

SNP Association Mapping across the Extended Major Histocompatibility Complex and Risk of B-Cell Precursor Acute Lymphoblastic Leukemia in Children

Kevin Y. Urayama^{1,2*}, Anand P. Chokkalingam¹, Catherine Metayer¹, Helen Hansen³, Suzanne May⁴, Patricia Ramsay⁴, Joseph L. Wiemels³, John K. Wiencke³, Elizabeth Trachtenberg⁵, Pamela Thompson⁶, Yasushi Ishida⁷, Paul Brennan⁸, Kent W. Jolly⁹, Amanda M. Termuhlen^{10,11}, Malcolm Taylor¹², Lisa F. Barcellos⁴, Patricia A. Buffler¹

1 School of Public Health, University of California, Berkeley, Berkeley, California, United States of America, **2** Center for Clinical Epidemiology, St. Luke's Life Science Institute, Tokyo, Japan, **3** Laboratory for Molecular and Neuroepidemiology, University of California San Francisco, San Francisco, California, United States of America, **4** Genetic Epidemiology and Genomics Laboratory, University of California, Berkeley, Berkeley, California, United States of America, **5** Center for Genetics, Children's Hospital Oakland Research Institute, Oakland, California, United States of America, **6** Cancer Immunogenetics, University of Manchester, St. Mary's Hospital, Manchester, United Kingdom, **7** Department of Pediatrics, Ehime Prefectural Center Hospital, Ehime, Japan, **8** Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon, France, **9** Department of Pediatrics, Kaiser Permanente, Roseville, California, United States of America, **10** Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **11** Jonathan Jaques Children's Cancer Center, Miller Children's Hospital, Long Beach, California, United States of America, **12** Handforth, Cheshire, United Kingdom

Abstract

The extended major histocompatibility complex (xMHC) is the most gene-dense region of the genome and harbors a disproportionately large number of genes involved in immune function. The postulated role of infection in the causation of childhood B-cell precursor acute lymphoblastic leukemia (BCP-ALL) suggests that the xMHC may make an important contribution to the risk of this disease. We conducted association mapping across an approximately 4 megabase region of the xMHC using a validated panel of single nucleotide polymorphisms (SNPs) in childhood BCP-ALL cases (n=567) enrolled in the Northern California Childhood Leukemia Study (NCCLS) compared with population controls (n=892). Logistic regression analyses of 1,145 SNPs, adjusted for age, sex, and Hispanic ethnicity indicated potential associations between several SNPs and childhood BCP-ALL. After accounting for multiple comparisons, one of these included a statistically significant increased risk associated with rs9296068 (OR=1.40, 95% CI=1.19-1.66, corrected p=0.036), located in proximity to *HLA-DOA*. Sliding window haplotype analysis identified an additional locus located in the extended class I region in proximity to *TRIM27* tagged by a haplotype comprising rs1237485, rs3118361, and rs2032502 (corrected global p=0.046). Our findings suggest that susceptibility to childhood BCP-ALL is influenced by genetic variation within the xMHC and indicate at least two important regions for future evaluation.

Citation: Urayama KY, Chokkalingam AP, Metayer C, Hansen H, May S, et al. (2013) SNP Association Mapping across the Extended Major Histocompatibility Complex and Risk of B-Cell Precursor Acute Lymphoblastic Leukemia in Children. PLoS ONE 8(8): e72557. doi:10.1371/journal.pone.0072557

Editor: Matthaïos Speletas, University of Thessaly, Faculty of Medicine, Greece

Received: February 26, 2013; **Accepted:** July 12, 2013; **Published:** August 22, 2013

Copyright: © 2013 Urayama et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the US National Institute of Environmental Health Sciences (R01ES09137, P42ES0470518), the National Cancer Institute (R03CA125823), and the Children with Cancer Foundation (2005/027, 2005/028, 2006/051, 2006/052), United Kingdom. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: kurayama@luke.or.jp

Introduction

Acute lymphoblastic leukemia (ALL) is a clonal disorder involving the dysregulated proliferation of genetically altered lymphoid progenitor cells that lack the ability for differentiation and maturation. In children, B-cell precursor (BCP) ALL is the most common ALL subtype and accounts for about 80% of

childhood ALL cases in most economically developed countries. BCP-ALL, which demonstrates a unique age-incidence peak between 2 and 5 years of age, is widely suspected to be caused by environmental exposures, though these have yet to be definitively identified [1]. Foremost among these are thought to be factors such as the effect of timing of exposure to infectious agents leading to inappropriate immune

responses, in conjunction with variants in genes of the immunological pathway and early lymphoid development [2].

The extended major histocompatibility complex (xMHC) region, spanning about 7.6 megabases (Mb) on the short arm of chromosome 6 (6p21.3), is densely populated with genes that are critical to both innate and adaptive immunity in humans [3]. The xMHC is divided into five sub-regions consisting of the extended class I region at the telomeric end, and successively the classical class I, III, and II clusters bounded by an extended class II region at the centromeric end. Historically, genetic association studies of the xMHC have focused on the classical human leukocyte antigen (HLA) genes of the class I (*HLA-A*, *B*, and *C*) and class II (*HLA-DP*, *DQ*, and *DR*) regions. These encode cell surface glycoproteins that selectively bind and present, in allele selective fashion, processed antigenic peptides to T lymphocytes that initiate T-cell responses. While *HLA* genes are among the most polymorphic in humans, they account for only a small proportion of over 250 expressed xMHC loci, which include genes encoding cytokines, complement factors, and various others involved in critical cellular processes.

Evidence of susceptibility associated with xMHC loci has been identified in several autoimmune, malignant and infectious diseases, including asthma, Hodgkin and non-Hodgkin lymphoma, hepatitis B and HIV infection and others [4,5,6,7]. Despite evidence of linkage between predisposition to retrovirus-induced leukemia and the murine MHC, attempts to identify associations between childhood ALL and classical HLA alleles have been inconsistent largely due to study design limitations [8,9,10,11,12]. Although these have been largely overcome by the application of high-resolution HLA molecular genotyping to carefully ascertained case-control series, strong and replicable associations have yet to emerge.

A previous analysis of MHC SNP data and imputed HLA class I and II alleles derived from a childhood ALL genome-wide association study (GWAS) suggested that MHC genetic variation is unlikely to be a major determinant of BCP-ALL [13]. Nonetheless, a modest SNP association noted in the HLA class II region, together with strong positive findings from the largest study to date utilizing directly typed HLA genotypes [9], suggested that further examination of the xMHC in childhood ALL was warranted in a well-defined case-control series. We report here the results of SNP association mapping across a 4 Mb stretch of the xMHC spanning all major class I, II, and III loci using a validated SNP panel in a large sample of non-Hispanic white and Hispanic BCP-ALL cases ($n=567$) and controls ($n=892$) enrolled in the Northern California Childhood Leukemia Study (NCCLS).

Materials and Methods

Ethics Statement

The study protocol was approved by the Institutional Review Boards of the University of California, Berkeley and all collaborating institutions (California Department of Public Health, University of California, Davis, University of California, San Francisco, Children's Hospital of Central California, Lucile Packard Children's Hospital, Children's Hospital and Research

Center, Oakland, Kaiser Permanente, Roseville, Kaiser Permanente, Santa Clara, Kaiser Permanente, San Francisco, Kaiser Permanente, Oakland), and written informed consent was obtained from the parents or guardians on behalf of the children participants involved in this study. This study was conducted in accordance with the Declaration of Helsinki.

Study participants

The current study was conducted within the NCCLS, an ongoing case-control study of childhood leukemia. Beginning in 1995, newly diagnosed childhood leukemia cases were ascertained at the time of diagnosis from major pediatric hospitals in a 17-county San Francisco Bay Area study region, expanded in 1999 to 35 counties in Northern and Central California, USA. Comparison with the California Cancer Registry (1997-2003) showed that the NCCLS case ascertainment protocol has captured about 95% of children diagnosed with leukemia in the participating study hospitals. For each eligible case, statewide birth records maintained by the California Office of Vital Records were utilized to generate a list of randomly selected controls that matched the case on child's date of birth, sex, Hispanic ethnicity (a biological parent who is Hispanic), and maternal race. Information obtained through the birth certificates and commercially available searching tools were used to trace and enroll one or two matched controls for each case.

