Supplementary Methods

Genome assembly

The common practice in the field of *de novo* sequence assembly is to select a global optimum k value that is a compromise between sensitivity and specificity. However, we find this limiting for cancer studies that aim to detect variations because whole genome sequencing experiments return fluctuating coverage redundancies. To compensate for that, we assemble our data at the single-end stage using three k-mer lengths (low/medium/high) and then merge the three assemblies, an approach that we refer to as a crucible assembly (Figure 1, below). The single-end stage is where a de Bruijn graph is formed out of reads, and unambiguous paths along the graph are joined after certain noise cleaning and error and variance removal stages. The approach allows us to "rescue" k-mers in regions of low coverage for an assembly at a higher k value. The low and medium k-mer single end contigs are then supplied to the assembly process along with the raw reads for a full assembly at a higher k value. The approach allows us to "rescue" k-mers in regions of low coverage for an assembly at a higher k value.

The contigs from assemblies (with k-mer values of 24, 44 and 64) were merged, aligned and post-processed using the standard Trans-ABySS pipeline (version 1.2.3)[1]. The optimal k-mer lengths are highly dependent on the coverage redundancy, read length, and the target genome. The values we used in this work were tuned on historical data with similar coverage and for read lengths of 100 bp. Contigs were aligned against the reference genome using BLAT with the following parameters: stepSize=5, repMatch=2253, minScore=0 and minIdentity=0. Potential fusion candidates were identified as contigs that could not be mapped to a single unique location for at least 95% of the sequence. Split alignments with the following characteristics were considered

candidates: the top two alignments have at least 98% identity; one alignment does not reside entirely within its partner in terms of genomic coordinates; alignments do not overlap by more than 5% in terms of contig coordinates; alignments do not overlap in terms of genome coordinates; and the two alignments together represent at least 90% of the contig. To further filter the candidate events, we leverage alignment of reads to both contigs and genome. Reads were aligned to the contigs using Bowtie 0.12.5 and to the genome with all exons juxtaposed with their neighbors using BWA. Reads mapped across exon junctions are repositioned to their original genomic positions. Candidate fusion cases were then filtered by requiring at least 2 reads spanning the contig breakpoint with at least 4 nucleotides on either side, and at least 4 read pairs flanking the genomic breakpoint.

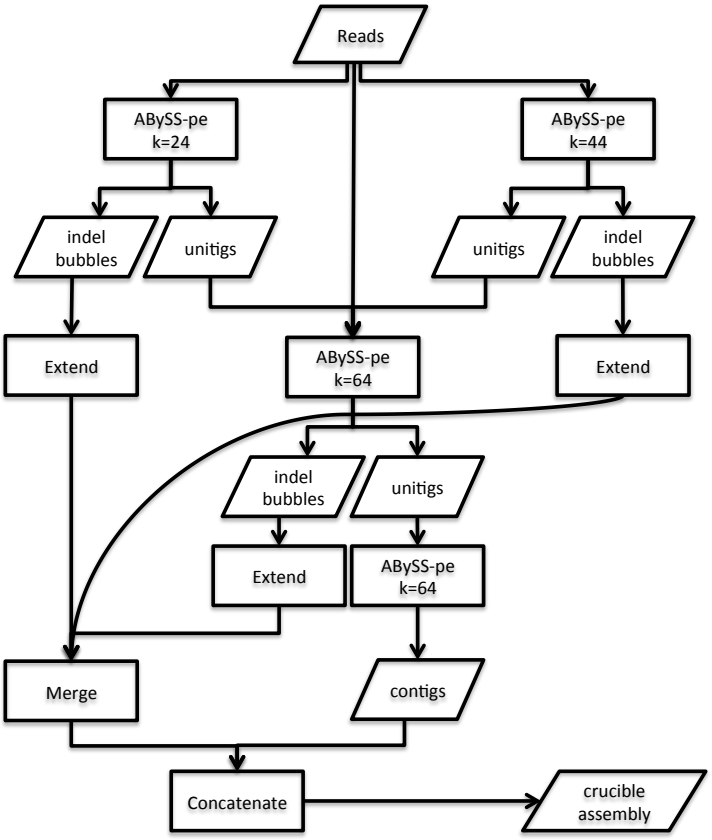


Figure 1: Flowchart representation of the crucible assembly process

Indel discovery by de novo assembly

Contigs were aligned against the reference genome using BLAT with the following parameters: stepSize=5, repMatch=2253, minScore=0, minIdentity=0. Contigs that were mapped to a single unique location for at least 90% of the sequence were retained for further analysis. Alignment blocks were determined for both the query (contig) and target (genome). Gaps between consecutive target blocks with absence of a corresponding query gap were identified as putative deletions. Vice versa, gaps between consecutive query blocks with absence of a corresponding absence of target gap were identified as putative insertions. Putative deletion gaps were screened for the presence of flanking canonical splicing motif to filter away intronic gaps. The resulting deletion cases were further filtered by requiring at least 3 reads spanning the contig breakpoint with at least 8 nucleotides in either direction. Candidate insertion cases were further filtered by requiring at least 3 reads harboring the insertion sequences with at least 8 flanking nucleotides on either side.

Cellularity estimates

Tumor cellularity was first estimated for each case by standard H&E staining. Cellularity estimates were then calculated directly from the WGS data using APOLLOH[2]. Cellularity estimates were revised when the value provided by APOLLOH was lower than the estimate from pathology. For comparison, we also derived tumor cellularity estimates in each case using the mean allele frequency of heterozygous SNVs detected.

Duplication timing

We used the approach described by Nik-Zainal et al[3] and Greenman et al[4] to compute the relative timing of duplication events. Conceptually, if both SNVs and large

chromosomal duplications are present, then the order in which these occurred can be inferred from the ratio of reads containing the SNV. For instance, if a chromosome is duplicated to result in three chromosomal copies then the mutations that occurred before the duplication will be present in 2/3 of the copies, and thus 2/3 of the reads; if mutations occur after the duplication, they will be present in 1/3 of the reads. Considering the ratios of reads of all mutations in a duplicated region, the relative timing (as a proportion of mutations) of the duplication can be inferred. This method requires taking into account mutations from the non-duplicated copy of the chromosome as well as any normal (stromal) contamination.

The first step in this analysis is to divide each tumor genome into continuous segments of the same copy number, and additionally, a consistent number of copies of each parental chromosome. Integer copy number was computed by comparing GC-corrected relative depth of tumor and normal reads in bins across the genome followed by Hidden Markov Model based segmentation, using cnaseq[5]. The number of copies of each parental chromosome in each segment was then inferred based on SNP allelic ratios using the package APOLLOH v0.1.1[2]. The output of APOLLOH and cnaseq were used to determine the major (M) and minor (m) chromosomal copy numbers for each segment of each chromosome in each DLBCL genome.

The second step is to determine the set of somatic single-base mutations, and for each mutation the most likely mutation ploidy or multiplicity (r), indicating the number of chromosomal copies on which that mutation resides. For timing analysis, all mutations across the genome, not coding mutations only, must be considered to provide sufficient numbers for accurate calculation. Only mutations of very high confidence, called by

SNVMix and subsequently given MutationSeq scores ≥ 0.5, were considered. To estimate the most likely ploidy for each mutation, the proportion of sampled cells thought to be normal cell contamination (π) was calculated as 1- (estimated tumor content). Given these values, the following probability derived from Greenman et al[4] was calculated for each possible ploidy of each mutation:

$$\Pr(r|M,m,n_s,n_w) \propto \left(\frac{(1-\pi)r}{(1-\pi)(M+m)+2\pi}\right)^{n_s}\left(1-\frac{(1-\pi)r}{(1-\pi)(M+m)+2\pi}\right)^{n_w}, r \in \{1,2,\dots,M\}$$

Where n_s represents the number of reads containing the somatic mutation, and n_w represents the number of reads representing the reference allele at that position. The ploidy of each mutation was chosen as that which gave the maximum probability using the above formula. It must be noted that due to the depth of whole genome sequencing, the sub-clonal structure of the entire genome cannot be determined accurately, and sub-clonal inference of individual mutations is not possible without further deeper sequencing as was done for a subset of mutations as part of mutation verification (see below).

The third step in this analysis is to compute, with confidence intervals, the relative timing of each duplication event observed in the DLBCL genomes with respect to mutation accumulation. For each segment, the number of mutations at each ploidy (N_p) are summed. Mutations may be present at ploidy 1 (N_1) or any ploidy up to M (N_M), the copy number of the major chromosome. For instance, in the simple case of a single duplication from two chromosomal copies to three, m = 1 and M = 2. Mutations may be present at ploidy 1, in which case they occurred after the duplication event or derive from the unduplicated chromosome, or at ploidy 2, in which case they occurred before the duplication on the duplicated copy of the chromosome. Based on the count of mutations

at each ploidy in a duplicated segment, for specific cases of the number of copies of each parental chromosome (m and M), the relative timing of the duplication event can be calculated. To compute timing directly the following must be true: (1) there is a duplication, i.e. $M \geq 2$ and (2) only one of the two chromosomal copies is duplicated, i.e. $m \leq 1$ and (3) the other chromosomal copy has been duplicated a maximum of two times, i.e. $M \leq 3$. In these situations, the fraction of mutational time at which the chromosomal gains occurred can be estimated using the methods described in Nik-Zainal et al[3], using the following sets of equations (P. Campbell, personal communication), where t_1 is the fraction of time before the first duplication and t_2 is the fraction of time after the first duplication. In the case of two duplications (M=3), t_3 is the fraction of time after the second duplication. As noted above, N_p is the number of mutations observed in the segment with ploidy *p*. Importantly, these equations take into account the mutations derived from the non-duplicated chromosomal copies, which will have ploidy 1 and could have occurred at any point in mutational time.

Case 1: M = 2, m = 0

Proportional time of duplication = $\frac{t_1}{t_1 + t_2} = \frac{N_2}{N_2 + \left(\frac{N_1}{2}\right)}$

Case 2: M = 2, m = 1

Proportional time of duplication = $\frac{t_1}{t_1 + t_2} = \frac{N_2}{N_2 + \left(\frac{N_1 - N_2}{3}\right)}$

Case 3: M = 3, m = 0

Proportional time of 1st duplication = $\frac{t_1}{t_1 + t_2 + t_3} = \frac{N_3}{N_3 + N_2 + \left(\frac{N_1 - N_2}{3}\right)}$

Proportional time of 2nd duplication = $\frac{t_1 + t_2}{t_1 + t_2 + t_3} = \frac{N_3 + N_2}{N_3 + N_2 + \left(\frac{N_1 - N_2}{3}\right)}$

Case 4: M = 3, m = 1

Proportional time of 1st duplication = $\frac{t_1}{t_1 + t_2 + t_3} = \frac{N_3}{N_3 + N_2 + \left(\frac{N_1 - N_3 - 2N_2}{4}\right)}$

Proportional time of 2nd duplication = $\frac{t_1 + t_2}{t_1 + t_2 + t_3} = \frac{N_3 + N_2}{N_3 + N_2 + \left(\frac{N_1 - N_3 - 2N_2}{4}\right)}$

The point estimate for timing of each duplication as calculated above will necessarily be in the range 0 (before all mutations occurred) to 1 (after all mutations occurred). Bootstrapping was performed by resampling the mutations in a segment with replacement 10,000 times, and re-computing the segment timing for each iteration. The distribution of timings calculated by bootstrapping was used to compute lower and upper bounds for 95% confidence intervals. For cases where there are multiple duplications (M>3 or m>1), such as in the *REL* amplicon, an exact point estimate cannot be determined, but by making assumptions about which chromosomal copy underwent duplication, the approximate time of each duplication can be estimated.

Validation of rearrangements and fusion transcripts

To validate fusion transcripts identified from RNA-seq, PCR primers were designed to flank the gene fusion breakpoints in the sequence contigs and were used to amplify cDNA prepared from 100 ng of total RNA for each sample using a different enzyme than that used for library construction. Structural rearrangements were validated using a similar approach. ABySS contigs supporting the event were extended to contain

600 nucleotides flanking the genomic positions of the aligned contig. PCR primers were designed against genomic DNA at the breakpoint position at an optimal Tm of 64°C with a desired size range of 200-400 bp. We selected 71 events identified by the Trans-ABySS pipeline of which we were able to design primers to amplify 66. In all cases, successful amplicons were purified then confirmed by Sanger capillary sequencing. Of the selected breakpoints, 54 were confirmed as somatic including 14 inversions, 8 translocations, 10 duplications and 22 deletions with the remaining primer sets failing to produce an amplicon (11) or a clean sequencing result (1).

CNA detection from SNP arrays

Tumor DNA from 96 DLBCLs, including the 40 that were sequenced, was analyzed using Affymetrix SNP 6.0 arrays. Data were normalized using CRMAv2, implemented in R package aroma.affymetrics with default settings[6]. Log ratios were computed by normalizing against a reference based on 270 HapMap samples. Segmentation employed a modified version of CNA-HMMer that was adapted for the Affymetrix SNP 6.0 platform. Another extension is the inclusion of additional copy number states (6 total), which offers a more intuitive interpretation of DNA dosage in cancer. The hidden Markov model performs segmentation of the log ratio data and predicts discrete copy number status for each resulting segment from the set of five somatic states (homozygous deletion, hemizygous deletion, gain, amplification, and high-level amplification) and one neutral copy number state.

Confirmation of somatic status for indels and rearrangements

Transcriptome fusion events were validated against the matching genome (tumor or normal) by finding read pairs mapped to the two breakpoint regions. A breakpoint

region was defined by adding a buffer of 200kb intronic to the boundary of the fused exon to allow for intronic sequences, and a buffer of 1kb to the other side of the breakpoint. Small-scale transcriptome indel events were validated against the genome by looking for similar indels reported by short-read alignments in the genomic region. In a lot of cases indels involve repeat sequences in which start coordinate can be reported at both the start and end of span of the event. In such cases, indel events were considered identical when the sequences flanking the events were identical. For insertions larger than a read length, we report from the genome BAM file the total number of reads with clipped bases that match part or the entirety of the insertion sequence, reads upstream of the breakpoint with unmapped mates that match part or the entirety of the insertion sequence, and reads downstream of the breakpoint with unmapped mates that match part or the entirety of insertion sequence. For deletions larger than a read length, we report the number of read pairs that mapped to the two breakpoint regions, where a breakpoint region is defined by the breakpoint coordinate with 2kb buffer on either side. An event is reported as 'somatic' where there is at least 2 read support (read pairs or spanning reads, depending on the scale of the event) from the tumor BAM file and 0 read support from the normal BAM file. An event is reported as 'germline' where there is at least 1 read support from the normal BAM file.

