Integrative Genomic and Transcriptomic Characterization of Pediatric Burkitt Lymphoma

Bruno Grande

January 23, 2018

Table of Contents

# Main Text

## Introduction

**Finish by: January 21st, 2018**

**Target length: 500 words**

To be done.

## Materials and methods

**Finish by: January 25th, 2018**

To be done.

## Results

**Target length: 2500 words**

### Clinical and molecular characteristics of the BLGSP pediatric discovery cohort

* *TODO: Update miRNA-seq sequencing statistics after including latest cases. Currently, only 83 are included.*
* *TODO: Carefully check the EBV type, esp. for the case with the deletion.*

The BLGSP pediatric discovery cohort consists of 74 endemic BL (eBL) cases from Uganda, 17 sporadic BL (sBL) cases from the United States, and 4 immunodeficiency-related cases due to HIV infection (HIV+ BL), amounting to 95 patients in total. Endemic and sporadic EBV infection rates are consistent with previous reports (92% and 18%, respectively) [REF]. Conversely, our cohort includes 9 cases that are either endemic and EBV-negative or sporadic and EBV-positive, which will assist in attributing biological differences to either clinical variant status or EBV infection status despite being correlated. Tumor samples underwent deep whole genome sequencing (WGS; mean depth 82X; range 55-96), ribo-depleted RNA sequencing (RNA-seq; mean 199M reads; range 132-255M), and microRNA sequencing (miRNA-seq; mean 13M reads; range 2.7-32M). Constitutive DNA was sequenced (mean depth 41X; range 30-51) for all patients to distinguish between somatic and germline mutations. Being the putative cells of origin for BL, centroblasts and centrocytes from six pediatric tonsil donors underwent identical RNA-seq and miRNA-seq protocols to act as normal comparators for gene expression analyses. We supplemented certain genomic analyses with re-analyzed WGS data for 17 pediatric sBL cases from the ICGC MMML-Seq project. Clinical and molecular characteristics are summarized in Table 1.

Table 1 Univariate table listing the clinical and molecular characteristics for the pediatric cohorts of the BLGSP discovery and ICGC MMML-Seq projects.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Level | BLGSP Discovery (n=95) | ICGC MMML-Seq (n=17) | Total (n=112) |
| Clinical variant | Sporadic | 17 (18%) | 17 (100%) | 34 (30%) |
|  | Endemic | 74 (78%) | 0 (0%) | 74 (66%) |
|  | HIV+ | 4 (4%) | 0 (0%) | 4 (4%) |
| EBV status | Negative | 21 (22%) | 17 (100%) | 38 (34%) |
|  | Positive | 74 (78%) | 0 (0%) | 74 (66%) |
| EBV type | Negative | 21 (22%) | 17 (100%) | 38 (34%) |
|  | Type 1 | 61 (64%) | 0 (0%) | 61 (54%) |
|  | Type 2 | 13 (14%) | 0 (0%) | 13 (12%) |
| Age group (years) | 0 - 5 | 24 (25%) | 6 (35%) | 30 (27%) |
|  | 6 - 10 | 50 (53%) | 6 (35%) | 56 (50%) |
|  | 11 - 15 | 18 (19%) | 3 (18%) | 21 (19%) |
|  | 16 - 20 | 3 (3%) | 2 (12%) | 5 (4%) |
| Tumor biopsy | FF | 89 (94%) | 17 (100%) | 106 (95%) |
|  | FFPE | 6 (6%) | 0 (0%) | 6 (5%) |
| MYC translocation | IGH-MYC | 77 (81%) | 13 (76%) | 90 (80%) |
|  | IGL-MYC | 9 (9%) | 3 (18%) | 12 (11%) |
|  | IGK-MYC | 7 (7%) | 1 (6%) | 8 (7%) |
|  | BCL6-MYC | 1 (1%) | 0 (0%) | 1 (1%) |
|  | IGH-GNA13-MYC | 1 (1%) | 0 (0%) | 1 (1%) |

### AID-mediated patterns of non-coding mutations

* *TODO: Update wavelet results once the PAX5 enhancer peak is back.*

Unlike previous genomic studies focused on BL, our dataset is sufficiently large to characterize the genome-wide landscape of non-coding mutations. We leveraged the Rainstorm method1 to discern regions enriched in silent mutations, *i.e.* intergenic, intronic or synonymous events. This method located 181 clusters of non-coding mutations with an average size of 884.9171271 bp (range 1- 15014 bp). Among these clusters, 61 (34%) overlap a gene, with the median distance from the nearest transcription start site (TSS) being 40638. In fact, if we only consider clusters that are mutated in 10 or more cases, 34 (72%) are within 3000 bp of a TSS (Supplementary Figure 15). The most recurrently mutated clusters are altered in up to 111 of cases and overlap established targets of physiologic (*i.e.* *IGH*, *IGK*, *IGL* and *MYC*) and aberrant (*e.g.* *BCL7A*, *BCL6*, *BACH2*, *TCL1A* and *BTG2*) somatic hypermutation (SHM) (Figure 1). Here, we consider *MYC* a target of physiologic SHM given its proximity to an immunoglobulin (Ig) chain locus in BL, which are natural targets of AID. Altogether, a majority of these clusters are consistent with AID-mediated SHM given their targets and proximity to TSSes [REF]. We observe similar clusters in other B-cell non-Hodgkin lymphomas such as diffuse large B-cell lymphoma (DLBCL)1 and follicular lymphoma (FL; unpublished). Physiologic SHM is observed in nearly all cases, especially at the *IGH* and *MYC* loci. On the other hand, signs of aberrant SHM appears to be more prevalent in endemic or EBV-positive cases (Supplementary Figure 16). This is consistent with the increased AID transcript levels in endemic or EBV-positive tumors (Figure 2) observed in the BLGSP RNA-seq data. The fold change in AID expression is greater according to EBV infection status (1.18) than clinical variant status (1.16), suggesting that EBV may be playing a role in modulating AID expression. We confirmed that AID expression is higher in EBV-positive tumors regardless of geographic origin (Supplementary Figure 17).

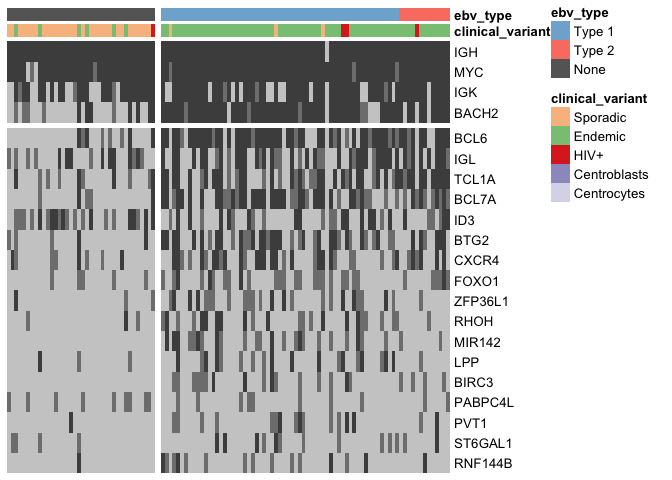


Figure 1 Discretized heat map showing the mutation rate of the most recurrently altered clusters of non-coding variants (altered in at least 15 cases). Clusters are labeled according to the nearest TSS (rows) across the entire BL cohort (columns).Clusters that are associated with the same TSS are combined for plotting. Light gray denotes no mutations, and darker shades of gray indicate the presence of up to two mutations or more than two mutations, respectively.

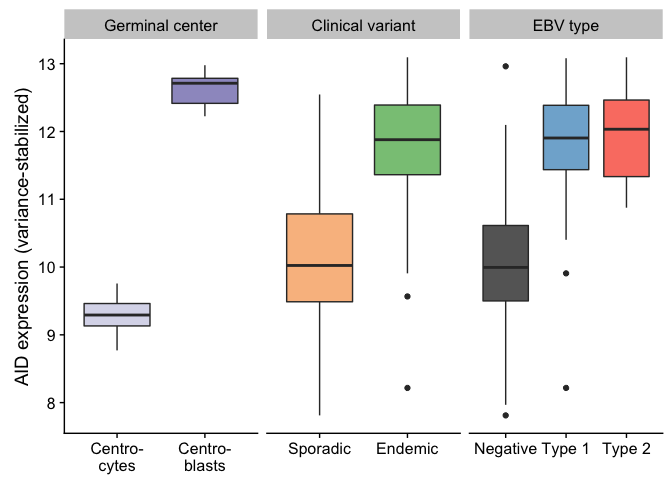


Figure 2 Box plots showing the variance-stabilized expression of AID in germinal center cells and tumor samples (comparing either clinical variant status or EBV infection status).

