Genetic variation near CXCL12 is associated with susceptibility to HIV-related non-Hodgkin lymphoma

Christian W Thorball, Tiphaine Oudot-Mellakh, Nava Ehsan, Christian Hammer, Federico A Santoni, Jonathan Niay, Dominique Costagliola, Cécile Goujard, Laurence Meyer, Sophia S Wang, et al.

ARTICLE

Genetic variation near *CXCL12* is associated with susceptibility to HIV-related non-Hodgkin lymphoma

Christian W. Thorball,[1,2] Tiphaine Oudot-Mellakh,[3] Nava Ehsan,[4,5] Christian Hammer,[6,7] Federico A. Santoni,[8] Jonathan Niay,[3] Dominique Costagliola,[9] Cécile Goujard,[10,11] Laurence Meyer,[12] Sophia S. Wang,[13] Shehnaz K. Hussain,[14] Ioannis Theodorou,[3] Matthias Cavassini,[15] Andri Rauch,[16] Manuel Battegay,[17] Matthias Hoffmann,[18] Patrick Schmid,[19] Enos Bernasconi,[20] Huldrych F. Günthard,[21,22] Pejman Mohammadi,[4,5] Paul J. McLaren,[23,24] Charles S. Rabkin,[25] Caroline Besson[25-27] and Jacques Fellay[1,2,28]

[1]School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; [2]Swiss Institute of Bioinformatics, Lausanne, Switzerland; [3]Centre de Génétique Moléculaire et Chromosomique, GH La Pitié Salpêtrière, Paris, France; [4]The Scripps Research Translational Institute, La Jolla, CA, USA; [5]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA; [6]Department of Cancer Immunology, Genentech, South San Francisco, CA, USA; [7]Department of Human Genetics, Genentech, South San Francisco, CA, USA; [8]Service of Endocrinology, Diabetology and Metabolism, Lausanne University Hospital, Lausanne, Switzerland; [9]Sorbonne Universités, INSERM, UPMC Université Paris 06, Institut Pierre Louis d'Épidémiologie et de Santé Publique (IPLESP UMRS 1136), Paris, France; [10]INSERM, CESP, U1018, Paris-Sud University, Le Kremlin-Bicêtre, France; [11]Department of Internal Medicine, Bicêtre Hospital, AP-HP, Le Kremlin-Bicêtre, France; [12]INSERM U1018, Centre de Recherche en Épidémiologie et Santé des Population, Paris-Sud University, Paris-Saclay University, Le Kremlin-Bicêtre, France; [13]Division of Health Analytics, City of Hope Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, CA, USA; [14]Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [15]Service of Infectious Diseases, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland; [16]Department of Infectious Diseases, Bern University Hospital, University of Bern, Switzerland; [17]Department of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, University of Basel, Basel, Switzerland; [18]Division of Infectious Diseases and Hospital Epidemiology, Kantonsspital Olten, Olten, Switzerland; [19]Division of Infectious Diseases, Cantonal Hospital of St. Gallen, St. Gallen, Switzerland; [20]Division of Infectious Diseases, Regional Hospital of Lugano, Lugano, Switzerland; [21]Department of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland; [22]Institute of Medical Virology, University of Zurich, Zurich, Switzerland; [23]JC Wilt Infectious Diseases Research Center, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada; [24]Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Manitoba, Canada; [25]Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA; [26]CESP, UVSQ, INSERM, Université Paris-Saclay, Villejuif, France; [27]Department of Hematology and Oncology, Hospital of Versailles, Le Chesnay, France and [28]Precision Medicine Unit, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland.

Ferrata Storti Foundation

Haematologica 2021
Volume 106(8):2233-2241

ABSTRACT

Human immunodeficiency virus (HIV) infection is associated with an increased risk of non-Hodgkin lymphoma (NHL). Even in the era of suppressive antiretroviral treatment, HIV-infected individuals remain at higher risk of developing NHL compared to the general population. In order to identify potential genetic risk loci, we performed case-control genome-wide association studies and a meta-analysis across three cohorts of HIV-infected patients of European ancestry, including a total of 278 cases and 1,924 matched controls. We observed a significant association with NHL susceptibility in the C-X-C motif chemokine ligand 12 (*CXCL12*) region on chromosome 10. A fine mapping analysis identified rs7919208 as the most likely causal variant (P=4.77e-11), with the G>A polymorphism creating a new transcription factor binding site for BATF and JUND. These results suggest a modulatory role of *CXCL12* regulation in the increased susceptibility to NHL observed in the HIV-infected population.

Correspondence:

JACQUES FELLAY
jacques.fellay@epfl.ch

Received: January 9, 2020.

Accepted: July 14, 2020.

Pre-published: July 16, 2020.

https://doi.org/10.3324/haematol.2020.247023

Introduction

Human immunodeficiency virus (HIV) infection is associated with a markedly increased risk of several types of cancer compared to the general population.[1–3] This elevated cancer risk can be attributed partly to viral-induced immunodeficiency, frequent co-infections with oncogenic viruses (*e.g.*, Epstein-Barr virus [EBV], hepatitis B and hepatitis C viruses, human herpesvirus 8 [HHV-8] and papillomavirus), and increased prevalence of traditional risk factors such as smoking.[4,5] However, all of these risk factors may not entirely explain the excess cancer burden seen in the HIV-infected (HIV+) population.[5,6]

A previous study performed in the Swiss HIV Cohort Study (SHCS) identified two AIDS-defining cancers, Kaposi sarcoma and non-Hodgkin lymphoma (NHL) as the main types of cancer found among HIV+ patients (NHL representing 34% of all identified cancers).[4] The relative risk of developing NHL in HIV+ patients was highly elevated compared to the general population (period-standardized incidence ratio [SIR] = 76.4).[4] High HIV plasma viral load, absence of antiretroviral therapy (ART) as well as low CD4+ T-cell counts are known predictive factors for NHL.[7,8] The introduction of ART into clinical practice has led to improved overall survival and restoration of immunity by decreasing viral load and increasing CD4+ T-cell counts, and has led to a decreased risk of developing NHL. However, the risk remains substantially elevated compared to the general population (suboptimal immune response [SIR] 9.1[range, 8.3–10.1])[9] and NHL still represents 20% of all cancers in people living with HIV in the ART era.[10] NHL associated with HIV are predominantly aggressive B-cell lymphomas. Although they are heterogeneous, they share several pathogenic mechanisms involving chronic antigen stimulation, impaired immune response, cytokine deregulation and reactivation of the oncogenic viruses EBV and HHV-8.[11]