Cases and controls were considered eligible if they were under 15 years of age at date of diagnosis for cases (or corresponding reference date for controls), residents of the study region, had a biological parent who spoke either English or Spanish, and had no prior history of malignancy. Approximately 85% of eligible cases and 86% of eligible controls consented to participate [14]. A detailed description of control selection in the NCCLS is reported elsewhere [14,15].

In the current study, non-Hispanic white and Hispanic children with ALL and control children, recruited between 1995 and 2008 (study phases 1-3), were included in the analysis. These are the two largest racial/ethnic groups which together comprise about 85% of enrolled subjects. Other ethnic groups were excluded due to the small number of subjects. Children were classified as Hispanic if at least one biological parent self-identified as Hispanic. Children were assigned to the non-Hispanic white group if both biological parents self-identified as non-Hispanic white. In a previous NCCLS analysis, genetic admixture was assessed using a series of 80 ancestry informative markers for a subset of the cases and controls [16], and estimates of genetic ancestry (percent of European, Amerindian, and African ancestry) were determined. Comparison of these estimates between cases and matched controls showed no significant differences [16]. For the current study, a total of six hundred and eighty-eight ALL cases and 1,012 controls were considered, of which 635 cases (92.3%) and 915 controls (90.4%) had a DNA sample available for genotyping.

DNA samples

Buccal cells as a source of DNA were obtained from case and control children using cytobrushes by trained interviewers.

Cytobrushes were processed within 48 hours of collection by heating in the presence of 0.5N NaOH. Isolated DNA was later re-purified either manually using Genra Puregene reagents (QIAGEN, USA, Valencia, CA) or an automated organic DNA extraction protocol (AutoGen, Holliston, MA). Whole genome amplification (WGA) was performed using GenomePlex reagents (Rubicon Genomics, Ann Arbor, MI) according to the manufacturer's protocol. WGA products were cleaned with a Montage PCR9 filter plate (Millipore, Billerica, MA). When buccal cytobrush DNA was inadequate or not available (26.6% of subjects), DNA was isolated from dried bloodspots collected at birth and archived at -20°C by the Genetic Disease Screening Program of the California Department of Public Health. After extraction using the QIAamp DNA Mini Kit (QIAGEN, USA, Valencia, CA), DNA samples were whole-genome amplified using REPLI-g reagents (QIAGEN, USA, Valencia, CA). Regardless of source, DNA specimens were quantified using human-specific Alu-PCR to confirm a minimum level of amplifiable human DNA [17].

MHC genotyping

Genotyping was conducted using the Illumina MHC Mapping Panel (Illumina Inc., San Diego, CA) which comprises 1,293 SNPs spanning an approximately 4 Mb region of the xMHC bounded by the tripartite motif containing protein 27 (*TRIM27*) and motilin (*MLN*) genes at the telomeric and centromeric ends, respectively (NCBI Build 36). There is an average 3.8 kilobase (kb) spacing between each SNP, covering all major regions of the xMHC, including the classical class I, II and III regions, the extended class II region and part of the extended class I region. The panel set was designed with a strong emphasis on haplotype tagging SNPs which are highly informative of SNPs in strong linkage disequilibrium (LD). The chance of detecting an association is significantly influenced by the ability of a SNP or combination of SNPs to adequately represent the haplotypic diversity of the region. Genotyping was performed utilizing the robust Golden-Gate technology in a 96-well format on a 1,536 Sentrix Array Matrix [18]. It was shown previously in NCCLS subjects that when analyzed using Golden-Gate genotyping, buccal cell WGA DNA yielded genotypes that are highly concordant with those from genomic DNA from peripheral blood [19].

Genotyping was conducted on 1,550 unique DNA samples (635 cases and 915 controls), in addition to 10 sets of *Centre d'Etude du Polymorphisme Humain* (CEPH) family trios and duplicates of 10% of study samples. For quality control purposes, 113 SNPs that successfully genotyped in less than 90% of samples were excluded, as well as 30 SNPs that deviated from Hardy Weinberg equilibrium ($p < 0.01$) in both non-Hispanic white and Hispanic controls, and 5 additional SNPs with a minor allele frequency of less than 0.01 in both non-Hispanic white and Hispanic controls.

Quality control metrics applied to the 1,550 samples also resulted in the exclusion of 17 samples (1.1%) with less than 95% overall genotyping success rate and 20 samples (1.3%) that showed questionable concordance between reported gender and gender prediction by the Illumina platform. There was 99.6% concordance of successfully genotyped SNPs in

Table 1. Characteristics of childhood BCP-ALL cases and controls.

	Cases		Controls	
	n	%	n	%
Total	567	100	892	100
Sex				
Male	298	52.6	495	55.5
Female	269	47.4	397	44.5
Age (years)				
0-1	51	9.0	111	12.4
2-5	340	60.0	460	51.6
6-10	114	20.1	210	23.5
11-14	62	10.9	111	12.4
Self-reported race/ethnicity				
White, non-Hispanic	241	42.5	426	47.8
Hispanic	326	57.5	466	52.2
BCP-ALL subtype^a				
cALL ^b	309	49.8	NA	NA
Non-cALL	258	41.5	NA	NA
High-hyperdiploid	178	28.7	NA	NA
<i>TEL-AML1</i>	96	15.5	NA	NA
Normal karyotype	58	9.3	NA	NA

Abbreviations: BCP-ALL, B-cell precursor acute lymphoblastic leukemia; cALL, common ALL; NA, not applicable

^a Categories of ALL subtypes are not mutually exclusive. Percentages do not equal to 100. Cytogenetic data were available for 87% of patients.

^b cALL is defined as CD10+ and CD19+ ALL diagnosed between age 2-5 years.

the duplicate series and a 0.2% Mendelian error rate was observed in the CEPH family trios. Application of these quality control criteria and a focus on BCP-ALL (54 T-cell or mixed lineage ALL cases excluded) resulted in the analysis of 1,145 SNPs in a total of 567 BCP-ALL cases and 892 controls (Table 1). Data are available on request in accordance with the policies and procedures of the NCCLS.

For a large subset of the BCP-ALL cases (87%), data on hyperdiploidy and *TEL-AML1* chromosomal translocation were available as described in detail previously [20]. Subtypes of the cases included 309 common ALL (cALL, defined as CD10+ and CD19+ ALL aged 2 to 5 years), 178 high-hyperdiploid (51-67 chromosomes), 96 positive for the *TEL-AML1* chromosomal translocation, and 58 BCP-ALL with normal karyotypes. These non-mutually exclusive subtype groupings were used to examine potential subtype-specific effects for the final set of associated SNPs.

Statistical analysis

Data analysis included a two-stage approach. First, we examined the contribution of 1,145 xMHC SNPs individually. We used logistic regression to calculate the odds ratio (OR) and 95% confidence intervals (CI) for each SNP adjusting for child's age, sex, and race/ethnicity (i.e. non-Hispanic white or Hispanic). Various genetic models of inheritance were considered including log-additive, dominant, and recessive models, in addition to an evaluation of the dominance deviation

from additivity. SNPs showing a nominal p-value of less than 0.01 in any of these analyses were considered potentially associated with childhood ALL and were subject to further analysis. Multivariable logistic regression was used to evaluate the independence of effect of multiple potentially associated SNPs within a region on childhood ALL risk. Stratified analyses by age (0-5 and 6-14 years) and gender were considered in sub-analyses of the data. To account for multiple comparisons in the presence of LD between SNPs, we calculated adjusted p-values based on 10,000 permutations of case-control status on 1,145 SNPs and considered adjusted p-values below a family-wise type I error rate threshold of 0.05 to be statistically significant.

Second, we conducted three-SNP sliding window haplotype analyses across candidate regions selected by the location of SNPs that showed nominal p-values of less than 0.01. This resulted in 3 broad regions (Figure 1) for the haplotype analysis: 1) a region bound by rs381808 and rs3117330 referred to as region A (~364 kb region, extended class I); 2) a region bound by rs1264419 and rs3828886 referred to as region B (~864 kb region, class I-class III); and 3) a region bound by rs516535 and rs210134 referred to as region C (~598 kb region, class II-extended class II). In total, the haplotype sliding window analysis was performed on 395 genotyped SNPs (393 3-SNP haplotypes) located within these 3 regions. Haplotypes were predicted and reconstructed for each individual and frequencies estimated using the expectation-maximization algorithm based on unphased genotype data. Haplotypes with a frequency of less than 0.01 were grouped into one category. For each three-SNP sliding window, a global likelihood ratio test of association was performed to test the null hypothesis of no effect of any haplotype at that position. The permutation method was used to adjust for multiple comparisons for the 393 haplotype windows tested. Haplotype-specific effects were evaluated by modeling individual haplotype probabilities in a logistic regression assuming a log-additive effect of the haplotype and adjusting for age, sex, and race/ethnicity. Analyses were conducted using PLINK, UNPHASED version 3.1.4, and Haploview [21,22,23].

