Identification of Familial Hodgkin Lymphoma Predisposing Genes Using Whole Genome Sequencing

Aayushi Srivastava1,2,3,4,*, Sara Giangiobbe1,4,#, Abhishek Kumar1, Dagmara Dymerska5, Wolfgang Behnisch6, Mathias Witzens-Harig4, Jan Lubinski5, Kari Hemminki1,7, Asta Först1,2,3,4, Obul Reddy Bandapalli1,2,3,4*

1 Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), D69120 Heidelberg, Germany
2 Hopp Children’s Cancer Center (KiTZ), Heidelberg, Germany
3 Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany
4 Heidelberg University, Medical Faculty, Heidelberg, Germany
5 Pomeranian Medical University, Szczecin, Poland
6 Department of Pediatric Oncology, Hematology and Immunology, University of Heidelberg, Heidelberg, Germany.
7 Faculty of Medicine and Biomedical Center in Pilsen, Charles University in Prague, 30605 Pilsen, Czech Republic.

# These authors contributed equally

* Correspondence: Obul Reddy Bandapalli
o.bandapalli@kitz-heidelberg.de

Keywords: Familial Hodgkin lymphoma, Whole genome sequencing, Predisposing genes, Germline variants, Variant prioritization, Next generation sequencing, Genetic predisposition to disease.

Abstract

Hodgkin lymphoma (HL) is a lymphoproliferative malignancy of B-cell origin that accounts for 10% of all lymphomas. Despite evidence suggesting strong familial clustering of HL, there is no clear understanding of the contribution of genes predisposing to HL. In this study, whole genome sequencing (WGS) was performed on 7 affected and 9 unaffected family members from three HL-prone families and variants were prioritized using our Familial Cancer Variant Prioritization Pipeline (FCVPPv2). WGS identified a total of 98564, 170550 and 113654 variants which were reduced by pedigree-based filtering to 18158, 465 and 26465 in families I, II and III, respectively. In addition to variants affecting amino acid sequences, variants in promoters, enhancers, transcription factor binding sites and microRNA seed sequences were identified from upstream, downstream, 5’ and 3’ untranslated regions. A panel of 565 cancer predisposing and other cancer-related genes and of 2383...
high-risk HL genes were also screened in these families to aid further prioritization. Pathway analysis of segregating genes with CADD (Combined Annotation Dependent Depletion Tool) scores > 20 was performed using Ingenuity Pathway Analysis software which implicated several candidate genes in pathways involved in B-cell activation and proliferation and in the network of “Cancer, Hematological disease and Immunological Disease”. We used the FCVPPv2 for further in silico analysis and prioritized 45 coding and 79 non-coding variants from the three families. Further literature-based analysis allowed us to constrict this list to one rare germline variant each in families I and II and two in family III. Functional studies were conducted on the candidate from family I in a previous study, resulting in the identification and functional validation of a novel heterozygous missense variant in the tumor suppressor gene Dicer1 as potential HL predisposition factor. We aim to identify the individual genes responsible for predisposition in the remaining two families and will functionally validate these in further studies.

1 Introduction

Hodgkin lymphoma (HL) is a lymphoproliferative malignancy originated in germinal center B-cells and is reported to account for about 10% of newly diagnosed lymphomas and 1% of all de novo neoplasms worldwide with an incidence of about 3 cases per 100,000 people in Western countries (Diehl et al., 2004). It is one of the most common tumors in young adults in economically developed countries, with one peak of incidence in the third decade of life and a second peak after 50 years of age.

Based on differences in the morphology and phenotype of the lymphoma cells and the composition of the cellular infiltrate, HL is subdivided into classical Hodgkin lymphoma (cHL) that accounts for about 95% of cases and nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL) that accounts for the remaining 5% of cases (Kuppers, 2009).

Although familial risk for HL is reported to be among the highest of all cancers (Kharazmi et al., 2015), not many genetic risk factors have been identified. An association between various HLA class I and class II alleles and increased risk of HL has been reported (Diepstra et al., 2005), while other non-HLA susceptibility loci have been detected through genome-wide association studies (Frampton et al., 2013; Cozen et al., 2014; Kushekhar et al., 2014). The identification of major predisposing genes is a more daunting task; however, rare germline variants in KLDHC8B, NPAT, ACAN, KDR, DICER1 and POT1 gene have been reported by different groups in high-risk HL families (Salipante et al., 2009; Saarinen et al., 2011; Ristolainen et al., 2015; Rotunno et al., 2016; Bandapalli et al., 2018; McMaster et al., 2018).

Here we report the results of whole genome sequencing (WGS) performed in three families with documented recurrence of HL. We used our Familial Cancer Variant Prioritization Pipeline (FCVPPv2) (Kumar et al., 2018) as well as two gene/variant panels based on cancer predisposing genes and variants prioritized in the largest familial HL cohort study to date in order to identify possible disease-causing high-penetrance germline variants in each family (Zhang et al., 2015; Rotunno et al., 2016). Pathway and network analyses using Ingenuity Pathway Analysis software also allowed us to gain insight into the molecular mechanisms of the pathogenesis of HL. We hope that these results can be used in the development of targeted therapy and in the screening of other individuals at risk of developing HL.

2 Materials and Methods
2.1 Patient samples

Three families with documented recurrence of HL were analyzed in this study, with a total number of 16 individuals (7 affected and 9 unaffected). HL family I and family III were recruited at the University Hospital of Heidelberg, Germany, while family II was recruited at the Pomeranian Medical University, Szczecin, Poland.

The study was approved by the Ethics Committee of the University of Heidelberg and Pomeranian Medical University, Poland. Collection of blood samples and clinical information from subjects was undertaken with a written informed consent in accordance with the tenets of the Declaration of Helsinki.

Germline DNA samples used for genome sequencing were isolated from peripheral blood using QIAamp® DNA Mini kit (Qiagen, Cat No. 51104) according to the manufacturer’s instructions.

2.2 Whole genome sequencing, variant calling, annotation and filtering

Whole genome sequencing (WGS) of available affected and unaffected members of the three HL families was performed using Illumina-based small read sequencing. Mapping to reference human genome (assembly version Hs37d5) was performed using BWA mem (version 0.7.8) (Li and Durbin, 2009) and duplicates were removed using biobambam (version 0.0.148). The SAMtools suite (Li, 2011) was used to detect single nucleotide variants (SNVs) and Platypus (Rimmer et al., 2014) to detect indels. Variants were annotated using ANNOVAR, 1000 Genomes, dbSNP and ExAC (Smigielski et al., 2000; Wang et al., 2010; The Genomes Project et al., 2015; Lek et al., 2016).

Variants with a quality score greater than 20 and a coverage greater than 5x, SNVs that passed the strand bias filter (a minimum one read support from both forward and reverse strand) and indels that passed all the Platypus internal filters were evaluated further for minor allele frequencies (MAFs) with respect to the 1000 Genomes Phase 3 and non-TCGA ExAC data. Variants with a MAF less than 0.1% were deduced from these two datasets. A pairwise comparison of shared rare variants was performed to check for sample swaps and family relatedness.

