 10 new cSNVs that had not been detected by RNA-seq.

The somatic mutations were distributed across *MLL2* (Fig. 3a). Of these, 37% ($n = 29/78$) were nonsense mutations, 46% ($n = 36/78$) were indels that altered the reading frame, 8% ($n = 6/78$) were point mutations at splice sites and 9% ($n = 7/78$) were non-synonymous amino acid substitutions (Table 2). Four of the somatic splice site mutations had effects on *MLL2* transcript length and structure. For example, two heterozygous splice site mutations resulted in the use of a novel splice donor site and an intron retention event.

Approximately half of the NHL cases we sequenced had two *MLL2* mutations (Supplementary Table 10). We used bacterial artificial

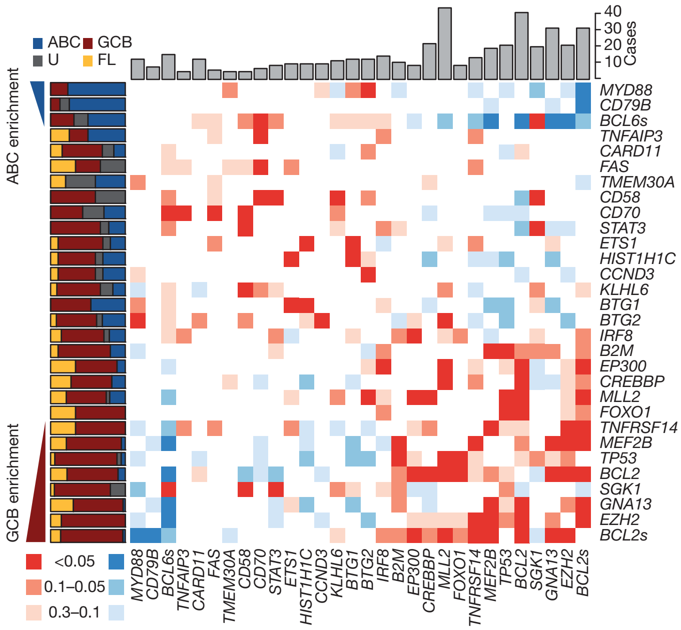

Figure 2 | Overview of mutations and potential cooperative interactions in NHL. This heat map displays possible trends towards co-occurrence (red) and mutual exclusion (blue) of somatic mutations and structural rearrangements. Colours were assigned by taking the minimum value of a left- and right-tailed Fisher's exact test. To capture trends a P-value threshold of 0.3 was used, with the darkest shade of the colour indicating those meeting statistical significance ($P \leq 0.05$). The relative frequency of mutations in ABC (blue), GCB (red), unclassifiable (black) DLBCLs and FL (yellow) cases is shown on the left. Genes were arranged with those having significant ($P < 0.05$, Fisher's exact test) enrichment for mutations in ABC cases (blue triangle) towards the top (and left) and those with significant enrichment for mutations in GCB cases (red triangle) towards the bottom (and right). The total number of cases in which each gene contained either cSNVs or confirmed somatic mutations is shown at the top. The cluster of blue squares (upper-right) results from the mutual exclusion of the ABC-enriched mutations (for example, *MYD88*, *CD79B*) from the GCB-enriched mutations (for example, *EZH2*, *GNA13*). Presence of structural rearrangements involving the two oncogenes *BCL6* and *BCL2* (indicated as *BCL6*s and *BCL2*s) was determined with FISH techniques using break-apart probes (Supplementary Methods).

chromosome (BAC) clone sequencing in eight FL cases to show that in all eight cases the mutations were in *trans*, affecting both *MLL2* alleles. This observation is consistent with the notion that there is a complete, or near-complete, loss of *MLL2* in the tumour cells of such patients.

With the exception of two primary FL cases and two DLBCL cell lines (Pfeiffer and SU-DHL-9), the majority of *MLL2* mutations seemed to be heterozygous. Analysis of Affymetrix 500k SNP array data from two FL cases with apparent homozygous mutations revealed that both tumours showed copy number neutral loss of heterozygosity (LOH) for the region of chromosome 12 containing *MLL2* (Supplementary Methods). Thus, in addition to bi-allelic mutation, LOH is a second, albeit less common mechanism by which *MLL2* function is lost.