Results

Case and control distributions for sex, age, and race/ethnicity were similar, as expected from the matched design of the NCCLS (Table 1). Hispanics comprised about 58% of cases and 52% of controls.

Analysis of 1,145 SNPs in childhood BCP-ALL, assuming a log-additive genetic model of inheritance, showed a quantile-quantile (Q-Q) plot of the expected versus observed $-\log_{10}$ p-value distribution that suggested little evidence of inflation in results caused by systematic error (Figure S1). Twenty SNPs were associated with a nominal p-value of less than 0.01 (log-additive, genotypic, dominant, and/or recessive genetic models), many of which were in LD (Figure 1 and Figure S2). These SNPs (Figure 1) appeared to cluster within three specific regions (designated A, B, and C) which served as the focus of the haplotype analysis. SNPs showing the strongest evidence of association based on p-value within each of the three

regions included rs7747023 (OR=0.73, 95% CI=0.60-0.89, $p=1.7 \times 10^{-3}$) of region A, rs3130785 (OR=1.45, 95% CI=1.16-1.82, $p=1.3 \times 10^{-3}$) of region B, and rs9296068 (OR=1.37, 95% CI=1.17-1.61, $p=1.2 \times 10^{-4}$) of region C. Multivariable analyses evaluating the independence of associations between the 20 SNPs on BCP-ALL risk (Table S1) resulted in 6 SNPs that maintained low p-values and minimally attenuated risk estimates (Table 2). Stratified analysis showed no remarkable gender- or age-specific associations (Figure S3). The final multivariable model including all 6 SNPs and correcting for multiple comparisons showed a statistically significant association between rs9296068 and BCP-ALL risk (OR=1.40, 95% CI=1.19-1.66, corrected $p=0.036$).

Further examination of SNP rs9296068 (Figure 2) showed little evidence of heterogeneity in effect between non-Hispanic white and Hispanic children ($p=0.503$), males and females ($p=0.356$), and children aged zero to five and aged six to fourteen ($p=0.906$). The risk estimates appeared consistently elevated for the two main cytogenetic subtypes (i.e. *TEL-AML1*-positive and high hyperdiploidy), but not for normal karyotype BCP-ALL (Figure 2, OR=1.26, 95% CI=0.83-1.91).

We performed a 3-SNP sliding window haplotype analysis that included 395 genotyped SNPs across each of the 3 candidate regions (regions A, B, and C) identified by the single SNP analysis. After adjusting for multiple comparisons, a statistically significant association was found for the rs1237485-rs3118361-rs2032502 haplotype (nominal global $p=3.2 \times 10^{-4}$; corrected $p=0.046$) in region A (Table 3 and Figure 3A). Specifically, haplotype G-A-G was associated with an increased risk of BCP-ALL compared to other haplotypes combined (OR=2.18, 95% CI=1.41-3.38) (Table 3). In multivariable analysis adjusting for the nearby region A SNP, rs7747023 (described above in the individual SNP analysis), evidence of association for this haplotype remained strong (nominal global $p=2.5 \times 10^{-3}$), while the effect for rs7747023 appeared to be attenuated. When rs7747023 was included in the haplotype, the global test for the 4-SNP haplotype (rs1237485-rs3118361-rs2032502-rs7747023) yielded stronger evidence of an association (nominal global $p=2.7 \times 10^{-5}$). Stratified analyses by race/ethnicity showed similar results in non-Hispanic white and Hispanic children.

Finally, haplotype rs423639-rs7754316-rs9296068 (Table 3 and Figure 3B) of region C was found to be statistically significant (nominal global $p=9.2 \times 10^{-5}$; corrected $p=0.014$), a locus also identified through the individual SNP analysis of rs9296068. Previously, we reported an association between childhood ALL risk and the *DP1* supertype of *HLA-DPB1* (comprised of *HLA-DPB1* alleles 01:01, 05:01, 50:01) [12], a class II gene located about 55 kb from the associated rs9296068 SNP. The two loci showed weak correlation ($r^2 < 0.1$) and the analysis adjusting for carriers of the *DP1* supertype indicated an independent effect for rs9296068 (OR=1.37, 95% CI=1.16-1.63, $p=3.0 \times 10^{-4}$).

Discussion

In this study, we conducted a SNP-association analysis of childhood BCP-ALL compared with controls across a 4 Mb

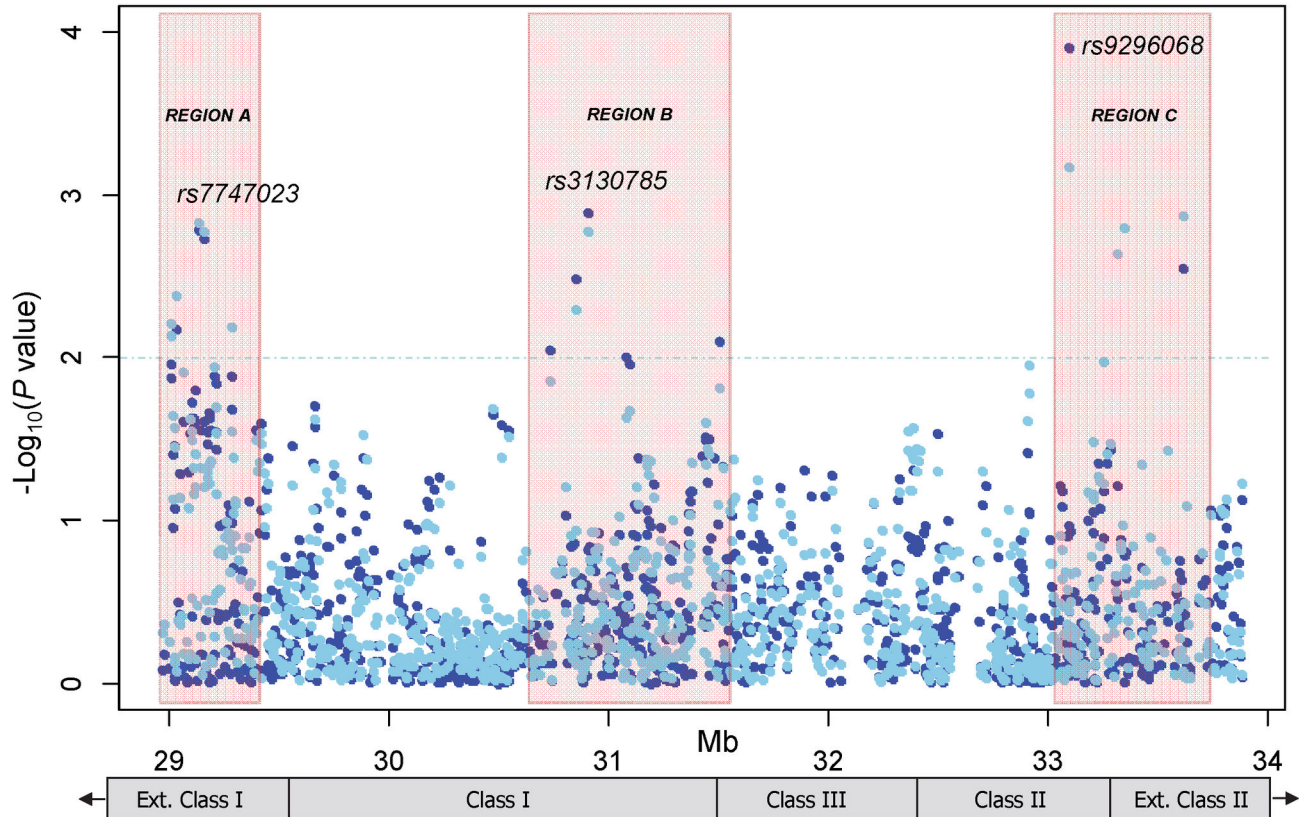


Figure 1. Analysis of 1,145 SNPs across a 4 Mb region of the extended major histocompatibility complex and risk of childhood BCP-ALL. Presented are $-\log_{10}(p\text{-values})$ resulting from the logistic regression analysis assuming log-additive (navy blue) and dominant (light blue) genetic models of inheritance and adjusting for child's age, sex, and race/ethnicity. Results plotted above the dotted line represent nominal p-values of less than 0.01. Analyses evaluating the recessive and genotypic genetic models were also performed (not plotted) resulting in five additional SNPs with a nominal p-value of less than 0.01 which were also located within one of the three designated regions (Regions A-C). A total of 20 SNPs with a p-value below this threshold were considered in further analyses.