2.3 Data analysis and variant prioritization

2.3.1 Prioritization of coding variants

Variant evaluation was performed using the criteria of our in-house developed variant prioritization pipeline (FCVPPv2) (Kumar et al., 2018). First, variants with MAF < 0.1% were filtered based on the pedigree data considering cancer patients as cases and unaffected persons as controls. The probability of an individual being a Mendelian case or true control was considered.

Variants were then ranked using the combined annotation dependent depletion (CADD) tool v1.3 (Kircher et al., 2014). Only variants with a scaled PHRED-like Combined Annotation Dependent Depletion (CADD) score greater than 10, i.e. variants belonging to the top 1% of probable deleterious variants in the human genome, were considered further. Genomic Evolutionary Rate Profiling (GERP) (Cooper et al., 2005), PhastCons (Siepel et al., 2005) and PhyloP (Pollard et al., 2010) were used to evaluate the evolutionary conservation of a particular variant. GERP scores > 2.0, PhastCons scores > 0.3 and PhyloP scores ≥ 3.0 were indicative of a good level of conservation and were therefore used as thresholds in the selection of potentially causative variants.

Next, all missense variants were assessed for deleteriousness using 10 tools accessed using dbNSFP (Liu et al., 2016), namely SIFT, PolyPhen V2-HDV, PolyPhen V2-HVAR, LRT, MutationTaster, Mutation Assessor, FATHMM, MetaSVM, MetLR and PROVEAN. Variants predicted to be
Germline Variants in Hodgkin Lymphoma

deleterious by at least 60% of these tools were analyzed further. Prediction scores for nonsense variants were attained via VarSome (Kopanos et al., 2018), the final verdict on pathogenicity offered by VarSome was based on the following tools: DANN, MutationTaster, FATHMM-MKL, FATHMM-XF, ALoFT, EIGEN, EIGEN PC and PrimateAI.

Lastly, three different intolerance scores derived from NHLBI-ESP6500 (Petrovski et al., 2013), ExAC (Lek et al., 2016) and a local dataset, all of which were developed with allele frequency data, were included to evaluate the intolerance of genes to functional mutations. However, these scores were used merely to rank the variants and not as cut-offs for selection. The ExAC consortium has developed two additional scoring systems using large-scale exome sequencing data including intolerance scores (pLI) for loss-of-function variants and Z-scores for missense and synonymous variants. These were used for nonsense and missense variants, respectively.

2.3.2 Analysis of non-coding variants

Variants located in the 3’ and 5’ untranslated regions (UTRs) were prioritized based on their location in regulatory regions. Putative miRNA targets at variant positions within 3’ UTRs and 1 kb downstream of transcription end sites were detected by scanning the entire dataset of the human miRNA target atlas from TargetScan 7.0 (Agarwal et al., 2015) using the intersect function of bedtools. Similarly, 5’UTRs and regions 1 kb upstream of transcription start sites were scanned for putative enhancers and promoters using merged enhancer and promoter data from the FANTOM5 consortium as well as super-enhancer data from the super-enhancer archive (SEA) and dbSUPER. These regions were also scanned for transcription factor binding sites using SNPnexus (Dayem Ullah et al., 2018).

The regulatory nature and the possible functional effects of non-coding variants were evaluated using CADD v1.3, HaploReg V4 (Ward and Kellis, 2012) and RegulomeDB (Boyle et al., 2012), primarily based on ENCODE data (Birney et al., 2007). Epigenomic data and marks from 127 cell lines from the NIH Roadmap Epigenomics Mapping Consortium were accessed via CADD v1.3, which gave us information on chromatin states from ChromHmm and Segway. CADD also provided mirSVR scores to rank predicted microRNA target sites by a down-regulation score. These scores are based on a new machine learning method based on sequence and contextual features extracted from miRanda-predicted target sites (Betel et al., 2010). Furthermore, SNPnexus was used to access non-coding scores for each variant and to identify regulatory variants located in CpG islands.

The final selection of 3’UTR and downstream variants was based on their CADD scores > 10 and whether or not they had predicted miRNA target site matches. Similarly, upstream and 5’ UTR variants in enhancers, promoters, super-enhancers or transcription factor binding sites with CADD scores > 10 were short-listed.

2.3.3 Presence of candidate variants in 565 cancer predisposing and other cancer-related genes

In a study on cancer predisposing genes (CPGs) in pediatric cancers, Zhang et al. compiled 565 CPGs based on review of the American College of Medical Genetics and Genomics (ACMG) and medical literature (Zhang et al., 2015). The categories included genes associated with autosomal dominant cancer-predisposition syndromes (60), genes associated with autosomal recessive cancer-predisposition syndromes (29), tumor-suppressor genes (58), tyrosine kinase genes (23) and other cancer genes (395). We investigated a list of genes corresponding to our shortlisted coding and non-coding variants for their presence in the list of genes in the aforementioned study.
2.3.4 Presence of candidate variants in prioritized HL genes from a large WES-based familial HL study

In a study by Rotunno et al., 2699 variants corresponding to 2383 genes were identified in 17 HL discovery families after filtering and prioritization (Rotunno et al., 2016). We intersected our list of candidate genes with this list of 2383 HL genes to identify coding and non-coding variants from our shortlist in potentially causative HL genes.

2.4 Variant validation

Specific variants of interest mentioned throughout the text were validated using specific primers for polymerase chain reaction amplification designed with Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) and Sanger sequencing on a 3500 Dx Genetic Analyzer (Life Technologies, CA, USA), using ABI PRISM 3.1 Big Dye terminator chemistry, according to the manufacturer’s instructions. The electrophoretic profiles were analyzed manually. Segregation analysis of the prioritized variants was performed in additional family members when DNA was available. Primer details are available on request.

2.5 Ingenuity pathway analysis (IPA)

IPA (Qiagen; http://www.qiagen.com/ingenuity; analysis date 15/10/2019) was used to perform a core analysis to identify enriched canonical pathways, diseases, biological functions and molecular networks among genes that passed the allele frequency cut-off, fulfilled family-based segregation criteria, met the CADD score cut-off and were not intergenic or intronic variants. Data were analyzed for all three families together. Top canonical pathways were identified from the IPA pathway library and ranked according to their significance to our input data. This significance was determined by p-values calculated using the right tailed Fisher’s exact test.

IPA was also used to generate gene networks in which upstream regulators were connected to the input dataset genes while taking advantage of paths that involved more than one link (i.e., through intermediate regulators). These connections represent experimentally observed cause-effect relationships that relate to expression, transcription, activation, molecular modification and transport as well as binding events. The networks were ranked according to scores that were generated by considering the number of focus genes (input data) and the size of the network to approximate the relevance of the network to the original list of focus genes.

3 Results

3.1 Whole genome sequencing results

In our study, we analyzed three families with reported recurrence of Hodgkin lymphoma. Their respective pedigrees are shown in Figure 1.
Germline Variants in Hodgkin Lymphoma

Figure 1. Pedigrees of the three HL families analyzed in this study. A: Family 1, B: Family 2, C: Family 3.