While several clusters of non-coding mutations overlap known targets of aberrant SHM, many affect genes that have not yet been linked to BL. One cluster overlaps a validated enhancer that is known to regulate the expression of *PAX5*, which plays an important role in B-cell differentiation. Mutations in this enhancer were first described in chronic lymphocytic leukemia (CLL)2, although we observe a higher mutation rate in BL (9.8% versus 11%). As in CLL, the mutation status of this enhancer is significantly associated with *PAX5* expression. Tumors with a mutated enhancer have higher *PAX5* expression, consistent with AID-mediated SHM of this presumably active enhancer (Supplementary Figure ??). Another target of non-coding mutations is the miR-142 locus. Variants affecting the mature sequence of miR-142 have been reported in DLBCL and FL3,4 in up to 20% of cases, but not in BL despite being investigated in one of those studies. We observe 3 cases (%) with mutations in either miR-142-3p or miR-142-5p. In our data, most mutations in this cluster are located within roughly 1 kb upstream from miR-142 on the negative strand. Altogether, these mutations affect 22% of our cases. Interestingly, Figure 1 shows that they are exclusive to EBV-positive cases (P-value 0.000007, Fisher’s exact test), resulting in a 33% mutation rate. Given they are mostly located outside of miR-142, it is possible that these variants disrupt a regulatory element that modulates the expression of the miRNA or other nearby genes. While we did not observe a difference in expression of either miR-142-5p (P-value 0.572) or miR-142-3p (P-value 0.521) based on the mutation status of this cluster, a number of nearby genes were differentially expressed, most being downregulated in cases with mutations (Supplementary Figure 18).

Figure 3 Non-coding mutations in *PAX5*.

Figure 3 Non-coding mutations in PAX5.

Figure 4 Non-coding mutations in *miR-142*.

Figure 4 Non-coding mutations in miR-142.

Lastly, a cluster of non-coding mutations seem to target the promoter region of the *PVT1* non-coding locus, which encodes the *PVT1* long non-coding RNA (lncRNA) and multiple miRNAs including miR-1204. *PVT1* is 60 kb downstream from the *MYC* locus and often harbours IG-MYC translocations due to this proximity. It is also a target of Myc-mediated upregulation5, which may account for its high expression in BL tumors: the median expression values in transcripts per million (TPM) for *MYC* and *PVT1* are in the 99th percentile. The cluster in question is mutated in 15% of cases, and its mutation status is significantly associated with higher *PVT1* expression (P-value 0.0159). Similar to the miR-142 cluster, mutations in this cluster are significantly associated with EBV-positive tumors (P-value 0.0103, Fisher’s exact test; Figure 1). To identify potential functional roles for these non-coding mutations, we looked for an enrichment of mutations in p53 binding sites, defined by chromatin immunoprecipitation paired-end tag sequencing (ChIP-PET) peaks. One such peak overlapped the *PVT1* promoter region and cluster of non-coding mutations and was deemed significantly mutated by these non-coding variants (adjusted P-value 0.0508, OncodriveFML). While the driver status for these mutations needs to be functionally validated, unpublished work has shown that *PVT1* acts as a tumor suppressor gene by lowering *MYC* expression despite published reports supporting an oncogene role in other cancer types6,7. These mutations suggest a potential role for AID in promoting BL formation through non-coding mutagenesis of biologically relevant targets.

Figure 5 Non-coding mutations in *PVT1*.

Figure 5 Non-coding mutations in PVT1.

### Differential AID expression is not related to cell-of-origin

* *TODO: Update the centroblast scores and add germinal center cells to centroblast score boxplot.*

Figure 2 shows that the difference in AID expression between EBV-positive and -negative tumors mirrors the difference between germinal center centroblasts and centrocytes. These differences in AID expression may be linked to differences in the tumors’ cell-of-origin, which may be linked to EBV infection. To test this hypothesis, a supervised approach was devised to compare the expression profile of each tumor with that of centroblasts and centrocytes using a weighted voting algorithm8,9. The algorithm used 128 genes that are differentially expressed between dark zone and light zone cells. This resulted in a centroblast score for each tumor, where positive and negative scores indicate a greater similarity to centroblasts and centrocytes, respectively. The majority of tumors (95%) are more similar to cells from the germinal center dark zone in terms of gene expression. This is consistent with previous reports that BL tumors derive from centroblasts10,11. However, unlike previous studies, we note the existence of a minority of BL tumors (5%) that are more similar to centrocytes. Heterogeneity in the cell-of-origin has been previously described where a majority of BL tumors were more similar to centrocytes9. The biological and clinical significance of this heterogeneity is not clear. One limitation of this approach is that it does not distinguish between B-cells undergoing differentiation in the germinal center for the first time and B-cells re-transiting through the germinal center for additional affinity maturation.

The difference in centroblast score between EBV-positive and -negative cases is small but significant (P-value 0.0052, Mann-Whitney U test). On the other hand, the difference based on clinical variant status is insignificant (P-value 0.0552). In other words, EBV-positive BL tumors distinguish themselves from their EBV-negative counterparts by resembling centroblasts to a greater degree, irrespective of clinical variant status (Supplementary Figure 21. This distinction may be caused by differences in cell-of-origin or might be a consequence of a centroblast-like expression program that EBV enforces. This result lends support to the notion that despite their correlated nature, EBV infection status results in more biologically meaningful differences than clinical variant status. However, the difference in centroblast score does not have the same magnitude as the difference between centroblasts and centrocytes. Therefore, it seems unlikely that cell-of-origin alone can account for the significantly greater AID expression in EBV-positive tumors. Recent work has shown that EBV is capable of inducing AID expression in lymphoblastic cell lines (LCLs)12. EBV achieves this through durable epigenetic changes of the *AICDA* gene locus modulated by the protein encoded by the latent gene, *EBNA3C*. Our findings seem to extend this observation to BL tumors *in vivo* and potentially implicate EBV in AID-mediated mutagenesis in EBV-positive BL.

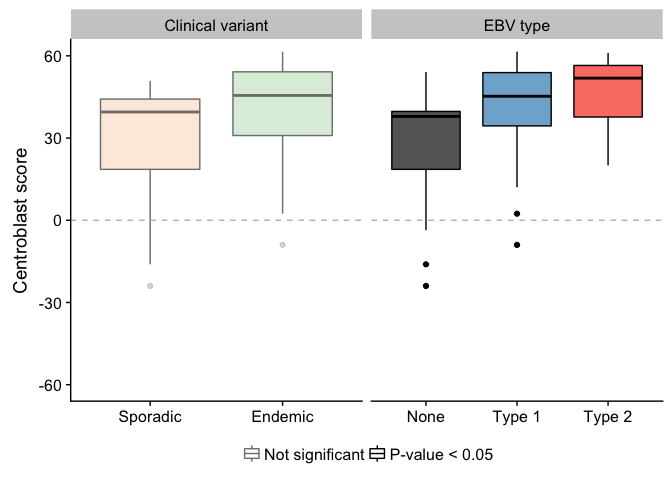


Figure 6 Box plots showing the centroblast score for germinal center cells and tumor samples (comparing either clinical variant status or EBV infection status). Positive and negative scores correspond to greater similarity to centroblasts and centrocytes, respectively.

### Structural deregulation of *MYC* expression

* *TODO: Circos plot showing rearrangements between MYC, IGH, IGK and IGL loci (one per pseudo-chromosome in Circos) with the links colored by clinical variant status or EBV status. Overlapping mutations can be shown to highlight SHM patterns. Make sure this plot shows the distribution of breakpoints in MYC.*
* *TODO: Look for any translocations near IG loci, MYC, BCL6 and BCL2. Now, I’m mostly focusing on translocations between IG loci and MYC, but I should expand my search to other recurrently translocated loci in lymphomas.*

Table 1 tabulates the partners involved in *MYC* translocations. In all but two cases, simple IG-*MYC* translocations were identified, with a majority involving *IGH* and the remainder associated with *IGL* and *IGK*. In another case (BL123), we detected two t(3;8)(q27;q24) translocations near *BCL6* and *MYC*, which have been previously described in BL and DLBCL13–16. Interestingly, the breakpoints for each translocation flank focal gains at each locus. Based on the orientation of the translocations, these regions may form part of a double minute consisting of the 3’ portion of *MYC* (after codon 79) and a portion of the *BCL6* super-enhancer. While *MYC* is not expressed as highly as in other tumors (percentile 33), its expression remains significantly higher than the normal germinal center cells (fold change 5.2).

The remaining case (BL277) harbors a complex three-way rearrangement involving *MYC*, *IGH* and a locus proximal to *GNA13* on chromosome 17, t(8;14;17)(q24;q32;q24). *MYC* expression is high (percentile 45), especially compared to centroblasts and centrocytes (fold change 6.4). We also inspected the expression of genes within 1 Mb of the breakpoint on chromosome 17 (Supplementary Figure 22). *GNA13* is the only gene in this region with an established role in BL, a *bona fide* tumor suppressor gene. While its expression is relatively high in this tumor (percentile 97), a somatic splice site mutation that causes the skipping of exon 2 has a variant allele fraction (VAF) of 86%, indicating that this region has underwent loss-of-heterozygosity. The RNA-seq data confirms that despite high expression, most *GNA13* transcripts lack exon 2 and are thus predicted to be non-functional.

Despite *MYC* being deregulated in all tumors by structural alterations, one endemic case harbored both an *IGH*-*MYC* translocation and a chromosome 11q aberration with the characteristic proximal gain and distal loss. Although, it is not possible to determine the chronology of these structural variations with the available data, which would determine whether the *MYC* translocation superseded the 11q aberration. The 11q abnormality has been previously associated with *MYC*-negative cases17 and prompted the creation of a new entity, Burkitt-like lymphoma with 11q aberration, in the 2016 revision of the World Health Organization (WHO) classification of lymphoid neoplasms18. Recent work has demonstrated the existence of sporadic BL tumors harboring both events19,20, and we extend this observation to an EBV-positive endemic case.