doi: 10.1371/journal.pone.0072557.g001

stretch of the xMHC in an attempt to pinpoint regions of potential involvement in susceptibility. The xMHC is a potentially strong candidate region for a role in genetic susceptibility to childhood ALL, a disease whose causation has been attributed to an inappropriate immune response to post-natal infection [2,24,25]. Using a validated panel of greater than 1,100 SNPs designed to capture the genetic diversity of this complex genomic region, we identified two loci associated with childhood BCP-ALL risk. After correction for multiple testing, we found a statistically significant increased risk associated with the minor allele of rs9296068 in proximity to the *HLA-DOA* gene. A second independently associated locus, represented by haplotypes comprised of SNPs rs1237485, rs3118361, and rs2032502, was identified using a haplotype sliding window analysis and is located in the extended class I region in proximity to the *TRIM27* gene.

The rs9296068 SNP is located in the 5' untranslated region of the *HLA-DOA* gene about 11.3 kb from the first exon and resides within a region that is predicted to have promoter

function [26,27]. *HLA-DOA* encodes the alpha subunit of the HLA-DO heterodimer and is selectively expressed in B-cells and thymic medullary epithelium [28]. HLA-DO interacts with HLA-DM to regulate peptide loading onto MHC class II molecules in a pH-dependent manner. While HLA-DM facilitates peptide binding by catalyzing the exchange between low and high affinity peptides, HLA-DO impedes this function by reducing class II-mediated presentation in general, and has the ability to skew the presented antigenic peptide repertoire in B cells [29]. Thus, a balanced expression between HLA-DM and DO is critical in controlling antigen presentation in B-cells.

Recently, HLA-DOA has been implicated in other disease association studies such as type 1 diabetes and chronic lymphocytic leukemia survival [30,31], and interestingly, the same rs9296068 SNP was reported in a multistage MHC association mapping study of pediatric liver transplant rejection [26]. Functional validation showed a nearly 3-fold higher intra-graft B-lymphocyte content in rejecting liver grafts among carriers of the risk allele compared to non-carrier rejecters. This

Table 2. Association between xMHC genetic variants and BCP-ALL in children.

SNP	Position	Region of xMHC	Minor allele	Frequency ^a		Single SNP			Mutually adjusted			
				Cases	Controls	OR	95% CI ^b	p-value	OR	95% CI ^b	p-value	p-value (FWE) ^c
rs7747023	29133659	Extended class I	G	0.17	0.21	0.73	(0.60-0.89)	1.7x10 ⁻³	0.72	(0.59-0.88)	1.4x10 ⁻³	0.518
rs3130785	30904717	Class I	A	0.14	0.11	1.45	(1.16-1.82)	1.3x10 ⁻³	1.37	(1.09-1.74)	7.9x10 ⁻³	0.973
rs1632856	31079715	Class I	A	0.25	0.29	0.80	(0.68-0.95)	9.9x10 ⁻³	0.79	(0.66-0.94)	8.6x10 ⁻³	0.898
rs2524279	31500885	Class I	G	0.11	0.15	0.73	(0.58-0.92)	7.9x10 ⁻³	0.70	(0.55-0.89)	3.0x10 ⁻³	0.749
rs9296068	33096673	Class II	C	0.42	0.36	1.37	(1.17-1.61)	1.2x10 ⁻⁴	1.40	(1.19-1.66)	5.7x10 ⁻⁵	0.036
rs213203 ^d	33346382	Extended class II	A	0.47	0.49	0.68	(0.55-0.84)	3.6x10 ⁻⁴	0.69	(0.55-0.86)	7.4x10 ⁻⁴	0.347

Abbreviations: CI, confidence interval; FWE, family-wise type I error; OR, odds ratio; SNP, single nucleotide polymorphism; xMHC, extended major histocompatibility complex

^a Frequency of the minor allele in case and control subjects

^b ORs and 95% CI for each SNP in the single SNP analysis were derived using logistic regression assuming a log-additive genetic model of inheritance and adjusting for child's age, sex, and race/ethnicity (non-Hispanic white versus Hispanic). The mutually adjusted analysis included additional adjustment for the effects of all other SNPs in the table.

^c Adjustment for multiple testing was performed with 10,000 permutations of case-control status on 1,145 SNPs and using FWE rate of 0.05. An adjusted p-value of less than 0.05 was considered statistically significant.

^d Evaluation of the genetic model of inheritance indicated a significant deviation from the log-additive model with an effect associated with heterozygotes. ORs and 95% CI were estimated for heterozygous genotypes compared to homozygous genotypes.

SNP was also associated in a recent GWAS of rheumatoid arthritis, but its independence from the known *HLA-DRB1* effect was not described [32]. Further biological relevance of this SNP locus is supported by publicly available expression quantitative trait loci data, with one source indicating an rs9296068 allelic-dependent association with *HLA-DOA* gene expression in lymphoblastoid cell lines [26,33], and another showing associations with *HLA-DPB1* gene expression in purified B-cells and monocytes [34].

Previous examinations of the xMHC in childhood ALL have mostly been candidate gene studies that focused on the classical HLA genes (i.e. *HLA-A*, *-B*, and *-C*, and *HLA-DP*, *-DQ*, and *-DR*). The most consistent evidence of an association has been for HLA class II loci, including *HLA-DPB1* [10,12] and *HLA-DR* [8,35], genes relatively close in proximity to rs9296068 and *HLA-DOA*. However, due to the lack of genetic characterization of the surrounding regions in these studies, it could not be unambiguously determined whether those associations indicated a causal link with the HLA gene or whether the associations were an effect of LD with an adjacent causal locus. The availability of directly genotyped *HLA-DPB1* data allowed us to confirm that the rs9296068 association of the class II region is independent of the *HLA-DPB1* allelic associations previously reported in the CCLS and elsewhere [10,12]. We were unable to confirm this for *HLA-DQ* and *HLA-DR*, but the presence of major recombination activity and the weak correlation between rs9296068 and SNPs immediately upstream make it unlikely that the association originates from the DQ/DR loci [36].

Authors of a recent report using data extracted from a prior GWAS analysis concluded no substantive support for a major role of MHC genetic variation on childhood BCP-ALL risk [13]. Among the results, the strongest single SNP association was observed for rs3135034 which exhibited a nominal p-value of 0.0017, but was not statistically significant after correction for multiple testing. SNP rs3135034 is located about 20 kb

downstream of *HLA-DOA*, and only 37 kb from rs9296068 in an intergenic region also in proximity to the bromodomain containing 2 (*BRD2*) gene. In our data, rs3135034 is weakly correlated ($r^2 < 0.1$) with rs9296068 in both non-Hispanic whites and Hispanics and showed no association with childhood BCP-ALL. However, rs9296068 and rs3135034 flank a well-characterized strong meiotic recombination hotspot, *DNA3* [37], and it has been noted that etiological variants within a recombination hot spot may be impossible to identify using standard association strategies [38]. This suggests that the identification of these two SNPs in close physical proximity in two independent studies of MHC association with childhood ALL could indicate a true association with *HLA-DOA*, partially masked by *DNA3*.

