

Supplemental Tables

Supplemental Table 1. Cohort level association statistics for genome-wide significant variants in the meta-analysis

SNP	ANRS BETA	ANRS P	SHCS BETA	SHCS P	MACS BETA	MACS P
rs7919208	0.32	2.78E-10	0.12	2.66E-03	-0.06	0.57
rs149399290	0.28	7.93E-07	0.13	2.30E-03	-0.03	0.79
rs17155463	0.28	7.93E-07	0.13	2.30E-03	-0.03	0.79
rs17155474	0.28	7.93E-07	0.13	2.30E-03	-0.03	0.79
rs17155478	0.28	7.93E-07	0.13	2.30E-03	-0.03	0.79
rs12249837	0.28	7.93E-07	0.13	2.30E-03	-0.03	0.79
rs10608969	0.28	7.93E-07	0.13	2.30E-03	-0.03	0.79

The analyses were performed using linear mixed models with GCTA within each cohort. No associations were seen in the MACS cohort. P-values and beta values are presented to show the level and direction of the association.

Supplemental Table 2. Top predicted changes associated with rs7919208 allelic variation in GM12878

Transcription factor	Effect (Log2fold change)	E-value	Normalized Prob. (Reference)	Normalized Prob. (Alternative)
BATF	3.27	0.00004	0.13	0.60
JUND	2.91	0.00009	0.12	0.50
MEF2A	2.06	0.00026	0.09	0.28
MEF2C	1.98	0.00022	0.08	0.26
BCL11A	2.19	0.00040	0.07	0.25
P300	1.74	0.00055	0.08	0.22
IRF4	2.12	0.00033	0.06	0.21

Significant changes induced by rs7919208 as predicted by DeepSEA for transcription factors with a normalized probability (Prob.) above 0.20 in the GM12878 lymphoblastoid cell line. The E-value is the expected proportion of variants with a larger predicted effect between the reference and alternative allele for a certain chromatin feature based on predicted effects calculated for variants in The 1000 Genomes Project.

Supplemental Table 3. Comparisons with genome-wide significant variants identified in GWAS of NHL in the general population

SNP	Gene	Publication	Subtype	META P	SHCS P	ANRS P	MACS P	OR	Mean MAF	POWER (P<0.05)	POP
rs116446171	EXOC2	Cerhan et al. ¹⁶	DLBCL	NA	NA	NA	0.08	2.20	0.01	58%	EUR
rs12195582	HLA region	Skibola et al. ²⁴	FL	0.33	0.47	0.51	NA	1.78	0.41	100%	EUR
rs12289961	LPXN	Vijai et al. (2013) ²³	DLBCL+FL	0.88	0.40	0.14	0.46	1.29	0.22	74%	EUR
rs13254990	PVT1	Skibola et al. ²⁴	FL	0.75	NA	0.58	0.82	1.18	0.32	48%	EUR
rs13255292	PVT1	Cerhan et al. ¹⁶	DLBCL	0.83	NA	0.72	0.90	1.22	0.32	62%	EUR
rs17203612	HLA class II	Skibola et al. ²⁴	FL	NA	0.82	NA	NA	1.44	0.38	98%	EUR
rs17749561	BCL2	Skibola et al. ²⁴	FL	0.74	NA	0.72	0.98	1.34	0.10	61%	EUR
rs2523607	HLA-B	Cerhan et al. ¹⁶	DLBCL	0.50	0.30	0.77	NA	1.32	0.08	49%	EUR
rs2922994	HLA-B	Vijai et al. (2015) ²⁵	MZL	0.55	0.30	0.66	NA	1.64	0.08	93%	EUR
rs3130437	HLA class I	Skibola et al. ²⁴	FL	0.33	0.47	NA	0.46	1.23	0.38	68%	EUR
rs4733601	PVT1	Cerhan et al. ¹⁶	DLBCL	0.29	NA	0.38	0.57	1.18	0.49	51%	EUR
rs4937362	ETS1	Skibola et al. ²⁴	FL	0.62	0.61	0.90	NA	1.17	0.45	47%	EUR
rs4938573	CXCR5	Skibola et al. ²⁴	FL	0.60	0.25	0.72	0.18	1.34	0.19	81%	EUR
rs6444305	LPP	Skibola et al. ²⁴	FL	NA	NA	NA	NA	1.21	NA	NA	EUR
rs6457327	HLA	Lim et al. ⁵⁵	DLBCL+FL	0.66	0.80	0.85	0.18	1.30	0.35	85%	EUR
rs6773854	BCL6	Tan et al. ²⁶	DLBCL	0.30	0.51	0.73	0.07	1.44	0.21	95%	CHN
rs79480871	NCOA1	Cerhan et al. ¹⁶	DLBCL	NA	NA	NA	NA	1.34	0.08	54%	EUR
rs9461741	BTNL2	Vijai et al. (2015) ²⁵	MZL	0.97	0.72	0.67	NA	2.66	0.03	99%	EUR