### IGH gene usage for productive allele

**Finish by: January 9th, 2018**

* *TODO: Plot clone fraction per sample to show how we chose dominant clones.*
* *TODO: Obtain P-values by comparing to healthy B-cell repertoires.*
* *TODO: Compare gene usage according to CV or EBV infection status.*
* *TODO: Look at combination in addition to individual gene usage.*

Given the important role for the B-cell receptor (BCR) in BL, we sought to characterize the Ig heavy (IGH) and light (IGK/IGL) chains of the productive (*i.e.* expressed) allele. MiXCR was applied to the RNA-seq data in sensitive mode in order to identify Ig clones21,22. Dominant clones were defined as those with a clonal fraction above 3%. For each dominant clone, we considered the top-scoring V, D, J and constant genes (Figure (fig:caption-barplot-ighv-freq)). The resulting gene usage distribution is non-uniform, with an enrichment for select Ig genes. Notably, we replicate a previously reported enrichment for *IGHV3-30*, *IGHV4-34* and *IGHV4-59* among the *IGHV* genes23,24. Perhaps the most striking result is the strong enrichment for *IGKV3-20* among *IGKV* genes, although with no association with clinical variant or EBV infection status. While *IGKV3-20* stereotypy has been described in other B-cell lymphoproliferations26, this is the first time it is linked to BL.

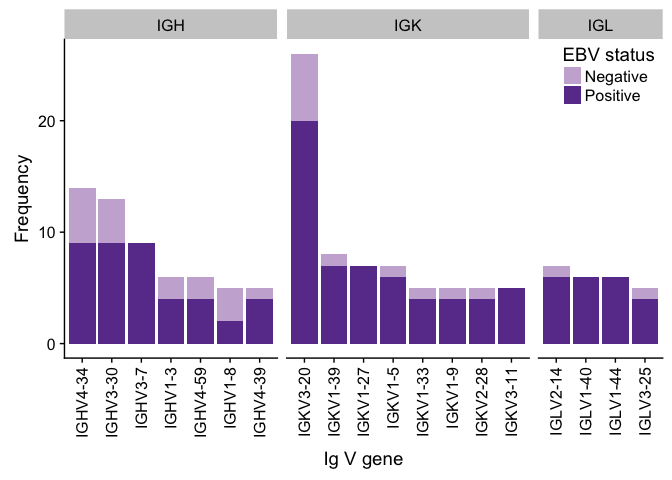


Figure 7 Frequency of VDJ and constant gene usage in dominant clones. Only displaying genes with at least a count of five.

*IGHM* is highly expressed in most tumors with no discernable difference based on clinical variant or EBV infection status (Figure 8. In the subset of samples with low *IGHM* expression, *IGHG1* displays a seemingly compensatory increase in expression. Indeed, IgG expression is anticorrelated with IgM expression (Supplementary Figure 25). One limitation of this observation is that we cannot attribute the Ig expression levels directly to the tumor due to the possibility of infiltrating B-cells. Inspection of the Ig clones detected by MiXCR reveals 4 cases with simultaneous IgM and IgG dominant clones. However, none share the same VDJ gene combination, indicating that the IgG expression is not deriving from a tumor subclone that has class-switched. Also, the clonal fraction for the IgG clonotypes is consistently low (Supplementary Figure 24). Lastly, we sought to establish whether IgM and IgG co-expression was restricted to cases with *MYC* translocations involving Ig light chains. Our hypothesis was that tumor cells could express both *IGH* alleles using different constant genes. Given high IgG expression in *IGH*-rearranged cases, this seemed unlikely to account for the Ig co-expression in our samples (Supplementary Figure 26)

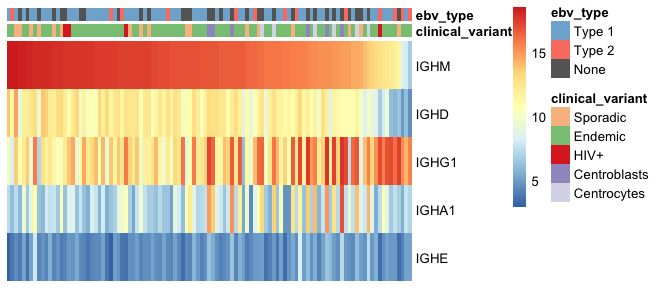


Figure 8 Expression of IGH constant genes (sorted by IGHM expression).

### Mutation signatures vary according to EBV infection status

**Finish by: January 10th, 2018**

* *TODO: Attempt to explain mutation burden using AID expression, mutation signatures and telomere content.*

The genome-wide mutation load is significantly higher in endemic or EBV-positive cases compared to their sporadic and EBV-negative cases, respectively (Figure (fig:boxplot-subtype-mutationcount-covariate-mutationtype-wilcoxon)). The same is true for non-synonymous mutations in all protein-coding genes. AID-mediated SHM might be the underlying cause for this difference. Interestingly, when we restrict to non-synonymous mutations in significantly mutated genes as described below, the direction of the difference is reversed and remains significant only according to EBV infection status, despite the generally higher mutation burden in EBV-positive cases. This result is consistent with an oncogenic role for EBV and reinforces the notion that EBV infection status may be a more biologically meaningful classification than clinical variant status.

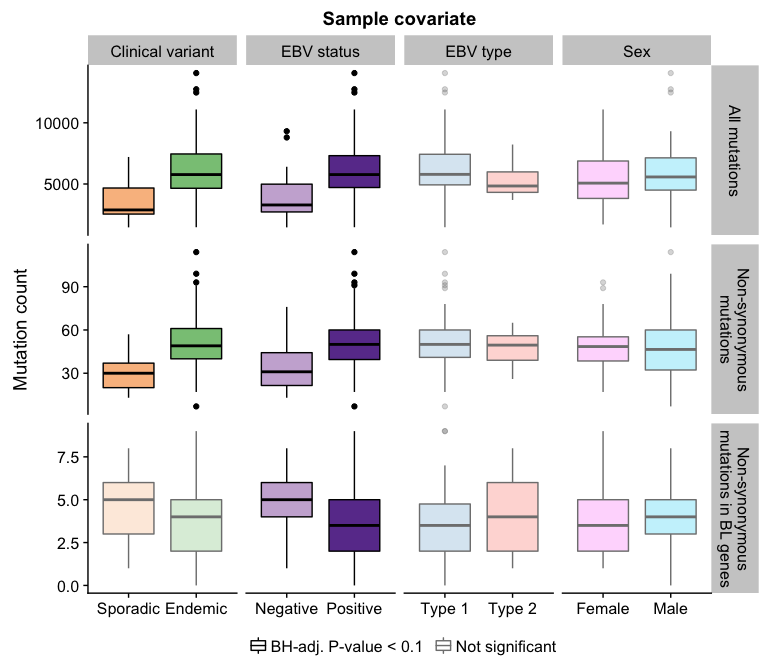


Figure 9 Frequency of all mutations, non-synonymous mutations and non-synonymous mutations in BL genes in various disease subtypes. Highlighted boxplots show a significant difference (Mann-Whitney U test).

Write this section once you have updated mutation signature results.

**Figures**

10

11

**Supplementary Figures**

29

28

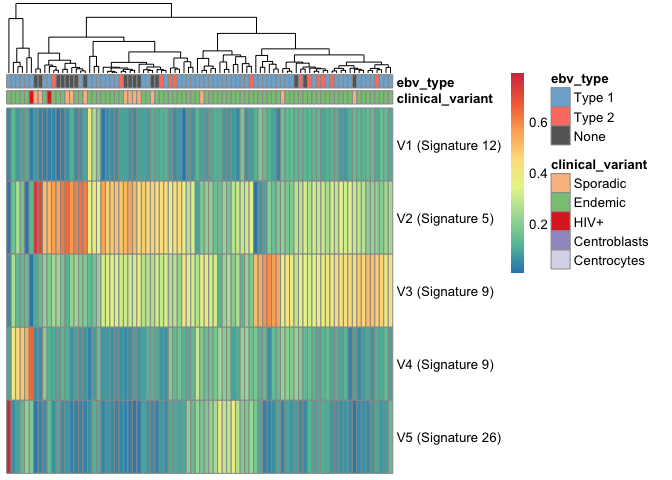


Figure 10 Exposure fractions across the cohort.

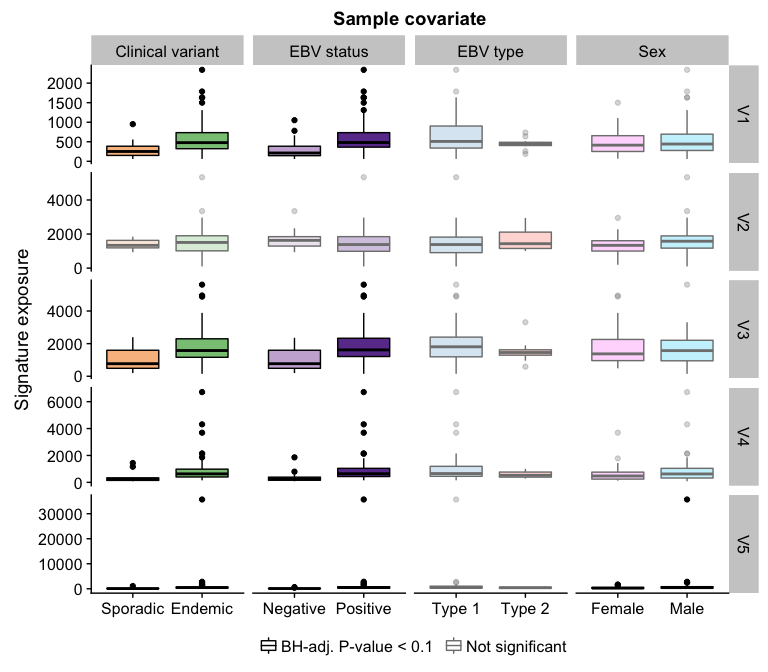


Figure 11 A figure showing the differential exposure of mutation signatures in BL subtypes (clinical variant status or EBV infection status).