The emergence of genome-wide approaches in human genomics has led to the discovery of many associations between common genetic polymorphisms and susceptibility to several diseases including HIV infection and multiple types of cancer.[12,13] Recent genome-wide association studies (GWAS) of NHL have identified multiple susceptibility loci in the European population.[14–22] These variants are located in the genes *LPXN*[21], *BTNL2*[23], *EXOC2*, *NCOA1*[14], *PVT1*[14,22], *CXCR5*, *ETS1*, *LPP*, and *BCL2*[22] for various subtypes of NHL, as well as *BCL6* in the Chinese population.[24] Strong associations with variation in human leukocyte antigen (HLA) genes have also been reported.[15,18,22] However, in the setting of HIV infection, no genome-wide analysis has been reported concerning the occurrence of NHL and the specific mechanisms driving their development remain largely unknown.

Here we report the results of the first genome-wide analysis of NHL susceptibility in individuals chronically infected with HIV. We combined three HIV cohort studies from France, Switzerland and the USA and searched for associations between >6 million single nucleotide polymorphisms (SNP) and a diagnosis of NHL. We identified a novel genetic locus near *CXCL12* to be associated with the development of NHL among HIV+ individuals.

Methods

Ethics statement

The SHCS, the Primo ANRS and ANRS CO16 Lymphovir cohorts (ANRS) and the Multicenter AIDS Cohort Study (MACS) cohorts have been approved by the competent Ethics Committees /Institutional Review Boards of all participating institutions. A written informed consent, including consent for human genetic testing, was obtained from all study participants.

Study participants and contributing centers

Genotyping and phenotypic data were obtained from a total of 2,202 HIV+ individuals enrolled in the SHCS, ANRS and MACS cohorts (278 cases and 1,924 controls) (Table 1). For details on inclusion criteria and cohorts, refer to the *Online Supplementary Appendix*.

Quality control and imputation of genotyping data

The genotyping data from each cohort was filtered and imputed in a similar way, with each genotyping array processed separately to minimize potential batch effects. All variants were first flipped to the correct strand orientation with BCFTOOLS (v1.8) using the human genome build GRCh37 as reference. Variants were removed if they had a larger than 20% minor allele frequency (MAF) deviation from the 1,000 genomes phase 3 EUR reference panel or if they showed a larger than 10% MAF deviation between genotyping chips in the same cohort.

The quality control (QC) filtered genotypes were phased with EAGLE2[25] and missing genotypes were imputed using PBWT[26] with the Sanger Imputation Service,[27] taking the 1,000 genomes project phase 3 panel as reference. Only high-quality variants with an imputation score (INFO >0.8) were retained for further analyses.

Genome-wide association testing and meta-analysis

In order to search for associations between human genomic variation and the development of HIV-related NHL, we first performed separate GWAS within each cohort (SHCS, ANRS and MACS) prior to combining the results in a meta-analysis.

For each cohort separately, the imputed variants were filtered out using PLINK (v2.00a2LM)[28] based on missingness (>0.1), MAF (<0.02) and deviation from Hardy-Weinberg Equilibrium (P_{HWE}<1e-6). Determination of population structure and calculation of principal components was done using EIGENSTRAT (v6.1.4)[29] and the HapMap3 reference panel.[30] All individuals not clustering with the European HapMap3 samples were excluded from further analyses. The samples were screened using KING (v2.1.3)[31] to ensure no duplicate or cryptic related samples were included. Single-marker case-control association analyses were performed using linear mixed models, with genetic relationship matrices calculated between pairs of individuals according to the leave-one-chromosome-out principle, as implemented in GCTA mlma-loco (v1.91.4beta).[32,33] Sex was included as a covariate, except in the MACS cohort, which only includes men.

The results of the three GWAS were combined across cohorts using a weighted Z-score-based meta-analysis in PLINK (v1.90b5.4), after exclusion of the variants that were not present in all three cohorts.

Other methods

The details of the cohorts and other methods used, *i.e.*, fine mapping, prediction of causal variants, long-range chromatin interactions, transcriptomic effects, comparisons to GWAS in

the general population and information on data sharing can be found in the *Online Supplementary Appendix*.

Results

Study participants and association testing

In order to identify human genetic determinants of HIV-associated NHL, we performed case-control GWAS in three groups of HIV+ patients of European ancestry (SHCS, ANRS and MACS). The characteristics of the study participants are presented in Table 1. In total, genotyping data were obtained for 278 cases (NHL+/HIV+) and 1,924 matched controls (NHL-/HIV+). With this sample size, we had 80% power to detect a common genetic variant (10% minor allele frequency) with a relative risk of 2.5, assuming an additive genetic model and using Bonferroni correction for multiple testing ($P_{threshold}$=5e-8).[34]

After genome-wide imputation and quality control, 6.2 million common variants were tested for association with the development of NHL using linear mixed models including sex as a covariate. Results were combined across cohorts using a weighted Z-score-based meta-analysis (Figure 1A). The genomic inflation factor (lambda) was in all cases very close to 1 [1.00–1.01], indicating an absence of systematic inflation of the association results (Figure 1B; *Online Supplementary Figure S2*).

Association results

We observed significant associations with the development of HIV-related NHL at a single locus on chromosome 10, downstream of *CXCL12* (Figure 1C). A total of seven SNP in this locus had *P*-values lower than the genome-wide significance threshold (*P*<5e-8), with rs7919208 displaying the strongest association (Table 2). This association was only detected in the SHCS and ANRS cohorts and not among MACS study participants (*Online Supplementary Table S1*).

