**Supplementary Material:**

Sample preparation and WGS: DNA was extracted from peripheral blood samples by standard procedures that generate high molecular weight DNA suitable for genome sequencing (organic extraction using phenol-chloroform, or Qiagen columns). While a skin biopsy is the preferred method for germline testing, peripheral blood samples are a recognized alternative for people without active cancer and were utilized due to limitations in feasibility. WGS was performed at HudsonAlpha Sequencing Core (Huntsville, Alabama). Burrows-Wheeler Aligner (BWA)[1] and the GATK Haplotype Caller[2] were used for alignment and genotype calling. More specifically, paired-end reads were aligned to the GRCh38 human reference (GCA_000001405.15_GRCh38_no_alt_analysis_set.fa) using BWA (BWA-ALN v0.7.12) followed by the GATK3 best-practices workflow implemented in the Genome Analysis Toolkit (GATK v3.4.0)[3]. All samples passed sequencing quality control (QC) with average 30X coverage. All genomic data have been submitted to dbGAP (study accession: phs001738.v1.p1).

Pedigree-informed quality control: After genotype calling, sex and relatedness checks were performed on all samples to identify potential sample swaps using standard approaches[4]. Non-mendelian transmission from parents to offspring was confirmed using PEDCHECK[5]. Principal Components Analysis (PCA)[6] was performed to determine genetic ancestry using the 1000 Genomes Project[7] as a reference cohort (Figure S1). The resulting ancestry estimates guided which gnomAD reference cohorts were used for minor allele frequency (MAF) filtering. A MAF threshold of 1% was applied to the total gnomAD frequency and major subpopulations (AFR and NFE) and a threshold of 5% for all minor subpopulations (AMR, ASJ, EAS, FIN, SAS, and OTH).

Candidate variant review and ACMG classification

REVEL scores were utilized to evaluate missense variants in coding regions, where REVEL ≤ 0.290 met supporting criteria for benignity (BP4) and REVEL ≥ 0.644 met supporting criteria for deleteriousness (PP3). Non-coding variants resulting in an indel or with a RegBase score >10 were considered to meet supporting evidence for deleteriousness based on *in silico* data (PP3). Genes with a missense z-score > 3.09 were considered to have low rates of missense variation (PP2). Rarity in population databases was defined as a variant having ≤6 alleles in gnomAD v3.1.2 (PM2_supporting). Co-segregation (PP1) was assessed by counting meioses of affected individuals and obligate carriers in and across pedigrees. This evidence was considered as supporting for ≥1 informative meiosis, moderate for ≥3, and strong for ≥5.

References:


1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
2. Ren S, Ahmed N, Bertels K, Al-Ars Z. GPU accelerated sequence alignment with traceback for GATK HaplotypeCaller. *BMC Genomics.* 2019;20(Suppl 2):184.
3. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43:11 10 11-11 10 33.
4. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. *Am J Hum Genet.* 2016;98(1):127-148.
5. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet.* 1998;63(1):259-266.
6. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904-909.
7. Siva N. 1000 Genomes project. *Nat Biotechnol.* 2008;26(3):256.


## Supplementary Table Legends

**Table S1. Candidate genes used during variant prioritization.** List of genes included in Supplemental Figure 2 including the source of contribution.

**Table S2. Segregating SNV/INDELs.** (A) C1-C4 Coding variants. (B) NC1 and NC2 Noncoding variants.

**Table S3. Additional Detail for Prioritized germline HL candidates.** Prioritized variants for each family based on segregation and predicted deleteriousness. Variant coordinate information is provided in terms of the hg38 reference genome. The total gnomAD v3.0 minor allele frequency and ethnically relevant frequency for each variant are provided in addition to the SNP-based inferred ethnicity of the pedigree. Number of variants meeting C1-C4 and NC1-NC8 for each pedigree are also provided. * where evidence applied based on hypothesis of genotype-phenotype correlation.

**Table S4. Noncoding variants that segregate in multiple families.** Prioritized noncoding variants that were observed segregating with HL affection status for more than one family. Variant position, genotype, gnomAD v3.0 genome minor allele frequency, RegBase prediction scores, and interpretation of FIMO difference between reference and alternative allele are included. Regbase scores greater than or equal to 15 are highlighted grey.