SNV Verification

PCR primers were directly tagged with portions of the Illumina adapters to circumvent the adapter ligation steps. The forward PCR primers were tagged with 5'-CGCTCTTCCGATCTCTG and the reverse PCR primers were tagged with 5'-TGCTCTTCCGATCTGAC. PCR amplification proceeded using Phusion high-fidelity

DNA polymerase kit (Fisher Scientific, cat# F540L) as per manufacturer's instructions. Reactions were set up in 96-well plates and comprised of 0.5 μM forward primer, 0.5 μM reverse primer, 1 ng of genomic DNA template, 5X Phusion HF Buffer, 0.2 μM dNTPs, 3% DMSO, and 0.4 units of Phusion DNA polymerase. Reaction plates were cycled on a MJR Peltier Thermocycler (model PTC-225) with cycling conditions comprising a denaturation step at 98 °C for 30 sec, followed by 7 touchdown cycles of [98°C for 10 sec, 72°C for 15 sec (-1°C per cycle), 72°C for 15 sec], 28 cycles of [98°C for 10 sec, 66°C for 15 sec, 72°C for 15 sec], and a final extension step at 72°C for 10 min. Amplicons from each template were pooled together into wells of a 96-well plate for tumor templates and a second 96-well plate for normal templates. To generate libraries, the remaining portion of the Illumina adaptor sequences were added by a second round of PCR using Phusion with the following conditions: Denaturation at 98•C for 30 sec, 6 cycles of 98•C for 10 sec, 65•C for 30 sec, 72•C for 30 sec, and a final extension step at 72•C for 5 min. Pooled amplicon libraries were sequenced on the Illumina MiSeq platform as paired-end indexed 150bp using v1 chemistry and MCS software version 1.2.3.

Integrative analysis using DriverNet

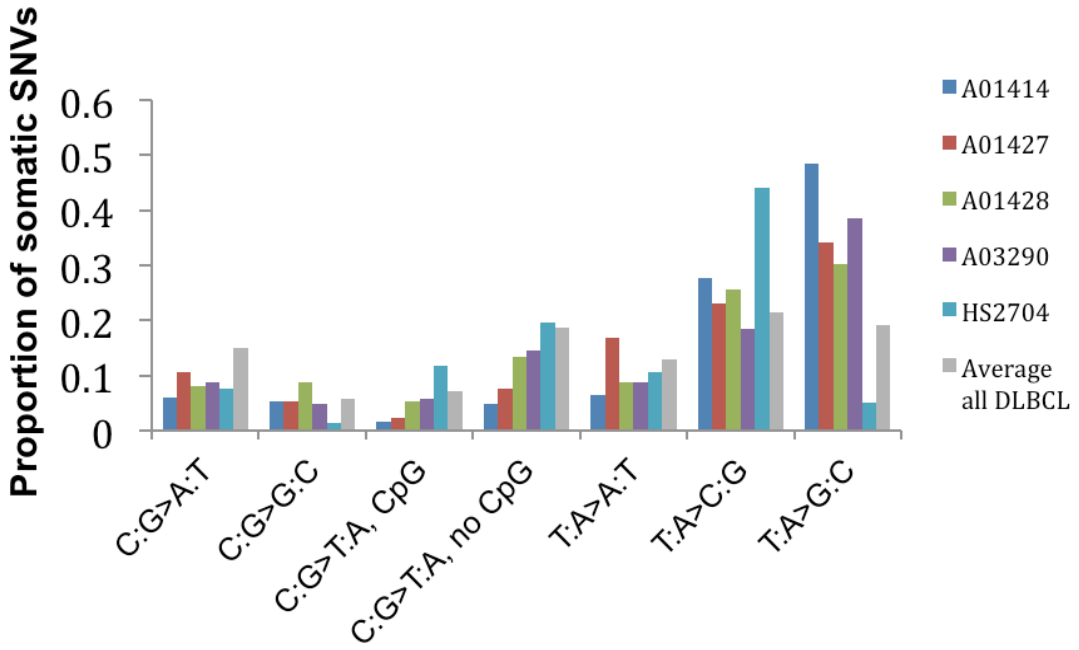
The three networks used by the DriverNet[7] model are STRING PPI network, Pathway Commons and wikipathway. The data from the latter is weighted based on the experiment-specific gene expression data. The methods assume that if a mutation in a tumor affects gene encoding protein A, and if protein A is connected to protein B in an interaction network and if the gene expression level for protein B in the same tumor is up- or down-regulated than those in the overall population level, such a change results in

a high functional impact score. Different types of mutations (e.g. truncating, nonsynonymous and synonymous) are weighted such that they will result in different functional impacts. We use the variable elimination algorithm to infer the functional probabilities of individual mutations as well as the population level functional probabilities of each gene.

References

1. Robertson, G. *et al.* De novo assembly and analysis of RNA-seq data. *Nat Meth* **7,** 909–912 (2010).
2. Ha, G. *et al.* 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 (2012).
3. Nik-Zainal, S. *et al.* The Life History of 21 Breast Cancers. *Cell* **149,** 994–1007 (2012).
4. Greenman, C. D. *et al.* Estimation of rearrangement phylogeny for cancer genomes. *Genome Res* (2011). doi:10.1101/gr.118414.110
5. Jones, S. J. *et al.* Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biol* **11,** R82–R82 (2010).
6. Bengtsson, H., Simpson, K., Bullard, J. & Hansen, K. aroma. affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. *Tech Report# 745* (2008).
7. Bashashati, A. *et al.* DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol* **13,** R124 (2012).

Supplemental Table S1: Overview of events detected from genomes and transcriptomes
Supplemental Table S2: Selective pressure estimates for SNVs and evidence for recurrence
Supplemental Table S3: All somatic SNVs identified from 40 genome pairs and 13 cell lines
Supplemental Table S4: All mutations identified in *GNA13* in the large DLBCL cohort
Supplemental Table S5: Structural variants detected
Supplemental Table S6: Fusion transcripts detected
Supplemental Table S7: Genes recurrently affected by any mutation type
Supplemental Table S8: Somatic SNVs affecting splice sites
Supplemental Table S9: Validated somatic indels
Supplemental Table S10: Timing results for all genomic amplifications
Supplemental Figure S1: Mutation spectrum in outlier cases compared to average (*embedded*)
Supplemental Figure S2: Mutations affecting histone H1 proteins in cell lines and patient samples *(attached)*
Supplemental Figure S3: Significantly mutated genes involved in cytokine response and B-cell homing and migration (*embedded*)
Supplemental Figure S4: Structural comparison of mutations affecting *GNAI2* to oncogenic *GNAS* mutations (*embedded*)
Supplemental Figure S5: Survival analysis of *GNA13* mutations in uniformly treated DLBCL (*embedded*)
Supplemental Figure S6: Examples of structural rearrangements detected in the genomes of individual cases (*attached*)
Supplemental Figure S7: Example of high-level amplification of *REL (embedded)*
Supplemental Figure S8: Overview of copy number variants detected in 92 DLBCLs *(embedded)*
Supplemental Figure S9: Overview of all fusion transcripts detected by RNA-seq (*attached*)
Supplemental Figure S10: Detailed representation of *ETV3*-IKZF3 fusion (*embedded*)
Supplemental Figure S11: Eight additional fusions with preserved reading frames (*embedded*)
Supplemental Figure S12. Deletions and fusions result in monoallelic expression of *TP63 (embedded)*
Supplemental Figure S13: Overview of genes commonly amplified, deleted or involved in rearrangements/fusions *(embedded)*
Supplemental Figure S14: Recurrence of deletions affecting *CDK11A* and *CDK11B (embedded)*
Supplemental Figure S15: Mutations with functional impact predicted by integrative analysis of mutation and expression data *(embedded)*
Supplemental Figure S16: Evidence for clonal and sub-clonal driver mutations *(embedded)*
Supplemental Figure S17: Timing of REL amplifications *(embedded)*
Supplemental Figure S18: Boxplots illustrating the variability in timing of most commonly duplicated chromosomes or chromosome arms *(embedded)*
Supplemental Figure S19: Examples of validated small deletions (*attached*)

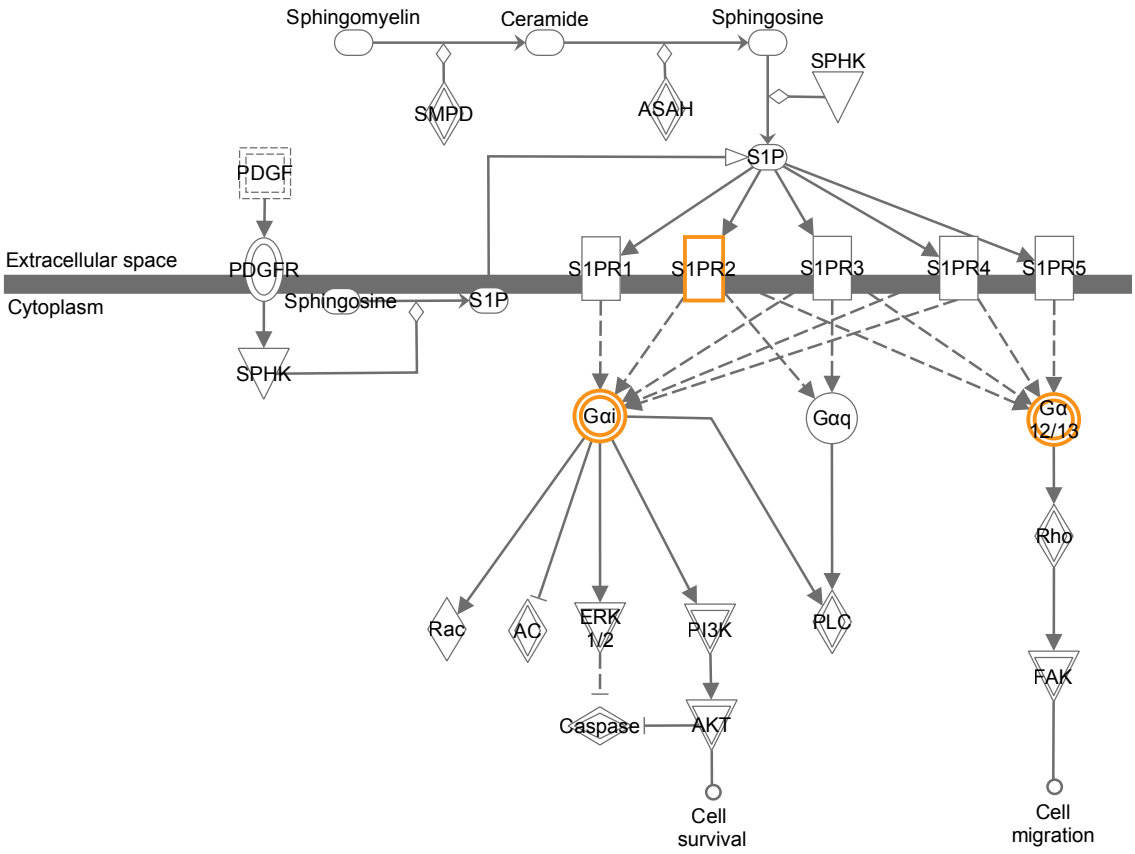


Supplemental Figure S1. Mutation spectrum in outlier cases compared to average. The mutation spectrum is shown for the five cases with the most discrepant patterns as well as the average (grey). The library with the most discrepant mutation spectrum contained an indel in *MSH3* as well as a nonsynonymous change in *MLH1*. This case as well as two of the additional outliers (A01427 and A01414) harbored mutations in one or more genes encoding DNA polymerase subunits including *POLE*, *POLA1* and *POLG* (Supplemental Table S1). There was not a strict correlation between the presence of mutations in POL genes and mutation spectrum but the five cases with mutations in these genes were all among the eight cases with the highest overall mutation load.

```
HIST1H1C        MSETAPAAPAAA--PPAEKAPVKKKA-AKKAGGTP--RKASGPPVSELITKAVAASKERS 55
HIST1H1D        MSETAPLAPTIP--APAEKTPVKKK--AKKAGATAGKRKASGPPVSELITKAVAASKERS 56
HIST1H1E        MSETAPAAPAAP--APAEKTPVKKKA-RKSAGAAK--RKASGPPVSELITKAVAASKERS 55
HIST1H1B        MSETAPAETATP--APVEKSPAKKKATKKAAGAGAAKRKATGPPVSELITKAVAASKERN 58
HIST1H1A        MSETVPPAPAAS--AAPEKPLAGKKAKKPAKAAAASKKKPAGPSVSELIVQAASSSKERG 58
HIST1H1T        MSETVPAASASAGVAAMEKLPTKKRG-RKPAGLISASRKVPNLSVSKLITEALSVSQERV 59
                ****.*  .: .  .. **  . *:      .     :* .. .**:**.:* : *:**

HIST1H1C        GVSLAALKKALAAAGYDVEKNNSRIKLGLKSLVSKGTLVQTKGTGASGSFKLNKKAASGE 115
HIST1H1D        GVSLAALKKALAAAGYDVEKNNSRIKLGLKSLVSKGTLVQTKGTGASGSFKLNKKAASGE 116
HIST1H1E        GVSLAALKKALAAAGYDVEKNNSRIKLGLKSLVSKGTLVQTKGTGASGSFKLNKKAASGE 115
HIST1H1B        GLSLAALKKALAAGGYDVEKNNSRIKLGLKSLVSKGTLVQTKGTGASGSFKLNKKAASGE 118
HIST1H1A        GVSLAALKKALAAAGYDVEKNNSRIKLGIKSLVSKGTLVQTKGTGASGSFKLNKKASSVE 118
HIST1H1T        GMSLVALKKALAAAGYDVEKNNSRIKLSLKSLVNKGILVQTRGTGASGSFKLSKKVIPKS 119
                *:**.********.**************.:****.** ****:**********.**. . .

HIST1H1C        AKPKVKKAGGTKPKKPVGAAKKPKKAAGGATPKKSAKKTPKKAKKPAAATVTKKVAKSPK 175
HIST1H1D        GKPKAKKAGAAKPRKPAGAAKKPKKVAGAATPKKSIKKTPKKVKKPATAAGTKKVAKSAK 176
HIST1H1E        AKPKAKKAGAAKAKKPAGAAKKPKKATGAATPKKSAKKTPKKAKKPAAAAGAKK-AKSPK 174
HIST1H1B        AKPKAKKAGAAKAKKPAGAT--PKKAKKAAGAKKAVKKTPKKAKKP-AAAGVKKVAKSPK 175
HIST1H1A        TKPGASKV--ATKTKATGASKKLKKATGAS--KKSVK-TPKKAKKP---AATRKSSKNPK 170
HIST1H1T        TRSKAKKSVSAKTKK--------LVLSRDSKSPKTAK-TNKRAKKP--RATTPKTVRSGR 168
                 :. ..*   :.  *              :   *: * * *:.***   : . *  :. :

HIST1H1C        KAKVA-KPKKAAKS--AAKAVK-----PKAAKP-----KVVKPKKAAPKKK 213
HIST1H1D        KVKTP-QPKKAAKSPAKAKAPK-----PKAAKPKSGKPKVTKAKKAAPKKK 221
HIST1H1E        KAKAA-KPKKAPKSPAKAKAVK-----PKAAKPKTAKPKAAKPKKAAAKKK 219
HIST1H1B        KAKAAAKPKKATKSPAKPKAVKPKAAKPKAAKPKAAKPKAAKAKKAAAKKK 226
HIST1H1A        KPKTV-KPKKVAKSPAKAKAVK-----PKAAKARVTKPKTAKPKKAAPKKK 215
HIST1H1T        KAKGA-KGKQQQKSPVKARASK----------SKLTQHHEVNVRKATSKK- 207
                * *   : *:  **   .:* *           .     : .: :**:.**
```