Fine mapping of the *CXCL12* locus

In order to identify the causal variant(s) among associated SNP and determine their potential functional effects, we used a multi-level fine mapping approach, combining the statistical fine mapping tool PAINTOR to obtain a 99% credible set and the deep learning framework DeepSEA to predict any effects on chromatin marks and transcription factor binding these variants may have.

Using PAINTOR, we identified a single variant, rs7919208, having a high posterior probability (=100%) of being causal among the 99% credible set based on the integration of the association results, linkage disequilibrium (LD) structure and enrichment of genomic features in this locus (Figure 2).

Consistent with the PAINTOR result, DeepSEA also identified rs7919208 as the sole variant, among the 99% credible set, predicted to have a functional impact by significantly increasing the probability of binding by the B-cell transcription factors BATF (log$_2$ fold-change=3.27) and JUND (log$_2$ fold-change=2.91) (*Online Supplementary Table S2*). Further analysis of the genomic sequence surrounding rs7919208 and the JASPAR transcription factor binding site (TFBS) motifs for BATF and JUND revealed that rs7919208 G->A polymorphism creates the TFBS motif required for the binding of these transcription factors (Figure 3A).

Long-range chromatin interactions

In order to assess the potential functional links between the TFBS created in the presence of the minor allele of rs7919208 and the nearby genes, we performed an analysis of promoter capture Hi-C data and topologically associating domains (TAD). We used the well-characterized GM12878 lymphoblastoid cell line produced by EBV transformation of B lymphocytes collected from a female European donor as a model.

First, in order to examine the interaction potential of the rs7919208 region with nearby promoters, we analyzed available promoter capture Hi-C data obtained from the GM12878 cell line. This analysis revealed a significant interaction between the rs7919208 region and the *CXCL12* promoter, suggesting a possible modulating impact of rs7919208 on the transcription of that gene (Figure 3B). Second, in order to further validate this observed genomic interaction, we analyzed available TAD calls from GM12878 cells,[35] using the 3D Genome Browser for visualization[36] (Figure 3C). We observed that rs7919208 is located within a large TAD together with *CXCL12*, signifying the interaction potential of the new TFBS at rs7919208 and *CXCL12*.

Transcriptomic effects of rs7919208

We did not observe any association between rs7919208 and mRNA expression levels of *CXCL12* in peripheral

Table 1. Summary of included samples and studies.

Cohort	Cases	Controls	Lambda	Genotyping chips	Years of NHL diagnosis	Control inclusion criteria
SHCS	145	1,090	1.00	Illumina HumanOmniExpress-24, Human1M, Human610, HumanHap550, HumanCore-12	2000 - 2017	HIV < 2005, no cancer diagnosis as of 2017 & matched with age
Age (median)	61	58				
Sex (male, %)	91%	80%				
ANRS	61	562	1.00	Illumina Human Omni5 Exome 4v 1-2, Illumina 300	2008 - 2015	No cancer diagnosis
Age (median)	50	34				
Sex (male, %)	89%	87%				
MACS	72	272	1.01	Illumina 1MV1, Human1M-Duo, HumanHap550	1985 - 2013	Matched to cases in terms of age, treatment & time of infection
Age (median)	69	68				
Sex (male, %)	100%	100%				

Cohort and patient characteristics for the Swiss HIV Cohort Study (SHCS), the Primo ANRS and ANRS CO16 Lymphovir cohorts (ANRS) and the Multicenter AIDS Cohort Study (MACS) cohorts. Lambda indicates the genomic inflation factor from the individual cohort genome-wide association studies (GWAS). NHL: non-Hodgkin lymphoma; HIV: human immunodeficiency virus.

blood or peripheral blood mononuclear cells (PBMC) from multiple publicly available datasets, including GTEx,[37] GEUVADIS[38] and the Milieu Intérieur Consortium[39] (*Online Supplementary Figure S3*). Of note, *CXCL12* expression levels were very low in all datasets (*Online Supplementary Figure S4A)*.

Using allele-specific expression analysis in the GTEx dataset, we observed a significant effect in individuals heterozygous of rs7919208, with increased allelic imbalance of *CXCL12* in fibroblasts (false discovery rate adjusted *P*=0.0006, one-sided rank sum test), which was not observed in other tissues (*Online Supplementary Figure 4B*).

HIV infection causes many profound transcriptomic changes.[40] Thus, in order to examine the effect of rs7919208 on *CXCL12* in the context of HIV infection, we extracted RNA from PBMC of 452 individuals in the SHCS with available genotyping data and sequenced them using the Bulk RNA Barcoding and sequencing (BRB-seq) approach.[41] However, the expression levels of *CXCL12* were below the limit of detection for most individuals, preventing an expression quantitative trait loci (eQTL) analysis.

Multiple isoforms of *CXCL12* exist, with variable degrees of expression and potency described in the context of HIV infection.[42] We observed a single significant correlation between rs7919208 and *CXCL12* transcript usage, which was restricted to visceral adipose tissue. The presence of the rs7919208 minor allele was associated with higher relative expression of the longer and rarer transcript isoform ENST00000374429.6 (*Online Supplementary Figure S5*).