MLL2 was the most frequently mutated gene in FL, and among the most frequently mutated genes in DLBCL (Fig. 2). We confirmed *MLL2* mutations in 31 of 35 FL patients (89%), in 12 of 37 DLBCL patients (32%), in 10 of 17 DLBCL cell lines (59%) and in none of the eight normal centroblast samples we sequenced. Our analysis predicted that the majority of the somatic mutations observed in *MLL2* were inactivating (91% disrupted the reading frame or were truncating point mutations), indicating to us that *MLL2* is a tumour suppressor of significance in NHL.

Recurrent point mutations in *MEF2B*

Our selective pressure analysis also revealed genes with stronger pressure for acquisition of amino acid substitutions than for nonsense mutations. One such gene was *MEF2B*, which had not previously been linked to lymphoma. We found that 20 (15.7%) cases had *MEF2B* cSNVs and 4 (3.1%) cases had *MEF2C* cSNVs. All cSNVs detected by RNA-seq affected either the MADS box or MEF2 domains. To determine the frequency and scope of *MEF2B* mutations, we Sanger-sequenced exons 2 and 3 in 261 primary FL samples; 259 DLBCL primary tumours; 17 cell lines; 35 cases of assorted NHL (IBL, composite FL and PBMCL); and eight non-malignant centroblast samples. We also used a capture strategy (Supplementary Methods) to sequence the entire *MEF2B* coding region in the 261 FL samples, revealing six additional variants outside exons 2 and 3. We thus identified 69 cases (34 DLBCL, 12.67%; and 35 FL, 15.33%) with *MEF2B* cSNVs or indels, failing to observe novel variants in other NHL and non-malignant samples. Of the variants 55 (80%) affected residues within the MADS box and MEF2 domains encoded by exons 2 and 3 (Supplementary Table 11; Fig. 3b). Each patient generally had a single *MEF2B* variant and we observed relatively few (eight in total, 10.7%) truncation-inducing SNVs or indels. Non-synonymous SNVs were by far the most common type of change observed, with 59.4% of detected variants affecting K4, Y69, N81 or D83. In 12 cases *MEF2B* mutations were shown to be somatic, including representative mutations at each of K4, Y69, N81 and D83 (Supplementary Table 12). We did not detect mutations in ABC cases, indicating that somatic mutations in *MEF2B* have a role unique to the development of GCB DLBCL and FL (Fig. 2).

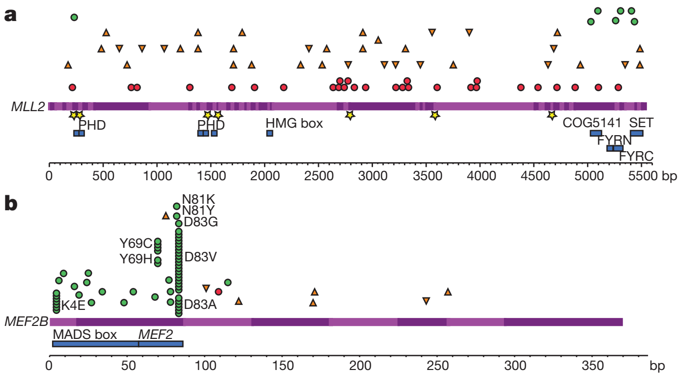

Figure 3 | Summary and effect of somatic mutations affecting *MLL2* and *MEF2B*. a, Re-sequencing the *MLL2* locus in 89 samples revealed mainly nonsense (red circles) and frameshift-inducing indel mutations (orange triangles; inverted triangles for insertions and upright triangles for deletions). A smaller number of non-synonymous somatic mutations (green circles) and point mutations or deletions affecting splice sites (yellow stars) were also observed. All of the non-synonymous point mutations affected a residue within either the catalytic SET domain, the FYRC domain (FY-rich carboxy-terminal domain) or PHD zinc finger domains. The effect of these splice-site mutations on *MLL2* splicing was also explored (Supplementary Figure 7). **b**, The cSNVs and somatic mutations found in *MEF2B* in all FL and DLBCL cases sequenced are shown with the same symbols. Only the amino acids with variants in at least two patients are labelled. cSNVs were most prevalent in the first two protein-coding exons of *MEF2B* (exons 2 and 3). The crystal structure of MEF2 bound to EP300 supports the idea that two of the mutated sites (L67 and Y69) are important in the interaction between these proteins (Supplementary Figure 8 and Supplementary Discussion)[50].

