Identification of Familial Hodgkin Lymphoma Predisposing Genes Using Whole Genome Sequencing

Aayushi Srivastava [1,2,3,4†], Sara Giangiobbe [1,4†], Abhishek Kumar [1], Nagarajan Paramasivam [5], Dagmara Dymerska [6], Wolfgang Behnisch [7], Mathias Witzens-Harig [4], Jan Lubinski [6], Kari Hemminki [1,8], Asta Försti [1,2,3] and Obul Reddy Bandapalli [1,2,3,4]*

[1] Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), 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] Medical Faculty, Heidelberg University, Heidelberg, Germany, [5] Computational Oncology, Molecular Diagnostics Program, National Center for Tumor Diseases (NCT), Heidelberg, Germany, [6] Department of Genetics and Pathology, International Hereditary Cancer Centre, Pomeranian Medical University, Szczecin, Poland, [7] Department of Pediatric Oncology, Hematology and Immunology, University of Heidelberg, Heidelberg, Germany, [8] Faculty of Medicine and Biomedical Center in Pilsen, Charles University in Prague, Pilsen, Czechia

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 98,564, 170,550, and 113,654 variants which were reduced by pedigree-based filtering to 18,158, 465, and 26,465 in families I, II, and III, respectively. In addition to variants affecting amino acid sequences, variants in promoters, enhancers, transcription factors 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 2,383 potential candidate HL genes were also screened in these families to aid further prioritization. Pathway analysis of segregating genes with Combined Annotation Dependent Depletion Tool (CADD) 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* analyses 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.

Keywords: familial Hodgkin lymphoma, whole genome sequencing, predisposing genes, germline variants, variant prioritization, next generation sequencing, genetic predisposition to disease

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.

MATERIALS AND METHODS

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.

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 Consortium et al., 2015; Lek et al., 2016). Variants with a quality score >20 and a coverage >5×, 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 1,000 Genomes Phase 3 and non-TCGA ExAC data. Variants with a MAF <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.

Data Analysis and Variant Prioritization
Prioritization of Coding Variants

Variant evaluation was performed using the criteria of our in-house developed variant prioritization pipeline (FCVPPv2) (Kumar et al., 2018). Shortly, variants with MAF < 0.1% were first 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 CADD tool v1.3 (Kircher et al., 2014). Only variants with a scaled PHRED-like CADD score >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 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.

Structural variants were analyzed using Canvas (version 1.40.0.1613) (https://academic.oup.com/bioinformatics/article/32/15/2375/1743834) program's SmallPedigree-WGS separately to detect the larger copy number variants. The joint genotyped VCF for all the samples in a family generated via Platypus was used as the b-allele input file along with the BAM files, and the rest of the parameters were kept default. Variants with "PASS" filters and present in all the cases in a family were processed further and variants overlapping common structural variants (AF > 1%) from gnomAD (version 2.1) were marked as common and removed. The remaining rare structural variants that affects the known cancer predisposition genes were selected for the manual inspection in IGV.

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.

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 checked 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.

Presence of Candidate Variants in Prioritized HL Genes From a Large WES-Based Familial HL Study

In a study by Rotunno et al. (2016) 2,699 variants corresponding to 2,383 genes were identified in 17 HL discovery families after filtering and prioritization. We intersected our list of candidate genes with this list of 2,383 HL genes to identify coding and non-coding variants from our shortlist in potentially causative HL genes.

Variant Validation

Specific variants of interest mentioned throughout the text (*DICER1, HLTF, LPP, PLK3, RAD51D, RELB, SH3GL2,* and *SPTAN1*) and highlighted as bold in the tables 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 3,500 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.

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.

RESULTS

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**.

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 98,564, 170,550, and 113,654 variants which were reduced by pedigree-based filtering to 18,158, 465, and 26,465 in families I, II, and III, respectively.

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**, **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 SNPNexus 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.

Analysis of structural variants resulted in identification of a large deletion in exons 9 and 10 (del5395) of Chek2 kinase gene (*CHEK2)* in family 1 that segregates with the disease.