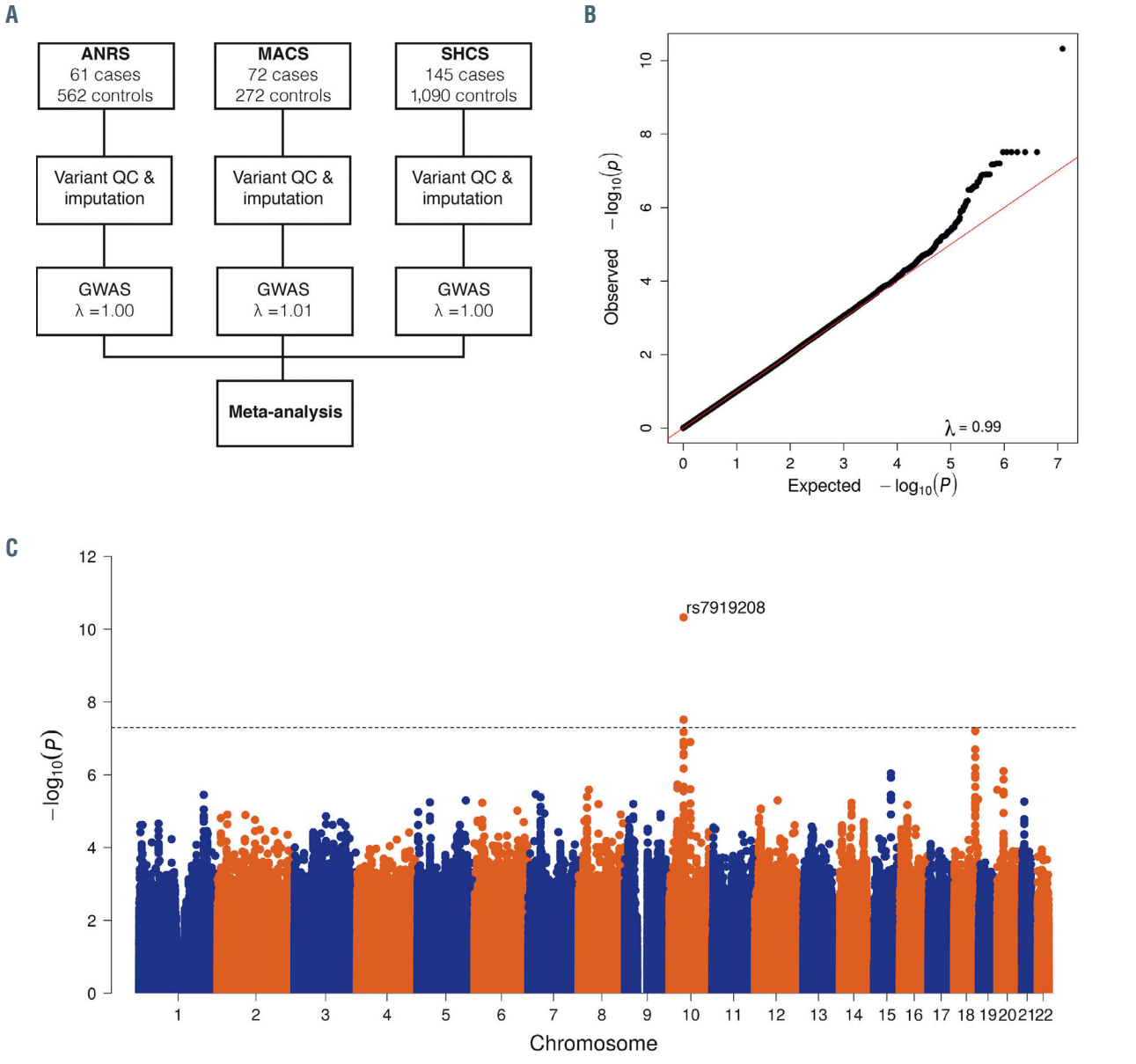

Figure 1. Genome-wide association analysis. (A) Schematic of analysis pipeline. (B) Quantile-quantile plot of the observed -log$_{10}$(*P*-value) (black dots, y-axis) *vs.* expected -log$_{10}$(*P*-values) under the null hypothesis (red line) to check for any genomic inflation of the observed *P*-values. No genomic inflation is observed, with the genomic inflation factor λ=0.99. (C) Manhattan plot of all obtained *P*-values for each variant included in the meta-analysis. The genome-wide threshold (*P*=5e-8) for significance is marked by a dotted line. Only variants at the *CXCL12* locus were found to be significant.

No replication of susceptibility loci found in the general population

In order to assess whether the genetic contribution to the risk of developing NHL is similar or distinct in the HIV+ population compared to the general population, we extracted the *P*-values of all variants found to be genome-wide significant in previous GWAS performed in the general population[14,21–24,43] and compared them to our results. We did not replicate any of the previously published genome-wide associated variants, even at nominal significance level ($P<0.05$), despite sufficient statistical power for many of the variants, thus indicating that the genetic susceptibility of NHL is distinct between the HIV+ and the general population (*Online Supplemental Table S3*). In order to further examine this possibility, we tested whether the NHL/HIV+ associated variant rs7919208 is associated

Table 2. Significant association with human immunodeficiency virus-related non-Hodgkin lymphoma.

Chr	Pos	SNP	Ref	Alt	*P*	OR
10	44673557	rs7919208	A	G	4.77e-11	1.23
10	44677967	rs149399290	T	C	3.09e-08	1.20
10	44678218	rs17155463	T	A	3.09e-08	1.20
10	44678262	rs17155474	C	T	3.09e-08	1.20
10	44678454	rs17155478	T	C	3.09e-08	1.20
10	44678898	rs12249837	G	A	3.09e-08	1.20
10	44680902	rs10608969	T	TAAAGA	3.09e-08	1.20

Variants significantly associated with human immunodeficiency virus (HIV)-related non-Hodgkin lymphoma in a weighted Z-score-based meta-analysis of all individuals included in the Swiss HIV Cohort Study (SHCS), the Primo ANRS and ANRS CO16 Lymphovir cohorts (ANRS) and the Multicenter AIDS Cohort Study (MACS) cohorts. Odds ratios (OR) were transformed from betas using the formula OR=exp(beta). Chr: chromosome; pos: position; SNP: single nucleotide polymorphisms; Ref: reference allele, Alt: alternative allele.