Table 2 | Summary of types of *MLL2* somatic mutations

Sample Type	FL	DLBCL	DLBCL cell-line	Centroblast
Truncation	18	4	7	0
Indel with frameshift	22	8	6	0
Splice site	4	2	0	0
SNV	3	2	2	0
Any mutation/ number of cases	31/35	12/37	10/17	0/8
Percentage	89	32	59	0

Discussion

In our study of genome, transcriptome and exome sequences from 127 B-cell NHL cases, we identified 109 genes with clear evidence of somatic mutation in multiple individuals. Significant selection seems to act on at least 26 of these for the acquisition of either nonsense or missense mutations. To the best of our knowledge, the majority of these genes had not previously been associated with any cancer type. We observed an enrichment of somatic mutations affecting genes involved in transcriptional regulation and, more specifically, chromatin modification.

MLL2 emerged from our analysis as a major tumour suppressor locus in NHL. It is one of six human H3K4-specific methyltransferases[29], all of which share homology with the *Drosophila* trithorax gene. Trimethylated H3K4 (H3K4me3) is an epigenetic mark associated with the promoters of actively transcribed genes. By laying down this mark, MLLs are responsible for the transcriptional regulation of developmental genes including the homeobox (*Hox*) gene family[30] which collectively control segment specificity and cell fate in the developing embryo[31,32]. Each MLL family member is thought to target different subsets of *Hox* genes[33] and in addition, MLL2 is known to regulate the transcription of a diverse set of genes[34]. Recently, *MLL2* mutations were reported in a small-cell lung cancer cell line[35] and in renal carcinoma[36], but the frequency of nonsense mutations affecting *MLL2* in these cancers was not established in these reports. Inactivating mutations were reported recently in *MLL2* or *MLL3* in 16% of medulloblastoma patients[37], further implicating *MLL2* as a cancer gene.

Our data link *MLL2* somatic mutations to B-cell NHL. The reported mutations are likely to be inactivating and in eight of the cases with multiple mutations, we confirmed that both alleles were affected, presumably resulting in essentially complete loss of MLL2 function. The high prevalence of *MLL2* mutations in FL (89%) equals the frequency of the t(14;18)(q32;q21) translocation, which is considered the most prevalent genetic abnormality in FL[3]. In DLBCL tumour samples and cell lines, *MLL2* mutation frequencies were 32% and 59%, respectively, also exceeding the prevalence of the most frequent cytogenetic abnormalities, such as the various translocations involving 3q27, which occur in 25–30% of DLBCLs and are enriched in ABC cases[38]. Importantly, we found *MLL2* mutated in both DLBCL subtypes (Fig. 2). Our analyses thus indicate that *MLL2* acts as a central tumour suppressor in FL and both DLBCL subtypes.

The *MEF2* gene family encodes four related transcription factors that recruit histone-modifying enzymes including histone deacetylases (HDACs) and HATs in a calcium-regulated manner. Although truncating variants were detected in our analysis of *MEF2* gene family members, our analysis suggests that, in contrast to *MLL2*, *MEF2* family members tend to selectively acquire non-synonymous amino acid substitutions. In the case of *MEF2B*, 59.4% of all the cSNVs were found at four sites within the protein (K4, Y69, N81 and D83), and all four of these sites were confirmed to be targets of somatic mutation. D83 is affected in 39% of the *MEF2B* alterations, resulting in replacement of the charged aspartate with any of alanine, glycine or valine. Although we cannot yet predict the consequences of these substitutions on protein function, it seems likely that their effect would have an impact on the ability of MEF2B to facilitate gene expression and thus have a role in promoting the malignant transformation of germinal centre B cells to lymphoma (Supplementary Discussion).