In family I (Figure 1A), the proband (III-1) and her mother (II-2) were diagnosed with two different histological subtypes of classical Hodgkin lymphoma (cHL) at the ages of 7 and 34, respectively. The daughter was diagnosed with nodular sclerosis cHL and the mother with lymphocyte-rich cHL. The sample of the unaffected father (II-1) was also sequenced. Family II (Figure 1B) is characterized by a strong recurrence of HL. Five family members were diagnosed with HL (II-3, II-4, III-3, III-4 and III-5), of which three (III-3, III-4 and III-5) underwent WGS. In addition, the family member (II-6), who was considered as an obligatory carrier of the mutation, was sequenced as were samples and four healthy family members (III-1, III-2, III-6 and III-8) and one family member diagnosed with uterine cancer (II-1) as controls. In family 3 (Figure 1C), II-1 and II-2 were diagnosed with cHL, at the age of 27 and 24 respectively. Their parents (I-1, I-2) were not affected, however one of them is expected to be a carrier and analyzed accordingly.

WGS of 7 affected and 9 unaffected members from the three studied families identified a total number of 98564, 170550 and 113654 variants which were reduced by pedigree-based filtering to 18158, 465 and 26465 in families I, II and III, respectively.

3.2 Prioritization of candidates according to the FCVPPv2

After pedigree-based filtering, 130, 7 and 196 exonic variants were left in families I, II and III, respectively, with a prevalence of non-synonymous and synonymous SNVs. The predominant type of substitution was the C>T transition. Among exonic variants fulfilling pedigree-based criteria, only
variants with CADD scores > 10 were taken into further consideration and prioritized according to deleteriousness, intolerance and conservational scores, as detailed in the methods section. At the end of this process, 37 potential missense variants and 9 potential nonsense mutations were prioritized for families I - III and are shown in Tables 1 and 2.

Pedigree-based filtering also reduced the number of potentially interesting variants located in the untranslated regions to 523 for 5’UTR variants (130 in family I, 5 in family II and 314 in family III) and 854 for 3’UTR variants (347 in family I, 10 in family II and 497 in family III). These variants were further prioritized based on their CADD score > 10 and their localization in known regulatory regions (Supplementary Table 1). 5’UTR variants were analyzed by the SNP Nexus tool, which allowed us to identify 4 variants located in transcription factors binding sites. In addition, the intersect function of bedtools was used to identify further 15 variants located in promoter regions and 4 located in super-enhancer regions. Among variants located in the 3’UTR region, 56 variants located in miRNA seed sequences were selected.

3.3 Candidate variants in 565 CPGs and 2383 potentially causative HL genes

Intersecting our prioritized list of candidate genes with the list of 565 CPGs identified 11 genes with variants in coding and selected non-coding regions (upstream and downstream variants, 3’ and 5’ UTRs) of the known CPGs. These include FUBP1, SEPT6, DICER1, EZR and NCOA1 from family 1 and BCL6, RAD51D, LPP and PTCH1 from family 3 (Table 3). DICER1 and PTCH1 are known in autosomal dominant cancer-predisposition syndromes, whereas the rest are categorized as being “other cancer genes”.

In addition to the identification of 11 variants in CPGs, we intersected our prioritized list of genes with a list of 2383 genes with potentially causative variants from a large WES-based familial HL study. We found 25 variants in the coding and non-coding regions in 23 of the HL genes, with 7 coming from family I and 18 from family III (Table 4).

3.4 Network and pathway analysis with IPA

Pathway analysis of the selected variants performed with IPA showed an enrichment of mutations in genes involved in pathways essential for B-cell proliferation and activation, specifically B-cell receptor signaling, and PI3K signaling in B lymphocytes and B cell activating factor signaling (Supplementary Table 2A, Supplementary Figures 1A and 1B).

Similarly, the IPA network analysis generated a comprehensive picture of possible gene interactions between our candidate genes (Supplementary Table 2B). The top network is related to cancer, hematological disease and immunological disease, which is in complete coherence with the pathogenesis of HL. Many genes from the prioritized list of top candidates are shown to play a role in the top networks (Figure 2).
Figure 2. The top three molecular networks identified by Ingenuity Pathway Analysis: (a) Network 1. Cancer, Hematological Disease, Immunological Disease; (b) Network 2. Developmental Disorder, Endocrine System Disorders, Hereditary Disorder; (c) Network 3. RNA Post-Transcriptional Modification, Cell Death and Survival, Cellular Movement. Genes from our input-data are shown in grey, genes from our prioritized candidate list are highlighted in peach.

3.5 Literature mining, consolidation of results and selection of candidates

With the aim of identifying high penetrance dominant variant per family, we used our pipeline results and literature-based mining to determine the genes’ link to Hodgkin lymphoma or immune-related processes. For family 1, we have short-listed 5 potential candidates (DICER1, HLTF, NOTCH3, PLK3 and RELB). Based on segregation, confirmation and functional validation, we identified DICER1 as a candidate HL predisposing gene by showing significant down-regulation of tumor suppressor miRNAs in DICER1-mutated family members (Bandapalli et al., 2018). The presence of DICER1 in the list of 565 CPGs also reinforces its status as the disease-causing variant in this family.

In family 2, three exonic variants made it to the final list (ALAD, CERCAM and SPTAN1) of which SPTAN1 was shown to be among the genes in one of the top IPA networks (Network 3; Figure 2C). No coding or non-coding variants intersected with the panel of CPGs or HL candidate genes.
Two genes stand out in family 3, namely \textit{LPP} and \textit{RAD51D}. Both genes were found in the list of 565 CPGs and \textit{LPP} was additionally found in the gene list from the large cohort of HL families. Three variants in \textit{LPP} were prioritized by the FCVPPv2 and made it to the shortlist including one stopgain variant (3\_188123978\_G\_T), one 3'UTR (3\_188608373\_A\_T) and one nonsynonymous missense variant (3\_188123979\_A\_T). \textit{LPP} (LIM domain containing preferred translocation partner in lipoma) is a member of the zyxin family of LIM proteins that is characterized as a promoter of mesenchymal/fibroblast cell migration. \textit{LPP} has been shown to be a critical inducer of tumor cell migration, invasion and metastasis by virtue of its ability to localize to adhesions and to promote invadopodia formation (Ngan et al., 2018). A genome-wide association study of 253 Chinese individuals with B-cell NHL also identified a new susceptibility locus between \textit{BCL6} and \textit{LPP} that was significantly associated with the increased risk of B-cell NHL (Tan et al., 2013). On the other hand, there are no reports of an association between \textit{RAD51D} and lymphomas; however, it is a well-established susceptibility gene in Breast-Ovarian Cancer, Familial 4 and Hereditary Breast Ovarian Cancer Syndrome (Loveday et al., 2011; Chen et al., 2018). The final selection of a candidate in this family will be based on further functional studies.