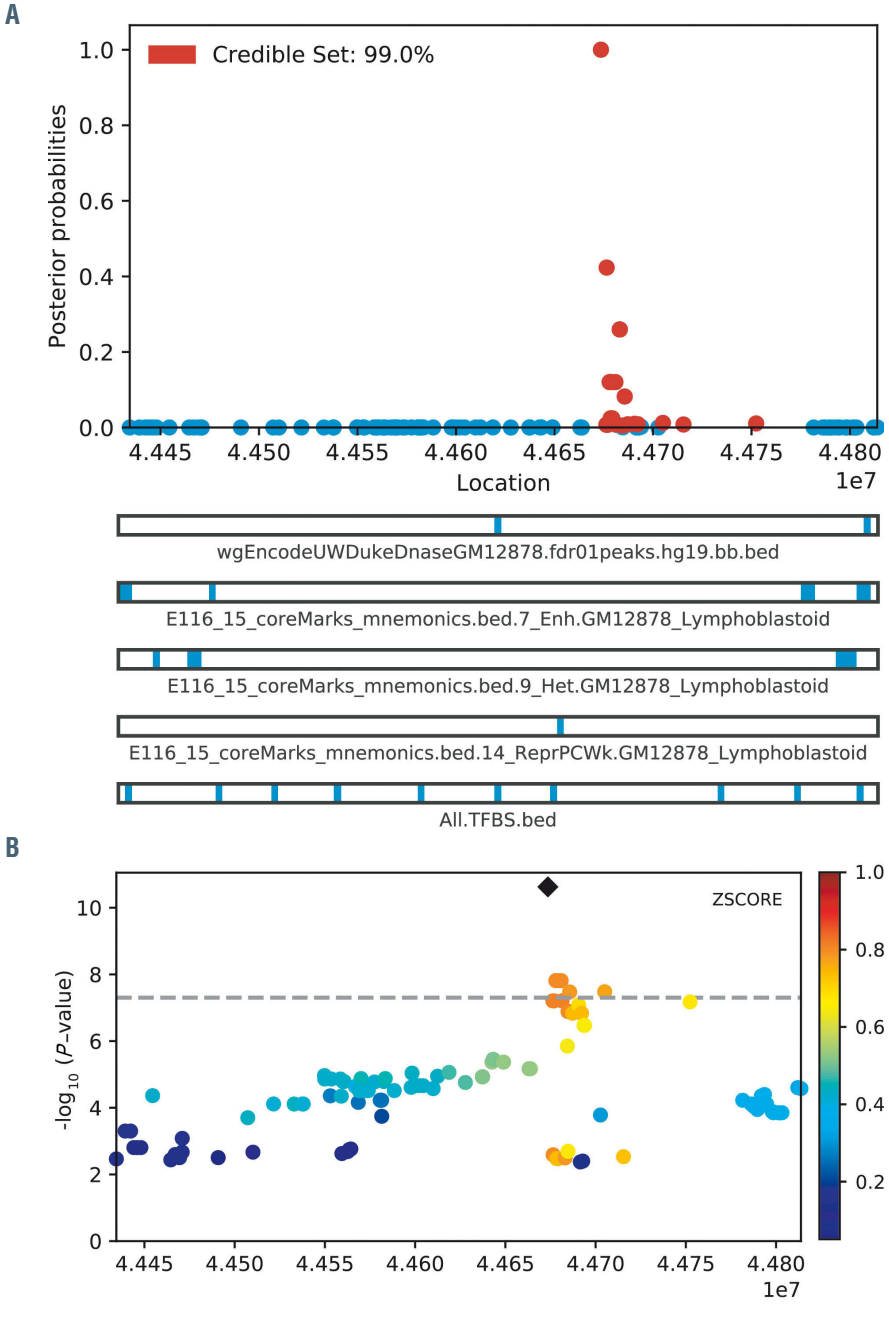

Figure 2. Fine mapping of genome-wide significant hits with PAINTOR. (A) The 99% credible set and posterior probabilities of being the causal variant. The genomic positions are listed on the x-axis. Bottom tracks represent DNAase and chromatin marks obtained from GM12878 cells as well as transcription factor binding site (TFBS) from the Roadmap Epigenomics Project and ENCODE in the region. (B) Locus plot of the associated variants, highlighting the LD relationship, based on the Swiss HIV Cohort Study cohort. The top variant rs7919208 is marked by a black diamond.