### Significantly mutated genes associate with EBV infection status

**Finish by: January 14th, 2018**

* *TODO: Look at all possible pairs for mutual exclusivity.*
* *TODO: Perform differential mutation rate analysis while considering other mutation types.*

We identified significantly mutated genes (SMGs) by employing an ensemble approach that leveraged four methods and expected support from at least two [REF]. Established BL genes figured on our list of SMGs, such as *DDX3X*, *ID3*, *TCF3*, *ARID1A*, *SMARCA4*, *TP53*, *FOXO1* and *CCND3*. We were also able to replicate recent reports that uncovered *TFAP4* and *KMT2D* as being recurrently mutated in BL. On the other hand, we note the absence of somatic mutations in *CCNF*, which has been described in endemic BL. We see the same hotspot mutation in two endemic cases but as germline events. It also appears as a rare allele in the African population in dbSNP [REF]. In addition, our dataset has allowed the identification of genes that have not yet been described as significantly mutated, namely *SIN3A*, *HIST1H1E*, *USP7*, *CHD8* and *RFX7*. Some of these novel genes carry out functions that are relevant to BL biology. Notably, *SIN3A* is a known antagonist of *MYC* activity, thus the diffuse pattern of truncating mutationsis consistent with a tumor suppressor role. *USP7* encodes a deubiquitinase that counteracts Mdm2-mediated ubiquitination and degradation of p53. Mutations in *USP7* are nearly exclusively truncating, consistent with its putative role as another tumor suppressor gene. Finally, *HIST1H1E* encodes a histone component and *CHD8* is involved in chromatin remodeling, reinforcing the critical role for chromatin organization in B-cell lymphomagenesis.

DDX3X

Figure 12 Oncoprint plot for BL cohort.

Figure 12 Oncoprint plot for BL cohort.

Figure 13 Clustering of mutations in GNAI2-encoded protein structure.

Figure 13 Clustering of mutations in GNAI2-encoded protein structure.

Many previous studies have compared the mutation rates of various BL genes between disease subtypes such as clinical variant status, EBV infection status and more recently, EBV strain type. However, these studies were limited by small cohort sizes and/or issues related to calling somatic variants from RNA-seq data. Our dataset is amenable for these types of analyses given its size and high-quality WGS data. Indeed, a number of SMGs are significantly differentially mutated when comparing either endemic and sporadic cases or EBV-positive and -negative cases (Fisher’s exact test). Notably, following multiple test correction, there were more differences on the basis of EBV infection status, with *SMARCA4*, *CCND3* and *TP53* (with and without *USP7*) being more commonly mutated in EBV-positive cases. We saw no differences in the number of mutations between EBV type 1 and type 2, nor between females and males.

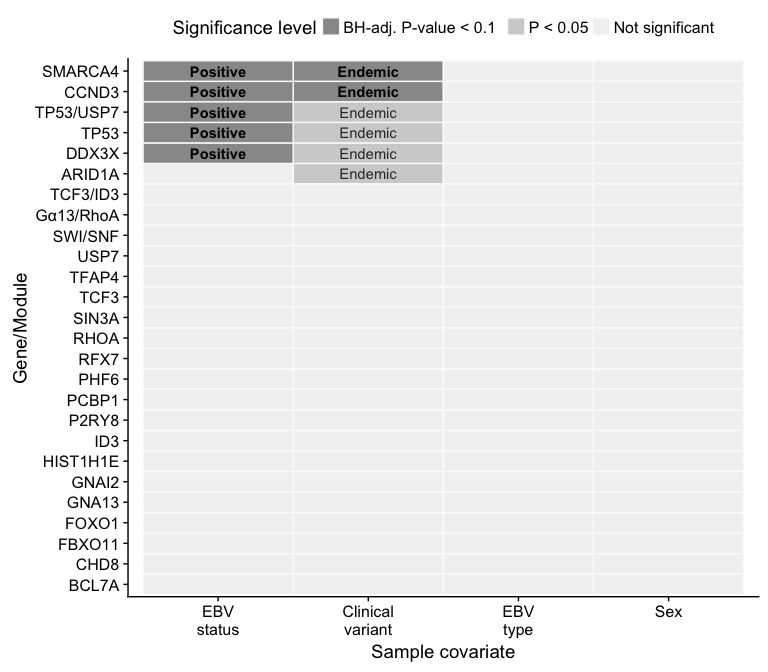


Figure 14 Differential mutation rates for BL SMGs.

## Discussion

**Finish by: February 2nd, 2018**

**Target length: 1000 words**

To be done.

## Acknowledgments

* ICGC MMML-Seq Project
* CIHR Travel Award
* CCSRI Travel Award

# Supplementary Material

## Supplementary Figures

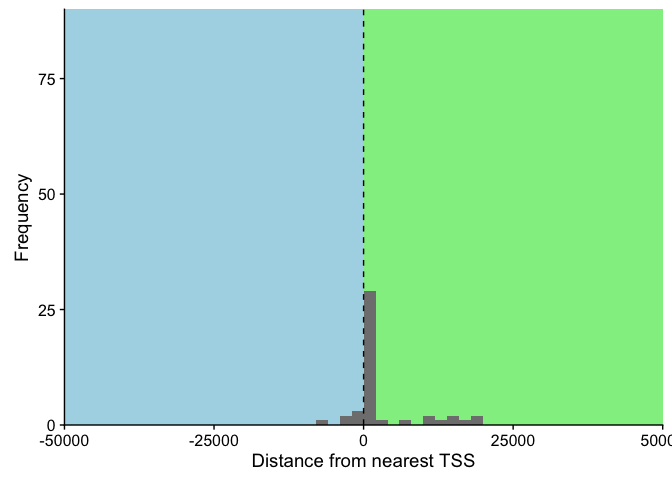


Figure 15 Distance between cluster of non-coding mutations and nearest transcription start site (TSS). Only considering clusters within 50000 bp of a TSS, which represents 96% of all clusters.

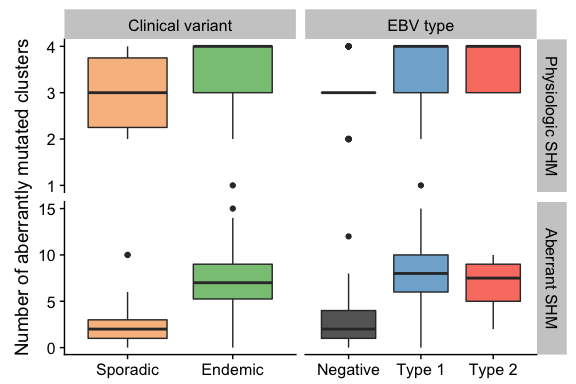


Figure 16 Number of mutated clusters of non-coding mutations per sample according to clinical variant or EBV infection status. HIV+ cases are omitted for simplicity.

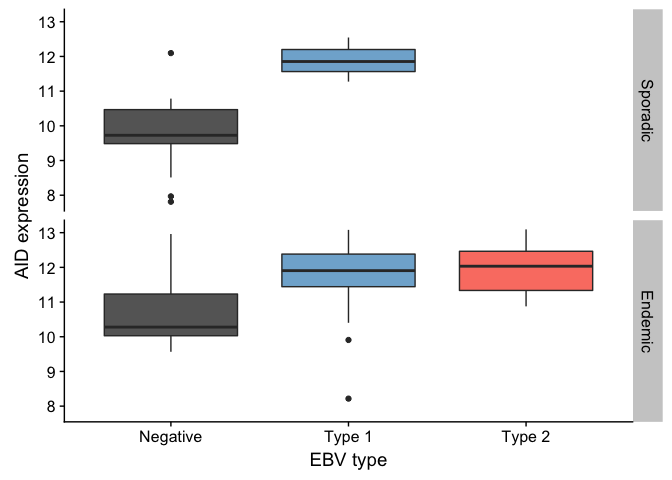


Figure 17 AID expression is higher in EBV-positive cases in both endemic and sporadic cases.