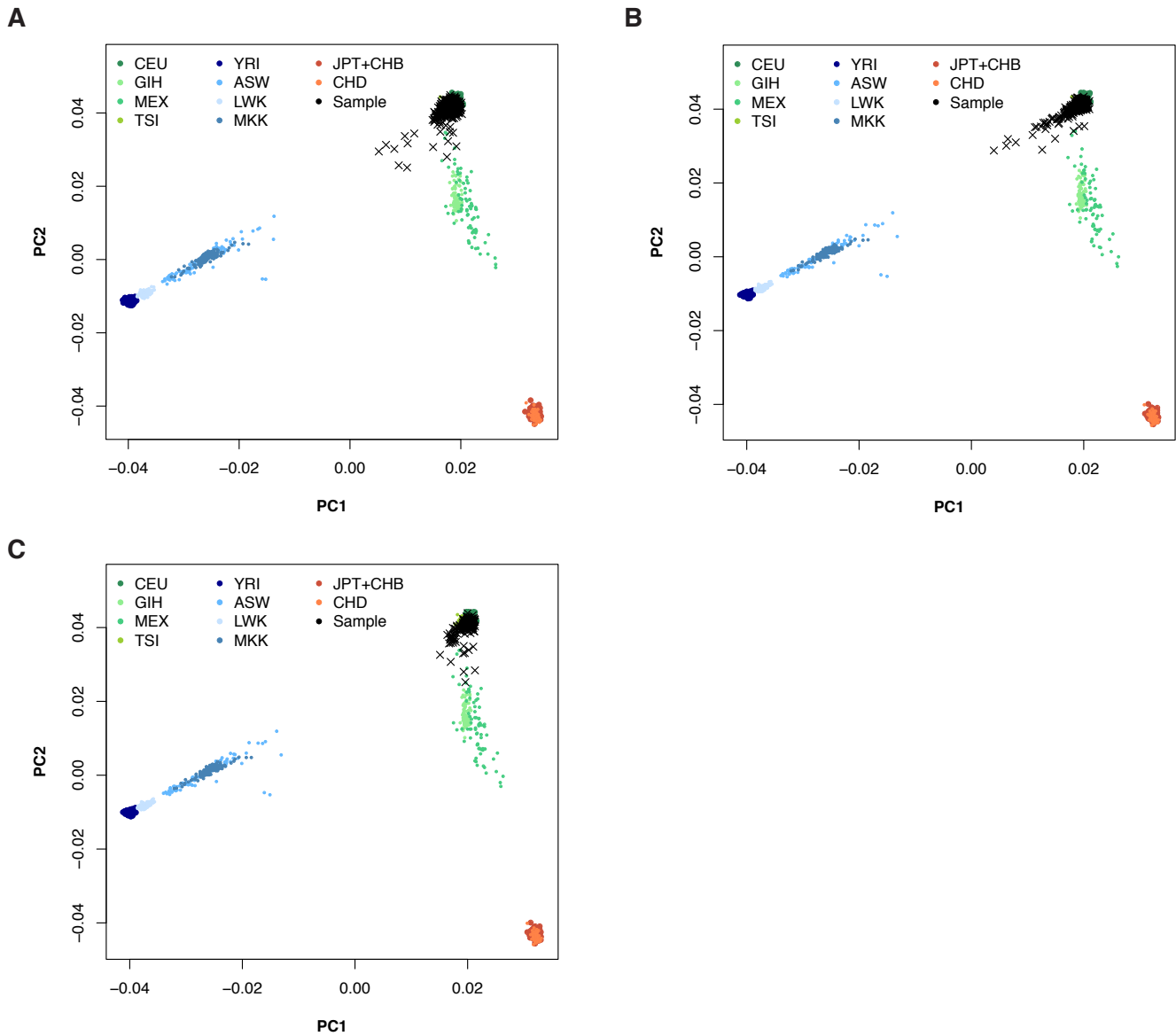
Comparisons with genome-wide significant variants identified in published GWAS of NHL in the general population. The NHL subtypes includes DLBCL (Diffuse large B-cell lymphoma), FL (Follicular lymphoma) and MZL (Marginal zone lymphoma). P-values for the HIV meta-analysis and the individual cohort GWAS are shown per variant. The statistical power to replicate the published variants under an additive model at $P < 0.05$, given their published odds ratios (OR) and the mean observed minor allele frequencies (MAF) in the HIV cohorts is also listed. Most of the published GWAS was on European (EUR) patients and Chinese (CHN).