**Table S5. Segregating CNVs.** (A) High ranking CNVs. (B) CNVs potentially impacting candidate genes.

## Supplemental Figure Legends

**Figure S1. Venn diagram of candidate genes included for prioritization.** Diagram depicts sources of various candidate genes included in our analysis.

**Figure S2**: **Quality control of dataset.** (A) Ethnicity estimation using PCA. The majority (207, 88.5%) of individuals were of European ancestry, followed by African (26, 11.1%), and South Asian ancestries (1, 0.4%) based on principal component analysis with the 1000 Genomes Project Phase 3 dataset. (B) Genetically determined relatedness among individuals included in the cohort.

**Figure S3. Non-coding variant frequency by priority.** Histogram of variant frequency per priority level (NC1-NC8) for each pedigree**.**

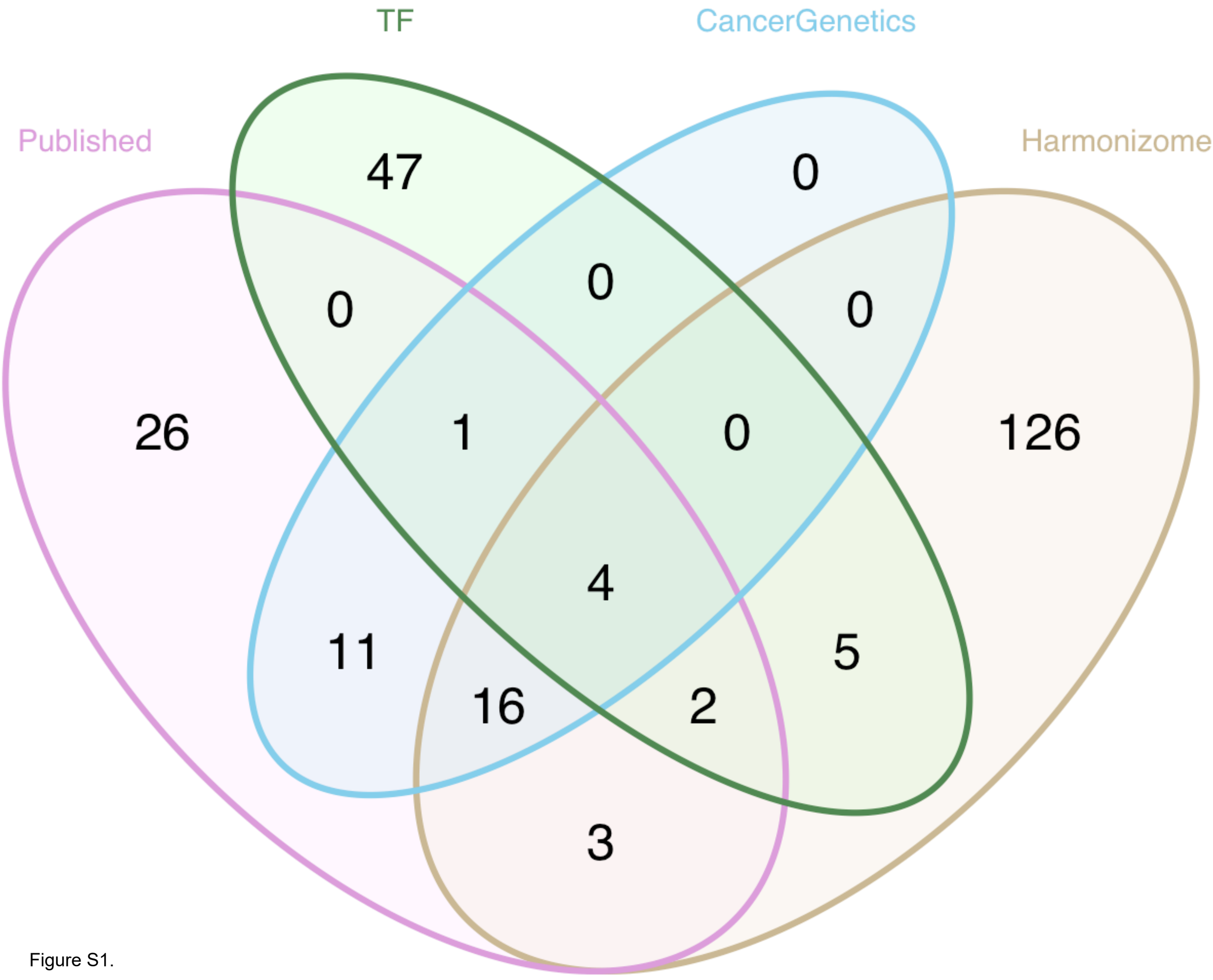
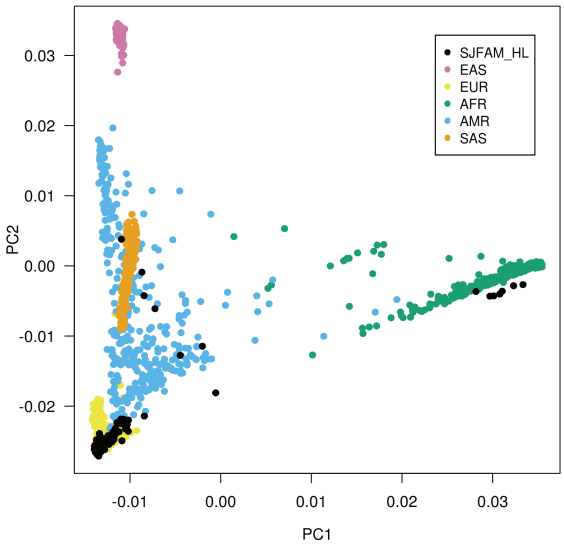