NULL

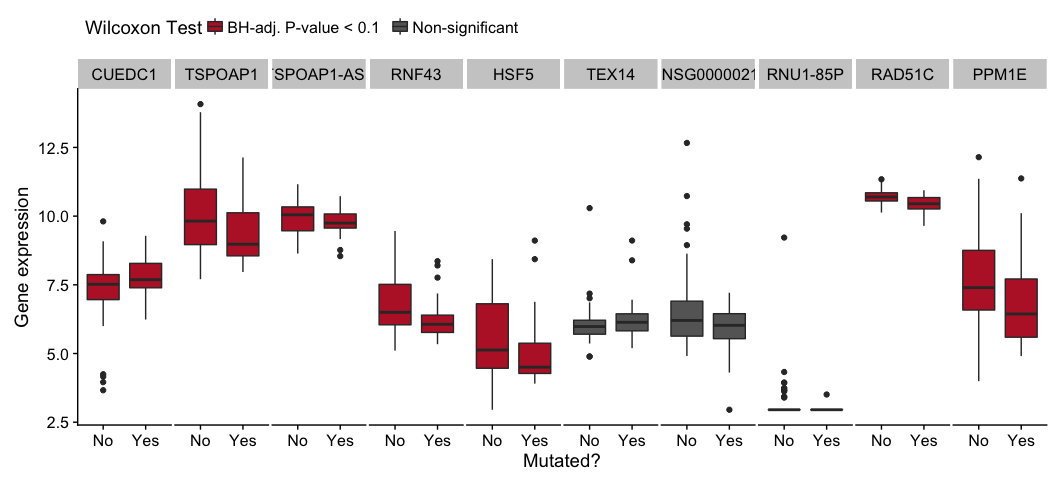


Figure 18 Differential expression of genes near the cluster of non-coding mutations that overlaps miR-142 as a function of its mutation status.



Figure 19 Differential expression of genes near the cluster of non-coding mutations that overlaps PVT1 as a function of its mutation status.

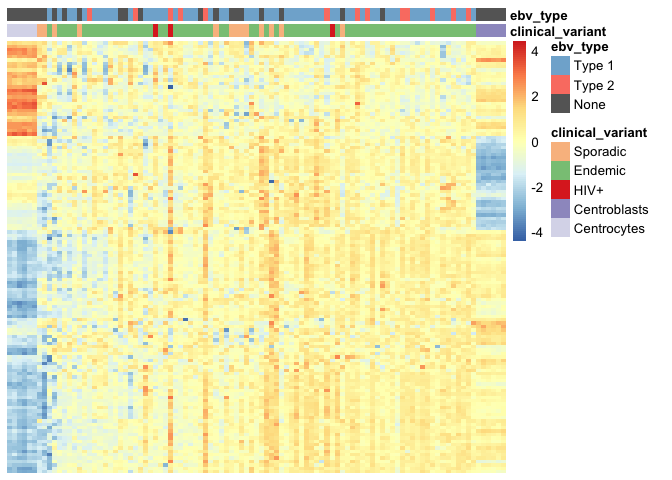


Figure 20 Heatmap showing expression of genes used to calculate centroblast score

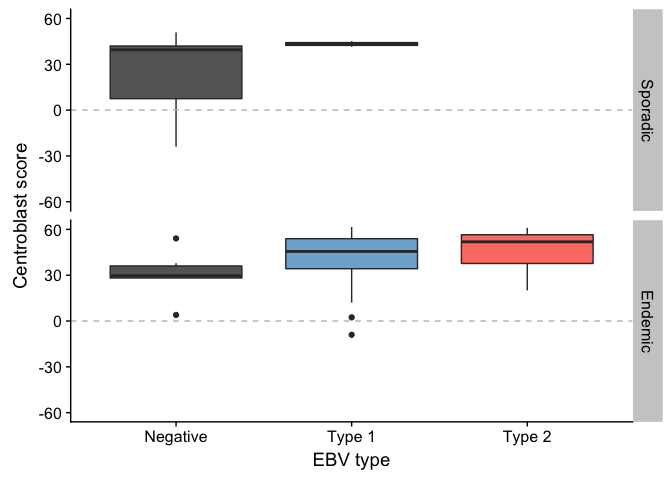


Figure 21 Centroblast score is higher in EBV-positive cases in both endemic and sporadic cases.

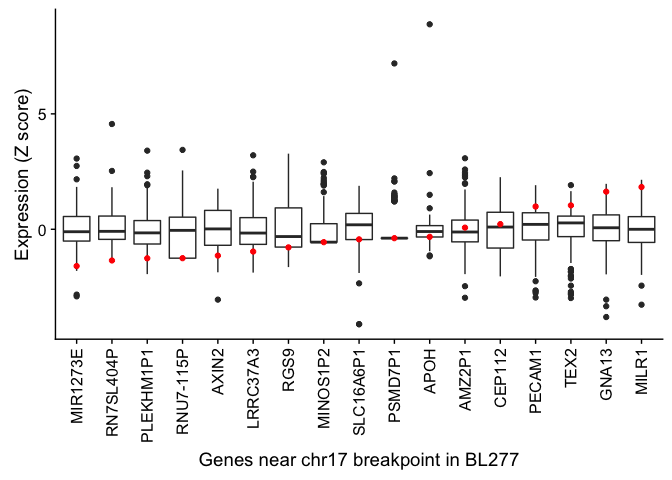


Figure 22 Expression (Z score) of genes within 1 Mb of the chr17 breakpoint in BL277.

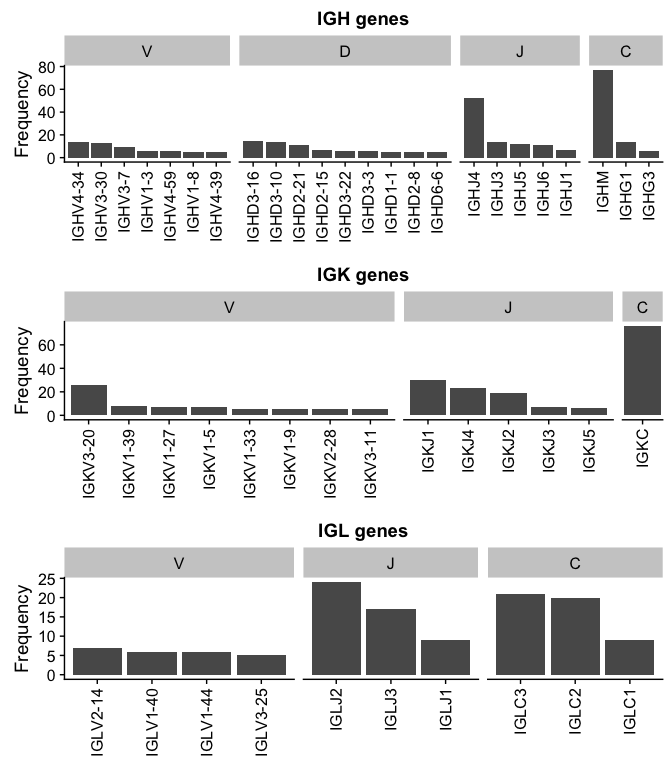


Figure 23 Ig gene usage distribution for all V, D, J and C genes.

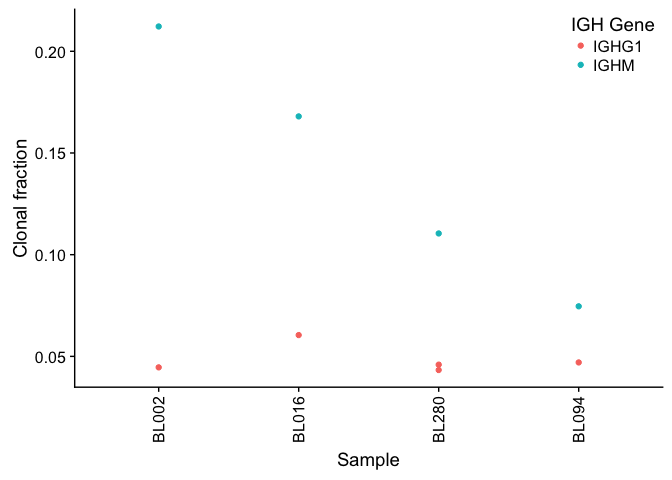


Figure 24 Clonal fraction for co-dominant IgM and IgG clones.

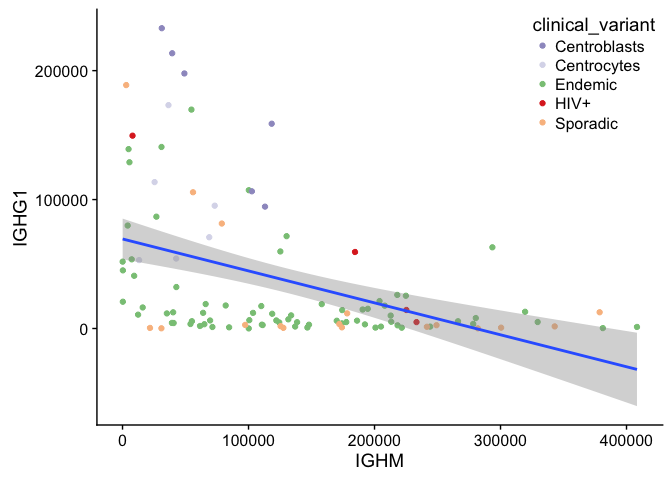


Figure 25 Anti-correlated IGHM and IGHG1 expression.