### 4 Discussion

In summary, WGS data analysis of three families with reported recurrence of HL allowed us to prioritize 45 coding and 79 non-coding variants from which we subsequently selected and validated one for family I, \textit{(DICER1)}, short-listed three and two for families II (ALAD, CERCAM and SPTAN1) and III (\textit{RAD51} and \textit{LPP}) respectively to investigate further with validation and functional studies. For family I we have already functionally validated \textit{DICER1} as the candidate predisposing gene in a previous study (Bandapalli et al., 2018). However, the family was included in this paper, especially with regard to the pathway and network analyses. We identified pathways related to B-cell proliferation and networks related to cancer, hematological disease, immunological disease, hereditary disorders, cell death and cell survival using IPA software, helping us to prioritize genes with functions in the pathogenesis of HL.

Personalized medicine is an upcoming and promising field of medicine in which medical decisions, practices, interventions and products are tailored to the individual patient based on their predicted response or risk of disease. The scope of this field has advanced rapidly with the advent of genomics and other omics and the possibility of implicating one gene or a set of genes in the pathogenesis of a particular disease. Thus, the identification of germline predisposing genes could be of great value in the screening of individuals at risk of developing HL, as well as in the development of personalized adjuvant therapies based on the affected pathways. In this aspect, delta-aminolevulinate dehydratase (\textit{ALAD}) from family 2 is interesting as it is involved in the catalysis of second step in the biosynthesis of heme and is also acts as an endogenous inhibitor of the 26S proteasome, a multi-catalytic ATP-dependent protease complex that functions as the degrading arm of the ubiquitin system: the major pathway for regulated degradation of proteins in all eukaryotes. Down regulation of \textit{ALAD} is shown to be associated with poor prognosis in patients with breast cancer (Ge et al., 2017) whereas the existing data on nonerythroid spectrin a1 (\textit{SPTAN1}) suggest that overexpression of \textit{SPTAN1} in tumor cells reflects neoplastic and tumor promoting activity or tumor suppressing effects by enabling DNA repair through interaction with DNA repair proteins (Ackermann and Brieger, 2019) and \textit{CERCAM} is known as an unfavourable prognostic marker in urothelial, renal and ovarian cancers implying the importance of the variants in these genes (Ma et al., 2016). \textit{RAD51D} from family III is particularly interesting since \textit{RAD51D} is involved in DNA repair through homologous recombination. Therefore, it is possible that carcinomas arising in patients carrying mutations in this gene will be sensitive to chemotherapeutic agents that target this pathway, such as cisplatin and the
PARP (poly (ADP-ribose) polymerase) inhibitor olaparib. This has already been demonstrated in BRCA1/2 mutation-carrier cancer patients (Banerjee et al., 2010; Loveday et al., 2011). This approach can also be applied to target pathways affected by the mutated genes. Several candidate genes were identified by IPA pathway analysis in B cell receptor pathways, offering a valuable target for other pharmaceutical drugs. The B cell receptor (BCR) signaling pathway, when dysregulated, is a potent contributor to lymphomagenesis and tumor survival (Valla et al., 2018). This pathway has been targeted in B-cell lymphomas and leukemias with several BCR-directed agents, such as inhibitors of Bruton’s tyrosine kinase (BTK), spleen tyrosine kinase (SYK) and phosphatidylinositol-3-kinase (PI3K) (Buggy and Elias, 2012; Dreyling et al., 2017; Liu and Mamorska-Dyga, 2017). In one study, excellent response rates could be demonstrated in certain non-Hodgkin lymphoma subtypes, however, issues related to the development of resistance to BTK inhibitors need to be addressed (Valla et al., 2018).

Advancements in the field of genomics have allowed WGS to become the state-of-the-art tool for the identification of novel cancer predisposing genes in Mendelian diseases. It is still a challenge to appropriately interpret the immense amount of data generated by WGS, especially with respect to non-coding variants. In our study, we have attempted to interpret a selection of non-coding variants using in-silico and bioinformatic tools, however, the adequate analysis of intronic and intergenic variants remains a challenge. There are several reports of WGS being successfully implemented to implicate rare, high-penetrance germline variants in cancer, for example POT1 mutations in familial melanoma and Hodgkin lymphoma (McMaster et al., 2018; Wong et al., 2019) and POLE and POLD1 mutations in colorectal adenomas or carcinomas (Palles et al., 2013). In a previous study, we have used our pipeline (FCVPPv2) to prioritize novel variants in non-medullary thyroid cancer prone families (Srivastava et al., 2019). We have also successfully combined our pipeline with literature review and functional studies to identify DICER1 as a candidate predisposing gene in one Hodgkin lymphoma family (Bandapalli et al., 2018). We aim to apply these methods in the remaining Hodgkin lymphoma families and hope that these results will facilitate personalized therapy in the studied families and contribute to the screening of other individuals at risk of developing HL.

5 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

O.R.B., A.F. and K.H. conceived and designed the study. W.B., M.W.-H., D.D. and J.L. provided the HL family samples. A.S., O.R.B., S.G. and A.K. analyzed the data. O.R.B. and S.G. performed the experiments. A.S. and O.R.B. wrote the first draft of the manuscript. All authors read, commented on and approved the manuscript.

7 Funding

The study was supported by the Harald Huppert Foundation and Transcan ERA-NET funding from the German Federal Ministry of Education and Research (BMBF).
8 Acknowledgments

The authors thank the Genomics and Proteomics Core Facility (GPCF) of the German Cancer Research Center (DKFZ) for providing excellent library preparation and sequencing services, the Omics IT and Data Management Core Facility (ODCF) of the DKFZ for the whole genome sequencing data management and Nagarajan Paramasivam and Matthias Schlesner for their help with analysis of the sequencing data.

9 Supplementary Material

The Supplementary Material for this article can be found online at:

10 Data Availability Statement

Unfortunately, we are not able to provide the sequencing data into a public data base. The data underlying the results presented in the study are available from the corresponding author or from Dr. Asta Försti (Email: a.foersti@kitz-heidelberg.de).

11 References

Ackermann, A., and Brieger, A. (2019). The Role of Nonerythroid Spectrin αII in Cancer. Journal of oncology 2019, 7079604-7079604.

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife 4.

Bandapalli, O.R., Paramasivam, N., Giangiobbe, S., Kumar, A., Benisch, W., Engert, A., Witzens-Harig, M., Schlesner, M., Hemminki, K., and Forsti, A. (2018). Whole genome sequencing reveals DICER1 as a candidate predisposing gene in familial Hodgkin lymphoma. Int J Cancer 143, 2076-2078.

Banerjee, S., Kaye, S.B., and Ashworth, A. (2010). Making the best of PARP inhibitors in ovarian cancer. Nat Rev Clin Oncol 7, 508-519.