Candidate Variants in 565 CPGs and 2383 Potentially Causative HL Genes

Intersecting our prioritized list of candidate genes with the list of 565 CPGs, we identified 11 variants in nine genes 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 2,383 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**).
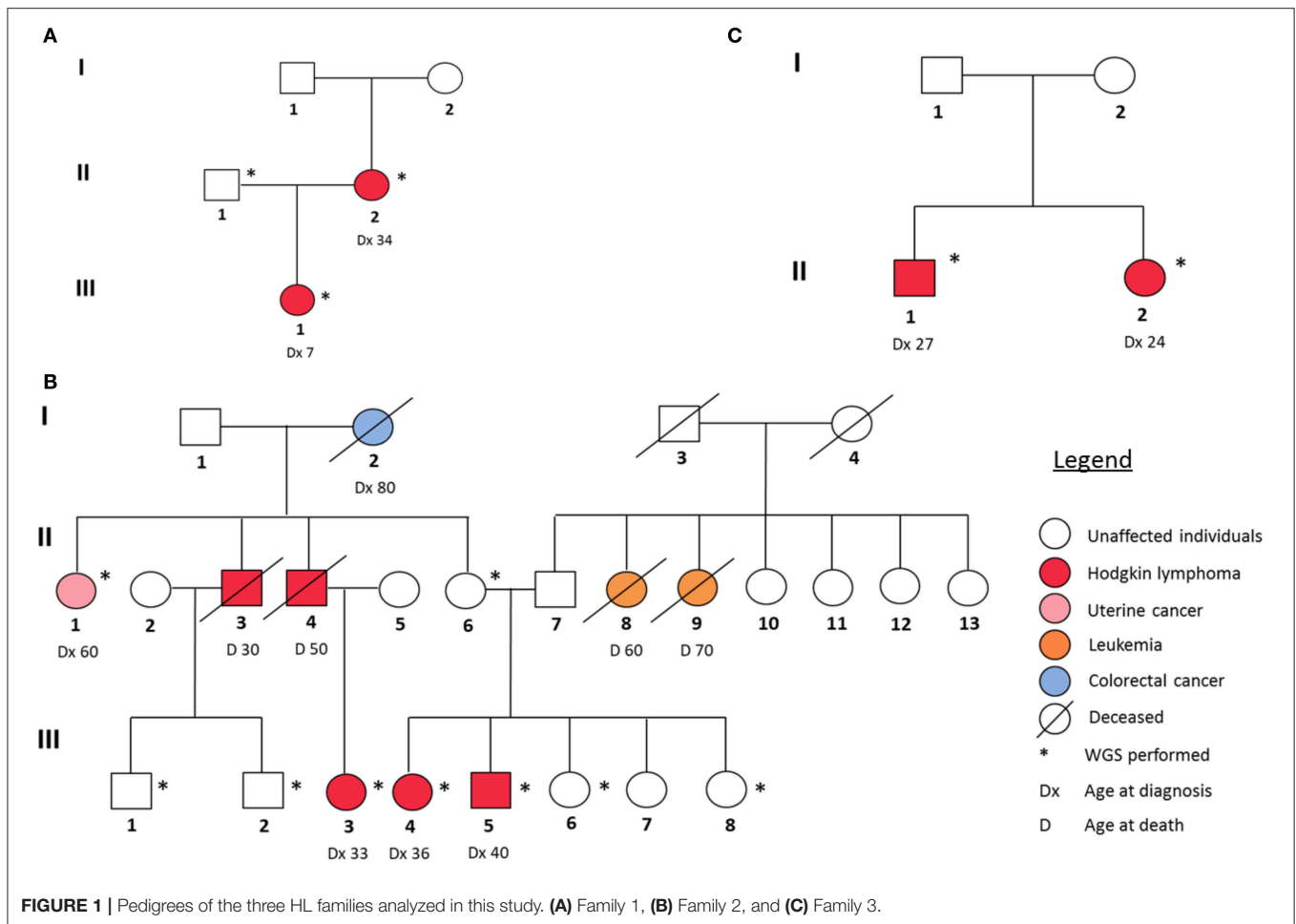
Network and Pathway Analysis With IPA

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,B**).

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**).

Literature Mining, Consolidation of Results, and Selection of Candidates

With the aim of identifying one highly penetrant 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

FIGURE 1 | Pedigrees of the three HL families analyzed in this study. **(A)** Family 1, **(B)** Family 2, and **(C)** Family 3.