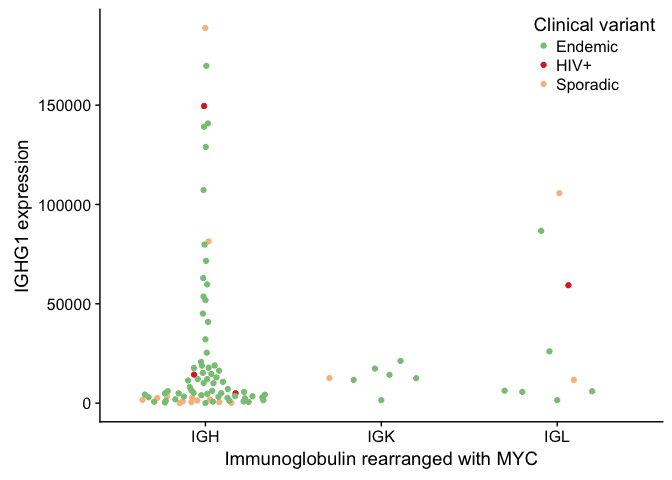


Figure 26 No association between IGG expression and which immunnoglobulin is rearranged with MYC.

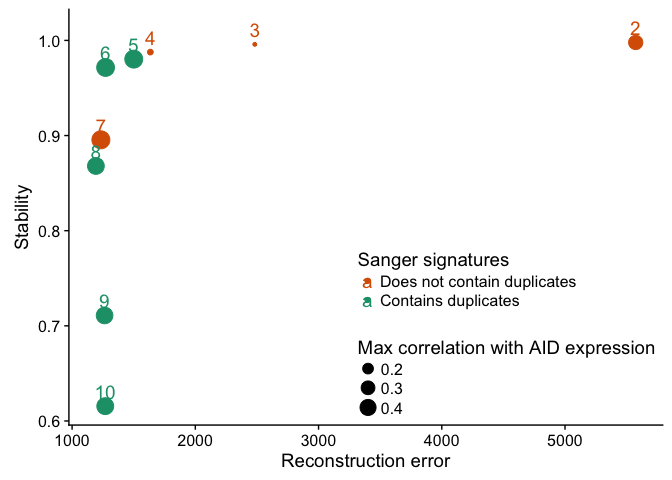


Figure 27 Selecting an optimal number of signatures.

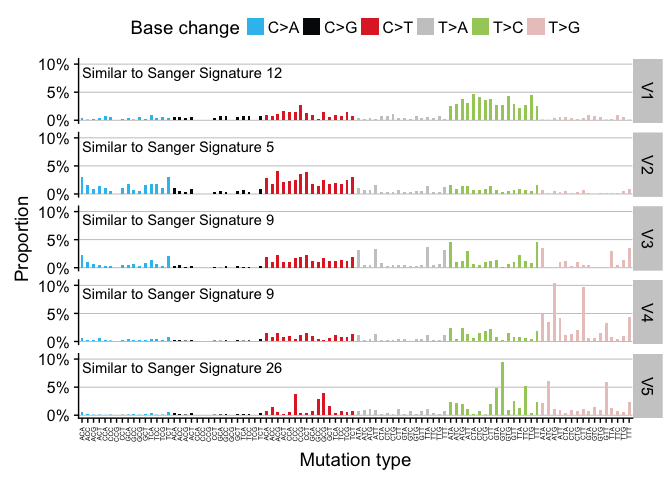


Figure 28 Mutation signatures breakdown.



Figure 29 There is minimal correlation amongst the WTSI signatures.

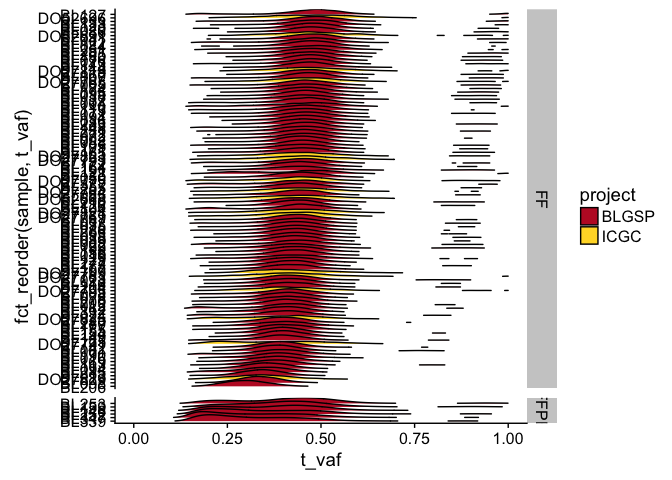


Figure 30 Variant allele fraction (VAF) distribution for SNVs and indels across all tumors.