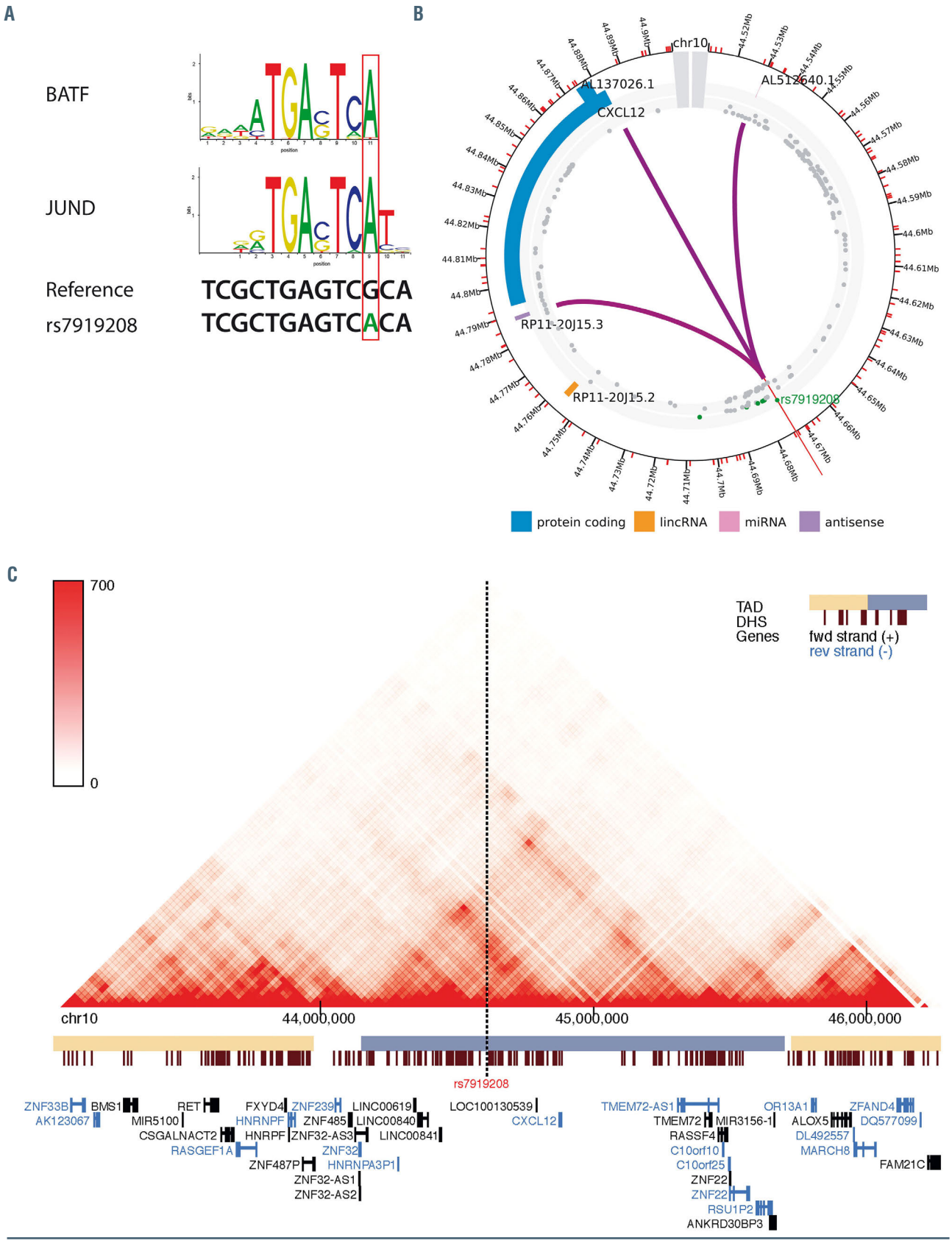

Figure 3. Novel transcription factor binding site and long-range interactions. (A) Canonical motifs of BATF and JUND with the underlying genomic reference sequence and the nucleotide change caused by rs7919208. (B) Promoter capture Hi-C analysis in the GM12878 cell line of the region with the predicted causal variant and *CXCL12*. Variants and their level of association in the meta-analysis are marked in the inner grey circle. Genome-wide significant variants are colored green. Purple lines indicate significant interactions between promoter and other genomic regions. (C) Topologically associating domains (TAD) in the GM12878 cell line in the region of *CXCL12*. The yellow and blue boxes indicate the called TAD from the Hi-C contact map above. The plot is centered on rs7919208. DHS: DNase I hypersensitive sites.

with an increased risk of NHL in the general population. We performed a series of case/control GWAS of four NHL subtypes (CLL, DLBCL, FL and MZL) as well as a combined GWAS with all NHL subtypes (*Online Supplementary Table S4*; *Online Supplementary Figure S6*) and assessed the association evidence at rs7919208. We found no association between rs7919208 and any of the subtypes in the general population, even at nominal significance.

Discussion

In this genome-wide analysis, including a total of 278 NHL HIV+ cases and 1,924 HIV+ controls from three independent cohorts, we identified a novel NHL susceptibility locus on chromosome 10 near the *CXCL12* gene. The strong signal observed in the meta-analysis was driven by the associations detected in the SHCS and ANRS cohorts and there was no evidence of association in the MACS cohort. Notably, most NHL cases in the MACS cohort date back to the pre-ART era, while only NHL cases diagnosed after the year 2000 were included in the SHCS and ANRS analyses. Conceivably, NHL occurring in the early years of the HIV pandemic may have been primarily driven by severe immunosuppression, which could have obscured any influence of human genetic variation among the cases in the MACS sample. Precise phenotype definition is crucial in designing large-scale genetic studies since any environmental noise tends to decrease the likelihood of identifying potential genetic influences.

NHL is a relatively rare cancer even among HIV infected individuals, making it difficult to collect the large numbers of cases that would typically be included in contemporary genome-wide genetic studies. Indeed, a recent study from the Data Collection on Adverse events of Anti-HIV Drugs (D:A:D) group showed an NHL incidence rate of 1.17/1,000 person-years of follow-up over the past 15 years (392 new cases in >40,000 HIV+ individuals).[8] Still, we were able to obtain clinical and genetic data from a total of 278 patients with confirmed NHL diagnosis. By matching them with a larger number of controls from the same cohorts, we had enough power to identify associated variants of relatively large effects in the *CXCL12* region.

Several groups have already suggested a potential role for *CXCL12* variation in HIV-related NHL. A prospective study correlated increased *CXCL12* expression with subsequent NHL development in HIV+ children but not in uninfected children.[44] The number of A alleles at the CXCL12-3' variant (rs1801157) has also previously been associated with an increased risk of developing HIV-related NHL during an 11.7 year follow-up period.[45] Thus, our data further support the role of CXCL12 as a critical modulator of the individual risk of developing NHL in the HIV+ population.

The role of CXCL12 and its receptor chemokine receptor 4 (CXCR4) in cancer in the general population is well established, with the levels of *CXCL12* and *CXCR4* found to be increased in multiple types of cancer and to be associated with tumor progression.[46,47] Furthermore, *in vivo* inhibition of either CXCR4 or CXCL12 signaling is capable of disrupting early lymphoma development in severe combined immunodeficient (SCID) mice transfused with EBV+ PBMC.[48] These results and others have already led to the development and testing of several small molecules targeting either CXCL12 or CXCR4 to inhibit tumor progression.[46]

We could not identify any significant relationship between rs7919208 and the expression levels of *CXCL12* in PBMC or EBV transformed lymphocytes. This could be due to the low expression levels of *CXCL12* in most tissues, apart from stromal cells. Still, our analysis of allele-specific expression showed a significant allelic imbalance for heterozygous carriers of rs7919208 for *CXCL12*. This signal was only observed in fibroblasts, the GTEx tissue most closely resembling stromal cells. Furthermore, we identified a significant association between rs7919208 and CXCL12 transcript isoform usage in the visceral adipose tissue, which is known to also contain a minority of stromal cells.[49] Combined these results underscore the potential importance of these cells in the development of HIV-related NHL and the ability of rs7919208 to modify transcription.