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 11, R90.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Stamatoyannopoulos, J.A., Thurman, R.E., Kuehn, M.S., Taylor, C.M., Nephe, S., Koch, C.M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J.A., Andrews, R.M., Flicek, P., Boyle, P.J., Cao, H., Carter, N.P., Clelland, G.K., Davis, S., Day, N., Dhahi, P., Dillon, S.C., Dorschner, M.O., Fiegler, H., Giresi, P.G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K.D., Johnson, B.E., Johnson, E.M., Frum, T.T., Rosenzweig, E.R., Karnani, N., Lee, K., Lefebvre, G.C., Navas, P.A., Neri, F., Parker, S.C.J., Sabo, P.J., Sandstrom, R., Shafier, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F.S., Dekker, J., Lieb, J.D., Tullius, T.D., Crawford, G.E., Sunyaev, S., Noble, W.S., Dunham, I., Dutta, A., Guigó, R., Denoëud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I.L., Baertsch, R., Keefe, D., Flicek, P., Dike, S., Cheng, J., Hirsch, H.A., Sekinger, E.A., Lagarde,
Germline Variants in Hodgkin Lymphoma

J., Abril, J.F., Shahab, A., Flamm, C., Fried, C., Hackermüller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J.S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816.

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., Cherry, J.M., and Snyder, M. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22, 1790-1797.

Buggy, J.J., and Elias, L. (2012). Bruton tyrosine kinase (BTK) and its role in B-cell malignancy. Int Rev Immunol 31, 119-132.

Chen, X., Li, Y., Ouyang, T., Li, J., Wang, T., Fan, Z., Fan, T., Lin, B., and Xie, Y. (2018). Associations between RAD51D germline mutations and breast cancer risk and survival in BRCA1/2-negative breast cancers. Ann Oncol 29, 2046-2051.

Cooper, G.M., Stone, E.A., Asimenos, G., Program, N.C.S., Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 15, 901-913.

Cozen, W., Timofeeva, M.N., Li, D., Diepstra, A., Hazelett, D., Delahaye-Sourdeix, M., Edlund, C.K., Franke, L., Rostgaard, K., Van Den Berg, D.J., Cortessis, V.K., Smedby, K.E., Glaser, S.L., Westra, H.J., Robison, L.L., Mack, T.M., Ghersiquieres, H., Hwang, A.E., Nieters, A., De Sanjose, S., Lightfoot, T., Becker, N., Maynadie, M., Foretova, L., Roman, E., Benavente, Y., Rand, K.A., Nathwani, B.N., Glimelius, B., Staines, A., Boffetta, P., Link, B.K., Kiemeney, L., Ansell, S.M., Bhatia, S., Strong, L.C., Galan, P., Vatten, L., Habermann, T.M., Duell, E.J., Lake, A., Veenstra, R.N., Visser, L., Liu, Y., Urayama, K.Y., Montgomery, D., Gaborieau, V., Weiss, L.M., Byrnes, G., Lathrop, M., Cocco, P., Best, T., Skol, A.D., Adami, H.O., Melbye, M., Cerhan, J.R., Gallagher, A., Taylor, G.M., Slager, S.L., Brennan, P., Coetze, G.A., Conti, D.V., Onel, K., Jarrett, R.F., Hjalgrim, H., Van Den Berg, A., and Mckay, J.D. (2014). A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. Nat Commun 5, 3856.

Dayem Ullah, A.Z., Oscanoa, J., Wang, J., Nagano, A., Lemoine, N.R., and Chelala, C. (2018). SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic acids research 46, W109-W113.

Diehl, V., Thomas, R.K., and Re, D. (2004). Part II: Hodgkin's lymphoma--diagnosis and treatment. Lancet Oncol 5, 19-26.

Diepstra, A., Niens, M., Vellenga, E., Van Imhoff, G.W., Nolte, I.M., Schaapveld, M., Van Der Steege, G., Van Den Berg, A., Kibbelaar, R.E., Te Meerman, G.J., and Poppema, S. (2005). Association with HLA class I in Epstein-Barr-virus-positive and with HLA class III in Epstein-Barr-virus-negative Hodgkin's lymphoma. Lancet 365, 2216-2224.

Dreyling, M., Santoro, A., Mollica, L., Leppa, S., Follows, G.A., Lenz, G., Kim, W.S., Nagler, A., Panayiotidis, P., Demeter, J., Ozcan, M., Kosinova, M., Bouabdallah, K., Morschhauser, F., Stevens, D.A., Trevarthen, D., Giurescu, M., Cupit, L., Liu, L., Kochert, K., Seidel, H., Pena, C., Yin, S., Hiemeyer, F., Garcia-Vargas, J., Childs, B.H., and Zinzani, P.L. (2017). Phosphatidylinositol 3-Kinase Inhibition by Copanlisib in Relapsed or Refractory Indolent Lymphoma. J Clin Oncol 35, 3898-3905.
Germline Variants in Hodgkin Lymphoma

Frampton, M., Da Silva Filho, M.I., Broderick, P., Thomesen, H., Forsti, A., Vijayakrishnan, J., Cooke, R., Enciso-Mora, V., Hoffmann, P., Nothen, M.M., Lloyd, A., Holroyd, A., Eisele, L., Jockel, K.H., Ponader, S., Von Strandmann, E.P., Lightfoot, T., Roman, E., Lake, A., Montgomery, D., Jarrett, R.F., Svendsen, A., and Houlston, R.S. (2013). Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. Nat Commun 4, 2549.

Ge, J., Yu, Y., Xin, F., Yang, Z.J., Zhao, H.M., Wang, X., Tong, Z.S., and Cao, X.C. (2017). Downregulation of delta-aminolevulinate dehydratase is associated with poor prognosis in patients with breast cancer. Cancer Sci 108, 604-611.

Kharazmi, E., Fallah, M., Pukkala, E., Olsen, J.H., Tryggvadottir, L., Sundquist, K., Tretli, S., and Hemminki, K. (2015). Risk of familial classical Hodgkin lymphoma by relationship, histology, age, and sex: a joint study from five Nordic countries. Blood 126, 1990-1995.

Kircher, M., Witten, D.M., Jain, P., O'roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46, 310-315.

Kopanos, C., Tsilokas, V., Kouris, A., Chapple, C.E., Albarca Aguilera, M., Meyer, R., and Massouras, A. (2018). VarSome: the human genomic variant search engine. Bioinformatics 35, 1978-1980.

Kumar, A., Bandapalli, O.R., Paramasivam, N., Giangiobbe, S., Diquigiovanni, C., Bonora, E., Eils, R., Schlesner, M., Hemminki, K., and Forsti, A. (2018). Familial Cancer Variant Prioritization Pipeline version 2 (FCVPPv2) applied to a papillary thyroid cancer family. Sci Rep 8, 11635.

Kuppers, R. (2009). The biology of Hodgkin's lymphoma. Nat Rev Cancer 9, 15-27.

Kushekhar, K., Van Den Berg, A., Nolte, I., Hekkema, B., Visser, L., and Diepstra, A. (2014). Genetic Associations in Classical Hodgkin Lymphoma: A Systematic Review and Insights into Susceptibility Mechanisms. Cancer Epidemiology Biomarkers & Prevention 23, 2789.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987-2993.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.

Liu, D., and Mamorska-Dyga, A. (2017). Syk inhibitors in clinical development for hematological malignancies. *J Hematol Oncol* 10, 145.

Liu, X., Wu, C., Li, C., and Boerwinkle, E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* 37, 235-241.