## Supplementary Tables

Table 2 Genes used in weighted voting algorithm for COO prediction.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| gene | t | df | p.value | cor | FDR | meanGroup1 | meanGroup0 | nGroup1 | nGroup0 |
| AC004381.6 | 17.99001 | 10 | 0 | 0.9848998 | 0 | 9.273472 | 7.688577 | 6 | 6 |
| ADAM8 | -16.67476 | 10 | 0 | -0.9824884 | 0 | 9.685855 | 11.585050 | 6 | 6 |
| ADAT2 | -16.52835 | 10 | 0 | -0.9821851 | 0 | 8.494233 | 9.724786 | 6 | 6 |
| AICDA | 18.40522 | 10 | 0 | 0.9855589 | 0 | 12.470056 | 9.127022 | 6 | 6 |
| ALG8 | -20.72813 | 10 | 0 | -0.9885620 | 0 | 9.732595 | 10.371011 | 6 | 6 |
| ANKRD36BP2 | 21.26383 | 10 | 0 | 0.9891218 | 0 | 11.503114 | 9.617483 | 6 | 6 |
| ANLN | 20.22239 | 10 | 0 | 0.9879932 | 0 | 11.956894 | 10.368462 | 6 | 6 |
| ASCC3 | -15.87857 | 10 | 0 | -0.9807400 | 0 | 11.548072 | 12.370395 | 6 | 6 |
| ATP6V0A2 | -16.72785 | 10 | 0 | -0.9825965 | 0 | 11.276957 | 11.994688 | 6 | 6 |
| ATP8B3 | 18.38185 | 10 | 0 | 0.9855230 | 0 | 9.509690 | 7.673152 | 6 | 6 |
| AURKA | 15.83935 | 10 | 0 | 0.9806472 | 0 | 11.289107 | 9.395377 | 6 | 6 |
| BUB1 | 23.36022 | 10 | 0 | 0.9909615 | 0 | 12.810233 | 11.654624 | 6 | 6 |
| CALM3 | 17.44642 | 10 | 0 | 0.9839670 | 0 | 14.399203 | 13.066351 | 6 | 6 |
| CCL22 | -20.85243 | 10 | 0 | -0.9886957 | 0 | 7.688588 | 10.847064 | 6 | 6 |
| CCNB2 | 22.90894 | 10 | 0 | 0.9906069 | 0 | 12.523602 | 10.767903 | 6 | 6 |
| CDC20 | 16.50972 | 10 | 0 | 0.9821459 | 0 | 12.916529 | 10.531248 | 6 | 6 |
| CDC25B | 19.00225 | 10 | 0 | 0.9864340 | 0 | 13.230491 | 10.908478 | 6 | 6 |
| CDCA3 | 26.45194 | 10 | 0 | 0.9929298 | 0 | 11.190156 | 9.395053 | 6 | 6 |
| CDK13 | 17.59437 | 10 | 0 | 0.9842292 | 0 | 13.574453 | 12.524293 | 6 | 6 |
| CDKN3 | 21.58777 | 10 | 0 | 0.9894408 | 0 | 10.547587 | 8.872735 | 6 | 6 |
| CENPA | 16.97987 | 10 | 0 | 0.9830964 | 0 | 9.942100 | 8.011155 | 6 | 6 |
| CENPE | 19.99240 | 10 | 0 | 0.9877204 | 0 | 12.511381 | 10.571742 | 6 | 6 |
| CENPF | 22.18668 | 10 | 0 | 0.9899947 | 0 | 13.995880 | 11.984171 | 6 | 6 |
| CEP128 | 20.56797 | 10 | 0 | 0.9883863 | 0 | 10.615381 | 8.934057 | 6 | 6 |
| CKAP2L | 18.18994 | 10 | 0 | 0.9852226 | 0 | 11.294683 | 9.828786 | 6 | 6 |
| CKAP5 | 20.83706 | 10 | 0 | 0.9886793 | 0 | 13.948876 | 12.872967 | 6 | 6 |
| CNNM3 | 18.70615 | 10 | 0 | 0.9860102 | 0 | 11.510351 | 10.514944 | 6 | 6 |
| CTIF | -20.29476 | 10 | 0 | -0.9880771 | 0 | 5.833588 | 7.234523 | 6 | 6 |
| DDX31 | -16.81421 | 10 | 0 | -0.9827703 | 0 | 8.924861 | 10.085787 | 6 | 6 |
| DEPDC1 | 20.77609 | 10 | 0 | 0.9886139 | 0 | 11.282544 | 9.275876 | 6 | 6 |
| DHODH | -15.60425 | 10 | 0 | -0.9800771 | 0 | 9.234021 | 9.751812 | 6 | 6 |
| DLGAP5 | 16.90778 | 10 | 0 | 0.9829556 | 0 | 11.800979 | 10.028697 | 6 | 6 |
| DNMT3B | 23.08312 | 10 | 0 | 0.9907462 | 0 | 9.247867 | 6.600045 | 6 | 6 |
| EBI3 | -28.93069 | 10 | 0 | -0.9940792 | 0 | 6.314655 | 11.045204 | 6 | 6 |
| ECT2 | 27.17753 | 10 | 0 | 0.9932986 | 0 | 11.527439 | 10.330274 | 6 | 6 |
| EEF2K | -15.78078 | 10 | 0 | -0.9805075 | 0 | 10.580837 | 11.406475 | 6 | 6 |
| FAM72A | 19.08746 | 10 | 0 | 0.9865524 | 0 | 10.129673 | 8.546356 | 6 | 6 |
| FBXO5 | 17.94219 | 10 | 0 | 0.9848210 | 0 | 10.677264 | 10.026028 | 6 | 6 |
| FHOD1 | -22.35500 | 10 | 0 | -0.9901426 | 0 | 11.107803 | 11.824779 | 6 | 6 |
| FOXM1 | 19.70410 | 10 | 0 | 0.9873653 | 0 | 12.354078 | 10.918852 | 6 | 6 |
| GALNT6 | -20.78716 | 10 | 0 | -0.9886258 | 0 | 9.876660 | 11.497068 | 6 | 6 |
| GCSAM | 18.43859 | 10 | 0 | 0.9856100 | 0 | 13.224285 | 11.538907 | 6 | 6 |
| GLS | -21.44352 | 10 | 0 | -0.9893005 | 0 | 13.045643 | 13.721683 | 6 | 6 |
| GPSM2 | 35.11818 | 10 | 0 | 0.9959703 | 0 | 11.480606 | 9.941918 | 6 | 6 |
| GRK3 | -17.08265 | 10 | 0 | -0.9832941 | 0 | 11.779250 | 12.788450 | 6 | 6 |
| GTSE1 | 16.54461 | 10 | 0 | 0.9822192 | 0 | 11.732077 | 10.157946 | 6 | 6 |
| HIVEP3 | -19.81442 | 10 | 0 | -0.9875030 | 0 | 11.403910 | 13.648372 | 6 | 6 |
| HMGB2 | 18.73157 | 10 | 0 | 0.9860473 | 0 | 13.802784 | 12.029731 | 6 | 6 |
| HMMR | 19.60711 | 10 | 0 | 0.9872424 | 0 | 11.801605 | 9.876689 | 6 | 6 |
| IPO11 | -27.08765 | 10 | 0 | -0.9932545 | 0 | 9.940090 | 10.556062 | 6 | 6 |
| IQGAP3 | 18.46188 | 10 | 0 | 0.9856455 | 0 | 11.277484 | 9.408517 | 6 | 6 |
| IQSEC1 | -21.91772 | 10 | 0 | -0.9897514 | 0 | 11.213566 | 12.914764 | 6 | 6 |
| IRF5 | -17.40281 | 10 | 0 | -0.9838885 | 0 | 9.806838 | 11.218324 | 6 | 6 |
| JAK3 | -22.67661 | 10 | 0 | -0.9904163 | 0 | 13.055447 | 14.096109 | 6 | 6 |
| KANK2 | 19.42436 | 10 | 0 | 0.9870059 | 0 | 12.575663 | 8.986882 | 6 | 6 |
| KIF14 | 20.24977 | 10 | 0 | 0.9880250 | 0 | 11.380907 | 9.375757 | 6 | 6 |
| KIF18A | 26.07303 | 10 | 0 | 0.9927251 | 0 | 10.925375 | 10.034092 | 6 | 6 |
| KIF18B | 17.07144 | 10 | 0 | 0.9832727 | 0 | 12.557645 | 10.874799 | 6 | 6 |
| KIF23 | 16.79354 | 10 | 0 | 0.9827289 | 0 | 11.704113 | 10.446041 | 6 | 6 |
| KIF2C | 17.99558 | 10 | 0 | 0.9849089 | 0 | 12.124035 | 10.812551 | 6 | 6 |
| KIF4A | 19.52967 | 10 | 0 | 0.9871429 | 0 | 11.629216 | 10.110359 | 6 | 6 |
| KIFC1 | 16.49513 | 10 | 0 | 0.9821152 | 0 | 12.973700 | 11.346191 | 6 | 6 |
| KMT2A | 18.29652 | 10 | 0 | 0.9853905 | 0 | 15.080104 | 13.660543 | 6 | 6 |
| LSM10 | 16.34758 | 10 | 0 | 0.9817997 | 0 | 11.179493 | 9.465388 | 6 | 6 |
| LY75 | -20.20807 | 10 | 0 | -0.9879765 | 0 | 12.085849 | 13.690896 | 6 | 6 |
| MAN2B1 | -18.97264 | 10 | 0 | -0.9863925 | 0 | 12.233487 | 13.096951 | 6 | 6 |
| MCCC2 | -16.30998 | 10 | 0 | -0.9817180 | 0 | 11.194589 | 11.829315 | 6 | 6 |
| MIOS | -17.91951 | 10 | 0 | -0.9847835 | 0 | 11.341148 | 12.573118 | 6 | 6 |
| MKI67 | 17.57007 | 10 | 0 | 0.9841866 | 0 | 15.688158 | 14.165188 | 6 | 6 |
| MS4A1 | -18.84718 | 10 | 0 | -0.9862145 | 0 | 15.289834 | 16.277485 | 6 | 6 |
| MTMR4 | -19.90047 | 10 | 0 | -0.9876088 | 0 | 11.981768 | 12.951977 | 6 | 6 |
| MXD3 | 16.25681 | 10 | 0 | 0.9816015 | 0 | 10.616526 | 8.716301 | 6 | 6 |
| NAA25 | -18.31013 | 10 | 0 | -0.9854118 | 0 | 11.081702 | 12.315264 | 6 | 6 |
| NCAPD2 | 21.07284 | 10 | 0 | 0.9889270 | 0 | 14.593813 | 13.226595 | 6 | 6 |
| NCAPH | 16.17730 | 10 | 0 | 0.9814252 | 0 | 11.872993 | 10.745387 | 6 | 6 |
| NDC80 | 18.42767 | 10 | 0 | 0.9855933 | 0 | 11.141136 | 10.059604 | 6 | 6 |
| NEIL3 | 17.50274 | 10 | 0 | 0.9840676 | 0 | 10.049762 | 8.465697 | 6 | 6 |
| NEK2 | 23.70452 | 10 | 0 | 0.9912187 | 0 | 10.580983 | 8.890369 | 6 | 6 |
| NLN | -17.13686 | 10 | 0 | -0.9833970 | 0 | 8.840430 | 10.392522 | 6 | 6 |
| NT5DC3 | -17.43905 | 10 | 0 | -0.9839538 | 0 | 8.814142 | 10.063417 | 6 | 6 |
| NUF2 | 23.35598 | 10 | 0 | 0.9909583 | 0 | 10.811161 | 9.173961 | 6 | 6 |
| NUGGC | 16.50703 | 10 | 0 | 0.9821403 | 0 | 14.423387 | 12.541249 | 6 | 6 |
| NUSAP1 | 18.78751 | 10 | 0 | 0.9861286 | 0 | 13.116843 | 11.639367 | 6 | 6 |
| PARP14 | -25.95541 | 10 | 0 | -0.9926597 | 0 | 12.417817 | 14.171772 | 6 | 6 |
| PDIA6 | -15.45224 | 10 | 0 | -0.9796951 | 0 | 12.549520 | 13.713688 | 6 | 6 |
| PHF3 | 18.43857 | 10 | 0 | 0.9856100 | 0 | 13.142487 | 12.488186 | 6 | 6 |
| PLEK | -18.67737 | 10 | 0 | -0.9859679 | 0 | 12.524681 | 14.054052 | 6 | 6 |
| PLK1 | 17.42683 | 10 | 0 | 0.9839318 | 0 | 13.032576 | 10.477862 | 6 | 6 |
| POLR1B | -25.14670 | 10 | 0 | -0.9921856 | 0 | 10.233438 | 11.434476 | 6 | 6 |
| PRC1 | 24.45833 | 10 | 0 | 0.9917451 | 0 | 12.496276 | 10.901093 | 6 | 6 |
| PRMT3 | -20.30800 | 10 | 0 | -0.9880924 | 0 | 8.983280 | 10.261198 | 6 | 6 |
| PRR11 | 21.26851 | 10 | 0 | 0.9891266 | 0 | 11.664168 | 10.187641 | 6 | 6 |
| PSRC1 | 18.65795 | 10 | 0 | 0.9859393 | 0 | 9.372250 | 6.518974 | 6 | 6 |
| PTMS | 18.08762 | 10 | 0 | 0.9850587 | 0 | 9.971545 | 8.078698 | 6 | 6 |
| RAB3GAP2 | -21.51565 | 10 | 0 | -0.9893710 | 0 | 12.504492 | 14.095900 | 6 | 6 |
| RACGAP1 | 18.21004 | 10 | 0 | 0.9852545 | 0 | 12.483429 | 11.153559 | 6 | 6 |
| RP11-486B10.3 | -16.42998 | 10 | 0 | -0.9819769 | 0 | 4.836816 | 6.983818 | 6 | 6 |
| RP11-486B10.4 | -18.74427 | 10 | 0 | -0.9860658 | 0 | 7.217801 | 9.781441 | 6 | 6 |
| RP3-512E2.2 | 15.39332 | 10 | 0 | 0.9795441 | 0 | 8.756973 | 7.106308 | 6 | 6 |
| SBNO2 | -16.31336 | 10 | 0 | -0.9817253 | 0 | 12.034426 | 12.733023 | 6 | 6 |
| SEMA4B | 19.53024 | 10 | 0 | 0.9871437 | 0 | 11.682386 | 9.818578 | 6 | 6 |
| SEMA7A | -18.08058 | 10 | 0 | -0.9850473 | 0 | 11.573854 | 13.994206 | 6 | 6 |
| SGO2 | 16.93514 | 10 | 0 | 0.9830092 | 0 | 11.233137 | 9.669418 | 6 | 6 |
| SH3BP2 | -14.83163 | 10 | 0 | -0.9780171 | 0 | 11.060950 | 12.206829 | 6 | 6 |
| SHQ1 | -17.58132 | 10 | 0 | -0.9842063 | 0 | 9.850923 | 10.578415 | 6 | 6 |
| SLC35F2 | -18.11323 | 10 | 0 | -0.9851000 | 0 | 9.053598 | 10.700651 | 6 | 6 |
| SMC4 | 20.28936 | 10 | 0 | 0.9880709 | 0 | 14.105521 | 13.116572 | 6 | 6 |
| SNX11 | -20.09753 | 10 | 0 | -0.9878462 | 0 | 10.875020 | 12.652434 | 6 | 6 |
| SPAG5 | 15.75173 | 10 | 0 | 0.9804376 | 0 | 12.714482 | 11.299931 | 6 | 6 |
| STAT6 | -17.76843 | 10 | 0 | -0.9845296 | 0 | 13.672719 | 14.681518 | 6 | 6 |
| SUGCT | 18.34658 | 10 | 0 | 0.9854684 | 0 | 10.056639 | 6.639025 | 6 | 6 |
| SWAP70 | -17.07365 | 10 | 0 | -0.9832769 | 0 | 14.691406 | 15.315681 | 6 | 6 |
| TANGO6 | -24.47070 | 10 | 0 | -0.9917533 | 0 | 10.038395 | 10.725384 | 6 | 6 |
| TET3 | -27.93057 | 10 | 0 | -0.9936517 | 0 | 11.002215 | 12.558617 | 6 | 6 |
| THADA | -17.54958 | 10 | 0 | -0.9841505 | 0 | 11.977110 | 12.415181 | 6 | 6 |
| TMEM120B | -17.95053 | 10 | 0 | -0.9848348 | 0 | 9.904027 | 10.724504 | 6 | 6 |
| TMEM192 | -23.48068 | 10 | 0 | -0.9910528 | 0 | 10.464384 | 10.949103 | 6 | 6 |
| TOP2A | 21.48368 | 10 | 0 | 0.9893398 | 0 | 14.549218 | 12.692722 | 6 | 6 |
| TPX2 | 18.01343 | 10 | 0 | 0.9849382 | 0 | 13.674886 | 11.955942 | 6 | 6 |
| TROAP | 16.21537 | 10 | 0 | 0.9815099 | 0 | 11.148048 | 9.322731 | 6 | 6 |
| UBE2C | 20.63188 | 10 | 0 | 0.9884569 | 0 | 11.031126 | 8.798615 | 6 | 6 |
| UTP20 | -20.60620 | 10 | 0 | -0.9884286 | 0 | 11.255062 | 12.341693 | 6 | 6 |
| VASP | -19.22721 | 10 | 0 | -0.9867433 | 0 | 12.671706 | 13.959636 | 6 | 6 |
| WDR3 | -19.62815 | 10 | 0 | -0.9872692 | 0 | 10.477528 | 11.464539 | 6 | 6 |
| WRN | -15.88713 | 10 | 0 | -0.9807601 | 0 | 10.235383 | 10.909290 | 6 | 6 |
| XPO5 | -23.81442 | 10 | 0 | -0.9912985 | 0 | 10.712007 | 11.693876 | 6 | 6 |
| YPEL3 | 22.20583 | 10 | 0 | 0.9900117 | 0 | 10.238074 | 8.945963 | 6 | 6 |
| ZNF101 | 17.67657 | 10 | 0 | 0.9843721 | 0 | 12.141379 | 11.117290 | 6 | 6 |