The new BATF and JUND binding site created by rs7919208 could act as an induced or dynamic eQTL, specifically triggered by HIV infection. Such eQTL can be found in regions deprived of regulatory annotations, since these have been mostly examined in static cell types.[50] Supporting this hypothesis, HIV has been shown to induce overexpression of BATF.[51] This would explain why rs7919208 is only a risk factor for HIV+ individuals and not in the general population.

Previous analyses in the general population have discovered both shared and distinct associations for NHL subtypes.[14,21–24,43] However, similar analyses were not possible in our sample since NHL subtype information was not available for many of our cases. Furthermore, information on serostatus for relevant co-infections with EBV or other oncogenic viruses was not available and could therefore not be assessed. In particular, EBV has been largely associated with the development of NHL and other lymphomas and is considered a driver of a subset of NHL in the general population.[52] Variants in the HLA region have consistently been associated with all NHL subtypes in HIV uninfected populations regardless of EBV serostatus, although different HLA associations have been observed for each NHL subtype. We did not find any evidence of HLA associations in our analyses of HIV-related NHL. This might be due to a lack of power, due to our limited sample size in comparison to the NHL GWAS performed in the general population. However, this lack of replication of HLA variants and all other risk variants previously identified in the general population strongly suggests that distinct genes or pathways influence susceptibility to NHL in the HIV+ population compared to the general population.[53] This distinction may be due to the unique pathogenic mechanisms involved in HIV-associated NHL, such as cytokine deregulation, chronic antigen stimulation and impaired immune response, among others.[11]

In summary, we have identified variants significantly associated with the development of NHL in the HIV+ population. Fine mapping of the associated locus and subsequent analyses of TAD, promoter capture Hi-C data as well as deep-learning models of mutational effects on transcription factor binding, points to a causative model involving the gain of a BATF and JUND transcription binding site downstream of *CXCL12* capable of physically interacting with the *CXCL12* promoter. These results suggest an important modulating role of CXCL12 in the development of HIV-related NHL.

Disclosure

CH is a full-time employee of F. Hoffmann–La Roche/Genentech; all other authors declare no conflicts of interest.

Contributions

CWT, JF, PJM, CSR, CB, CH and TOM contributed to the conception and design of the study; CWT, JF, PJM, FAS, DC, LM, CG, IT, SKH, MC, AR, MB, MH, PS, EB, HFG, CSR and CB contributed to the acquisition of data; CWT, TOM, CH, FAS, CB, CSR, NE, PM, and JF contributed to the analysis and interpretation of data; CWT, JF, CSR, CB and SW contributed to drafting the article and revising it critically for important intellectual content; all authors critically reviewed and approved the final manuscript.

Acknowledgments

The data are gathered by the Five Swiss University Hospitals, two Cantonal Hospitals, 15 affiliated hospitals and 36 private physicians (listed in http://www.shcs.ch/180-health-care-providers). The datasets have been accessed through the National Institutes of Health (NIH) database for Genotypes and Phenotypes (dbGaP) under accession # phs000801. A full list of acknowledgements can be found in the supplementary note (Berndt SI et al., Nature Genet., 2013, PMID: 23770605).

Funding

This study has been financed within the framework of the Swiss HIV Cohort Study, supported by the Swiss National Science Foundation (grant #177499), by SHCS project #789 and by the SHCS research foundation. This work further benefited from the ANRS funding of both the Primo and Lymphovir cohorts. Foundation Monahan and Fulbright funded the stay of CB at the National Cancer Institute (NCI). The Genome-Wide Association Study (GWAS) of Non-Hodgkin Lymphoma (NHL) project was supported by the intramural program of the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI), NIH.