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 *LPP* and *RAD51D*. Both genes were found in the list of 565 CPGs and *LPP* was additionally found in the gene list from the large cohort of HL families. Three variants in *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 non-synonymous missense variant (3_188123979_A_T). 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. 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 *BCL6* and *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 *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.

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 (*DICER1*), short-listed three in family II (*ALAD*, *CERCAM*, and *SPTAN1*) and two in family III (*RAD51* and *LPP*), to investigate further with validation and functional studies. For family I we have already functionally validated *DICER1* as the candidate predisposing gene in a previous study (Bandapalli et al., 2018). However, it was important to include the family in this paper, especially with regard to the integrity of the pathway and network analyses. We identified pathways related to

TABLE 1 | Top missense variants prioritized using the FCVPPv2.

Family ID	Position (Hg 19)	Gene	Effect	CADD PHRED	Int (n/3)	Del (n/10)	VarSome Score
Family_1	17_48746518_C_T	ABCC3	p.P652L	22.6	2	7	Uncertain significance
Family_1	1_49052793_G_A	AGBL4	p.R384C	35	2	7	Uncertain significance
Family_1	5_139909090_A_G	ANKHD1, ANKHD1-EIF4EBP3	p.N2187D	25.2	.	6	Uncertain significance
Family_1	1_160164884_T_C	CASQ1	p.I183T	26.5	2	10	Likely Benign
Family_1	**14_95560456_A_C**	**DICER1**	**p.I1711M**	**24**	**3**	**7**	**Uncertain significance**
Family_1	6_159206584_G_A	EZR	p.P75L	32	3	9	Uncertain significance
Family_1	12_8192537_G_A	FOXJ2	p.G37R	29.9	3	9	Uncertain significance
Family_1	14_88729713_C_T	KCNK10	p.A79T	27.1	3	6	Uncertain significance
Family_1	10_88705360_G_A	MMRN2	p.P58L	29.8	2	7	Uncertain significance
Family_1	5_36962227_G_A	NIPBL	p.R154Q	27.5	3	8	Uncertain significance
Family_1	2_206614449_A_G	NRP2	p.D596G	23.1	2	9	Uncertain significance
Family_1	**1_45268632_C_T**	**PLK3**	**p.T252M**	**25.3**	**3**	**6**	**Uncertain significance**
Family_1	**19_45515485_T_C**	**RELB**	**p.I152T**	**26**	**3**	**6**	**Uncertain significance**
Family_1	6_52372363_G_C	TRAM2	p.A205G	29.8	3	10	Uncertain significance
Family_1	22_18613830_C_T	TUBA8	p.A450V	24.7	3	10	Uncertain significance
Family_1	X_47272364_G_A	ZNF157	p.G298R	27.6	2	6	Uncertain significance
Family_2	9_116151739_G_C	ALAD	p.I243M	22.9	2	6	Uncertain significance
Family_2	9_131196759_G_T	CERCAM	p.A468S	24.5	2	7	Uncertain significance
Family_2	**9_131367689_C_T**	**SPTAN1**	**p.R1327C**	**34**	**3**	**6**	**Uncertain significance**
Family_3	9_139917418_C_T	ABCA2	p.G83S	26.3	2	7	Uncertain significance
Family_3	17_40971572_G_C	BECN1	p.P85R	23.4	3	6	Uncertain significance
Family_3	8_67968830_G_T	COPS5	p.P131T	23.9	3	9	Uncertain significance
Family_3	3_5246773_C_T	EDEM1	p.T160M	34	3	9	Uncertain significance
Family_3	6_131191103_G_A	EPB41L2	p.S736F	22.1	3	7	Uncertain significance
Family_3	8_28575243_G_A	EXTL3	p.R172H	23	3	6	Likely Benign
Family_3	**3_188123979_A_T**	**LPP**	**p.E24V**	**32**	**2**	**6**	**Uncertain significance**
Family_3	14_74970734_C_T	LTBP2	p.G1493R	27.7	3	10	Uncertain significance
Family_3	3_196730925_C_A	MFI2	p.D662Y	34	3	6	Uncertain significance
Family_3	17_27441099_G_A	MYO18A	p.A843V	24.3	3	6	Uncertain significance
Family_3	19_14584756_A_G	PTGER1	p.L126P	25.9	2	6	Uncertain significance
Family_3	3_49138083_G_A	QARS	p.R301C	34	2	9	Uncertain significance
Family_3	**17_33428327_G_A**	**RAD51D,RAD51L3-RFFL**	**p.R266C**	**27**	**3**	**8**	**Benign**
Family_3	11_9838541_C_T	SBF2	p.R1275H	33	2	9	Likely Benign
Family_3	**9_17761502_A_G**	**SH3GL2**	**p.N14S**	**26.1**	**3**	**9**	**Uncertain significance**
Family_3	20_35467682_G_A	SOGA1	p.R46C	32	3	7	Uncertain significance
Family_3	1_43891311_G_A	SZT2	p.A96T	31	3	6	Uncertain significance