Using a haplotype sliding window analysis, we identified a second independently associated locus which is localized to the extended class I region and is represented by a haplotype comprised of SNPs rs1237485, rs3118361, and rs2032502. This haplotype maps to the 5' untranslated region of the *TRIM27* gene about 2.5 kb from the coding region and is greater than 1Mb from the nearest classical HLA locus (*HLA-A*). *TRIM27* (also known as Ret finger protein, RFP) belongs to an expanding family of proteins that are characterized by a tripartite motif comprising Really Interesting New Gene (RING) and B-box zinc-binding domains, and a variable coiled-coil region [39]. They participate in a variety of critical biological processes including cell growth, tumor suppression, DNA damage signaling, senescence, apoptosis, stem cell differentiation, and immune response to infections. Recent evidence demonstrated that *TRIM27* (and other *TRIM* proteins) may contribute to a repertoire of pathways through its function as both a small ubiquitin-like modifier (SUMO) protein and ubiquitin E3 ligase important in post-translational modification [40,41]. The p53 tumor suppressor and its principal antagonist murine double minute 2 (*Mdm2*) oncogene are among the several substrates of *TRIM27* SUMO and ubiquitin E3 ligase

MHC Class II rs9296068 (A>C)	Ca	Co	OR	95%CI
B-cell ALL: log-additive	565	888	1.40	1.19-1.65
Genetic Models				
C/A (vs A/A)	306	445	1.48	1.16-1.89
C/C (vs A/A)	88	97	1.91	1.34-2.72
Dominant	394	542	1.55	1.24-1.94
Recessive	88	97	1.51	1.10-2.07
By Race/Ethnicity (P homogeneity=0.503)				
Non-Hispanic white	240	424	1.34	1.04-1.73
Hispanic	325	464	1.48	1.18-1.86
By Sex (P homogeneity=0.356)				
Male	296	493	1.31	1.05-1.63
Female	269	395	1.50	1.18-1.91
By Age (P homogeneity=0.906)				
0 - 5 years	389	569	1.44	1.17-1.77
6 - 14 years	176	319	1.36	1.02-1.81
Major B-cell ALL subtypes				
cALL (CD10+/CD19+, 2-5 yrs)	308	888	1.50	1.22-1.84
Non-cALL	257	888	1.35	1.08-1.69
Hyperdiploid	178	888	1.57	1.22-2.02
TEL-AML1	95	888	1.74	1.25-2.42
Normal karyotype	58	888	1.26	0.83-1.91

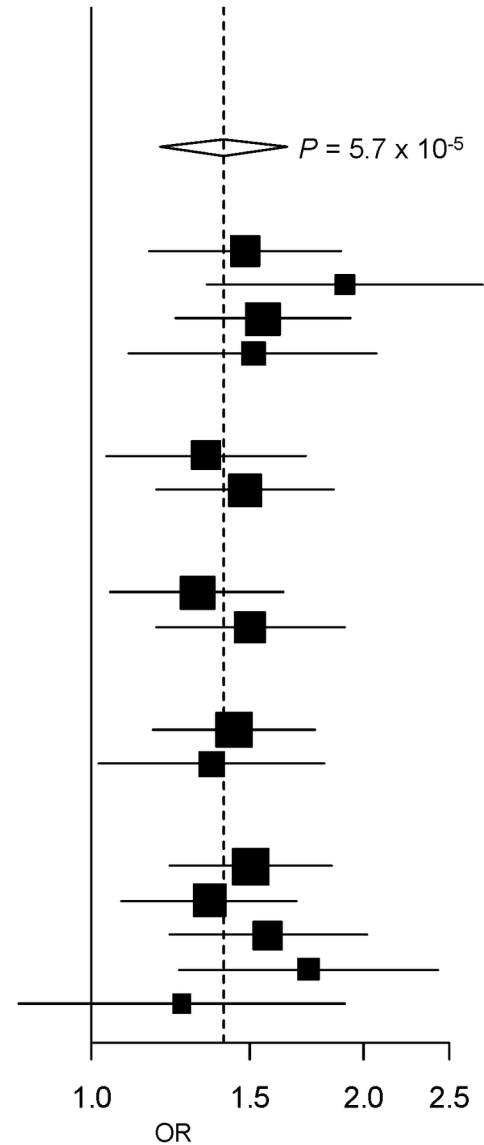


Figure 2. Stratified analysis of childhood BCP-ALL and the SNP rs9296068 by race/ethnicity, sex, and age group, and subgroup analyses by major subtypes. Odds ratios (ORs, represented by boxes with the area of each box inversely proportional to the variance of the estimate) and 95% CI (error bars) were derived using logistic regression assuming a log-additive genetic model and adjusting for rs7747023, rs3130785, rs1632856, rs2524279, and rs213203 (other potentially associated SNPs presented in Table 1) and additionally for child's age, sex, and race/ethnicity based on the stratification variable. The dashed vertical line represents the OR of the SNP in the analysis of BCP-ALL among all subjects and the width of the diamond is the corresponding 95% CI. $P_{\text{homogeneity}}$ was on the basis of the Cochran's Q test statistic. Abbreviations: Ca, number of case; cALL, common acute lymphoblastic leukemia; Co, number of controls.

doi: 10.1371/journal.pone.0072557.g002

activity. With respect to immune regulation, TRIM27 is thought to down-regulate the immune response at multiple levels, including inhibition of toll-like receptor activation of nuclear factor-kappaB (NF- κ B) and interferon regulatory factor 3 (IRF3) [42], and the ability to negatively regulate CD4+ T cells through inhibitory effects on KCa3.1 (calcium-activated potassium channel) protein activity [43,44].

We did not impute genotypes for additional xMHC SNP loci or classical HLA alleles because certain features of the current study made it suboptimal for implementation of imputation including, 1) uncertainty in the use of currently available reference panels for imputation in a recently admixed population [45,46], 2) a focus on the xMHC, a region of complex LD that shows varying degrees of heterogeneity even

Table 3. Results for two associated regions based on a 3-SNP haplotype sliding window analysis of BCP-ALL cases and controls.

Haplotype	Frequency		Compare to reference haplotype		Compared to all other haplotypes			
	Cases (%)	Controls (%)	OR	95% CI ^a	OR	95% CI ^a	p-value	p-value _(FWE) ^b
Region A								
rs1237485-rs3118361-rs2032502 ^c								
A-A-G	0.06	0.06	1.03	(0.72-1.48)	1.16	(0.82-1.62)	0.406	
A-G-A	0.10	0.13	0.84	(0.65-1.08)	0.74	(0.58-0.94)	0.015	
A-G-G	0.37	0.33	1.22	(1.02-1.45)	1.14	(0.97-1.34)	0.109	
G-A-G	0.06	0.03	2.44	(1.54-3.88)	2.18	(1.41-3.38)	2.9 x10 ⁻⁴	
G-G-G	0.41	0.45	1.00	Ref	0.89	(0.76-1.04)	0.136	
						Global p-value ^d	3.2x10 ⁻⁴	0.046
Region C								
rs423639-rs7754316-rs9296068 ^c								
A-A-A	0.07	0.08	0.82	(0.61-1.11)	0.93	(0.69-1.23)	0.576	
G-A-A	0.27	0.32	0.77	(0.63-0.93)	0.80	(0.67-0.95)	9.1x10 ⁻³	
G-A-C	0.39	0.34	1.00	Ref	1.25	(1.07-1.46)	6.0x10 ⁻³	
G-G-A	0.24	0.25	0.82	(0.67-1.00)	0.91	(0.76-1.10)	0.331	
G-G-C	0.04	0.02	1.98	(1.08-3.63)	2.47	(1.38-4.43)	1.7x10 ⁻³	
						Global p-value ^d	9.2 x10 ⁻⁵	0.014

Abbreviations: CI, confidence interval; FWE, family-wise type I error; OR, odds ratio; SNP, single nucleotide polymorphism

^a ORs and 95% CI for each haplotype were derived using logistic regression modeling haplotype probabilities and adjusting for child's age, sex, and race/ethnicity (non-Hispanic white versus Hispanic).

^b Adjustment for multiple testing was performed with 10,000 permutations of case-control status on 393 haplotype windows and using FWE rate of 0.05. An adjusted p-value of less than 0.05 was considered statistically significant.

^c The associated region A haplotype is located in the extended class I region and the SNPs are at chromosomal positions 29,002,323 (rs1237485), 29,006,266 (rs3118361), and 29,009,544 (rs2032502). The associated region C haplotype is located in the class II region and the SNPs are chromosomal positions 33,095,752 (rs423639), 33,095,976 (rs7754316), and 33,096,673 (rs9296068).