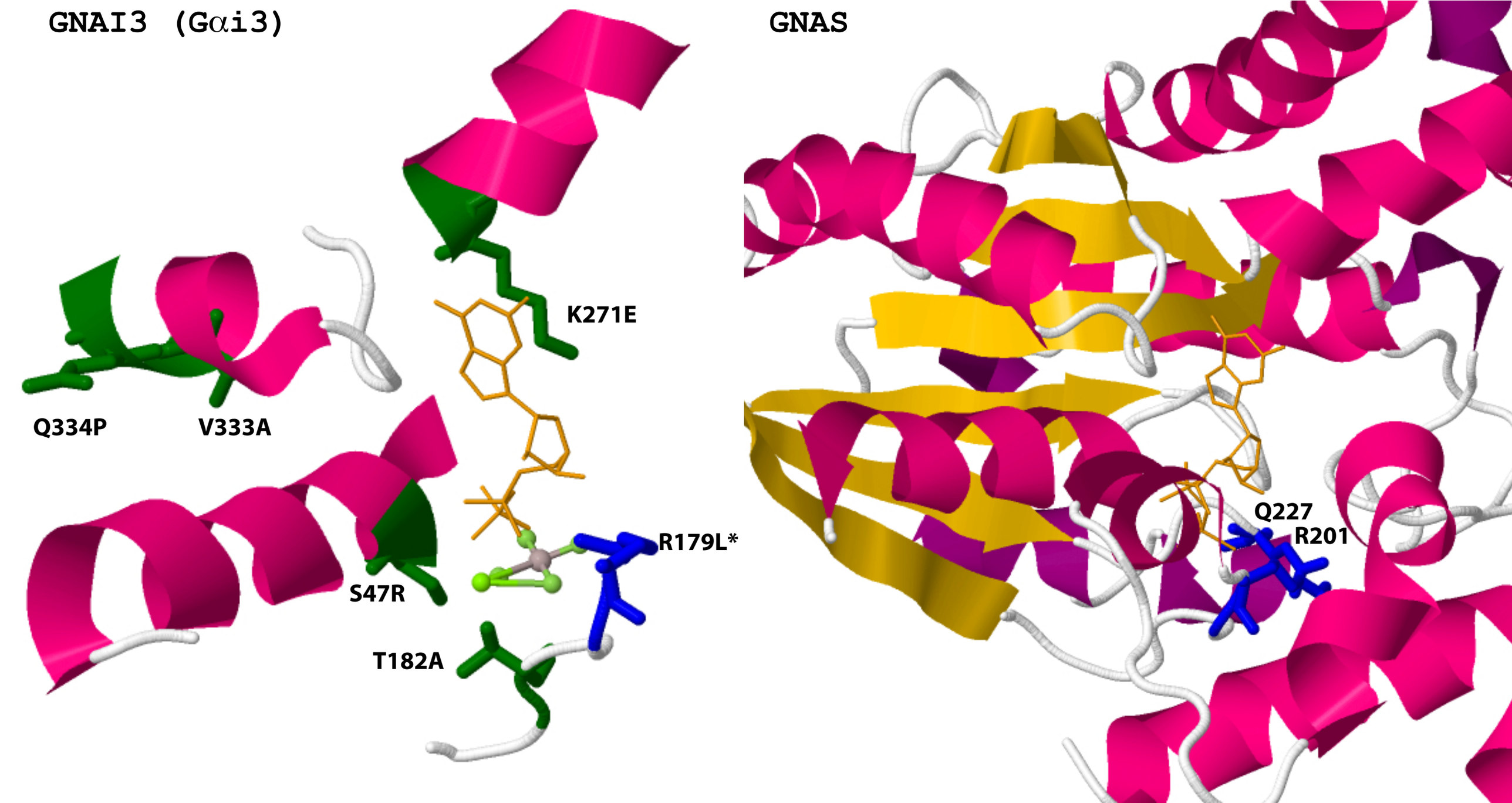
Supplemental Figure S2. Mutations affecting histone H1 proteins in cell lines and patient samples.

The histone proteins were a novel class of genes that was found significantly mutated in this study. We previously reported that *HIST1H1C* was significantly mutated in NHL but WGS revealed mutations in additional genes that encode variants of histone protein H1 including each of *HIST1H1C*, *D* and *E*. Mutations identified in patient samples by WGS are shown in blue and additional mutations identified by RNA-seq and subsequently confirmed to be somatic are shown in purple. Mutations identified in DLBCL cell lines are indicated in red (see Supplemental Table S3 for more detail). Some residues orthologous between these proteins were mutated (boxed). None of the mutated amino acids are known to be targets of post-transcriptional modifications. A small number of candidate mutations in the less conserved *HIST1H1A*, *HIST1H1B* and *HIST1H1T* genes were observed in patient samples and cell lines but none have been found to be somatic to date. The recurrence of mutations affecting histone proteins is particularly notable considering the emerging importance of epigenetic regulation in lymphoma[1,2]. Mutations affecting histone protein H3 have been described in pediatric brain cancers[3,4] but unlike that report, these mutations were typically non-recurrent and none of the observed mutations affect regions of these histones targeted by post-translational modifications. Mutations affecting histone genes have been reported in DLBCL[2,5-7] but their potential role in lymphomagenesis has not been appreciated.

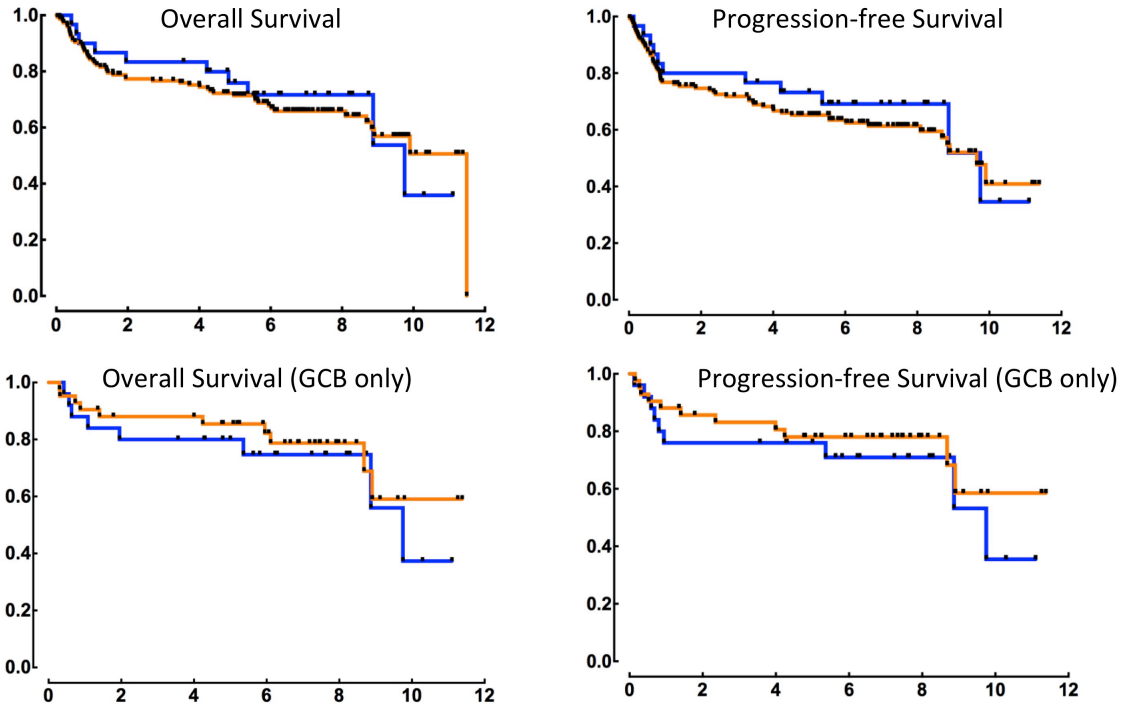


Supplemental Figure S3. Significantly mutated genes involved in cytokine response and B-cell homing and migration.

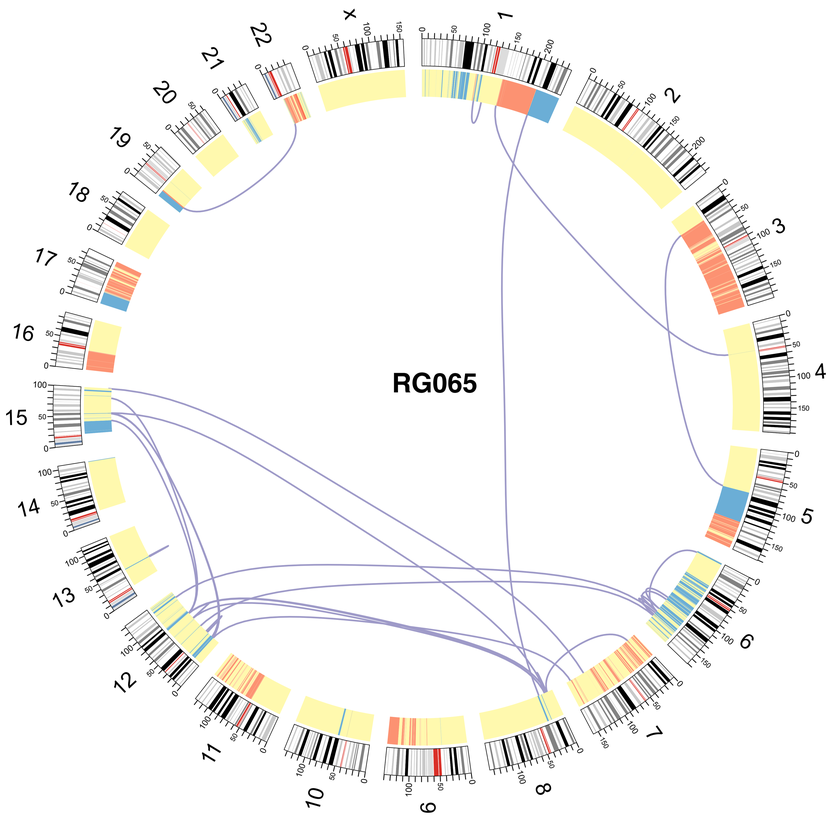
Response to cytokines present in the microenvironment is important in the regulation of B cell fate and localization within/homing to lymphoid tissues. We and others previously identified recurrence of mutations in *S1PR2*[7,8], which encodes $S1P_2$, a receptor for sphingosine-1-phosphate (S1P), a lipid that can promote migration of lymphocytes. Depending on the receptor expressed (such as $S1P_1$, $S1P_2$ or $S1P_3$), S1P signaling can either impact Rho or Rac activity and either promote retention of B cells in germinal centers or the egress from lymphoid tissues into circulation[9,10]. $S1P_2$ is highly expressed in germinal center B cells and, coupled with $G\alpha_{12}$ and $G\alpha_{13}$ promotes retention in the germinal centers. Of note, no mutations were observed in *GNA12*, which encodes $G\alpha_{12}$. $S1P_2$ can also signal through $G\alpha_{i2}$[1,2], which is encoded by *GNAI2*, a gene that we found significantly mutated in this study. Knock out of *S1PR2* in mice was demonstrated to result in increased Akt activity[3,4] and results in a GCB-like form of DLBCL[2,5-7]. In accordance with its role in transducing signals from activated $S1P_2$, $G\alpha_{13}$ has accordingly been shown to inhibit Akt activity[7,8]. $G\alpha_{13}$ typically couples with $S1P_2$ and, when stimulated by S1P, can induce motility by activating Rho and also suppress PI(3)K/Akt signaling[9,10]. In contrast, $G\alpha_i$ proteins including $G\alpha_{i2}$ couples more commonly with $S1P_1$ and promotes (rather than inhibits) PI(3)k/Akt activity[9].