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. Variants of interest that were validated by Sanger sequencing in the provided family samples have been shown in bold.

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. Interestingly, several genes in our gene list were related to DNA repair (e.g., *NOTCH3*, *RAD51*, and *SPTAN1*).

In the current study, we also identified a deletion of exon 9 and 10 in *CHEK2* in family 1. The same deletion has been reported in several unrelated patients with breast cancer of Polish origin. In that study the deletion of exon 9 and 10 in *CHEK2* was shown

to lead to a premature protein truncation at codon 381 and to evoke a 2-fold increase in the risk of prostate cancer and a 4-fold increase in the risk of familial prostate cancer (Cybulski et al., 2006). The detection of mRNA of abnormal length suggests that the deletion does not lead to complete transcript loss and therefore, the effect of this truncating mutation on cancer risk may differ or work in tandem with another genetic effect, may be with *DICER1* in this family but warrants further experiments. Personalized medicine is an upcoming and promising field of medicine in which medical decisions, practices, interventions,

TABLE 2 | Top non-sense variants prioritized using the FCVPPv2.

Family ID	Position (Hg 19)	Gene	Exonic classification	Effect	CADD	Int (n/3)	VarSome score [I]
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

[I]VarSome Scores.
PM2, Pathogenic Moderat; PP3, Pathogenic Supporting (no. of scores predicting pathogenicity); Uncertain Significance: No scores could be found for the variant in question, PVS1, Pathogenic Very Strong.
Chromosomal positions, classifications, PHRED-like CADD scores, protein changes, the number of positive intolerance (Int) and VarSome prediction scores are included for each variant. Variants of interest that were validated by Sanger sequencing in the provided family samples have been shown in bold.

TABLE 3 | Variants corresponding to genes present in the panel of 565 known cancer predisposition genes from a study by Zhang et al. (2015).

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_TCCTCCTCCC	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	Non-synonymous SNV	Dicer 1, ribonuclease type III	24	DICER1 syndrome, Familial Multinodular Goiter	Autosomal Dominant
7430	1	EZR	6_159206584_G_A	SNVs	Non-synonymous 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	Non-synonymous SNV	LIM domain containing preferred translocation partner in lipoma	32		Other CancerGene
	3	LPP	3_188608373_A_T	SNVs	UTR3	LIM domain containing preferred translocation partner in lipoma	10.5		Other CancerGene
5892	3	RAD51D	17_33428327_G_A	SNVs	Non-synonymous SNV	RAD51 paralog D	27		Other CancerGene
5727	3	PTCH1	9_98270531_C_A	SNVs	Non-synonymous SNV	Patched 1	20.4	Gorlin syndrome	Autosomal Dominant

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 (*ALAD)* from family 2 is interesting, as it is

TABLE 4 | Variants corresponding to genes intersecting with the list of 2,383 high-risk HL genes from a study by Rotunno et al. (2016).