^d Global p-values were derived based on a likelihood ratio test of association based on the null hypothesis of no effect of any haplotype at that position.

across sub-strata of individuals of European descent [47,48], and 3) a sample size comparably large for a study of a rare disease, but not statistically robust to a multiple comparisons burden that would be elevated by close to 10-fold with the additional loci. Thus, it is possible that associations were missed due to limited SNP coverage in certain regions.

The associations reported from these analyses were not identified in the previous GWAS [49,50,51,52,53]. Notably, our study was performed on a sample size comparable to that of the previous GWAS of childhood ALL, but with a substantially smaller multiple comparisons burden on statistical power due to its focus only on the xMHC. However, we acknowledge that statistical power may have been affected by combining non-Hispanic white and Hispanic children for the analysis. The success of the association mapping approach is highly dependent on the degree of LD between the genotyped SNP and the causal locus. A loss in precision would be expected if this LD between the SNP and causal locus differed across populations included in the analysis [54]. While described as a limitation, an advantage of our approach is that the detectable associations would likely only be those that showed a relationship in both race/ethnicity populations, which may add to the confidence in results. Accordingly, the associations reported in the current analysis showed consistency between non-Hispanic white and Hispanic children in stratified analyses.

Certain characteristics of the association mapping approach, namely the dependence on SNP coverage and the effect of multiple comparisons on statistical power, may have contributed to inconsistencies between results of the current study and associations reported in previous studies based on the candidate gene approach. As reviewed previously [55], the two approaches should be viewed as complementary strategies for identification of disease associated loci as they both have their respective strengths and weaknesses depending on the study being conducted. A review of the literature identified six xMHC childhood ALL candidate gene studies of non-classical HLA loci, and statistically significant associations have been reported for SNPs of the *HFE*, *HSPA1B* and *BAT3* genes [56,57,58]. None of the SNPs specifically examined in these studies were genotyped as part of the current mapping panel which precluded our ability to directly evaluate these previous associations. Indirect assessment of previously associated *BAT3* SNPs was possible through identification of proxy SNPs ($r^2 > 0.8$ in HapMap CEU) using publicly available resources [59], but evidence of an association was not observed in our study.

Any substantial effect of population stratification is likely to be minimal in the NCCLS due to the careful and detailed account of race and ethnicity obtained from the subjects and statistical adjustment. As described earlier, this is further supported by our previous report showing estimates of genetic

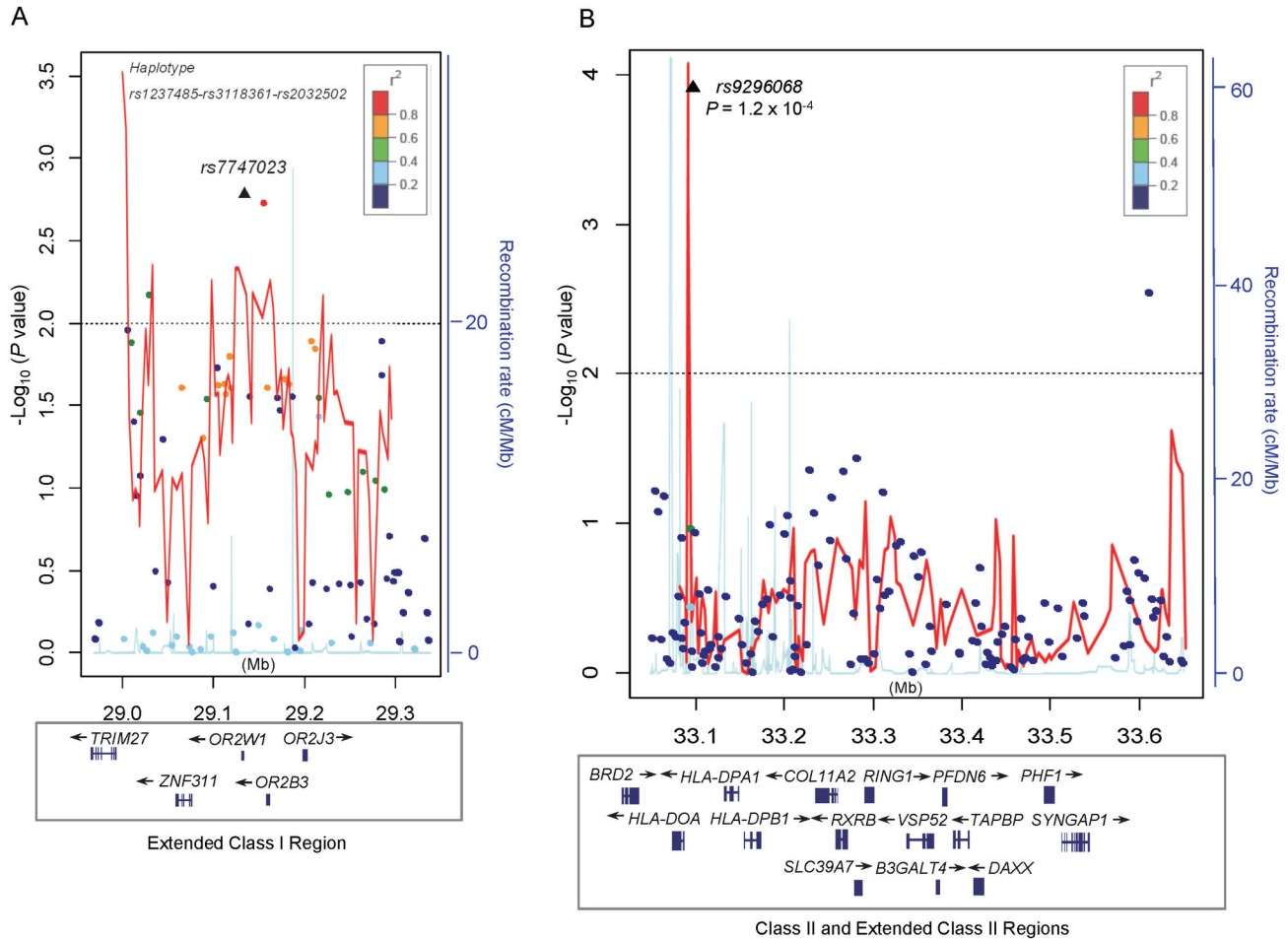


Figure 3. Plots of the two associated loci showing the results for the analysis of childhood BCP-ALL and individual SNPs (points) and three-SNP sliding window haplotypes (red lines). The $-\log_{10}$ (p-values) for each SNP (y-axis) are plotted against their chromosomal position (x-axis, Mb). The colors of the points indicate the degree of linkage disequilibrium (based on r^2) in relation to the index SNP (indicated by a black triangle). Results of the global likelihood ratio test of each three-SNP sliding window haplotype analysis are plotted and connected by the red lines. The plotted lines in blue are recombination rates (cM/Mb) based on HapMap Phase I and II data (<http://hapmap.ncbi.nlm.nih.gov>). A) Region A is indexed by rs7747023 as the strongest associated SNP. A statistically significant haplotype comprising rs1237485, rs3118361, and rs2032502 located adjacent to *TRIM27* was found to be associated with childhood BCP-ALL. B) Region C is indexed by a statistically significant SNP, rs9296068, located near *HLA-DOA*. A three-SNP haplotype containing rs9296068 was significantly associated with childhood BCP-ALL.

doi: 10.1371/journal.pone.0072557.g003

ancestry (percent of European, Amerindian, and African ancestry) to be similar between cases and matched controls [16]. However, the effects of any potential difference in localized genetic ancestry within the MHC between cases and controls cannot be ruled out.

In this comprehensive examination of genetic variation across the xMHC, we provide evidence localizing potential disease susceptibility loci for childhood BCP-ALL to two regions, the extended class I near *TRIM27* and class II near *HLA-DOA*. Confirmation of these findings in future studies through fine-mapping and replication in other populations is warranted.

Supporting Information

Figure S1. Quantile-quantile plot of the expected versus observed $-\log_{10}$ (p-value) distribution in the analysis of 1,145 xMHC SNPs and childhood BCP-ALL risk.

Association results were derived by logistic regression assuming a log-additive genetic model and adjusting for child's age, sex, and race/ethnicity. The red line represents the plot where the observed distribution of the $-\log_{10}$ (p-value) is same as the expected distribution given the number of SNPs tested. (TIF)

Figure S2. Linkage disequilibrium (LD) plot of the twenty SNPs associated with BCP-ALL with a p-value of less than 0.01.