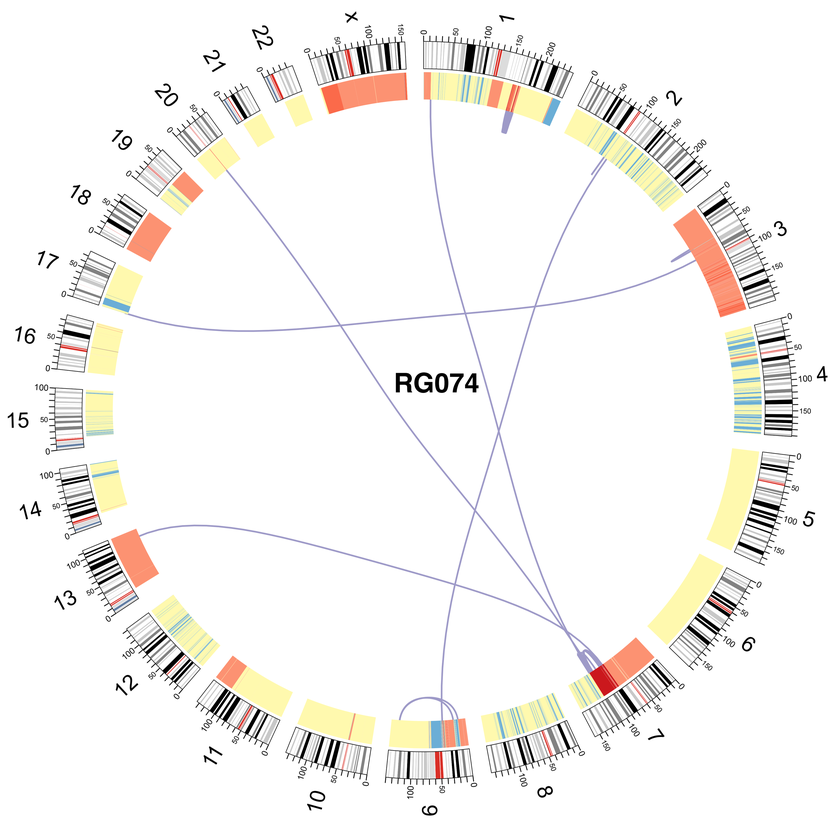

GNAI3 (Gαi3)
GNAS
K271E
Q334P
V333A
R179L*
S47R
T182A
Q227
R201

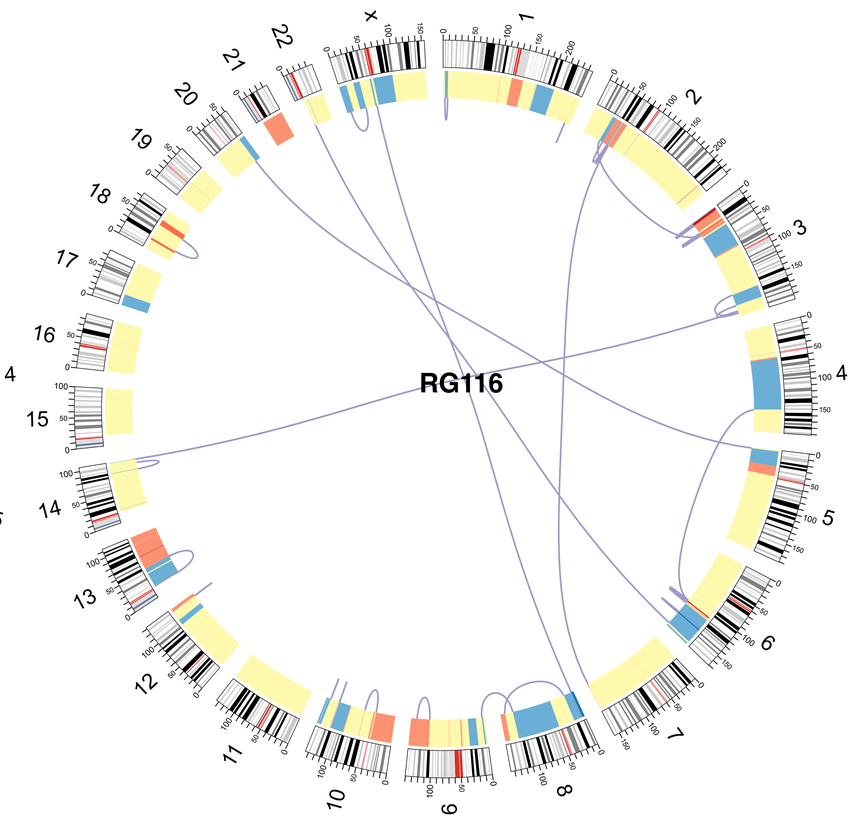
Supplementary Figure S4. Structural investigation of mutations affecting *GNAI2* and comparison to constitutively activating *GNAS* mutations.
GNAI2 was also significantly mutated in our cohort albeit less frequently than *GNA13*. In contrast to the mutation pattern we observed in *GNA13*, the *GNAI2* mutations were solely non-synonymous. Recurrent mutations in *GNAS*, *GNAO1*, *GNAQ* and *GNAI2* have all been reported in various human tumours with residues 201 and 227 in GNAS both common hot spots[11]. Mutations affecting R179 in *GNAI2* (orthologous to R201 in *GNAS*) have been reported in human endocrine tumours, indicating that dominant activating mutations of this protein may also be oncogenic in certain settings. The above images show the positioning of mutations observed in DLBCL mapped on the relevant regions of the solved structure of human GNAI3 (PDB: 2V4Z), a close homologue of *GNAI2*. The mutations in the DLBCL patients are shown in green, and R179, the residue orthologous to R201, is shown in blue. According to the solved structure of another human homologue, GNAI1, S47 and T181 are crucial residues that interact directly with Magnesium and the mutated residues Q334 and V333 reside within the switch III region of the protein[12]. One of these mutations V332A in GNAI1 (which corresponds to V333 in GNAI2) has been studied previously and the mutant enzyme shows a 14-fold increase in basal (receptor-independent) nucleotide exchange relative to wild type GNAI1, resulting in a constitutively active protein[12]. Interestingly, both S47 and T181 were noted to display considerable alterations in the GNAI1 structure with a separate constitutively active mutation in the switch III region (T329). Overall, these data indicate the potential for the observed mutations to result in constitutive activation of $G\alpha_{i2}$ in DLBCL.

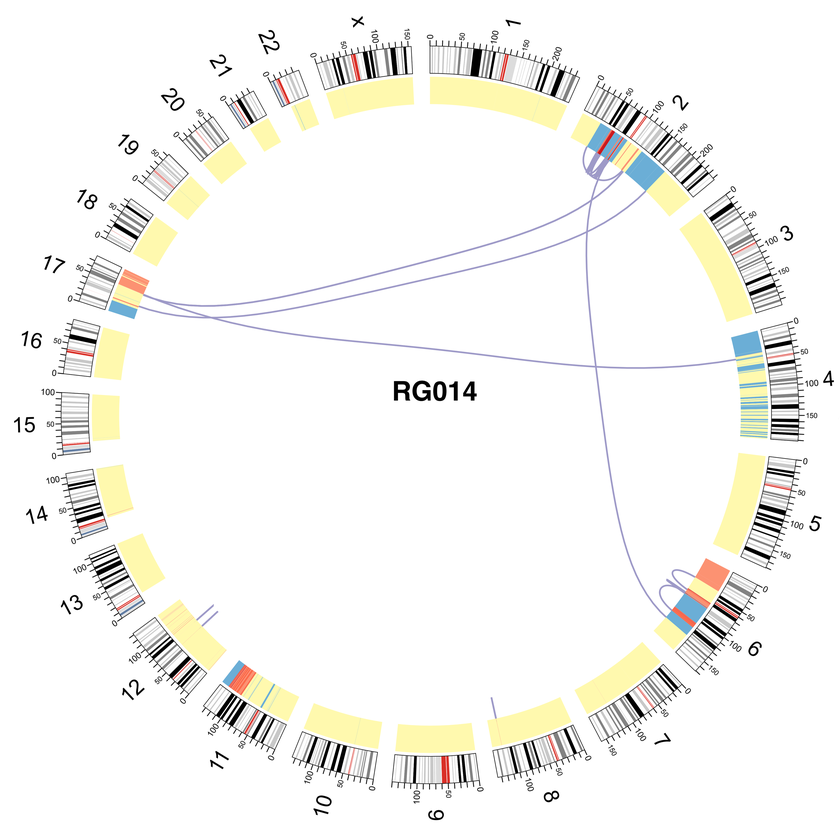


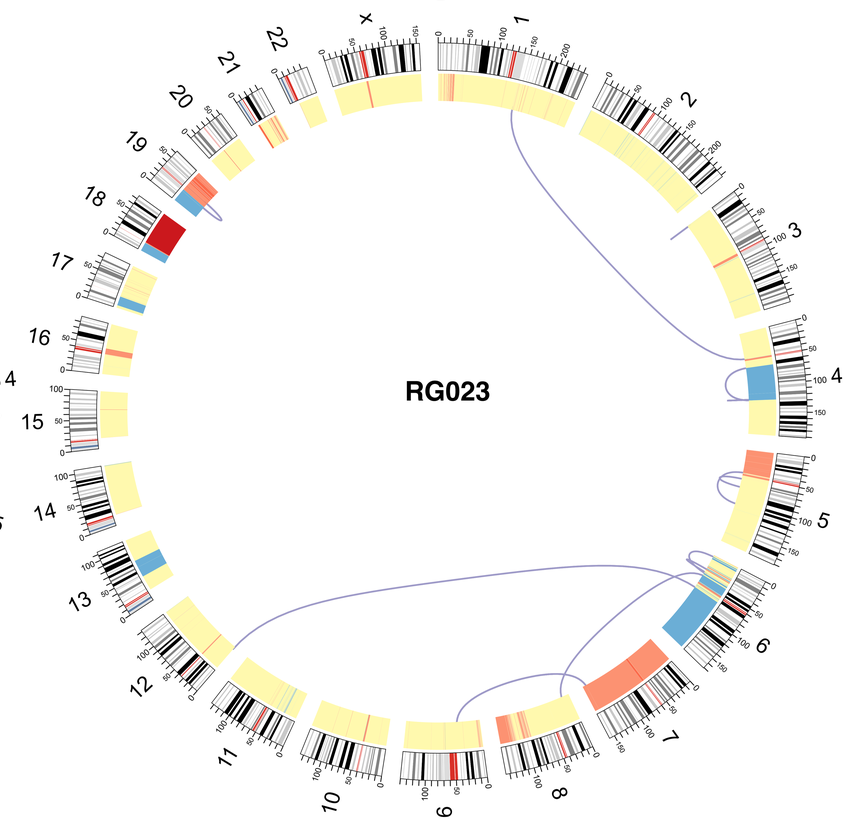
Supplemental Figure S5. Survival analysis of GNA13 mutations in uniformly treated DLBCL.
The overall survival (OS) and progression-free survival (PFS) was compared for patients with (blue) and without (orange) *GNA13* mutations detected as described in a recent