HL variants from Rotunno et al. (2016)			Variant in matched gene from present study					
Gene symbol	Variant position	IDS	HL family	Variant position	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	Non-synonymous SNV	p.T160M
EIF4G3	1_21494519_T_C		3	1_21267855_C_T	14.54	SNVs	stopgain SNV	p.W7X
EPB41L2	6_131202023_A_G		3	6_131191103_G_A	22.1	SNVs	Non-synonymous SNV	p.S736F
EXTL3	8_28609630_G_A	rs191528081	3	8_28575243_G_A	23	SNVs	Non-synonymous 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		1	16_70501193_C_T	10.08	SNVs	UTR3	-
HLTF	3_148757909_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	3_188123979_A_T	32	SNVs	Non-synonymous SNV	p.E24V
LRP6	12_12419973_G_T		3	12_12272924_AATA TATATAT ATATATATATATATA TATATATA TATAT_A	12.55	Indel	UTR3	–
LTBP2	14_74983553_G_A	rs145851939	3	14_74970734_C_T	27.7	SNVs	Non-synonymous SNV	p.G1493R
MAPKAP1	9_128199718_AT_		3	9_128199770_TAA_T	14.3	Indel	UTR3	–
MARCH10	17_60827878_G_A	rs112201730	3	17_60885673_G_A	12.03	SNVs	UTR5	–
MYO18A	17_27448659_C_T	rs371862120	3	17_27441099_G_A	24.3	SNVs	Non-synonymous SNV	p.A843V
NCAM1	11_113113556_A_G		1	11_113134920_C_A	11.07	SNVs	UTR3	–
NIPBL	5_36876673_G_A		1	5_36962227_G_A	27.5	SNVs	Non-synonymous SNV	p.R154Q
PHC2	1_33820711_G_A		3	1_33896663_C_A	15.12	SNVs	upstream	–
RCN1	11_32126524_C_T		3	11_32112681_C_T	10.66	SNVs	UTR5	–
SBF2	11_9985135_TAAT_		3	11_9838541_C_T	33	SNVs	Non-synonymous SNV	p.R1275H
SLMAP	3_57914019_A_C	rs191613999	1	3_57742023_C_G	13.51	SNVs	UTR5	–
SZT2	1_43885320_C_T		3	1_43891311_G_A	31	SNVs	Non-synonymous SNV	p.A96T

Variant details from both databases the present study and the study by Rotunno et al. (2016) are shown.