The values displayed in the plot are correlation coefficients (r^2) and the intensity of shading corresponds to the D' measure for each marker pair. The plot and LD measures were generated separately among non-Hispanic white control children (A) and Hispanic control children (B) using Haploview (<http://www.broad.mit.edu/mpg/haploview/>). (TIFF)

Figure S3. Stratified analysis of childhood BCP-ALL and the potentially associated SNPs presented in Table 1 by race/ethnicity, sex, and age group, and subgroup analyses by major subtypes.

Odds ratios (ORs, represented by boxes with the area of each box inversely proportional to the variance of the estimate) and 95% confidence intervals (CIs, error bars) were derived using logistic regression adjusting for child's age, sex, and race/ethnicity depending on the stratification variable. The dashed vertical line represents the OR of the SNP in the analysis of BCP-ALL among all subjects and the width of the diamond is the corresponding 95% CI. $P_{\text{homogeneity}}$ was on the basis of the Cochran's Q test statistic. Abbreviations: Ca, number of case; cALL, common acute lymphoblastic leukemia; Co, number of controls. (PDF)

References

- Greaves MF, Colman SM, Beard ME, Bradstock K, Cabrera ME et al. (1993) Geographical distribution of acute lymphoblastic leukaemia subtypes: second report of the collaborative group study. *Leukemia* 7: 27-34. PubMed: 8418376.
- Greaves M (2006) Infection, immune responses and the aetiology of childhood leukaemia. *Nat Rev Cancer* 6: 193-203. doi:10.1038/nrc1816. PubMed: 16467884.
- Horton R, Wilming L, Rand V, Lovering RC, Bruford EA et al. (2004) Gene map of the extended human MHC. *Nat Rev Genet* 5: 889-899. doi:10.1038/nrg1489. PubMed: 15573121.
- Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA et al. (2009) High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. *PLOS Genet* 5: e1000696. PubMed: 19851445.
- Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K et al. (2011) Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet* 43: 893-896. doi:10.1038/ng.887. PubMed: 21804548.
- Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N et al. (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* 42: 210-215. doi:10.1038/ng.531. PubMed: 20139978.
- Urayama KY, Jarrett RF, Hjalgrim H, Diepstra A, Kamatani Y et al. (2012) Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. *J Natl Cancer Inst* 104: 240-253. doi:10.1093/jnci/djr516. PubMed: 22286212.
- Dorak MT, Lawson T, Machulla HK, Darke C, Mills KI et al. (1999) Unravelling an HLA-DR association in childhood acute lymphoblastic leukemia. *Blood* 94: 694-700. PubMed: 10397736.
- Klitz W, Gragert L, Trachtenberg E (2012) Spectrum of HLA associations: the case of medically refractory pediatric acute lymphoblastic leukemia. *Immunogenetics* 64: 409-419. doi:10.1007/s00251-012-0605-5. PubMed: 22350167.
- Taylor GM, Dearden S, Ravetto P, Ayres M, Watson P et al. (2002) Genetic susceptibility to childhood common acute lymphoblastic leukaemia is associated with polymorphic peptide-binding pocket profiles in HLA-DPB1*0201. *Hum Mol Genet* 11: 1585-1597. doi:10.1093/hmg/11.14.1585. PubMed: 12075003.
- Taylor GM, Richards S, Wade R, Hussain A, Simpson J et al. (2009) Relationship between HLA-DP supertype and survival in childhood acute lymphoblastic leukaemia: evidence for selective loss of immunological control of residual disease? *Br J Haematol* 145: 87-95. doi:10.1111/j.1365-2141.2008.07571.x. PubMed: 19183185.
- Urayama KY, Chokkalingam AP, Metayer C, Ma X, Selvin S et al. (2012) HLA-DP genetic variation, proxies for early life immune modulation, and childhood acute lymphoblastic leukemia risk. *Blood* 120: 3039-3047. doi:10.1182/blood-2012-01-404723. PubMed: 22923493.
- Hosking FJ, Leslie S, Dilthey A, Moutsianas L, Wang Y et al. (2011) MHC variation and risk of childhood B-cell precursor acute lymphoblastic leukemia. *Blood* 117: 1633-1640. doi:10.1182/blood-2010-08-301598. PubMed: 21059899.
- Bartley K, Metayer C, Selvin S, Ducore J, Buffler P (2010) Diagnostic X-rays and risk of childhood leukaemia. *Int J Epidemiol* 39: 1628-1637. doi:10.1093/ije/dyq162. PubMed: 20889538.
- Ma X, Buffler PA, Layefsky M, Does MB, Reynolds P (2004) Control selection strategies in case-control studies of childhood diseases. *Am J Epidemiol* 159: 915-921. doi:10.1093/aje/kwh136. PubMed: 15128601.
- Chokkalingam AP, Aldrich MC, Bartley K, Hsu L, Metayer C et al. (2011) Matching on race and ethnicity in Case-control studies as a means of control for population stratification. *Epidemiol.* p. 101
- Hansen HM, Wiemels JL, Wrensch M, Wiencke JK (2007) DNA quantification of whole genome amplified samples for genotyping on a multiplexed bead array platform. *Cancer Epidemiol Biomarkers Prev* 16: 1686-1690. doi:10.1158/1055-9965.EPI-06-1024. PubMed: 17684147.
- Oliphant A, Barker DL, Stuelpnagel JR, Chee MS (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *BioTechniques Suppl.* 56-58: 60-51. PubMed: 12083399.
- Paynter RA, Skibola DR, Skibola CF, Buffler PA, Wiemels JL et al. (2006) Accuracy of multiplexed Illumina platform-based single-nucleotide polymorphism genotyping compared between genomic and

Table S1. Evaluation of the independent effects in the multivariable analysis of 20 xMHC SNPs potentially associated with childhood BCP-ALL (p-value < 0.01 in the singles SNP analysis).

(PDF)

Acknowledgements

We would like to thank the staff of the University of California, Berkeley Genetic Epidemiology and Genomics Laboratory, the Northern California Childhood leukemia Study, and the Survey Research Center, and the participating children and their families for their important contributions to this study. We also thank the clinical collaborators and participating hospitals: University of California Davis, University of California San Francisco, Children's Hospital of Central California, Lucile Packard Children's Hospital, Children's Hospital and Research Center Oakland, Kaiser Permanente Roseville, Kaiser Permanente Santa Clara, Kaiser Permanente San Francisco, Kaiser Permanente Oakland.

Author Contributions

Conceived and designed the experiments: KYU APC CM JLW JKW ET PB MT LFB PAB. Performed the experiments: KYU APC HH SM PR KJ AT. Analyzed the data: KYU APC SM PR. Contributed reagents/materials/analysis tools: KYU HH SM PR PT YI. Wrote the manuscript: KYU APC.