Loveday, C., Turnbull, C., Ramsay, E., Hughes, D., Ruark, E., Frankum, J.R., Bowden, G., Kalmyrzaev, B., Warren-Perry, M., Snape, K., Adlard, J.W., Barwell, J., Berg, J., Brady, A.F., Brewer, C., Brice, G., Chapman, C., Cook, J., Davidson, R., Donaldson, A., Douglas, F., Greenhalgh, L., Henderson, A., Izatt, L., Kumar, A., Laloo, F., Miedzybrodzka, Z., Morrison, P.J., Paterson, J., Porteous, M., Rogers, M.T., Shanley, S., Walker, L., Eccles, D., Evans, D.G., Renwick, A., Seal, S., Lord, C.J., Ashworth, A., Reis-Filho, J.S., Antoniou, A.C., and Rahman, N. (2011). Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat Genet* 43, 879-882.

Ma, L.J., Wu, W.J., Wang, Y.H., Wu, T.F., Liang, P.I., Chang, I.W., He, H.L., and Li, C.F. (2016). SPOCK1 Overexpression Confers a Poor Prognosis in Urothelial Carcinoma. *J Cancer* 7, 467-476.

Mcmaster, M.L., Sun, C., Landi, M.T., Savage, S.A., Rotunno, M., Yang, X.R., Jones, K., Vogt, A., Hutchinson, A., Zhu, B., Wang, M., Hicks, B., Thirunavukkarasan, A., Stewart, D.R., Koutros, S., Goldstein, A.M., Chanock, S.J., Caporaso, N.E., Tucker, M.A., Goldin, L.R., and Liu, Y. (2018). Germline mutations in Protection of Telomeres 1 in two families with Hodgkin lymphoma. *British journal of haematology* 181, 372-377.

Ngan, E., Kiepas, A., Brown, C.M., and Siegel, P.M. (2018). Emerging roles for LPP in metastatic cancer progression. *Journal of cell communication and signaling* 12, 143-156.

Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I., Sherborne, A., Chubb, D., Carvajal-Carmona, L.G., Ma, Y., Kaur, K., Dobbins, S., Barclay, E., Gorman, M., Martin, L., Kovac, M.B., Humphray, S., Lucassen, A., Holmes, C.C., Bentley, D., Donnelly, P., Taylor, J., Petridis, C., Roylan, R., Sawyer, E.J., Kerr, D.J., Clark, S., Grimes, J., Kearsey, S.E., Thomas, H.J., Mcvean, G., Houlston, R.S., and Tomlinson, I. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 45, 136-144.

Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* 9, e1003709.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20, 110-121.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., Consortium, W.G.S., Wilkie, A.O.M., Mcvean, G., and Lunter, G. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet* 46, 912-918.

Ristolainen, H., Kilpivaara, O., Kamper, P., Taskinen, M., Saarinen, S., Leppa, S., D'amore, F., and Aaltonen, L.A. (2015). Identification of homozygous deletion in ACAN and other candidate
variants in familial classical Hodgkin lymphoma by exome sequencing. *Br J Haematol* 170, 428-431.

Rotunno, M., McMaster, M.L., Boland, J., Bass, S., Zhang, X., Burdett, L., Hicks, B., Ravichandran, S., Luke, B.T., Yeager, M., Fontaine, L., Hyland, P.L., Goldstein, A.M., Group, N.D.C.S.W., Laboratory, N.D.C.G.R., Chanock, S.J., Caporaso, N.E., Tucker, M.A., and Goldin, L.R. (2016). Whole exome sequencing in families at high risk for Hodgkin lymphoma: identification of a predisposing mutation in the KDR gene. *Haematologica* 101, 853-860.

Saarinen, S., Aavikko, M., Aittomaki, K., Launonen, V., Lehtonen, R., Franssila, K., Lehtonen, H.J., Kaasinen, E., Broderick, P., Tarkkalan, J., Bain, B.J., Bauduer, F., Unal, A., Swerdlow, A.J., Cooke, R., Makinen, M.J., Houlston, R., Vahteristo, P., and Aaltonen, L.A. (2011). Exome sequencing reveals germline NPAT mutation as a candidate risk factor for Hodgkin lymphoma. *Blood* 118, 493-498.

Salipante, S.J., Mealiffe, M.E., Wechsler, J., Krem, M.M., Liu, Y., Namkoong, S., Bhagat, G., Kirchhoff, T., Offit, K., Lynch, H., Wiernik, P.H., Roshal, M., McMaster, M.L., Tucker, M., Fromm, J.R., Goldin, L.R., and Horwitz, M.S. (2009). Mutations in a gene encoding a midbody kelch protein in familial and sporadic classical Hodgkin lymphoma lead to binucleated cells. *Proc Natl Acad Sci U S A* 106, 14920-14925.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., and Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15, 1034-1050.

Smigielski, E.M., Sirotkin, K., Ward, M., and Sherry, S.T. (2000). dbSNP: a database of single nucleotide polymorphisms. *Nucleic acids research* 28, 352-355.

Srivastava, A., Kumar, A., Giangiobbe, S., Bonora, E., Hemminki, K., Forsti, A., and Bandapalli, O.R. (2019). Whole Genome Sequencing of Familial Non-Medullary Thyroid Cancer Identifies Germline Alterations in MAPK/ERK and PI3K/AKT Signaling Pathways. *Biomolecules* 9.

Tan, D.E.K., Foo, J.N., Bei, J.-X., Chang, J., Peng, R., Zheng, X., Wei, L., Huang, Y., Lim, W.Y., Li, J., Cui, Q., Chew, S.H., Ebstein, R.P., Kuperan, P., Lim, S.T., Tao, M., Tan, S.H., Wong, A., Wong, G.C., Tan, S.Y., Ng, S.B., Zeng, Y.-X., Khor, C.C., Lin, D., Seow, A.L.H., Jia, W.-H., and Liu, J. (2013). Genome-wide association study of B cell non-Hodgkin lymphoma identifies 3q27 as a susceptibility locus in the Chinese population. *Nature Genetics* 45, 804.
Kingsbury, Z., Lehrach, H., Sudbrak, R., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68.

Valla, K., Flowers, C.R., and Koff, J.L. (2018). Targeting the B cell receptor pathway in non-Hodgkin lymphoma. *Expert opinion on investigational drugs* 27, 513-522.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38, e164.

Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 40, D930-934.

Wong, K., Robles-Espinoza, C.D., Rodriguez, D., Rudat, S.S., Puig, S., Potrony, M., Wong, C.C., Hewinson, J., Aguilera, P., Puig-Butille, J.A., Bressac-De Paillerets, B., Zattara, H., Van Der Weyden, L., Fletcher, C.D.M., Brenn, T., Arends, M.J., Quesada, V., Newton-Bishop, J.A., Lopez-Otin, C., Bishop, D.T., Harms, P.W., Johnson, T.M., Durham, A.B., Lombard, D.B., and Adams, D.J. (2019). Association of the POT1 Germline Missense Variant p.I78T With Familial Melanoma. *JAMA Dermatol* 155, 604-609.

Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D.A., Wilkinson, M.R., Vadodaria, B., Chen, X., Mcgee, R.B., Hines-Dowell, S., Nuccio, R., Quinn, E., Shurtleff, S.A., Rusch, M., Patel, A., Becksfort, J.B., Wang, S., Weaver, M.S., Ding, L., Mardis, E.R., Wilson, R.K., Gajjar, A., Ellison, D.W., Pappo, A.S., Pui, C.-H., Nichols, K.E., and Downing, J.R. (2015). Germline Mutations in Predisposition Genes in Pediatric Cancer. *New England Journal of Medicine* 373, 2336-2346.
### Table 1. Top missense variants prioritized using the FCVPPv2. Chromosomal positions, classifications, PHRED-like CADD scores, protein changes and the number of positive intolerance (Int) and deleteriousness (Del) scores are shown for each variant.

| Family ID | Position (Hg 19) | Gene | Effect | CADD PHRED | Int (n/3) | Del (n/10) |
|-----------|------------------|------|--------|------------|-----------|------------|
| Family_1  | 17_48746518_C_T  | ABCC3 | p.P652L | 22.6       | 2         | 7          |
| Family_1  | 1_49052793_G_A   | AGBL4 | p.R384C | 35         | 2         | 7          |
| Family_1  | 5_139909090_A_G  | ANKH1D1, ANKH1D1-EIF4B3 | p.N2187D | 25.2       | .         | 6          |
| Family_1  | 1_160164884_T_C  | CASQ1 | p.L183T | 26.5       | 2         | 10         |
| Family_1  | 14_95560456_A_C  | Dicer1 | p.I1711M | 24         | 3         | 7          |
| Family_1  | 6_159206584_G_A  | EZR   | p.P75L  | 32         | 3         | 9          |
| Family_1  | 12_8192537_G_A   | FOXJ2 | p.G37R  | 29.9       | 3         | 9          |
| Family_1  | 14_88729713_C_T  | KCNk10 | p.A79T | 27.1       | 3         | 6          |
| Family_1  | 10_88705360_G_A  | MMN2  | p.P58L  | 29.8       | 2         | 7          |
| Family_1  | 5_36962227_G_A   | NIPBL | p.R154Q | 27.5       | 3         | 8          |
| Family_1  | 2_206614449_A_G  | NRP2  | p.D596G | 23.1       | 2         | 9          |
| Family_1  | 1_45268632_C_T   | PLK3   | p.T252M | 25.3       | 3         | 6          |
| Family_1  | 19_45515485_T_C  | RELB   | p.I152T | 26         | 3         | 6          |
| Family_1  | 6_52372363_G_C   | TRAM2  | p.A205G | 29.8       | 3         | 10         |
| Family_1  | 22_18613830_C_T  | TUBA8  | p.A450V | 24.7       | 3         | 10         |
| Family_2  | X_47272364_G_A   | ZNF157 | p.G298R | 27.6       | 2         | 6          |
| Family_2  | 9_116151739_G_C  | ALAD   | p.J243M | 22.9       | 2         | 6          |
| Family_2  | 9_131196759_G_T  | CERCAM | p.A468S | 24.5       | 2         | 7          |
| Family_2  | 9_131367889_C_T  | SPTAN1 | p.R1327C | 34         | 3         | 6          |
| Family_3  | 9_139917418_C_T  | ABCA2  | p.G83S  | 26.3       | 2         | 7          |
| Family_3  | 17_40971572_G_C  | BECN1  | p.P85R  | 23.4       | 3         | 6          |
| Family_3  | 8_67968830_G_T   | COPS5  | p.P131T | 23.9       | 3         | 9          |
| Family_3  | 3_5246773_C_T    | EDEM1  | p.T160M | 34         | 3         | 9          |
| Family_3  | 6_131191103_G_A  | EPB41L2 | p.S736F | 22.1       | 3         | 7          |
| Family_3  | 8_28575243_G_A   | EXT3   | p.R172H | 23         | 3         | 6          |
| Family_3  | 3_188123979_A_T  | LPP    | p.E24V  | 32         | 2         | 6          |
| Family_3  | 14_74970734_C_T  | LTBP2  | p.G1493R | 27.7       | 3         | 10         |
| Family_3  | 3_196730925_C_A  | MFI2   | p.D662Y | 34         | 3         | 6          |
| Family_3  | 17_27441093_G_A  | MYO18A | p.A843V | 24.3       | 3         | 6          |
| Family_3  | 19_14584756_A_G  | PTGER1 | p.L1162P | 25.9       | 2         | 6          |
| Family_3  | 3_49138083_G_A   | QARS   | p.R301C | 34         | 2         | 9          |
| Family_3  | 17_34423827_G_A  | RAD51D, RAD51L3-RFFL | p.R266C | 27         | 3         | 8          |
| Family_3  | 11_9838541_C_T   | SBF2   | p.R1275H | 33         | 2         | 9          |
| Family_3  | 9_17761502_A_G   | SH3GL2 | p.N14S  | 26.1       | 3         | 9          |
| Family_3  | 20_35467682_G_A  | SOGA1  | p.R46C  | 32         | 3         | 7          |
| Family_3  | 1_43891311_G_A   | SZT2   | p.A96T  | 31         | 3         | 6          |
Table 2. Top non-sense variants prioritized using the FCVPPv2. Chromosomal positions, classifications, PHRED-like CADD scores, protein changes, the number of positive intolerance (Int) and VarSome prediction scores are included for each variant.

| Family ID | Position (Hg 19) | Gene    | Exonic Classification | Effect   | CADD | Int (n/3) | VarSome Score [II] |
|-----------|-----------------|---------|-----------------------|----------|------|----------|-------------------|
| Family_1  | 10_88911115_AGT_A | FAM35A  | frameshift deletion   | p.2_2del | 25.8 | 2        | PM2               |
| Family_1  | 3_148802664_C_T  | HLTF    | stopgain SNV          | p.W11X   | 37   | 2        | PP3 (4)           |
| Family_1  | 1_177923437_CTG_C | SEC16B  | frameshift deletion   | p.481_481del | 36   | 0        | Uncertain Significance |
| Family_1  | 15_91546350_TG_T | VPS33B  | frameshift deletion   | p.P321fs | 36   | 3        | PVS1               |
| Family_3  | 7_31683260_AT_A  | CCDC129 | frameshift deletion   | p.D611fs | 34   | 0        | Uncertain Significance |
| Family_3  | 1_21267855_C_T   | EIF4G3  | stopgain SNV          | p.W7X    | 14.54| 2        | PVS1, PP3 (1)     |
| Family_3  | 3_188123978_G_T  | LPP     | stopgain SNV          | p.E24X   | 40   | 2        | PM2, PP3 (4)      |
| Family_3  | 15_24921469_G_A  | NPAP1   | stopgain SNV          | p.W152X  | 24.8 | 0        | PM2, PP3 (3)      |
| Family_3  | 1_241958547_CAG_C | WDR64   | frameshift deletion   | p.836_836del | 37   | 0        | Uncertain Significance |