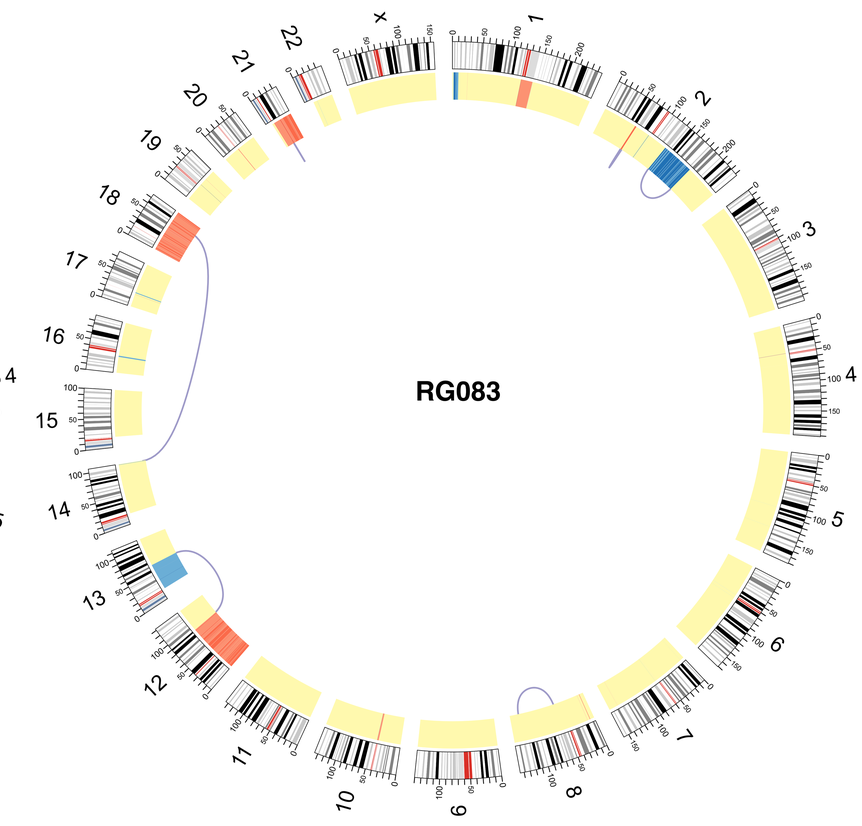

RG065

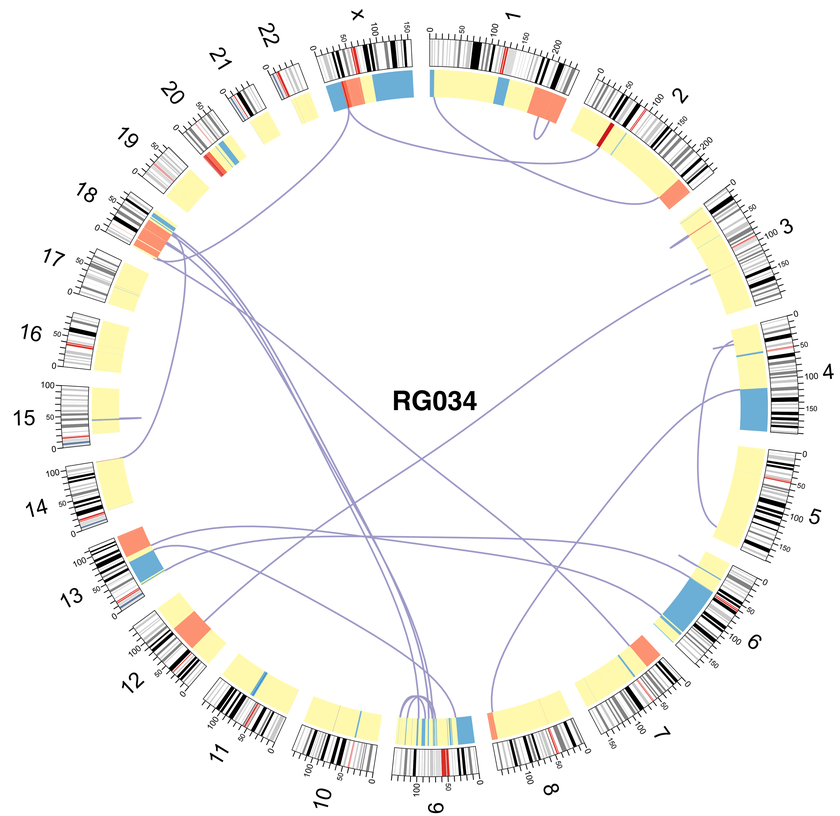

RG074

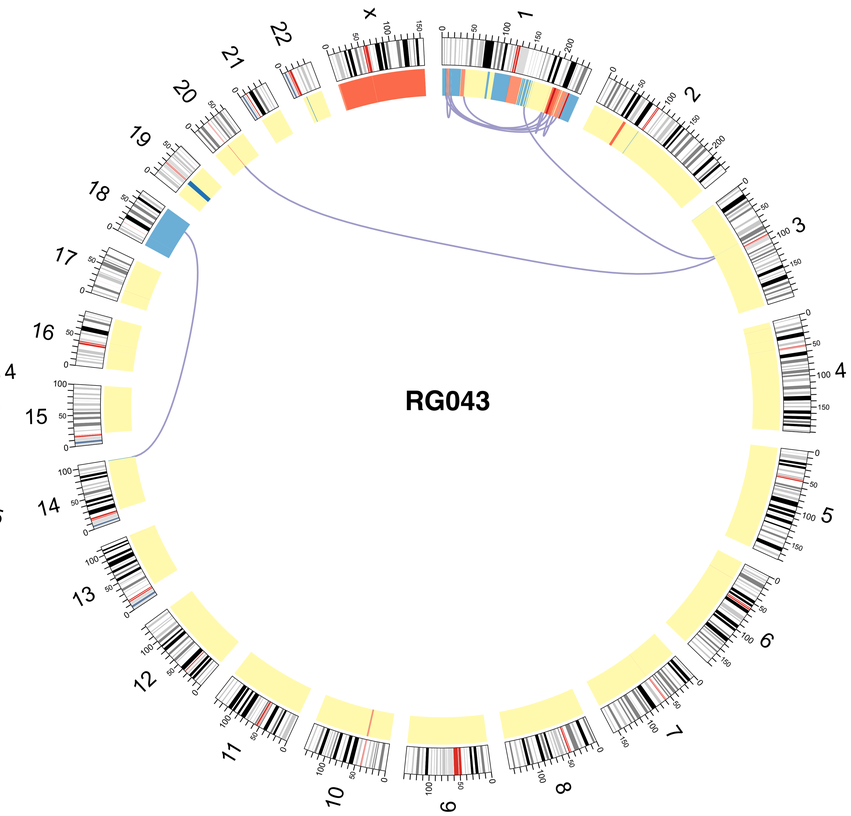

RG116

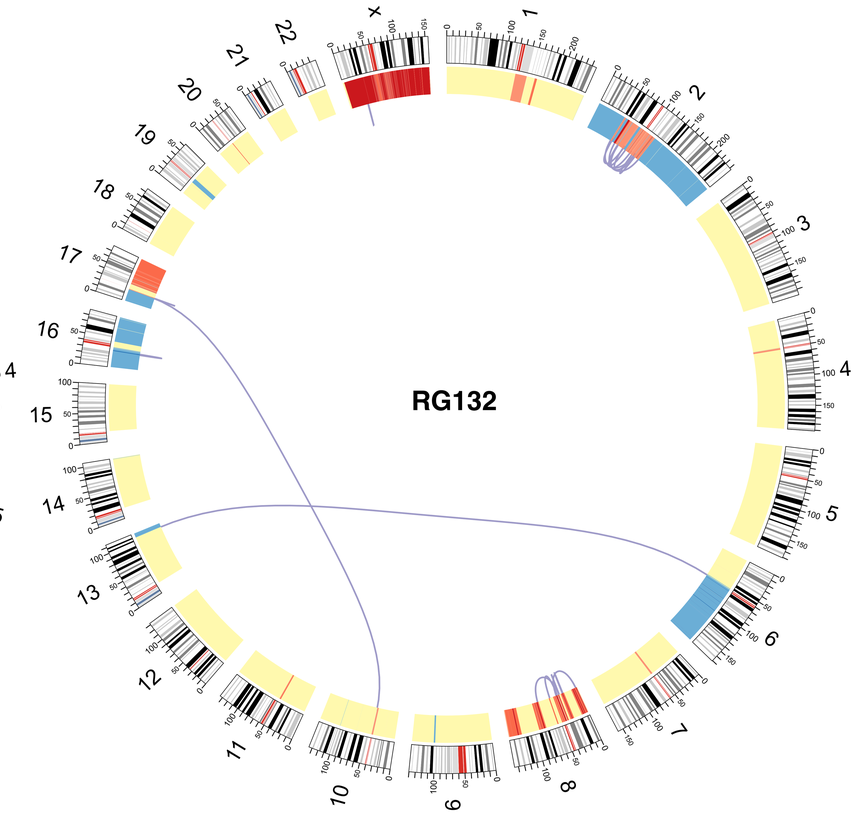

RG014


RG023


RG083


RG034

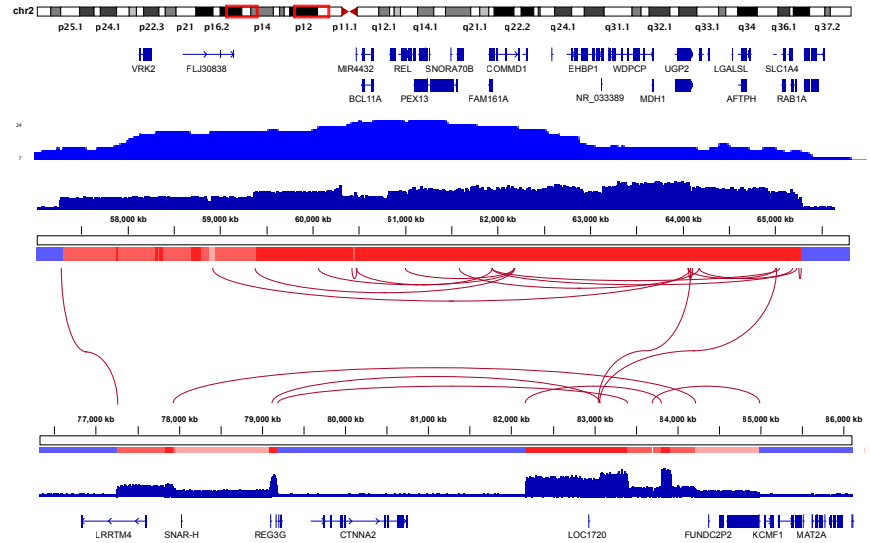

RG043


RG132

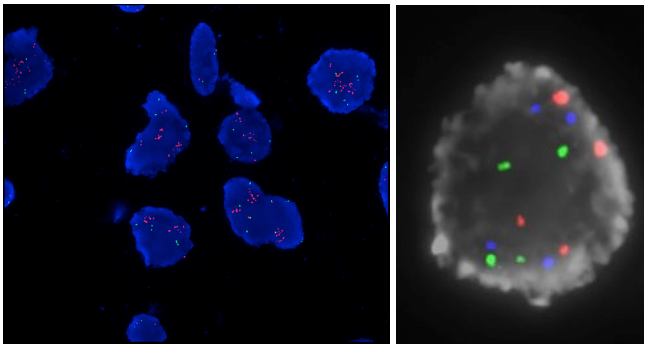
study of *FOXO1* mutations[13]. No significant differences were observed for PFS or OS in the full cohort (P = 0.58 and 0.77, respectively, log-rank test). Mutation status was also not correlated with PFS or OS in the GCB cases (P = 0.45 and 0.49, respectively, log-rank test).

Supplemental Figure S6. Examples of structural rearrangements detected in the genomes of individual cases.
Arcs indicate rearrangements detected by *de novo* assembly. Regions colored in red or blue were amplified or deleted, respectively. RG065 contains rearrangements and concomitant loss of genetic material consistent with chromothripsis. RG014, RG043, RG083, RG034, RG132 and RG116 each contain amplification of the REL locus (chromosome 2), often with multiple breaks detected.

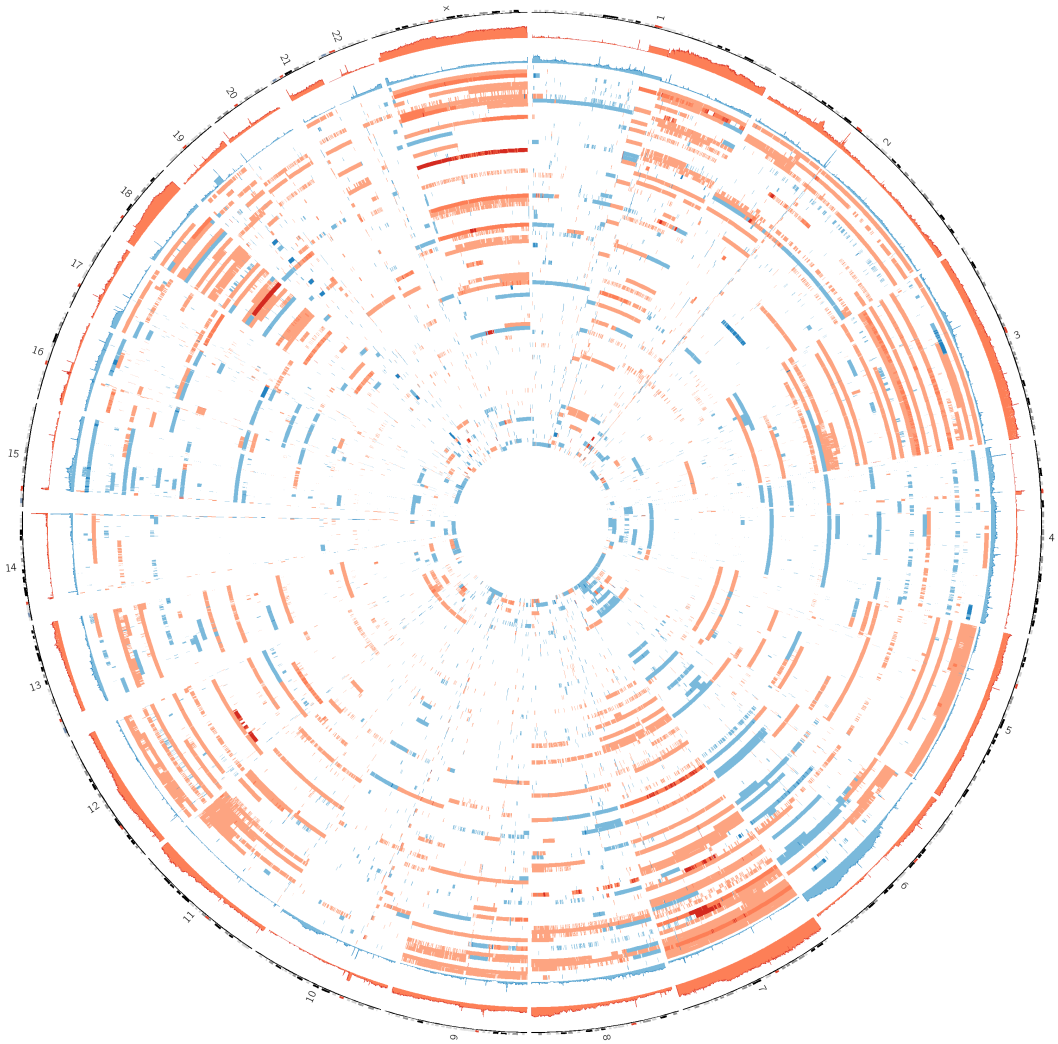
A



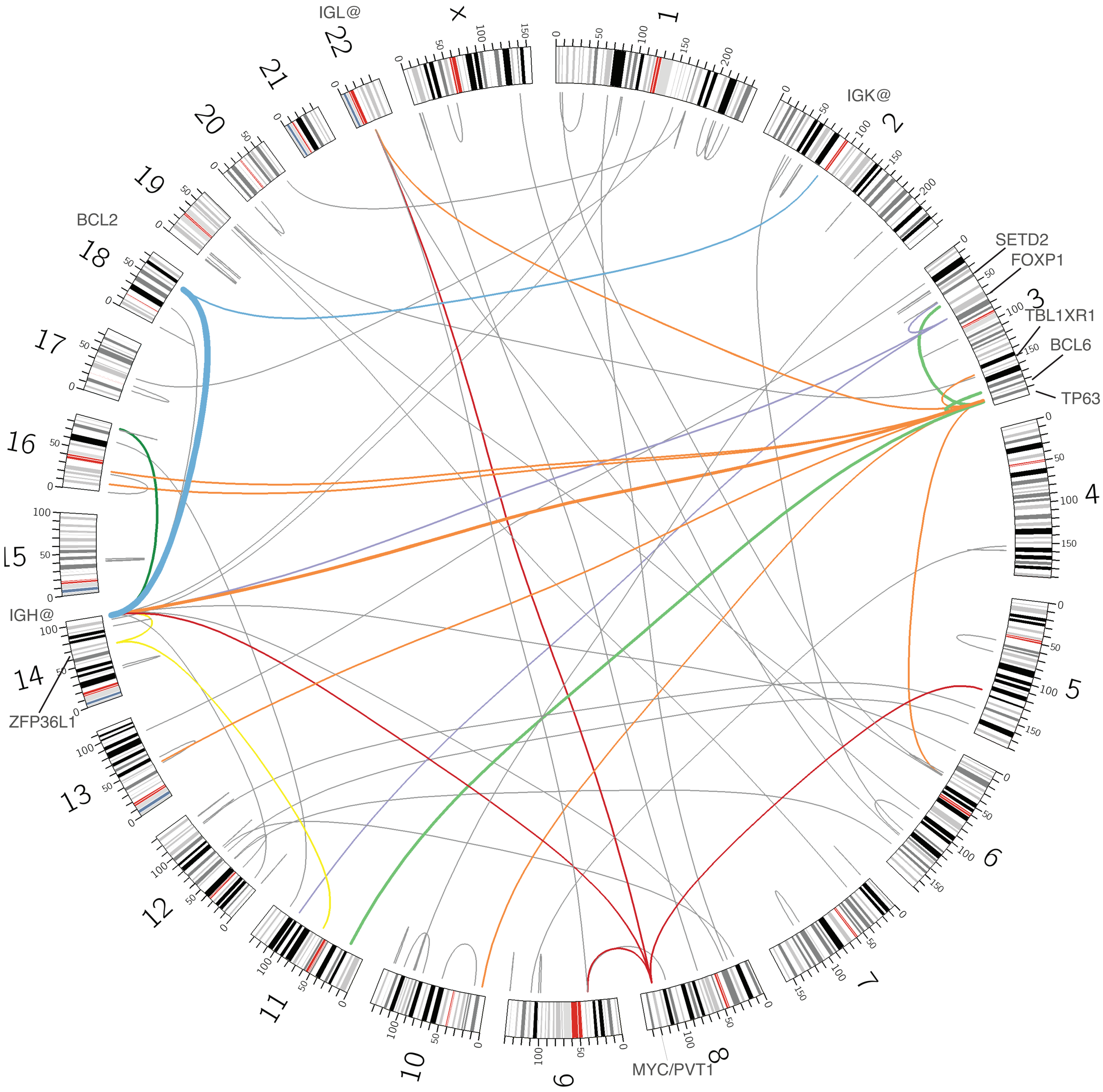
B



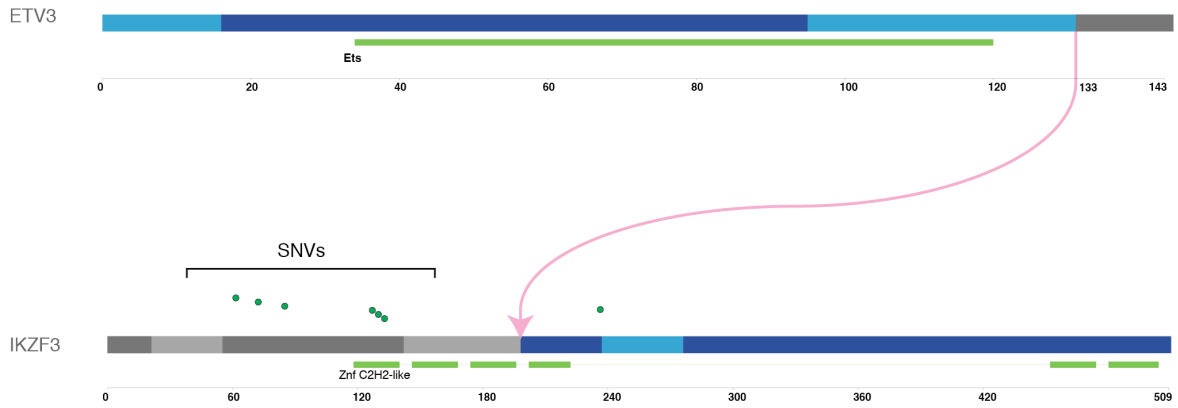
Supplemental Figure S7. Example of high-level amplification of *REL*
The genomic region surrounding REL (2p15-16) was the only recurrent locus focally amplified in our cohort of 92 DLBCLs. **(A)** The upper light blue bar plot shows the pooled copy number state from all 92 cases with a clear peak affecting REL and neighboring genes. The dark blue plot shows the WGS read density in a single case with high-level REL amplification identified from SNP array and WGS data. Interestingly, this case also contained two other regions of high-level amplification on the same chromosome (below). Assembly of this genome revealed contigs (red arcs) linking the REL locus to each of these two amplified regions, indicating that these regions were co-amplified. The mechanism of amplification is unclear but is not due to chromothripsis as it would require a stepwise accumulation of genetic material that is not possible in a single cell cycle. **(B)** FISH analysis performed on nuclei isolated from formalin-fixed paraffin-embedded tissue[14] confirms amplification of the REL locus (red signal) in this case and the diffuse signal does not support that the amplification results from tandem duplication. The presence of 4 signals for the control probe (green signal) that encompasses the SUMO1 gene, a region that was identified as copy neutral in analysis of SNP 6.0 data, suggests that the entire genome was tetraploid. This is supported by FISH using a commercial probe set (LSI IGH/MYC, CEP8 tricolor; panel B, right), which shows four signals for each locus on chromosomes 3, 8 and 14, respectively.