involved in the catalysis of the second step in the biosynthesis of heme and also acts as an endogenous inhibitor of the 26 S proteasome, a multi-catalytic ATP-dependent protease complex that functions as the degrading arm of the ubiquitin system, which is the major pathway for regulated degradation of proteins in all eukaryotes. Down regulation of *ALAD* is shown to be associated with poor prognosis in patients with breast cancer (Ge et al., 2017) whereas the existing data on non-erythroid spectrin αII (*SPTAN1*) suggest that overexpression of *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). *CERCAM* is known as an unfavorable prognostic marker in urothelial, renal, and ovarian cancers implying the importance of the variants in these genes (Ma et al., 2016). RAD51D from family III is particularly interesting since it 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
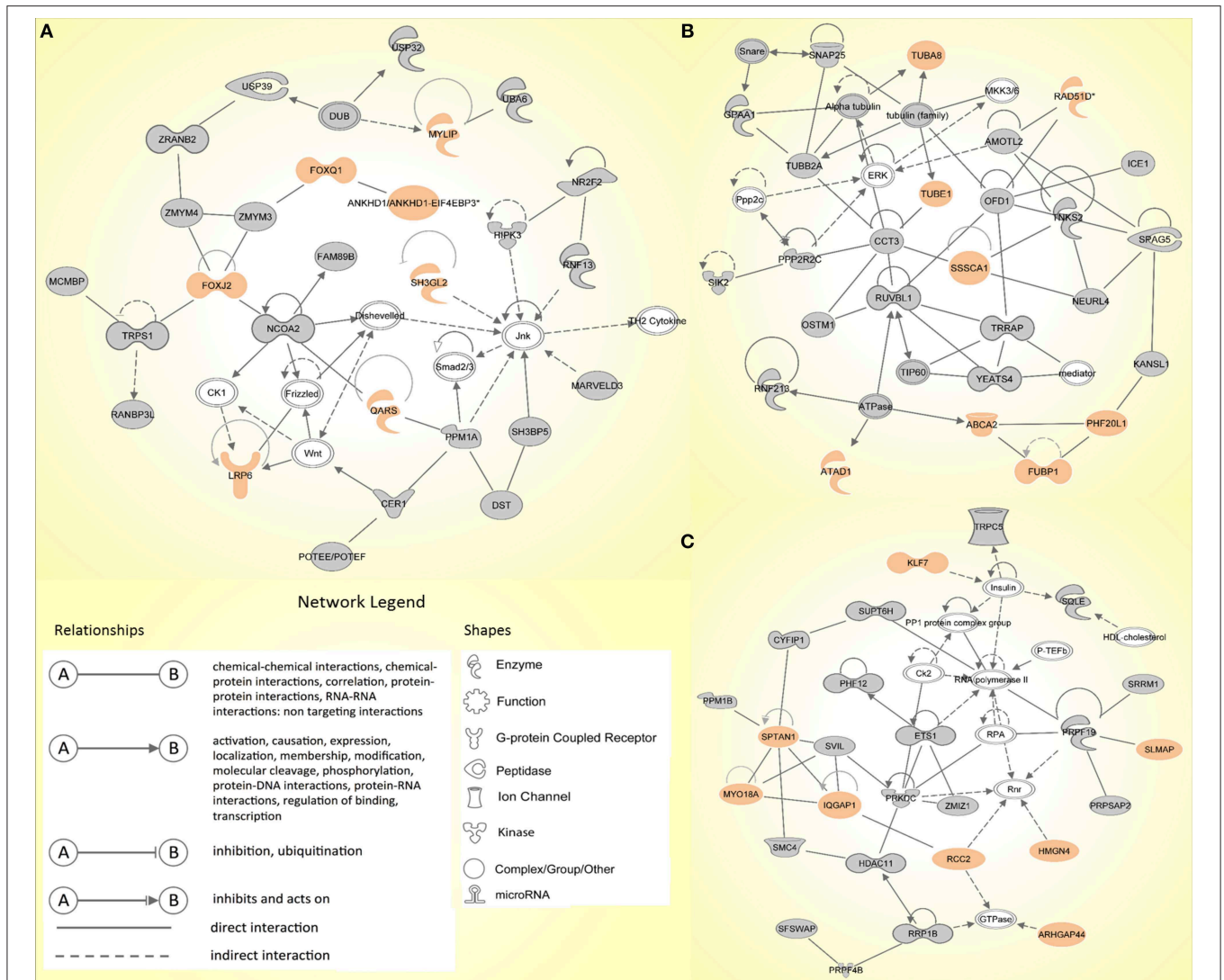
FIGURE 2 | The top three molecular networks identified by Ingenuity Pathway Analysis: **(A)** Network 1. Cancer, hematological disease, immunological disease; **(B)** Network 2. developmental disorder, endocrine system disorders, hereditary disorder; **(C)** Network 3. RNA post-transcriptional modification, cell death and survival, cellular movement. Genes from our input-data are shown in gray, genes from our prioritized candidate list are highlighted in peach.

(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 lympomagenesis 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 (BTK9), 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.

DATA AVAILABILITY STATEMENT

Unfortunately, for reasons of ethics and patient confidentiality 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).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the University of Heidelberg, Germany & Ethics Committee of the Pomeranian Medical University, Poland. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

OB, AF, and KH conceived and designed the study. WB, MW-H, DD, and JL provided the HL family samples. NP ran WGS pipeline and CNVs analysis. AS, OB, SG, and AK analyzed the data. OB and SG performed the experiments. AS and OB wrote the first draft of the manuscript. All authors read, commented on, and approved the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00179/full#supplementary-material

REFERENCES

Ackermann, A., and Brieger, A. (2019). The role of nonerythroid spectrin αII in cancer. *J. Oncol.* 2019:7079604. doi: 10.1155/2019/7079604

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4, 1–14. doi: 10.7554/eLife.05005.028

Bandapalli, O. R., Paramasivam, N., Giangiobbe, S., Kumar, A., Benisch, W., Engert, A., et al. (2018). Whole genome sequencing reveals DICER1 as a candidate predisposing gene in familial Hodgkin lymphoma. *Int J Cancer* 143, 2076–2078. doi: 10.1002/ijc.31576

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. doi: 10.1038/nrclinonc.2010.116

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. doi: 10.1186/gb-2010-11-8-r90

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi: 10.1038/nature05874

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790–1797. doi: 10.1101/gr.137323.112

Buggy, J. J., and Elias, L. (2012). Bruton tyrosine kinase (BTK) and its role in B-cell malignancy. *Int. Rev. Immunol.* 31, 119–132. doi: 10.3109/08830185.2012.664797