References

1. Patel P, Hanson DL, Sullivan PS, et al. Incidence of types of cancer among HIV-infected persons compared with the general population in the United States, 1992–2003. Ann Intern Med 2008;148(10):728-736.
2. Vogel M, Friedrich O, Lüchters G, et al. Cancer risk in HIV-infected individuals on HAART is largely attributed to oncogenic infections and state of immunocompetence. Eur J Med Res. 2011;16(3):101.
3. Robbins HA, Pfeiffer RM, Shiels MS, Li J, Hall HI, Engels EA. Excess cancers among HIV-infected people in the United States. J Natl Cancer Inst. 2015;107(4):dju503.
4. Clifford GM, Polesel J, Rickenbach M, et al. Cancer risk in the Swiss HIV Cohort Study: associations with immunodeficiency, smoking, and highly active antiretroviral therapy. J Natl Cancer Inst. 2005;97(6):425-432.
5. Engels EA. Non-AIDS-defining malignancies in HIV-infected persons: etiologic puzzles, epidemiologic perils, prevention opportunities. AIDS. 2009;23(8):875-885.
6. Borges ÁH, Dubrow R, Silverberg MJ. Factors contributing to risk for cancer among HIV-infected individuals, and evidence that earlier combination antiretroviral therapy will alter this risk: Curr Opin HIV AIDS. 2014;9(1):34-40.
7. Guiguet M, Boué F, Cadranel J, Lang J-M, Rosenthal E, Costagliola D. Effect of immunodeficiency, HIV viral load, and antiretroviral therapy on the risk of individual malignancies (FHDH-ANRS CO4): a prospective cohort study. Lancet Oncol. 2009;10(12):1152-1159.
8. Shepherd L, Ryom L, Law M, et al. Differences in virological and immunological risk factors for non-Hodgkin and Hodgkin lymphoma. J Natl Cancer Inst. 2018;110(6):598-607.
9. Hleyhel M, Belot A, Bouvier AM, et al. Risk of AIDS-defining cancers among HIV-1–infected patients in France between 1992 and 2009: results from the FHDH-ANRS CO4 Cohort. Clin Infect Dis. 2013;57(11):1638-1647.
10. Robbins HA, Pfeiffer RM, Shiels MS, Li J, Hall HI, Engels EA. Excess cancers among HIV-infected people in the United States. J Natl Cancer Inst. 2015;107(4):dju503.
11. Swerdlow SH. WHO classification of tumours of haematopoietic and lymphoid tissues. International Agency for Research on Cancer; 2017.
12. McLaren PJ, Carrington M. The impact of host genetic variation on infection with HIV-1. Nat Immunol. 2015;16(6):577-583.
13. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. Nat Rev Cancer. 2017;17(11):692-704.
14. Cerhan JR, Berndt SI, Vijai J, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. Nat Genet. 2014;46(11):1233-1238.
15. Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. Nat Genet. 2010;42(8):661-664.
16. Frampton M, da Silva Filho MI, Broderick P, et al. Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. Nat Commun. 2013;4:2549.
17. Kumar V, Matsuo K, Takahashi A, et al. Common variants on 14q32 and 13q12 are associated with DLBCL susceptibility. J Hum Genet. 2011;56(6):436-439.
18. Moutsianas L, Enciso-Mora V, Ma YP, et al. Multiple Hodgkin lymphoma–associated loci within the HLA region at chromosome 6p21.3. Blood. 2011;118(3):670-674.
19. Skibola CF, Bracci PM, Halperin E, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet. 2009;41(8):873-875.
20. Urayama KY, Jarrett RF, Hjalgrim H, et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein–Barr virus status–defined subgroups. J Natl Cancer Inst. 2012;104(3):240-253.
21. Vijai J, Kirchhoff T, Schrader KA, et al. Susceptibility loci associated with specific and shared subtypes of lymphoid malignancies. PLoS Genet. 2013;9(1):e1003220.
22. Skibola CF, Berndt SI, Vijai J, et al. Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. Am J Hum Genet. 2014;95(4):462-471.
23. Vijai J, Wang Z, Berndt SI, et al. A genome-wide association study of marginal zone lymphoma shows association to the HLA region. Nat Commun. 2015;6:5751.
24. Tan DEK, Foo JN, Bei J-X, et al. Genome-wide association study of B cell non-Hodgkin lymphoma identifies 3q27 as a susceptibility locus in the Chinese population. Nat Genet. 2013;45(7):804-807.
25. Loh P-R, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016;48(11):1443-1448.
26. Durbin R. Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). Bioinformatics. 2014;30(9):1266-1272.
27. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279-1283.
28. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):1-16.
29. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904-909.
30. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52-58.
31. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867-2873.
32. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82.
33. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. 2014;46(2):100-106.
34. Johnson JL, Abecasis GR. GAS Power Calculator: web-based power calculator for genetic association studies. bioRxiv. 2017;164343.
35. Rao SSP, Huntley MH, Durand NC, et al. A 3D Map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665-1680.
36. Wang Y, Song F, Zhang B, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. 2018;19(1):151.
37. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204-213.

38. Lappalainen T, Sammeth M, Friedländer MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501(7468):506-511.
39. Piasecka B, Duffy D, Urrutia A, et al. Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. PNAS. 2018;115(3):e488-E497.
40. Mohammadi P, Desfarges S, Bartha I, et al. 24 hours in the life of HIV-1 in a T cell line. PLOS Pathogens. 2013;9(1):e1003161.
41. Alpern D, Gardeux V, Russeil J, et al. BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. Genome Biol. 2019;20(1):71.
42. Janssens R, Struyf S, Proost P. The unique structural and functional features of CXCL12. Cell Mol Immunol. 2018;15(4):299-311.
43. Lim U, Kocarnik JM, Bush WS, et al. Pleiotropy of cancer susceptibility variants on the risk of non-Hodgkin lymphoma: The PAGE Consortium. PLoS One. 2014;9(3):e89791.
44. Sei S, O'Neill DP, Stewart SK, et al. Increased level of stromal cell-derived factor-1 mRNA in peripheral blood mononuclear cells from children with AIDS-related lymphoma. Cancer Res. 2001;61(13):5028-5037.
45. Rabkin CS, Yang Q, Goedert JJ, Nguyen G, Mitsuya H, Sei S. Chemokine and chemokine receptor gene variants and risk of non-Hodgkin's lymphoma in human immunodeficiency virus-1–infected individuals. Blood. 1999;93(6):1838-1842.
46. Meng W, Xue S, Chen Y. The role of CXCL12 in tumor microenvironment. Gene. 2018;641:105-110.
47. Peled A, Klein S, Beider K, Burger JA, Abraham M. Role of CXCL12 and CXCR4 in the pathogenesis of hematological malignancies. Cytokine. 2018;109:11-16.
48. Piovan E, Tosello V, Indraccolo S, et al. Chemokine receptor expression in EBV-associated lymphoproliferation in hu/SCID mice: implications for CXCL12/CXCR4 axis in lymphoma generation. Blood. 2005;105(3):931-939.
49. Vijay J, Gauthier M-F, Biswell RL, et al. Single-cell analysis of human adipose tissue identifies depot- and disease-specific cell types. Nat Metab. 2020;2(1):97-109.
50. Strober BJ, Elorbany R, Rhodes K, et al. Dynamic genetic regulation of gene expression during cellular differentiation. Science. 2019;364(6447):1287-1290.
51. Quigley M, Pereyra F, Nilsson B, et al. Transcriptional analysis of HIV-specific CD8+ T cells shows that PD-1 inhibits T cell function by upregulating BATF. Nat Med. 2010;16(10):1147-1151.
52. Gasser O, Bihl FK, Wolbers M, Loggi E, Steffen I, Hirsch HH. HIV patients developing primary CNS lymphoma lack EBV-specific CD4þ T cell function irrespective of absolute CD4þ T cell counts. PLoS Med. 2007;4(3):6.
53. Swerdlow SH, Campo E, Harris NL, et al. WHO classification of tumours of haematopoietic and lymphoid tissues.