TF
CancerGenetics
Published
Harmonizome
47
0
0
0
0
26
1
0
126
11
4
5
16
2
3

Figure S1.

A


SJFAM_HL
EAS
EUR
AFR
AMR
SAS
PC2
PC1
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.01
0.00
0.01
0.02
0.03

B

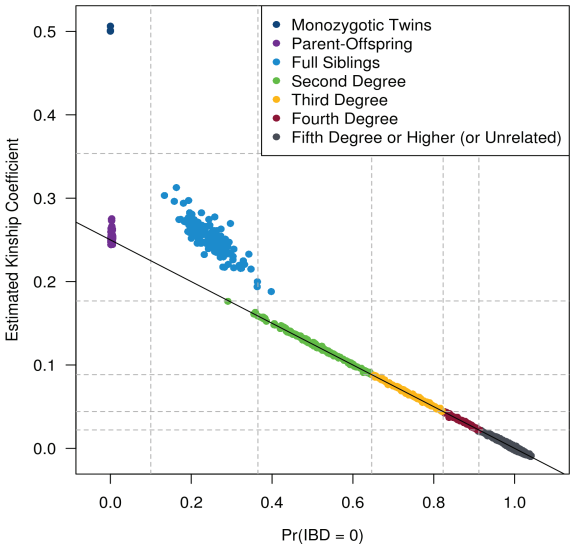

Monozygotic Twins
Parent-Offspring
Full Siblings
Second Degree
Third Degree
Fourth Degree
Fifth Degree or Higher (or Unrelated)
Estimated Kinship Coefficient
Pr(IBD = 0)
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Figure S2.