Chen, X., Li, Y., Ouyang, T., Li, J., Wang, T., Fan, Z., et al. (2018). Associations between RAD51D germline mutations and breast cancer risk

and survival in BRCA1/2-negative breast cancers. *Ann. Oncol.* 29, 2046–2051. doi: 10.1093/annonc/mdy338

Cooper, G. M., Stone, E. A., Asimenos, G., Program, N. C. S., Green, E. D., Batzoglou, S., et al. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 15, 901–913. doi: 10.1101/gr.3577405

Cozen, W., Timofeeva, M. N., Li, D., Diepstra, A., Hazelett, D., Delahaye-Sourdeix, M., et al. (2014). A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. *Nat. Commun.* 5:3856. doi: 10.1038/ncomms4856

Cybulski, C., Wokolorczyk, D., Huzarski, T., Byrski, T., Gronwald, J., Gorski, B., et al. (2006). A large germline deletion in the Chek2 kinase gene is associated with an increased risk of prostate cancer. *J. Med. Genet.* 43, 863–866. doi: 10.1136/jmg.2006.044974

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 Res.* 46, W109–W113. doi: 10.1093/nar/gky399

Diehl, V., Thomas, R. K., and Re, D. (2004). Part II: Hodgkin's lymphoma–diagnosis and treatment. *Lancet Oncol.* 5, 19–26. doi: 10.1016/S1470-2045(03)01320-2

Diepstra, A., Niens, M., Vellenga, E., Van Imhoff, G. W., Nolte, I. M., Schaapveld, M., et al. (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. doi: 10.1016/S0140-6736(05)66780-3

Dreyling, M., Santoro, A., Mollica, L., Leppa, S., Follows, G. A., Lenz, G., et al. (2017). Phosphatidylinositol 3-kinase inhibition by copanlisib in relapsed or refractory indolent lymphoma. *J. Clin. Oncol.* 35, 3898–3905. doi: 10.1200/JCO.2017.75.4648

Frampton, M., Da Silva Filho, M. I., Broderick, P., Thomsen, H., Forsti, A., Vijayakrishnan, J., et al. (2013). Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. *Nat. Commun.* 4:2549. doi: 10.1038/ncomms3549

Ge, J., Yu, Y., Xin, F., Yang, Z. J., Zhao, H. M., Wang, X., et al. (2017). Downregulation of delta-aminolevulinate dehydratase is associated with poor prognosis in patients with breast cancer. *Cancer Sci.* 108, 604–611. doi: 10.1111/cas.13180

Kharazmi, E., Fallah, M., Pukkala, E., Olsen, J. H., Tryggvadottir, L., Sundquist, K., et al. (2015). Risk of familial classical Hodgkin lymphoma by relationship, histology, age, and sex: a joint study from five Nordic countries. *Blood* 126, 1990–1995. doi: 10.1182/blood-2015-04-639781

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. doi: 10.1038/ng.2892

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., et al. (2018). VarSome: the human genomic variant search engine. *Bioinformatics* 35, 1978–1980. doi: 10.1093/bioinformatics/bty897

Kumar, A., Bandapalli, O. R., Paramasivam, N., Giangiobbe, S., Diquigiovanni, C., Bonora, E., et al. (2018). Familial cancer variant prioritization pipeline version 2 (FCVPPv2) applied to a papillary thyroid cancer family. *Sci. Rep.* 8:11635. doi: 10.1038/s41598-018-29952-z

Kuppers, R. (2009). The biology of Hodgkin's lymphoma. *Nat. Rev. Cancer* 9, 15–27. doi: 10.1038/nrc2542

Kushekhar, K., Van Den Berg, A., Nolte, I., Hepkema, B., Visser, L., and Diepstra, A. (2014). Genetic associations in classical hodgkin lymphoma: a systematic review and insights into susceptibility mechanisms. *Cancer Epidemiol. Biomarkers Prev.* 23:2737–2747. doi: 10.1158/1055-9965.EPI-14-0683

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291. doi: 10.1038/nature19057

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. doi: 10.1093/bioinformatics/btr509

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Liu, D., and Mamorska-Dyga, A. (2017). Syk inhibitors in clinical development for hematological malignancies. *J. Hematol. Oncol.* 10:145. doi: 10.1186/s13045-017-0512-1