MEF2B mutations can be linked to *CREBBP* and *EP300* mutations, and to recurrent Y641 mutations in *EZH2* (ref. 13). One target of CREBBP/EP300 HAT activity is H3K27, which is methylated by EZH2 to repress transcription. There is evidence that the action of EZH2 antagonizes that of CREBBP/EP300 (ref. 39). One function of MEF2 is to recruit either HDACs or CREBBP/EP300 to target genes[40], and it has been suggested that HDACs compete with CREBBP/EP300 for the same binding site on MEF2 (ref. 41). Under normal Ca^{2+} levels, MEF2 is bound by type IIa HDACs, which maintain the tails of histone proteins in a deacetylated repressive chromatin state[42]. Increased cytoplasmic Ca^{2+} levels induce the nuclear export of HDACs, enabling the recruitment of HATs such as CREBBP/EP300, facilitating transcription at MEF2 target genes. Mutation of *CREBBP*, *EP300* or *MEF2B* may have an impact on the expression of MEF2 target genes owing to reduced acetylation of nucleosomes near these genes (Supplementary Figure 5; Supplementary Discussion). In light of the recent finding that heterozygous *EZH2* Y641 mutations enhance overall H3K27 trimethylation activity of PRC2 (refs 43, 44), it is possible that mutation of both *MLL2* and *EZH2* could cooperate in reducing the expression of some of the same target genes. Our data indicate that (1) post-transcriptional modification of histones is of key importance in germinal centre B cells and (2) deregulated histone modification due to these mutations is likely to result in reduced acetylation and enhanced methylation, and acts as a core driver event in the development of NHL (Supplementary Figure 5).

METHODS SUMMARY

All samples analysed contained at least 50% tumour cells. Genomes, exomes and transcriptomes were sequenced using a combination of Illumina GAIIx and HiSeq 2000 instruments to read lengths of between 36 and 100 nucleotides. Exome capture was performed using the Agilent SureSelect Target Enrichment System Protocol (Version 1.0, September 2009). Alignment was accomplished using BWA[45] and variants were identified using SNVmix[46]. Variants were manually reviewed in IGV and were confirmed (where applicable) by PCR followed by either Sanger sequencing or Illumina re-sequencing. Structural rearrangements in genomes and transcriptomes were identified using ABySS[47]. Gene expression values used for subtype assignment were calculated as reads per kilobase gene model per million mapped reads (RPKM) values[48] and subtypes were assigned using an adaptation of the method developed for data from Affymetrix expression arrays[49] trained with samples previously classified by this standard approach.

Received 13 November 2010; accepted 7 July 2011.
Published online 27 July 2011.

1. Anderson, J. R., Armitage, J. O., Weisenburger, D. D., Non-Hodgkin's Lymphoma Classification Project. Epidemiology of the non-Hodgkin's lymphomas: distributions of the major subtypes differ by geographic locations. *Ann. Oncol.* **9,** 717–720 (1998).
2. Lenz, G. & Staudt, L. M. Aggressive lymphomas. *N. Engl. J. Med.* **362,** 1417–1429 (2010).
3. Horsman, D. E. *et al.* Follicular lymphoma lacking the t(14;18)(q32;q21): identification of two disease subtypes. *Br. J. Haematol.* **120,** 424–433 (2003).
4. Iqbal, J. *et al.* *BCL2* translocation defines a unique tumor subset within the germinal center B-cell-like diffuse large B-cell lymphoma. *Am. J. Pathol.* **165,** 159–166 (2004).
5. Lenz, G. *et al.* Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc. Natl Acad. Sci. USA* **105,** 13520–13525 (2008).
6. Pasqualucci, L. *et al.* Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J. Exp. Med.* **203,** 311–317 (2006).
7. Kato, M. *et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* **459,** 712–716 (2009).
8. Compagno, M. *et al.* Mutations of multiple genes cause deregulation of NF-κB in diffuse large B-cell lymphoma. *Nature* **459,** 717–721 (2009).
9. Davis, R. E. *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463,** 88–92 (2010).
10. Ngo, V. N. *et al.* Oncogenically active *MYD88* mutations in human lymphoma. *Nature* **470,** 115–119 (2011).
11. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361,** 1058–1066 (2009).
12. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
13. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature Genet.* **42,** 181–185 (2010).
14. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4,** 177–183 (2004).
15. Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471,** 189–195 (2011).
16. Yusuf, I., Zhu, X., Kharas, M. G., Chen, J. & Fruman, D. A. Optimal B-cell proliferation requires phosphoinositide 3-kinase-dependent inactivation of FOXO transcription factors. *Blood* **104,** 784–787 (2004).
17. Saito, M. *et al.* BCL6 suppression of BCL2 via Miz1 and its disruption in diffuse large B cell lymphoma. *Proc. Natl Acad. Sci. USA* **106,** 11294–11299 (2009).
18. Lenz, G. *et al.* Oncogenic *CARD11* mutations in human diffuse large B cell lymphoma. *Science* **319,** 1676–1679 (2008).