[II] = VarSome Scores: PM2: Pathogenic Moderate, PP3: Pathogenic Supporting (no. of scores predicting pathogenicity);
Uncertain Significance: No scores could be found for the variant in question, PVS1: Pathogenic Very Strong.
Table 3. Variants corresponding to genes present in the panel of 565 known cancer predisposition genes from a study by Zhang et al.

| Gene ID | HL Family | HL Gene | HL Variant | Variant Type | Variant Classification | HGNC Approved Name | CADD_PHRED | Familial Syndrome | Category |
|---------|-----------|---------|------------|--------------|------------------------|-------------------|------------|-------------------|----------|
| 10499   | 1         | NCOA2   | 8_71316112_T_TCCTCTCCCTCC | Indel | upstream | nuclear receptor coactivator 2 | 15.56 | | Other CancerGene |
| 8880    | 1         | FUBP1   | 1_78414225_A_G | SNVs | UTR3 | far upstream element (FUSE) binding protein 1 | 13.59 | | Other CancerGene |
| 23157   | 1         | SEPT6   | X_118751062_CGTGT_C | Indel | UTR3 | septin 6 | 10.56 | | Other CancerGene |
| 23405   | 1         | DICER1  | 14_95560456_A_C | SNVs | nonsynonymous SNV | dicer 1, ribonuclease type III | 24 | DICER1 syndrome, Familial Multinodular Goiter | Autosomal Dominant |
| 7430    | 1         | EZR     | 6_159206584_G_A | SNVs | nonsynonymous SNV | ezrin | 32 | | Other CancerGene |
| 604     | 3         | BCL6    | 3_187463568_C_A | SNVs | upstream; downstream | B-cell CLL/lymphoma 6 | 13 | | Other CancerGene |
| 4026    | 3         | LPP     | 3_188123978_G_T | SNVs | stopgain SNV | LIM domain containing preferred translocation partner in lipoma | 40 | | Other CancerGene |
| 3       | LPP       | 3_188123979_A_T | SNVs | nonsynonymous SNV | LIM domain containing preferred translocation partner in lipoma | 32 | | Other CancerGene |
| 3       | LPP       | 3_188083873_A_T | SNVs | UTR3 | LIM domain containing preferred translocation partner in lipoma | 10.5 | | Other CancerGene |
| 5892    | 3         | RAD51D  | 17_33428327_G_A | SNVs | nonsynonymous SNV | RAD51 paralog D | 27 | | Other CancerGene |
| 5727    | 3         | PTCH1   | 9_98270531_C_A | SNVs | nonsynonymous SNV | patched 1 | 20.4 | Gorlin syndrome | Autosomal Dominant |
Table 4. Variants corresponding to genes intersecting with the list of 2383 high-risk HL genes from a study by Rutunno et al. Variant details from both databases (the present study and the study by Rutunno et al.) are shown.

| Gene Symbol | Variant Position IDS | HL Family | Variant Position IDS | CADD_PHRED Score | Variant Type | Variant Consequence | Protein Effect |
|-------------|----------------------|-----------|----------------------|-------------------|--------------|----------------------|---------------|
| **HL Variants From Rutunno et al.** | | | | | | | |
| **Variant in Matched Gene from Present Study** | | | | | | | |
| **Gene Symbol** | **Variant Position** | **IDS** | **HL Family** | **Variant Position** | **IDS** | **CADD_PHRED Score** | **Variant Type** | **Variant Consequence** | **Protein Effect** |
| ABHD16A | 6_31670740_A_T | 3 | 6_31671105_G_A | 13.23 | SNVs | UTR5 | - |
| C6orf62 | 6_24719009_T_C | rs147402940 | 3 | 6_24705773_T_C | 12.31 | SNVs | UTR3 | - |
| CEP120 | 5_122758609_G_T | rs141808885 | 1 | 5_122681069_C_T | 12.31 | SNVs | UTR3 | - |
| EDEM1 | 3_5257909_A_G | rs139745426 | 3 | 3_5246773_C_T | 34 | SNVs | nonsynonymous SNV | p.T160M |
| EIF4G3 | 1_21494519_T_C | rs147402940 | 3 | 1_21267855_C_T | 14.54 | SNVs | stopgain SNV | p.W7X |
| EPB41L2 | 6_131202023_A_G | rs191528081 | 3 | 6_131191103_G_A | 22.1 | SNVs | nonsynonymous SNV | p.S736F |
| EXTL3 | 8_28609630_G_A | rs191528081 | 3 | 8_28575243_G_A | 23 | SNVs | nonsynonymous SNV | p.R172H |
| FAM35A | 10_88917757_A_G | rs371636091 | 1 | 10_88911115_AGT_A | 25.8 | Indel | frameshift deletion | p.2_2del |
| FUK | 16_70507153_G_A | rs147402940 | 1 | 16_70501193_C_T | 10.08 | SNVs | UTR3 | - |
| HTLF | 3_14875790_A_G | rs61750365 | 1 | 3_148802664_C_T | 37 | SNVs | stopgain SNV | p.W11X |
| LPP | 3_188464149_C_T | 3 | 3_188608373_A_T | 10.5 | SNVs | UTR3 | - |
| LPP | 3_188464149_C_T | 3 | 3_188123978_G_T | 40 | SNVs | stopgain SNV | p.E24X |
| LPP | 3_188464149_C_T | 3 | 3_188123979_A_T | 32 | SNVs | nonsynonymous SNV | p.E24V |
| LRP6 | 12_12419973_G_T | 3 | 12_12272924_AATATATATATATATATATATA | 12.55 | Indel | UTR3 | - |
| LTBP2 | 14_74983553_G_A | rs145851939 | 3 | 14_74970734_C_T | 27.7 | SNVs | nonsynonymous SNV | p.G1493R |
| MAPKAP1 | 9_128199718_AT | rs112201730 | 3 | 9_128199770_TAA_T | 14.3 | Indel | UTR3 | - |
| MARCH1O | 17_60827878_G_A | rs112201730 | 3 | 17_60885673_G_A | 12.03 | SNVs | UTR5 | - |
### Germline Variants in Hodgkin Lymphoma

| Gene | rsID | POS (chromosome:position) | nSNVs | Type | Effect | Protein Change |
|------|------|---------------------------|-------|------|--------|----------------|
| MYO18A | rs371862120 | 17:27448659_C>T | 3 | SNVs | Nonsynonymous SNV | p.A843V |
| NCAM1 | | 11:11311356_A>G | 1 | SNVs | UTR3 | |
| NIPBL | | 5:36876673_G>A | 1 | SNVs | UTR3 | |
| PHC2 | | 1:33820711_G>A | 3 | SNVs | Upstream | |
| RCN1 | | 11:32126524_C>T | 3 | SNVs | UTR5 | |
| SBF2 | | 11:9985135_TAAT_ | 3 | SNVs | Nonsynonymous SNV | p.R1275H |
| SLMAP | rs191613999 | 3:57914019_A>C | 1 | SNVs | UTR5 | |
| SZT2 | | 1:43885320_C>T | 3 | SNVs | Nonsynonymous SNV | p.A96T |