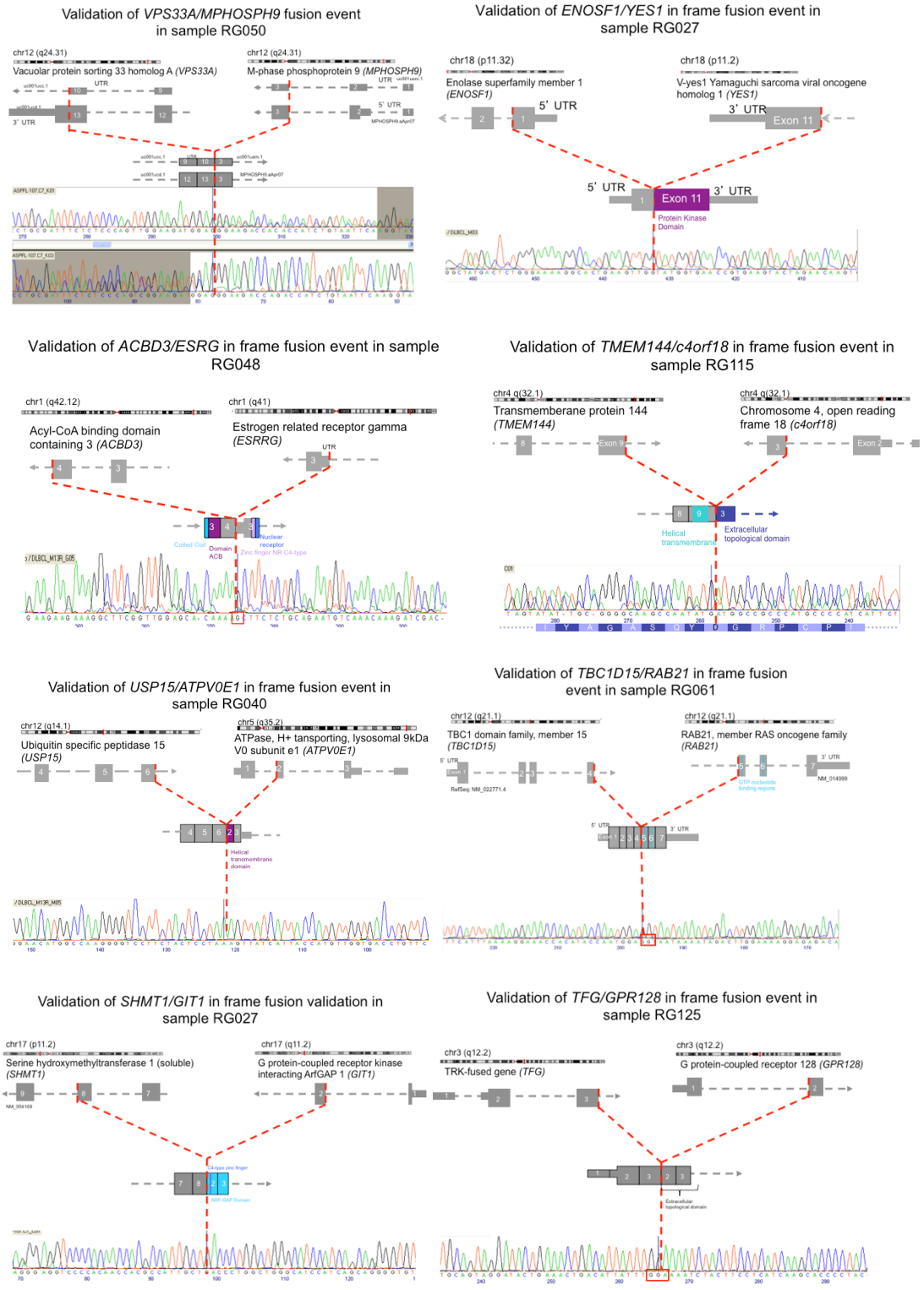
Supplemental Figure S8. Overview of copy number variants detected in 92 DLBCLs. The 40 tumour/normal pairs were searched for copy number variants using HMMCopy and the remaining cases were analyzed with Affymetrix SNP6.0 arrays (tumour tissue only). Dark red indicates high-level amplifications and light red indicates low-level gains. Light blue indicates heterozygous deletions and dark blue indicates homozygous deletions. The outer two tracks summarize the overlap of regions gained or lost in the 92 genomes. Gain of chromosome 3, 7, 12, 18, 21 and X in their entirety were common. Arm-level events such as loss of 1p and 6q and gain of 1q were also common. Besides whole chromosome (or arm-level) events, the REL locus (on chromosome 2) was only recurrent high-level amplicon detected.


IGL@
X
1
IGK@
2
BCL2
SETD2
FOXP1
3
TBL1XR1
BCL6
TP63
4
5
6
7
8
MYC/PVT1
9
10
11
12
13
14
ZFP36L1
IGH@
15
16
17
18
19
20
21
22

Supplemental Figure S9. Overview of all fusion transcripts detected by RNA-seq. All rearrangements detected from the RNA-seq data (Supplemental Table S6) are shown with recurrent events and promiscuous genes indicated in color and the remainder shown in grey. The identity of genes involved in such events is also indicated. For example, rearrangements involving *MYC* or *PVT1* are red and those involving *BCL6* are orange. The thickness of arcs is proportional to the number of times the fusion was observed in 96 DLBCLs. Eight of these involved *BCL6* with a distinct partner gene (*ST6GAL1*, *CIITA*, *IL21R*, *PIM1*, *MBNL1*, *EIF4A2*, *LCP1*, *FBXO18*) while 23 involved immunoglobulin regions and *BCL6* (7) or *BCL2* (16). All but one of the *BCL6* partners (*FBXO18*) have been previously described in NHL and, on further inspection of the corresponding genome, this event results from a three-way rearrangement also involving the region encoding immunoglobulin lambda light chain gene, likely resulting in induction of *BCL6* expression. Beyond those involving these genes, additional fusions indicative of rearrangements with immunoglobulin loci were detected including *PVT1*, *GTF2E2*, *DUSP22*, *IRF8*, *ZFP36L1*, *FOXP1*, *NOL4*, and *MYC*. We suspect the rearrangement involving *DUSP22* also results in deregulated expression of the neighboring gene *IRF4*. Immunoglobulin-induced deregulation of *MYC, IRF4* and *FOXP1* and rearrangements involving *ZFP36L1* have been described in NHL[15] but the role of the latter in lymphomagenesis has not been elucidated. Of note, we observed additional fusions involving each of *ZFP36L1* and *FOXP1* with other partner genes. The novel fusions were largely singleton events with two exceptions: *TBL1XR1-TP63,* which was observed in four cases[16] and *TFG-GPR128,* which occurs as the result of a polymorphic copy number variation[17]. The presence of fusions involving any of these genes in the 13 DLBCL cell lines is indicated in Supplemental Table S6. Where possible, we noted whether the breakpoints that underlie each fusion event were detectable in the tumor and normal genomes (Supplemental Table S5). Some events in the 40 cases with matched genome data were observed in both tumor and normal genomes and others were not supported by the genome data. Thus, each fusion reported in this table should be considered a candidate event unless indicated otherwise.

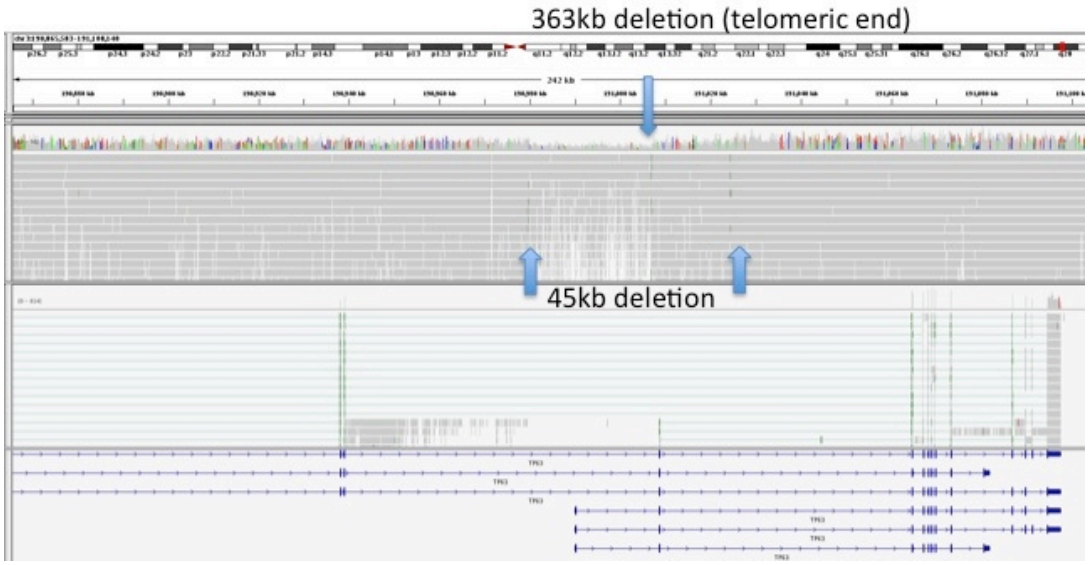


Supplemental Figure S10. Detailed representation of *ETV3*-IKZF3 fusion. Excluding the known fusions, only eight of the remaining events detected were predicted to result in retained reading frames with one example, involving *ETV3* and *IKZF3* (the B-cell specific transcription factor Aiolos), shown here. Fusions involving other *ETS*-family transcription factors are commonly found in in prostate tumors. Mutations resulting in dominant-negative isoforms of IKZF1 (Ikaros), the related transcription factor, are common in acute lymphoblastic leukemias[18]. IKZF1 and IKZF3 are known to repress *MYC* during B-cell development and are both regulated by IRF4 and IRF8[19], two transcription factors found here and previously to be commonly mutated and involved in rare fusions in DLBCL[2]. The fusion transcript includes three N-terminal exons of *ETV3* (including the Ets domain) and the three C-terminal exons of *IKZF3*. Similar to the IKZF1 dominant-negative isoforms seen in ALL, this fusion appears to retain the dimerization domain of IKZF3 but lack the N-terminal DNA binding domain. *ETV3* encodes a STAT3-regulated ETS-family transcription factor with poorly understood function. *IKZF3* was also a target of somatic SNVs and harbored six individual variants in a separate case (green circles) indicating that this locus, in some cases, may be targeted by aSHM.

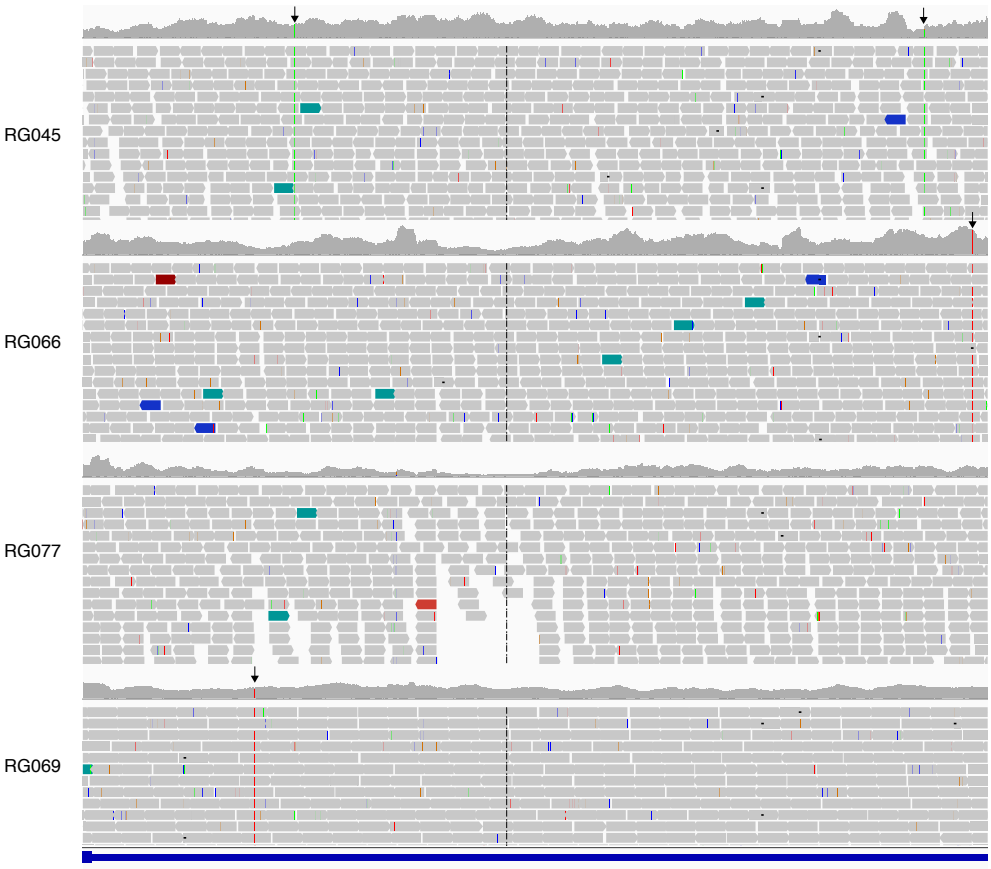


Supplemental Figure S11. Eight additional fusions with preserved reading frames.

A

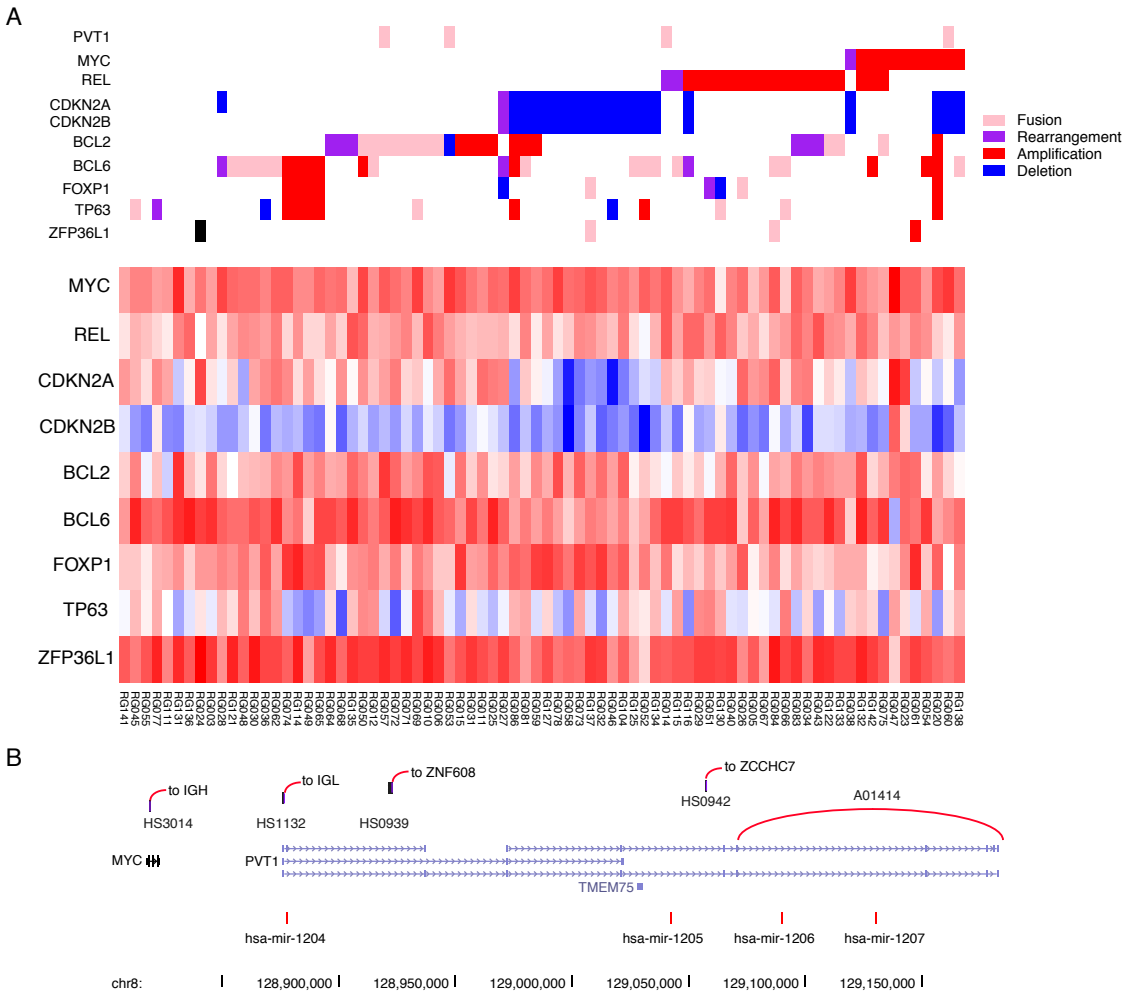


B



Supplemental Figure S12. Deletions and fusions result in monoallelic expression of *TP63*.