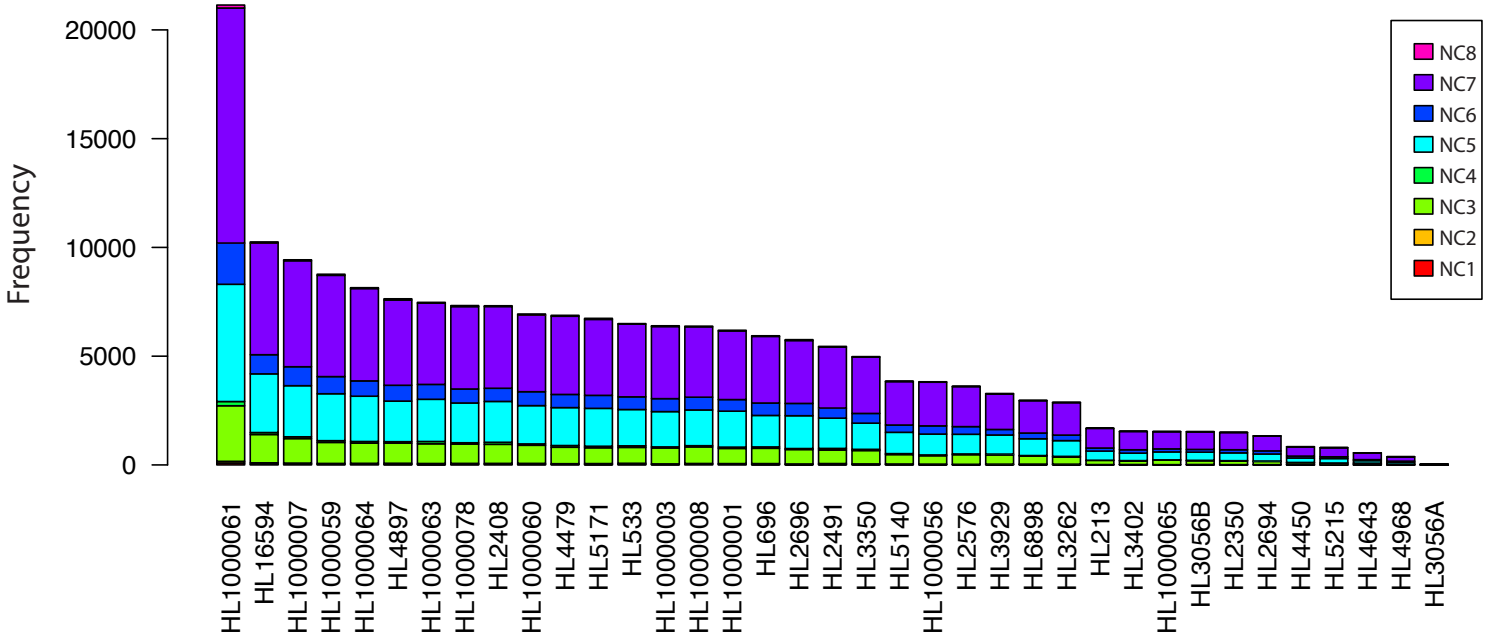

Frequency
20000
15000
10000
5000
0
NC8
NC7
NC6
NC5
NC4
NC3
NC2
NC1
HL1000061
HL16594
HL1000007
HL1000059
HL1000064
HL4897
HL1000063
HL1000078
HL2408
HL1000060
HL4479
HL5171
HL533
HL1000003
HL1000008
HL1000001
HL696
HL2696
HL2491
HL3350
HL5140
HL1000056
HL2576
HL3929
HL6898
HL3262
HL213
HL3402
HL1000065
HL3056B
HL2350
HL2694
HL4450
HL5215
HL4643
HL4968
HL3056A

Figure S3.