- whole genome amplified DNA collected from multiple sources. *Cancer Epidemiol Biomarkers Prev* 15: 2533-2536. doi: 10.1158/1055-9965.EPI-06-0219. PubMed: 17164381.
20. Aldrich MC, Zhang L, Wiemels JL, Ma X, Loh ML et al. (2006) Cytogenetics of Hispanic and White children with acute lymphoblastic leukemia in California. *Cancer Epidemiol Biomarkers Prev* 15: 578-581. doi:10.1158/1055-9965.EPI-05-0833. PubMed: 16537719.
 21. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263-265. doi:10.1093/bioinformatics/bth457. PubMed: 15297300.
 22. Dudbridge F (2008) Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered* 66: 87-98. doi:10.1159/000119108. PubMed: 18382088.
 23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575. doi:10.1086/519795. PubMed: 17701901.
 24. Chang JS, Tsai CR, Tsai YW, Wiemels JL (2012) Medically diagnosed infections and risk of childhood leukaemia: a population-based case-control study. *Int J Epidemiol* 41: 1050-1059. doi:10.1093/ije/dys113. PubMed: 22836110.
 25. Crouch S, Lightfoot T, Simpson J, Smith A, Ansell P et al. (2012) Infectious illness in children subsequently diagnosed with acute lymphoblastic leukemia: modeling the trends from birth to diagnosis. *Am J Epidemiol* 176: 402-408. doi:10.1093/aje/kws180. PubMed: 2289827.
 26. Sindhi R, Higgs BW, Weeks DE, Ashokkumar C, Jaffe R et al. (2008) Genetic variants in major histocompatibility complex-linked genes associate with pediatric liver transplant rejection. *Gastroenterology* 135: 830-839.e831-810. doi:10.1053/j.gastro.2008.05.080. PubMed: 18639552.
 27. Zhao T, Chang LW, McLeod HL, Stormo GD (2004) PromoLign: a database for upstream region analysis and SNPs. *Hum Mutat* 23: 534-539. doi:10.1002/humu.20049. PubMed: 15146456.
 28. Douek DC, Altmann DM (1997) HLA-DO is an intracellular class II molecule with distinctive thymic expression. *Int Immunol* 9: 355-364. doi:10.1093/intimm/9.3.355. PubMed: 9088974.
 29. van Ham M, van Lith M, Lillemeier B, Tjin E, Grüneberg U et al. (2000) Modulation of the major histocompatibility complex class II-associated peptide repertoire by human histocompatibility leukocyte antigen (HLA)-DO. *J Exp Med* 191: 1127-1136. doi:10.1084/jem.191.7.1127. PubMed: 10748231.
 30. Santin I, Castellanos-Rubio A, Aransay AM, Gutierrez G, Gaztambide S et al. (2009) Exploring the diabetogenicity of the HLA-B18-DR3 CEH: independent association with T1D genetic risk close to HLA-DOA. *Genes Immun* 10: 596-600. doi:10.1038/gene.2009.41. PubMed: 19458622.
 31. Souwer Y, Chamuleau ME, van de Loosdrecht AA, Tolosa E, Jorritsma T et al. (2009) Detection of aberrant transcription of major histocompatibility complex class II antigen presentation genes in chronic lymphocytic leukaemia identifies HLA-DOA mRNA as a prognostic factor for survival. *Br J Haematol* 145: 334-343. doi: 10.1111/j.1365-2141.2009.07625.x. PubMed: 19245431.
 32. Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF et al. (2009) REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat Genet* 41: 820-823. doi:10.1038/ng.395. PubMed: 19503088.
 33. Ge D, Zhang K, Need AC, Martin O, Fellay J et al. (2008) WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res* 18: 640-643. doi:10.1101/gr.071571.107. PubMed: 18256235.
 34. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S et al. (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* 44: 502-510. doi:10.1038/ng.2205. PubMed: 22446964.
 35. Dorak MT, Oguz FS, Yalman N, Diler AS, Kalayoglu S et al. (2002) A male-specific increase in the HLA-DRB4 (DR53) frequency in high-risk and relapsed childhood ALL. *Leuk Res* 26: 651-656. doi:10.1016/S0145-2126(01)00189-8. PubMed: 12008082.
 36. Kauppi L, Sajantila A, Jeffreys AJ (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum Mol Genet* 12: 33-40. doi:10.1093/hmg/ddg008. PubMed: 12490530.
 37. Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29: 217-222. doi:10.1038/ng1001-217. PubMed: 11586303.
 38. Kauppi L, Jeffreys AJ, Keeney S (2004) Where the crossovers are: recombination distributions in mammals. *Nat Rev Genet* 5: 413-424. doi:10.1038/nrg1346. PubMed: 15153994.
 39. Reymond A, Meroni G, Fantozzi A, Merla G, Cairo S et al. (2001) The tripartite motif family identifies cell compartments. *EMBO J* 20: 2140-2151. doi:10.1093/emboj/20.9.2140. PubMed: 11331580.
 40. Chu Y, Yang X (2011) SUMO E3 ligase activity of TRIM proteins. *Oncogene* 30: 1108-1116. doi:10.1038/ncr.2010.462. PubMed: 20972456.
 41. Ozato K, Shin DM, Chang TH, Morse HC 3rd (2008) TRIM family proteins and their emerging roles in innate immunity. *Nat Rev Immunol* 8: 849-860. doi:10.1038/nri2413. PubMed: 18836477.
 42. Zha J, Han KJ, Xu LG, He W, Zhou Q et al. (2006) The Ret finger protein inhibits signaling mediated by the noncanonical and canonical IkkappaB kinase family members. *J Immunol* 176: 1072-1080. PubMed: 16393995.
 43. Cai X, Srivastava S, Sun Y, Li Z, Wu H et al. (2011) Tripartite motif containing protein 27 negatively regulates CD4 T cells by ubiquitinating and inhibiting the class II PI3K-C2beta. *Proc Natl Acad Sci U S A* 108: 20072-20077. doi:10.1073/pnas.1111233109. PubMed: 22128329.
 44. Di L, Srivastava S, Zhdanova O, Sun Y, Li Z et al. (2010) Nucleoside diphosphate kinase B knock-out mice have impaired activation of the K + channel KCa3.1, resulting in defective T cell activation. *J Biol Chem* 285: 38765-38771. doi:10.1074/jbc.M110.168070. PubMed: 20884616.
 45. Huang L, Li Y, Singleton AB, Hardy JA, Abecasis G et al. (2009) Genotype-imputation accuracy across worldwide human populations. *Am J Hum Genet* 84: 235-250. doi:10.1016/j.ajhg.2009.01.013. PubMed: 19215730.
 46. Paşaniuc B, Avinery R, Gur T, Skibola CF, Bracci PM et al. (2010) A generic coalescent-based framework for the selection of a reference panel for imputation. *Genet Epidemiol* 34: 773-782. doi:10.1002/gepi.20505. PubMed: 21058333.
 47. Alper CA, Larsen CE, Dubey DP, Awdeh ZL, Fici DA et al. (2006) The haplotype structure of the human major histocompatibility complex. *Hum Immunol* 67: 73-84. doi:10.1016/j.humimm.2006.08.104. PubMed: 16698428.
 48. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T et al. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 38: 1166-1172. doi: 10.1038/ng1885. PubMed: 16998491.
 49. Ellinghaus E, Stanulla M, Richter G, Ellinghaus D, te Kronnie G et al. (2012) Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia* 26: 902-909. doi:10.1038/leu.2011.302. PubMed: 22076464.
 50. Han S, Lee KM, Park SK, Lee JE, Ahn HS et al. (2010) Genome-wide association study of childhood acute lymphoblastic leukemia in Korea. *Leuk Res* 34: 1271-1274. doi:10.1016/j.leukres.2010.02.001. PubMed: 20189245.
 51. Orsi L, Rudant J, Bonaventure A, Goujon-Bellec S, Corda E et al. (2012) Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE). *Leukemia* 26: 2561-2564. doi:10.1038/leu.2012.148. PubMed: 22660188.
 52. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B et al. (2009) Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* 41: 1006-1010. doi: 10.1038/ng.430. PubMed: 19684604.
 53. Treviño LR, Yang W, French D, Hunger SP, Carroll WL et al. (2009) Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* 41: 1001-1005. doi:10.1038/ng.432. PubMed: 19684603.
 54. de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R et al. (2006) Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet* 38: 1298-1303. doi:10.1038/ng1899. PubMed: 17057720.
 55. Urayama KY, Chokkalingam AP, Manabe A, Mizutani S (2012) Current evidence for an inherited genetic basis of childhood acute lymphoblastic leukemia. *Int J Hematol* 97: 3-19. PubMed: 23239135.
 56. Do TN, Ucisik-Akkaya E, Davis CF, Morrison BA, Dorak MT (2009) TP53 R72P and MDM2 SNP309 polymorphisms in modification of childhood acute lymphoblastic leukemia susceptibility. *Cancer Genet Cytogenet* 195: 31-36. doi:10.1016/j.cancergencyto.2009.05.009. PubMed: 19837266.
 57. Dorak MT, Mackay RK, Relton CL, Woodward M, Parker L et al. (2009) Hereditary hemochromatosis gene (HFE) variants are associated with birth weight and childhood leukemia risk. *Pediatr Blood Cancer* 53: 1242-1248. doi:10.1002/psc.22236. PubMed: 19711434.
 58. Ucisik-Akkaya E, Davis CF, Gorodezky C, Alaez C, Dorak MT (2010) HLA complex-linked heat shock protein genes and childhood acute

lymphoblastic leukemia susceptibility. *Cell Stress Chaperones* 15: 475-485. doi:10.1007/s12192-009-0161-6. PubMed: 20012387.

59. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ et al. (2008) SNAP: a web-based tool for identification and annotation of

proxy SNPs using HapMap. *Bioinformatics* 24: 2938-2939. doi:10.1093/bioinformatics/btn564. PubMed: 18974171.