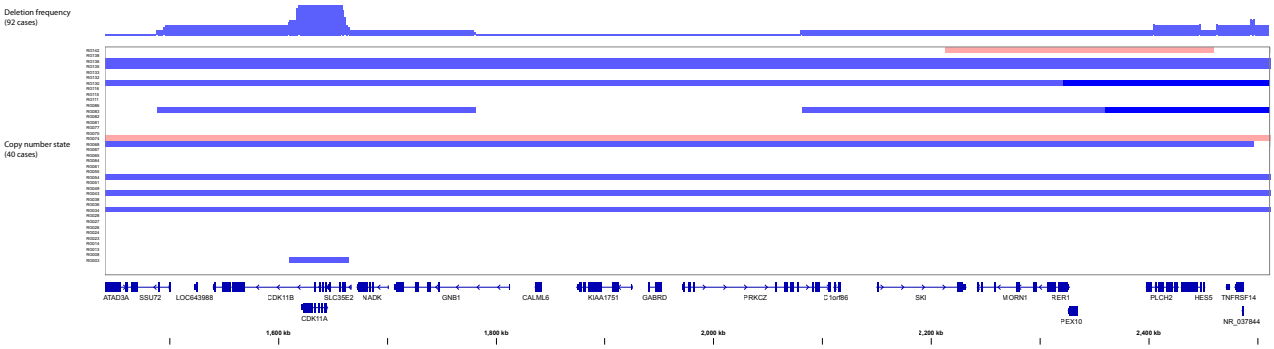
(A) A single genome contained two deletions affecting the *TP63* locus. The smaller deletion removes the first two exons of the shorter *TP63* isoform. The smaller deletion results in an isoform lacking the two deleted exons but with conserved reading frame (shortened by 85 amino acids). Unlike the *TBL1XR1-TP63* fusions, this isoform is not expected to lack the transactivation domain and thus may not be equivalent to the dominant negative deltaN isoforms of *TP63*[20]. Though an isoform lacking these exons has been annotated (UniProt: Q9H3D4-9), its existence in normal cells lacks experimental confirmation. In the RNA-seq data from this case we observed monoallelic expression of one of the larger isoforms but with an in-frame loss of exon 4, which was removed by the smaller deletion. Unexpectedly, this case had the highest expression of *TP63* besides those samples containing a fusion transcript. It is unclear whether this isoform and the fusions involving *TP63* have *de facto* oncogenic function or whether general loss of *TP63* activity is a driver event. **(B)** Based on imbalanced allelic ratios for SNPs in the 3' UTR of TP63 (indicated with black arrows), monoallelic expression is detectable in 3 of the 4 cases with *TBL1XR1-TP63* fusions. Only RG077 lacked a heterozygous SNP in the exons of *TP63*. Monoallelic expression is consistent with the fusion event either greatly increasing or suppressing expression of *TP63* mRNA. When considered with our observation that *TP63* expression overall is higher in cases with this fusion, we infer that enhanced expression of *TP63* is the more likely result of the fusion event.



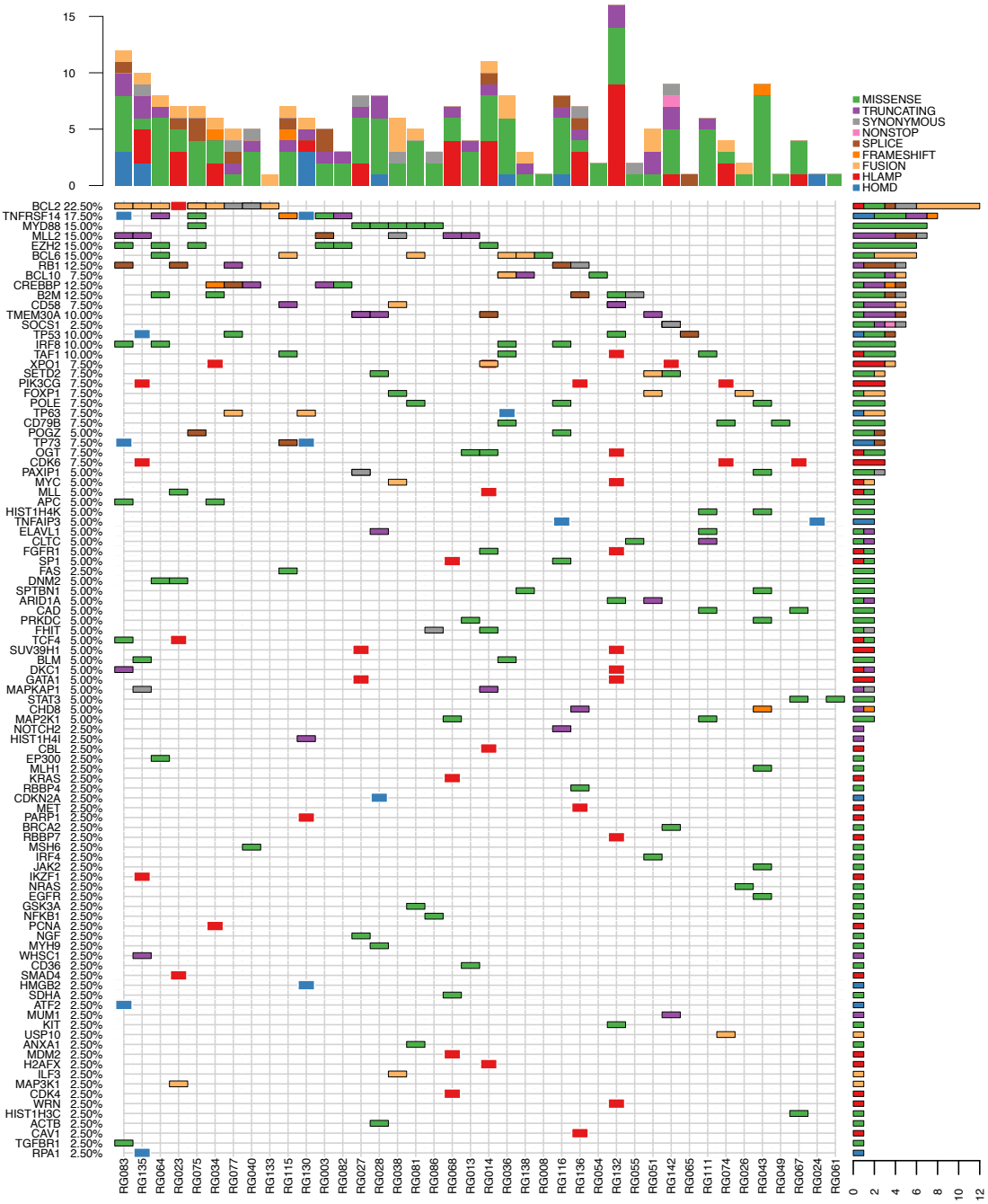
Supplemental Figure S13. Overview of genes commonly amplified, deleted or involved in rearrangements/fusions.

(A) The status of chromosomal alterations including amplifications, deletions, proximal (<250kb) rearrangements or fusion transcripts is shown for these genes in all 96 DLBCL cases. Columns (cases) are ordered by presence of mutations in the more commonly involved genes. Mutual exclusivity between events affecting *MYC*, *REL* and *CDKN2A* are apparent. Events affecting *BCL2* and *BCL6* are also largely mutually exclusive. The lower heat map shows the gene expression levels, determined from RNA-seq, for the full DLBCL cohort with red indicating high expression and blue indicating low expression. Deletions of *CDKN2A* correspond with low expression (darker blue) but there is otherwise no clear correspondence between gene expression and these events. Notably, 91% of the cases have genomic alterations involving one of these five genes. The expression of *TP63* was variable across the cohort but significantly higher in cases with fusions or deletions affecting this gene (see, for example, RG036, RG045 and RG077). *ZFP36L1* expression was uniformly high. **(B)** Three additional genes were seen fused with multiple partners, namely *RABGAP1*, *FOXP1* and the noncoding *PVT1* gene. The *PVT1* locus neighbors *MYC* and rearrangements involving *PVT1* have been documented in Burkitt lymphoma[21], multiple myeloma[22], medulloblastoma[23] and breast cancer[24]. This locus is also a common site of retroviral integration in mouse models of T-cell

lymphoma[25]. Despite its proximity to *MYC*, there is accumulating evidence that rearrangements affecting this locus as well as amplifications of *PVT1* may themselves be pathogenic[26]. The *PVT1* transcript is thought to encode multiple miRNAs[25] but we found no evidence for the mature form of any of these in the small RNA libraries from these cases (not shown). Nonetheless, a recent study provided evidence that PVT1-encoded miRNAs are expressed at low levels in medulloblastoma[23] and thus it may be that these miRNAs are also relevant in DLBCL. The noncoding RNA gene *PVT1*, which is approximately 50kb from *MYC*, was found in multiple fusion transcripts detected by analysis of the RNA-seq data from our large DLBCL cohort. The fusion partner of each of the four events detected is indicated. In a single case, a small tandem amplification of a 3' segment of *PVT1* was detected (A01414, red arc). *MYC* expression is known to be deregulated by translocations involving the IGH locus but it is unclear whether rearrangements involving *PVT1* impact *MYC* expression. Fusions involving *PVT1* (pink, top of panel A) did not correspond to significantly higher expression of *MYC* (top of heat map) or *PVT1* (not shown).

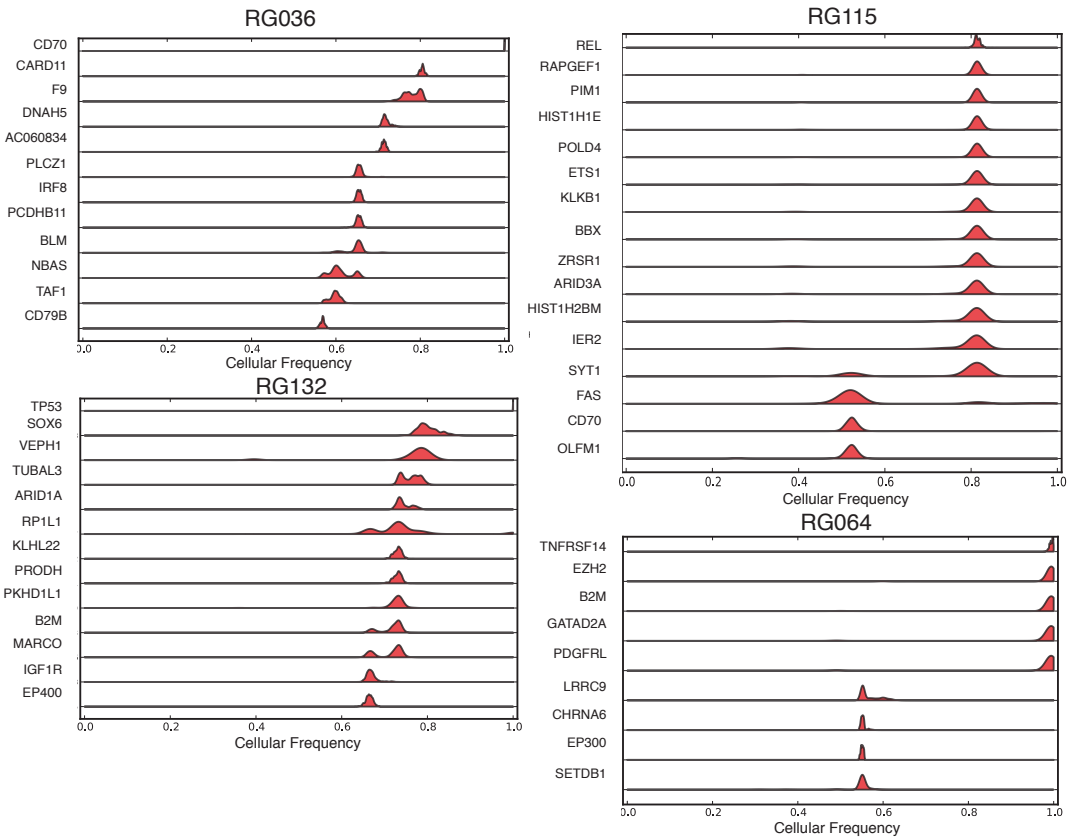


Supplemental Figure S14. Recurrence of deletions affecting *CDK11A* and *CDK11B*. Deletions affecting 1p36.33 are common in both FL and DLBCL. *TNFRSF14* was previously reported to be the relevant target of these deletions[27]. In the copy number data from 96 DLBCL cases, we observed focal and large-scale deletions affecting *TNFRSF14* (far right) and a stronger focus of deletion around the *CDK11A/CDK11B* locus, indicating those genes may also be relevant targets of 1p36.33 deletions. The upper blue track shows the cumulative signal from all DLBCL patients in the large DLBCL cohort (96 cases). The lower heat map shows the somatic copy number state derived from the 40 WGS cases. Two focal deletions are visible, with the smallest affecting only affecting three genes.