## Supplemental Table 2. Recurrent non-coding variants.

| Gene | Type | Location | Genotype | MAF** | NC priority | REG PHRED | CAN PHRED | PAT PHRED | FIMO |
|---|---|---|---|---|---|---|---|---|---|
| *DDB2* | intronic | chr11:47233910 | C/A | 0.0012 | 1 | 14.417 | 5.955 | 1.592 | GAIN - CRX |
| *GATA3* | intronic | chr10:8061430 | G/A | 0.0079 | 1 | 15.393 | 15.328 | 19.140 | LOSS - TCF12 & TAL1::TCF3 |
| *RUNX3* | intronic | chr1:24912674 | A/G | 0.0028 | 1 | 15.622 | 4.950 | 6.282 | COMPLEX - EBF1 gain, Esrra loss |
| *MC1R* | intergenic | chr16:89915484 | G/C | 0.0024 | 2 | 16.559 | 2.413 | 7.198 | COMPLEX - GLIS3 loss, ZBTB7B loss |
| *EZH2* | intronic | chr7:148863895 | G/A | 0.0036 | 1 | 15.084 | 6.122 | 3.693 | COMPLEX - GRHL1 loss, GRHL2 loss, SOX8 gain |
| *HLA-A* | intronic | chr6:29944961 | T/A | 9.33E-05 | 1 | 11.223 | 8.922 | 1.050 | COMPLEX - HSF2 loss, STAT1 gain, PAX1 gain |
| *PTPN1* | intergenic | chr20:50463747 | C/A | 0.0049 | 2 | 4.558 | 10.977 | 2.283 | COMPLEX - MGA loss |
| *KLHDC8B* | UTR5 | chr3:49171662* | C/T | 0.0048 | 1 | 27.085 | 2.526 | 21.838 | COMPLEX - NFIC gain, NFIX gain |
| *PAX5* | intronic | chr9:37025905 | T/C | 0.0011 | 1 | 18.930 | 9.022 | 1.457 | COMPLEX - NFYB loss |
| *IRF8* | intronic | chr16:85906807 | A/G | 0.0083 | 1 | 10.177 | 4.758 | 6.532 | COMPLEX - Pax2 gain |
| *RUNX3* | intronic | chr1:24960209 | G/A | 0.0032 | 1 | 11.402 | 2.559 | 4.655 | COMPLEX - PLAG1 gain, INSM1 gain |
| *PDGFRA* | intergenic | chr4:54226125 | G/A | 0.0049 | 1 | 18.510 | 7.552 | 4.690 | COMPLEX - PLAG1 loss, THAP1 gain |
| *CDKN2C* | UTR5 | chr1:50970305 | C/T | 6.00E-04 | 1 | 34.180 | 14.824 | 22.060 | COMPLEX - Pparg::Rxra gain, ZNF263 loss, IRF3 gain |
| *PAX5* | intronic | chr9:36969034 | G/A | 3.49E-05 | 1 | 10.300 | 4.690 | 10.977 | COMPLEX - XBP1 loss, PAX9 loss, Atf3 loss, Atf1 loss |
| *TBX21* | intergenic | chr17:47758718 | TC/T | 0.0036 | 1 | N/A | N/A | N/A | COMPLEX - YY1 gain |
| *LPXN* | intronic | chr11:58570075 | G/A | 0.0056 | 2 | 7.893 | 18.568 | 0.991 | COMPLEX - ZEB1 gain, SNAI2 gain, ID4 gain |
| *IL10* | downstream | chr1:206766846 | A/C | 0.0054 | 1 | 9.310 | 2.983 | 10.724 | COMPLEX - ZNF410 gain, RREB1 gain, ONECUT3 loss |
| *CDH1* | intronic | chr16:68765049 | C/T | 0.0032 | 1 | 18.049 | 13.980 | 1.276 | NOCHANGE |
| *CMIP* | intronic | chr16:81543016 | C/T | 0.0039 | 2 | 7.355 | 7.317 | 11.132 | NOCHANGE |
| *ERCC6* | intergenic | chr10:49574203 | T/C | 0.0051 | 2 | 10.127 | 6.376 | 1.714 | NOCHANGE |
| *KLHDC8B* | intronic | chr3:49174099 | G/A | 0.0054 | 2 | 17.523 | 5.651 | 6.577 | NOCHANGE |
| *LYN* | intronic | chr8:55918648 | CG/C | 0.0039 | 2 | N/A | N/A | N/A | NOCHANGE |
| *LYN* | intronic | chr8:55960828 | A/G | 0.0034 | 2 | 22.905 | 20.128 | 18.055 | NOCHANGE |
| *MAP3K7* | intergenic | chr6:90473200 | T/TTTTTA | 0.0063 | 1 | N/A | N/A | N/A | NOCHANGE |
| *MYB* | intronic | chr6:135202691 | A/G | 0.006 | 2 | 10.887 | 4.985 | 5.071 | NOCHANGE |
| *PAX5* | intronic | chr9:37005648 | T/C | 0.0011 | 1 | 12.491 | 5.920 | 3.280 | NOCHANGE |
| *RB1* | intronic | chr13:48318338 | C/T | 0.006 | 2 | 18.849 | 5.405 | 11.154 | NOCHANGE |
| *SDHC* | intronic | chr1:161352118 | ACT/A | 0.0054 | 2 | N/A | N/A | N/A | NOCHANGE |

*ClinVar designated pathogenic and linked to Hodgkin Lymphoma[21]

**Based on gnomAD v3.0 genome

Abbreviations: MAF=minor allele frequency; NC = non-coding; REG PHRED= phred-scaled RegBase functional score; CAN PRED= phred-scaled RegBase cancer driver score; PAT PHRED= phred-scaled RegBase pathogenic score; FIMO = *in silico* motif testing of REF vs ALT allele