19. Greenman, C., Wooster, R., Futreal, P. A., Stratton, M. R. & Easton, D. F. Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* **173,** 2187–2198 (2006).
20. Cheung, K. J. *et al.* Acquired *TNFRSF14* mutations in follicular lymphoma are associated with worse prognosis. *Cancer Res.* **70,** 9166–9174 (2010).
21. Du, M. Q. *et al. BCL10* gene mutation in lymphoma. *Blood* **95,** 3885–3890 (2000).
22. Kreutz, B., Hajicek, N., Yau, D. M., Nakamura, S. & Kozasa, T. Distinct regions of Gα13 participate in its regulatory interactions with RGS homology domain-containing RhoGEFs. *Cell. Signal.* **19,** 1681–1689 (2007).
23. Bhattacharyya, R. & Wedegaertner, P. Gα13 requires palmitoylation for plasma membrane localization, Rho-dependent signaling, and promotion of p115-RhoGEF membrane binding. *J. Biol. Chem.* **275,** 14992–14999 (2000).
24. Manganello, J. M., Huang, J., Kozasa, T., Voyno-Yasenetskaya, T. A. & Le Breton, G. C. Protein kinase A-mediated phosphorylation of the Gα13 switch I region alters the Gαβγ13-G protein-coupled receptor complex and inhibits Rho activation. *J. Biol. Chem.* **278,** 124–130 (2003).
25. Brunet, A. *et al.* Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). *Mol. Cell. Biol.* **21,** 952–965 (2001).
26. Tai, D. J. C., Su, C.-C., Ma, Y.-L. & Lee, E. H. Y. SGK1 phosphorylation of IκB kinase α and p300 Up-regulates NF-κB activity and increases *N*-methyl-D-aspartate receptor NR2A and NR2B expression. *J. Biol. Chem.* **284,** 4073–4089 (2009).
27. Mo, J. *et al.* Serum- and glucocorticoid-inducible kinase 1 (SGK1) controls Notch1 signaling by downregulation of protein stability through Fbw7 ubiquitin ligase. *J. Cell Sci.* **124,** 100–112 (2011).
28. Young, K. H. *et al.* Structural profiles of *TP53* gene mutations predict clinical outcome in diffuse large B-cell lymphoma: an international collaborative study. *Blood* **112,** 3088–3098 (2008).
29. Shilatifard, A. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr. Opin. Cell Biol.* **20,** 341–348 (2008).
30. Milne, T. *et al.* MLL targets SET domain methyltransferase activity to *Hox* gene promoters. *Mol. Cell* **10,** 1107–1117 (2002).
31. Krumlauf, R. *Hox* genes in vertebrate development. *Cell* **78,** 191–201 (1994).
32. Canaani, E. *et al.* ALL-1//MLL1, a homologue of *Drosophila* TRITHORAX, modifies chromatin and is directly involved in infant acute leukaemia. *Br. J. Cancer* **90,** 756–760 (2004).
33. Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell. Biol.* **29,** 6074–6085 (2009).
34. Issaeva, I. *et al.* Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol. Cell. Biol.* **27,** 1889–1903 (2007).
35. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463,** 184–190 (2010).
36. Dalgliesh, G. L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463,** 360–363 (2010).
37. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331,** 435–439 (2011).
38. Iqbal, J. *et al.* Distinctive patterns of BCL6 molecular alterations and their functional consequences in different subgroups of diffuse large B-cell lymphoma. *Leukemia* **21,** 2332–2343 (2007).
39. Pasini, D. *et al.* Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* (2010).
40. Giordano, A. & Avantaggiati, M. p300 and CBP: partners for life and death. *J. Cell. Physiol.* **181,** 218–230 (1999).
41. Han, A., He, J., Wu, Y., Liu, J. O. & Chen, L. Mechanism of recruitment of class II histone deacetylases by myocyte enhancer factor-2. *J. Mol. Biol.* **345,** 91–102 (2005).
42. Youn, H. & Liu, J. Cabin1 represses MEF2-dependent Nur77 expression and T cell apoptosis by controlling association of histone deacetylases and acetylases with MEF2. *Immunity* **13,** 85–94 (2000).
43. Yap, D. B. *et al.* Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* **117,** 2451–2459 (2011).
44. Sneeringer, C. J. *et al.* Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc. Natl Acad. Sci. USA* **107,** 20980–20985 (2010).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
46. Goya, R. *et al.* SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* **26,** 730–736 (2010).
47. Robertson, G. *et al. De novo* assembly and analysis of RNA-seq data. *Nature Methods* **7,** 909–912 (2010).
48. Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5,** 621–628 (2008).
49. Wright, G. *et al.* A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl Acad. Sci. USA* **100,** 9991–9996 (2003).
50. He, J. *et al.* Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly. *Nucleic Acids Res.* (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was funded in part by funding from the National Cancer Institute Office of Cancer Genomics (Contract No. HHSN261200800001E), the Terry Fox Foundation (grant 019001, Biology of Cancer: Insights from Genomic Analyses of Lymphoid Neoplasms) and Genome Canada/Genome British Columbia Grant Competition III (Project Title: High Resolution Analysis of Follicular Lymphoma Genomes) to J.M.C., R.D.G. and M.A.M. We acknowledge support from NIH grants P50CA130805-01 ''SPORE in Lymphoma, Tissue Resource Core (PI Fisher)'' and 1U01CA114778 ''Molecular Signatures to Improve Diagnosis and Outcome in Lymphoma (PI Chan)''. A.J.M. is a Career Development Program Fellow of the Leukemia and Lymphoma Society. N.A.J. was a research fellow of the Terry Fox Foundation (award NCIC 019005) and the Michael Smith Foundation for Health Research (ST-PDF-01793). M.A.M. is a Terry Fox Young Investigator and a Michael Smith Senior Research Scholar. R.D.M. is a Vanier Scholar (CIHR) and holds a MSFHR senior graduate studentship. M.M.-L. acknowledges support from a Postdoctoral Fellowship from the Spanish Ministry of Education, under the ''Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I-D+i 2008-2011''. D.W.S. was supported by the Terry Fox Foundation Strategic Health Research Training Program in Cancer Research at Canadian Institutes of Health Research (Grant No. TGT-53912). J.J.S. acknowledges funding from The Canadian Cancer Society and the Canadian Institutes of Health Research. R.G. is supported by a UBC Four Year Fellowship. I.M.M. acknowledges the Canadian Foundation for Innovation for a Leaders Opportunity Fund. The laboratory work for this study was undertaken at the Genome Sciences Centre, British Columbia Cancer Research Centre and the Centre for Translational and Applied Genomics, a program of the Provincial Health Services Authority Laboratories. The authors would like to thank C. Greenman for supplying his software and also acknowledge D. Gerhard and S. Aparicio for discussions and guidance. Special thanks to C. Suragh, R. Roscoe, A. Troussard and A. Drobnies for expert project management assistance, and to the Library Construction, Sequencing and Bioinformatics teams at the Genome Sciences Centre. The content of this publication does not necessarily reflect the views of policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Author Contributions M.A.M., R.D.G., D.E.H., M.H. and J.M.C. conceived of the study and led the design of the experiments. R.D.M. performed the analysis of sequence data, identified mutations and, with M.M.-L., A.J.M. and M.A.M., produced figures and wrote the manuscript. M.M.-L., A.J.M., D.L.T., S.Chan, S.Chittarajan, D.S., H.M., J.S., M.M., T.Z., A.D., K.T., Y.B., M.R.F., J.T.-W. and T.M.S. designed and performed experiments to amplify, discover and validate mutations. R.G., M.G. and I.M.M. contributed to analyses and reviewed the manuscript. N.A.J., M.B., B.W. and B.M. prepared the samples, performed sample sorting and COO analysis and contributed to the text. A.B.-W. and J.J.S. collected and prepared constitutional DNA samples. K.L.M., R.C., S.L., M.F. and S.J. generated *de novo* assemblies and identified mutations. M.K., S.R., M.G., O.Y. and E.Y.Z. wrote software and contributed to figures. R.D.C. performed copy number analysis and produced a figure and S.B.-N. performed confirmatory FISH experiments. Y.Z. and A.T. produced the sequencing libraries. I.B., R.H., S.J.M.J., R.M., J.S. and M.H. contributed to the development of experimental and analytical protocols. L.R. provided materials and reviewed the manuscript.

Author Information The SRA accession number for the submission of the data not included in previous publications is SRP001599, which is linked to the dbGAP study accession phs000235.v2.p1. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.A.M. (mmarra@bcgsc.ca).