# References

1. Arthur S, Jiang A, Grande B, et al. Genome-wide discovery of somatic coding and regulatory variants in diffuse large b-cell lymphoma. *bioRxiv*. 2017;225870.

2. Puente XS, Beà S, Valdés-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526(7574):519–524.

3. Kwanhian W, Lenze D, Alles J, et al. MicroRNA-142 is mutated in about 20% of diffuse large b-cell lymphoma. *Cancer Med.* 2012;1(2):141–155.

4. Hezaveh K, Kloetgen A, Bernhart SH, et al. Alterations of microRNA and microRNA-regulated messenger RNA expression in germinal center b-cell lymphomas determined by integrative sequencing analysis. *Haematologica*. 2016;101(11):1380–1389.

5. Carramusa L, Contino F, Ferro A, et al. The PVT-1 oncogene is a myc protein target that is overexpressed in transformed cells. *J. Cell. Physiol.* 2007;213(2):511–518.

6. 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.

7. Tseng Y-Y, Moriarity BS, Gong W, et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature*. 2014;512(7512):82–86.

8. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*. 1999;286(5439):531–537.

9. Victora GD, Dominguez-Sola D, Holmes AB, et al. Identification of human germinal center light and dark zone cells and their relationship to human b-cell lymphomas. *Blood*. 2012;120(11):2240–2248.

10. Dave SS, Fu K, Wright GW, et al. Molecular diagnosis of burkitt’s lymphoma. *N. Engl. J. Med.* 2006;354(23):2431–2442.

11. Caron G, Le Gallou S, Lamy T, Tarte K, Fest T. CXCR4 expression functionally discriminates centroblasts versus centrocytes within human germinal center B cells. *J. Immunol.* 2009;182(12):7595–7602.

12. Kalchschmidt JS, Bashford-Rogers R, Paschos K, et al. Epstein-Barr virus nuclear protein EBNA3C directly induces expression of AID and somatic mutations in B cells. *J. Exp. Med.* 2016;213(6):921–928.

13. Sandlund JT, Kastan MB, Kennedy W, et al. A subtle t(3;8) results in plausible juxtaposition of MYC and BCL6 in a child with burkitt lymphoma/leukemia and ataxia-telangiectasia. *Cancer Genet. Cytogenet.* 2006;168(1):69–72.

14. Wang H-Y, Bossler AD, Schaffer A, et al. A novel t(3;8)(q27;q24.1) simultaneously involving both the BCL6 and MYC genes in a diffuse large b-cell lymphoma. *Cancer Genet. Cytogenet.* 2007;172(1):45–53.

15. Bertrand P, Bastard C, Maingonnat C, et al. Mapping of MYC breakpoints in 8q24 rearrangements involving non-immunoglobulin partners in b-cell lymphomas. *Leukemia*. 2007;21(3):515–523.

16. Huret JL. T(3;8)(q27;q24). 2009;

17. Salaverria I, Martin-Guerrero I, Wagener R, et al. A recurrent 11q aberration pattern characterizes a subset of MYC-negative high-grade b-cell lymphomas resembling burkitt lymphoma. *Blood*. 2014;123(8):1187–1198.

18. Swerdlow S, Campo E, Harris NL, et al. WHO classification of tumours of haematopoietic and lymphoid tissues (IARC WHO classification of tumours). World Health Organization; 2017.

19. Grygalewicz B, Woroniecka R, Rymkiewicz G, et al. The 11q-Gain/Loss aberration occurs recurrently in MYC-Negative burkitt-like lymphoma with 11q aberration, as well as MYC-Positive burkitt lymphoma and MYC-Positive High-Grade B-Cell lymphoma, NOS. *Am. J. Clin. Pathol.* 2017;149(1):17–28.

20. Rymkiewicz G, Grygalewicz B, Chechlinska M, et al. A comprehensive flow-cytometry-based immunophenotypic characterization of burkitt-like lymphoma with 11q aberration. *Mod. Pathol.* 2018;

21. Bolotin DA, Poslavsky S, Davydov AN, et al. Antigen receptor repertoire profiling from RNA-seq data. *Nat. Biotechnol.* 2017;35(10):908–911.

22. Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods*. 2015;12(5):380–381.

23. Lombardo KA, Coffey DG, Morales AJ, et al. High-throughput sequencing of the b-cell receptor in african burkitt lymphoma reveals clues to pathogenesis. *Blood Adv*. 2017;1(9):535–544.

24. Amato T, Abate F, Piccaluga P, et al. Clonality analysis of immunoglobulin gene rearrangement by Next-Generation sequencing in endemic burkitt lymphoma suggests antigen drive activation of BCR as opposed to sporadic burkitt lymphoma. *Am. J. Clin. Pathol.* 2016;145(1):116–127.

25. Baptista MJ, Calpe E, Fernandez E, et al. Analysis of the IGHV region in burkitt’s lymphomas supports a germinal center origin and a role for superantigens in lymphomagenesis. *Leuk. Res.* 2014;38(4):509–515.

26. Kostareli E, Gounari M, Janus A, et al. Antigen receptor stereotypy across b-cell lymphoproliferations: The case of IGHV4-59/IGKV3-20 receptors with rheumatoid factor activity. *Leukemia*. 2012;26(5):1127–1131.