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. doi: 10.1002/humu.22932

Loveday, C., Turnbull, C., Ramsay, E., Hughes, D., Ruark, E., Frankum, J. R., et al. (2011). Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat. Genet.* 43, 879–882. doi: 10.1038/ng.893

Ma, L. J., Wu, W. J., Wang, Y. H., Wu, T. F., Liang, P. I., Chang, I. W., et al. (2016). SPOCK1 overexpression confers a poor prognosis in urothelial carcinoma. *J. Cancer* 7, 467–476. doi: 10.7150/jca.13625

Mcmaster, M. L., Sun, C., Landi, M. T., Savage, S. A., Rotunno, M., Yang, X. R., et al. (2018). Germline mutations in protection of telomeres 1 in two families with Hodgkin lymphoma. *Br. J. Haematol.* 181, 372–377. doi: 10.1111/bjh.15203

Ngan, E., Kiepas, A., Brown, C. M., and Siegel, P. M. (2018). Emerging roles for LPP in metastatic cancer progression. *J. Cell Commun. Signal.* 12, 143–156. doi: 10.1007/s12079-017-0415-5

Palles, C., Cazier, J. B., Howarth, K. M., Domingo, E., Jones, A. M., Broderick, P., et al. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* 45, 136–144. doi: 10.1038/ng.2503

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. doi: 10.1371/annotation/32c8d343-9e1d-46c6-bfd4-b0cd3fb7a97e

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. doi: 10.1101/gr.097857.109

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R. F., Consortium, W. G. S., et al. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* 46, 912–918. doi: 10.1038/ng.3036

Ristolainen, H., Kilpivaara, O., Kamper, P., Taskinen, M., Saarinen, S., Leppa, S., et al. (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. doi: 10.1111/bjh.13295

Rotunno, M., Mcmaster, M. L., Boland, J., Bass, S., Zhang, X., Burdett, L., et al. (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. doi: 10.3324/haematol.2015.135475

Saarinen, S., Aavikko, M., Aittomaki, K., Launonen, V., Lehtonen, R., Franssila, K., et al. (2011). Exome sequencing reveals germline NPAT mutation as a candidate risk factor for Hodgkin lymphoma. *Blood* 118, 493–498. doi: 10.1182/blood-2011-03-341560

Salipante, S. J., Mealiffe, M. E., Wechsler, J., Krem, M. M., Liu, Y., Namkoong, S., et al. (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. doi: 10.1073/pnas.0904231106

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050. doi: 10.1101/gr.3715005

Smigielski, E. M., Sirotkin, K., Ward, M., and Sherry, S. T. (2000). dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.* 28, 352–355. doi: 10.1093/nar/28.1.352

Srivastava, A., Kumar, A., Giangiobbe, S., Bonora, E., Hemminki, K., Forsti, A., et al. (2019). Whole genome sequencing of familial non-medullary thyroid cancer identifies germline alterations in MAPK/ERK and PI3K/AKT signaling pathways. *Biomolecules* 9:E605. doi: 10.3390/biom9100605

Tan, D. E. K., Foo, J. N., Bei, J.-X., Chang, J., Peng, R., Zheng, X., et al. (2013). Genome-wide association study of B cell non-Hodgkin lymphoma identifies 3q27 as a susceptibility locus in the Chinese population. *Nat. Genet.* 45:804. doi: 10.1038/ng.2666

The Genomes Project Consortium, Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

Valla, K., Flowers, C. R., and Koff, J. L. (2018). Targeting the B cell receptor pathway in non-Hodgkin lymphoma. *Expert Opin. Investig. Drugs* 27, 513–522. doi: 10.1080/13543784.2018.1482273

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603

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–D934. doi: 10.1093/nar/gkr917

Wong, K., Robles-Espinoza, C. D., Rodriguez, D., Rudat, S. S., Puig, S., Potrony, M., et al. (2019). Association of the POT1 germline missense variant p.I78T with familial melanoma. *JAMA Dermatol.* 155, 604–609. doi: 10.1001/jamadermatol.2018.3662

Zhang, J., Walsh, M. F., Wu, G., Edmonson, M. N., Gruber, T. A., Easton, J., et al. (2015). Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* 373, 2336–2346. doi: 10.1056/NEJMoa1508054