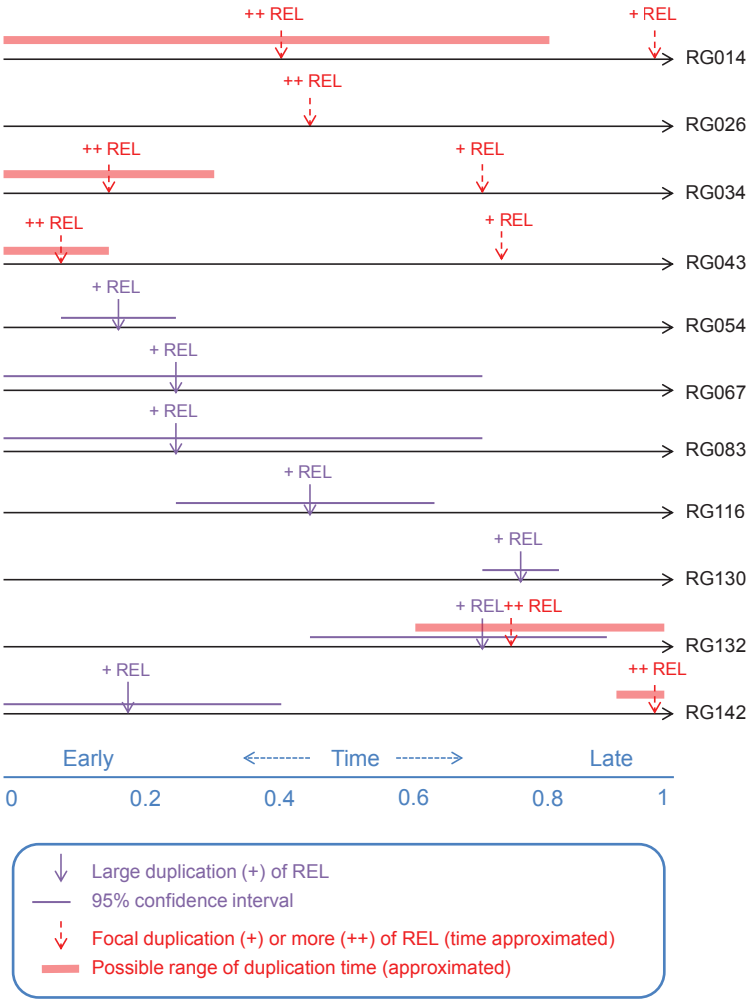


Supplemental Figure S15. Mutations with functional impact predicted by integrative analysis of mutation and expression data.
Individual mutations predicted to be functional from the 40 DLBCL based on analysis with matched expression data (RNA-seq) are color-coded by mutation type. The genes are ordered on the total number of mutational events observed in them in the cohort (totaled on the right, by mutation type) and the percentage of cases in which the gene had any of the displayed mutation types is shown next to its name. (TRUNCATING: truncating mutation, SYNONYMOUS: synonymous mutation, SPLICE: splice mutation, MISSENSE: missense mutation, HOMD: homozygous deletion, FRAMESHIFT: indel

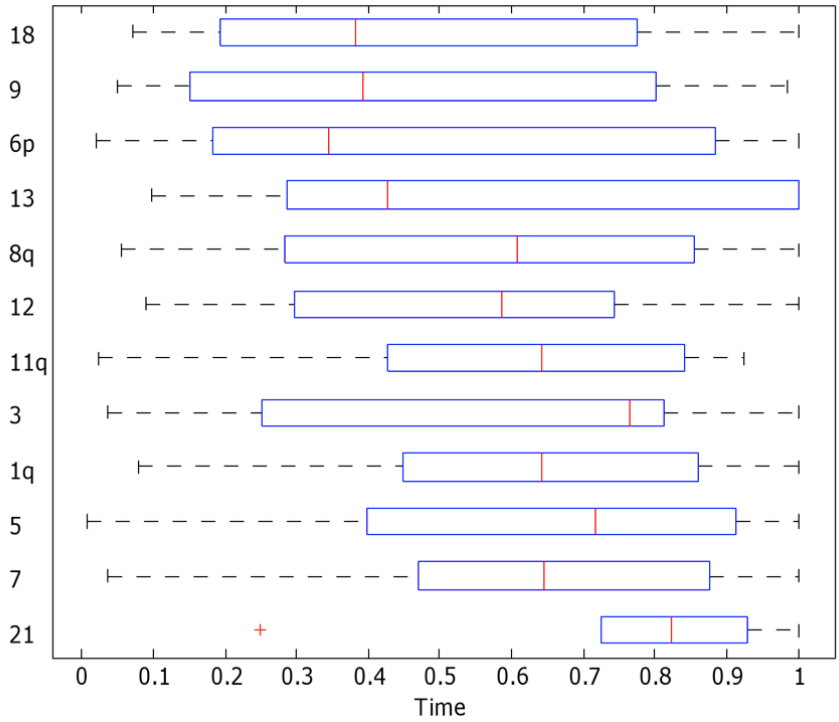
resulting in a frameshift, HLAMP: high-level amplification, NONSTOP: SNV resulting in loss of a stop codon, FUSION: fusion transcript involving this gene detected by RNA-seq). Heterozygous deletions and low-copy amplifications are not shown. Genes are displayed if their functional probability threshold, determined by DriverNet, exceeded 0.79 (Methods). The upper plot summarizes the mutations of each class affecting any of these genes within each patient. The plot on the right-hand side summarizes the mutations affecting each individual gene across the entire cohort. *TNFRSF14* is a known tumor suppressor gene and in this cohort was affected by SNVs resulting in missense nonsense changes as well as frame-shifting indels and homozygous deletions. TAF1 phosphorylates TP53 and induces its turnover. In addition to the high-level amplifications shown, TAF1 was amplified in 16 additional cases in the cohort of 92 patients, often due to aneuploidy of X.



Supplemental Figure S16. Evidence for clonal and sub-clonal driver mutations. The amplicon sequencing method utilized for SNV verification facilitates accurate determination of the allele frequencies of each variant. By integrating this data with copy number and LOH state, one can estimate the cellular frequency of each mutation in the entire tumor population. The estimated cellular frequency of each SNV tested is plotted (red peaks). This offers a separate approach for determining the approximate relative acquisition time of mutations[24]. Mutations with higher cellular frequency (right/top) for each case are likely to be early driver events whereas those with lower cellular frequency exist in a sub-clone and were thus likely acquired later in tumor development. These examples imply early acquisition of certain driver mutations including those in *TNFRSF14*, *PIM1*, *TP53*, *EZH2* and later acquisition of other driver mutations such as those in *EP300*, *FAS* and *CD70*. Counter-examples in which known drivers appear in sub-clones are also notable such as *CARD11* and *CD79B* in RG036. Additional images for all cases tested are included separately.

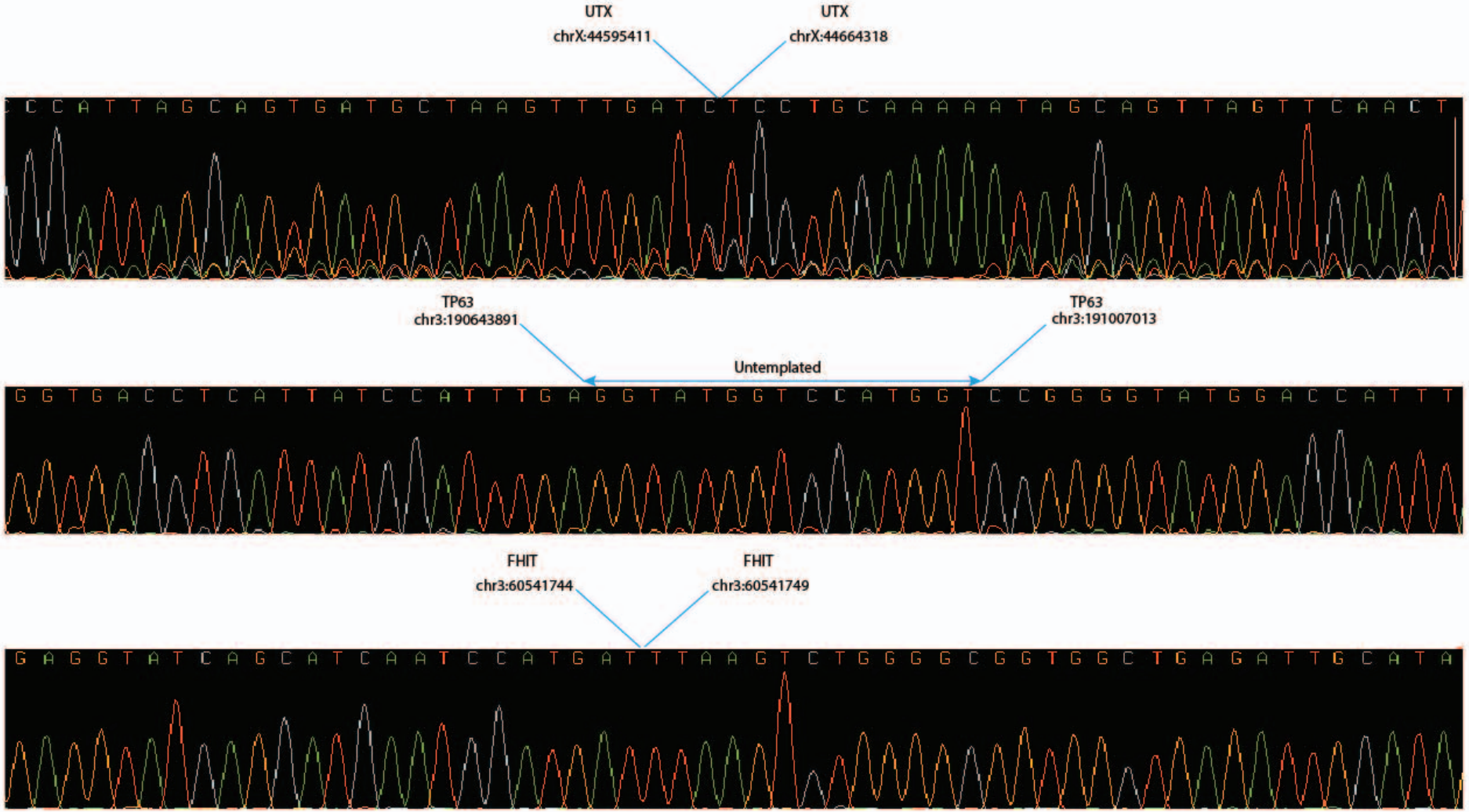


Supplemental Figure S17. Timing of *REL* amplifications.
Duplications specifically resulting in *REL* amplicons were individually timed with increasing time represented on the X-axis. 95% confidence intervals were determined by bootstrapping.



Supplemental Figure S18. Boxplots illustrating the variability in timing of most commonly duplicated chromosomes or chromosome arms.
The events are ordered by increasing average duplication time; red bars indicate median, blue boxes indicate quartiles, whiskers represent furthest point within 1.5 IQR.

Supplemental Figure S19: Examples of validated small deletions.
Representative Sanger sequence data from the validation of three of the small deletions shown in Figure 3 is shown. We observed untemplated sequence additions at the breakpoints of some of the rearrangements and deletions.


UTX
chrX:44595411
UTX
chrX:44664318
TP63
chr3:190643891
TP63
chr3:191007013
Untemplated
FHIT
chr3:60541744
FHIT
chr3:60541749

References:

1. Sinha RK, Park C, Hwang I-Y, Davis MD, Kehrl JH. ScienceDirect.com - Immunity - B Lymphocytes Exit Lymph Nodes through Cortical Lymphatic Sinusoids by a Mechanism Independent of Sphingosine-1-Phosphate-Mediated Chemotaxis. *Immunity*. 2009;30(3):434–446.
2. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011;476(7360):298–303.
3. Green JA, Suzuki K, Cho B, et al. The sphingosine 1-phosphate receptor S1P2 maintains the homeostasis of germinal center B cells and promotes niche confinement. *Nat Immunol*. 2011;12(7):672–680.
4. Schwartzentruber J, Korshunov A, Liu X-Y, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*. 2012;482(7384):226–231.
5. Cattoretti G, Cattoretti G, Mandelbaum J, et al. Targeted Disruption of the S1P2 Sphingosine 1-Phosphate Receptor Gene Leads to Diffuse Large B-Cell Lymphoma Formation. *Cancer Res*. 2009;69(22):8686–8692.
6. Pasqualucci L, Dominguez-Sola D, Chiarenza A, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*. 2011;471(7337):189–195.
7. Lohr JG, Stojanov P, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci USA*. 2012;109(10):3879–3884.
8. Wu EHT, Tam BHL, Wong YH. Constitutively active α subunits of Gq/11 and G12/13 families inhibit activation of the pro-survival Akt signaling cascade. *FEBS Journal*. 2006;273(11):2388–2398.
9. Takuwa N, Du W, Kaneko E, et al. Tumor-suppressive sphingosine-1-phosphate receptor-2 counteracting tumor-promoting sphingosine-1-phosphate receptor-1 and sphingosine kinase 1 - Jekyll Hidden behind Hyde. *Am J Cancer Res*. 2011;1(4):460–481.
10. Cyster JG, Schwab SR. Sphingosine-1-phosphate and lymphocyte egress from lymphoid organs. *Annu Rev Immunol*. 2012;30:69–94.
11. Lyons J, Landis CA, Harsh G, et al. Two G protein oncogenes in human endocrine tumors. *Science*. 1990;249(4969):655–659.
12. Kapoor N, Menon ST, Chauhan R, Sachdev P, Sakmar TP. Structural Evidence for a Sequential Release Mechanism for Activation of Heterotrimeric G Proteins. *J Mol Biol*. 2009;393(4):882–897.
13. Trinh DL, Scott DW, Morin RD, et al. Analysis of FOXO1 mutations in diffuse large B-cell lymphoma. *Blood*. 2013.
14. Paternoster SF, Brockman SR, McClure RF, et al. A new method to extract nuclei from paraffin-embedded tissue to study lymphomas using interphase fluorescence in situ hybridization. *Am J Pathol*. 2002;160(6):1967–1972.
15. Pospisilova H, Baens M, Michaux L, et al. Interstitial del(14)(q) involving IGH: a novel recurrent aberration in B-NHL. *Leukemia*. 2007;21(9):2079–2083.
16. Scott DW, Mungall KL, Ben-Neriah S, et al. TBL1XR1/TP63: a novel recurrent gene fusion in B-cell non-Hodgkin lymphoma. *Blood*. 2012.
17. Chase A, Ernst T, Fiebig A, et al. TFG, a target of chromosome translocations in

lymphoma and soft tissue tumors, fuses to GPR128 in healthy individuals. *Haematologica*. 2010;95(1):20–26.
18. Mullighan CG, Miller CB, Radtke I, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature*. 2008;453(7191):110–114.
19. Ma S, Pathak S, Mandal M, et al. Ikaros and Aiolos inhibit pre-B-cell proliferation by directly suppressing c-Myc expression. *Mol Cell Biol*. 2010;30(17):4149–4158.
20. Petitjean A, Ruptier C, Tribollet V, et al. Properties of the six isoforms of p63: p53-like regulation in response to genotoxic stress and cross talk with Np73. *Carcinogenesis*. 2007;29(2):273–281.
21. Zeidler R, Joos S, Delecluse HJ, et al. Breakpoints of Burkitt's lymphoma t(8;22) translocations map within a distance of 300 kb downstream of MYC. *Genes Chromosom. Cancer*. 1994;9(4):282–287.
22. Nagoshi H, Taki T, Hanamura I, et al. Frequent PVT1 rearrangement and novel chimeric genes PVT1-NBEA and PVT1-WWOX occur in multiple myeloma with 8q24 abnormality. *Cancer Res*. 2012.
23. Northcott PA, Shih DJH, Peacock J, et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature*. 2012;488(7409):49–56.
24. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012.
25. Huppi K, Volfovsky N, Runfola T, et al. The identification of microRNAs in a genomically unstable region of human chromosome 8q24. *Mol Cancer Res*. 2008;6(2):212–221.
26. Guan Y, Kuo W-L, Stilwell JL, et al. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clin Cancer Res*. 2007;13(19):5745–5755.
27. Cheung K-JJ, Johnson NA, Affleck JG, et al. Acquired TNFRSF14 mutations in follicular lymphoma are associated with worse prognosis. *Cancer Res*. 2010;70(22):9166–9174.