Supplemental Table 4. The effect of rs7919208 in GWAS in the general population

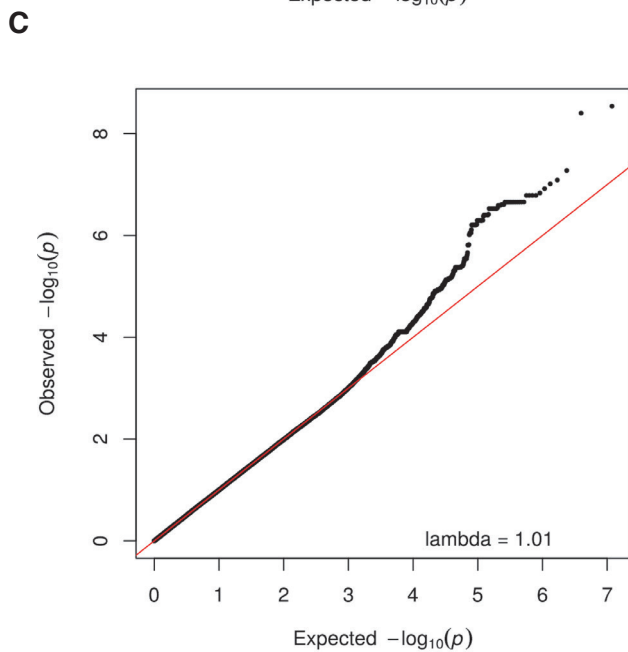
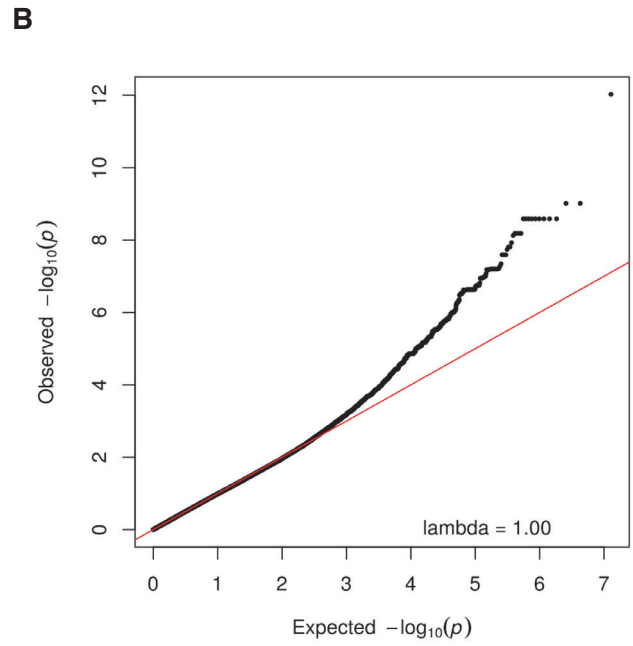
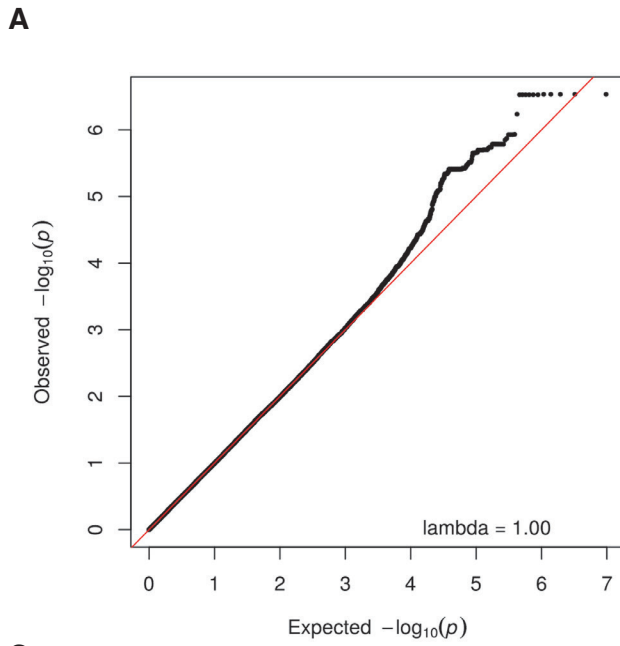
Subtype	Cases	Controls	Lambda	P (rs7919208)	OR (rs7919208)
CLL	1033	2635	1.04	0.29	0.97
DLBCL	2173	2635	1.04	0.65	1.01
FL	1753	2635	1.03	0.36	0.97
MZL	617	2635	1.01	0.56	0.98
Combined	5556	2635	1.04	0.60	0.99

The association of rs7919208 in the general (non-HIV) population across NHL subtypes. Lambda indicates the genome-wide inflation factor for the GWAS performed for each subtype to ensure the test-statistics observed are valid. The calculated p-values and odds ratios for rs7919208 are listed for each GWAS.

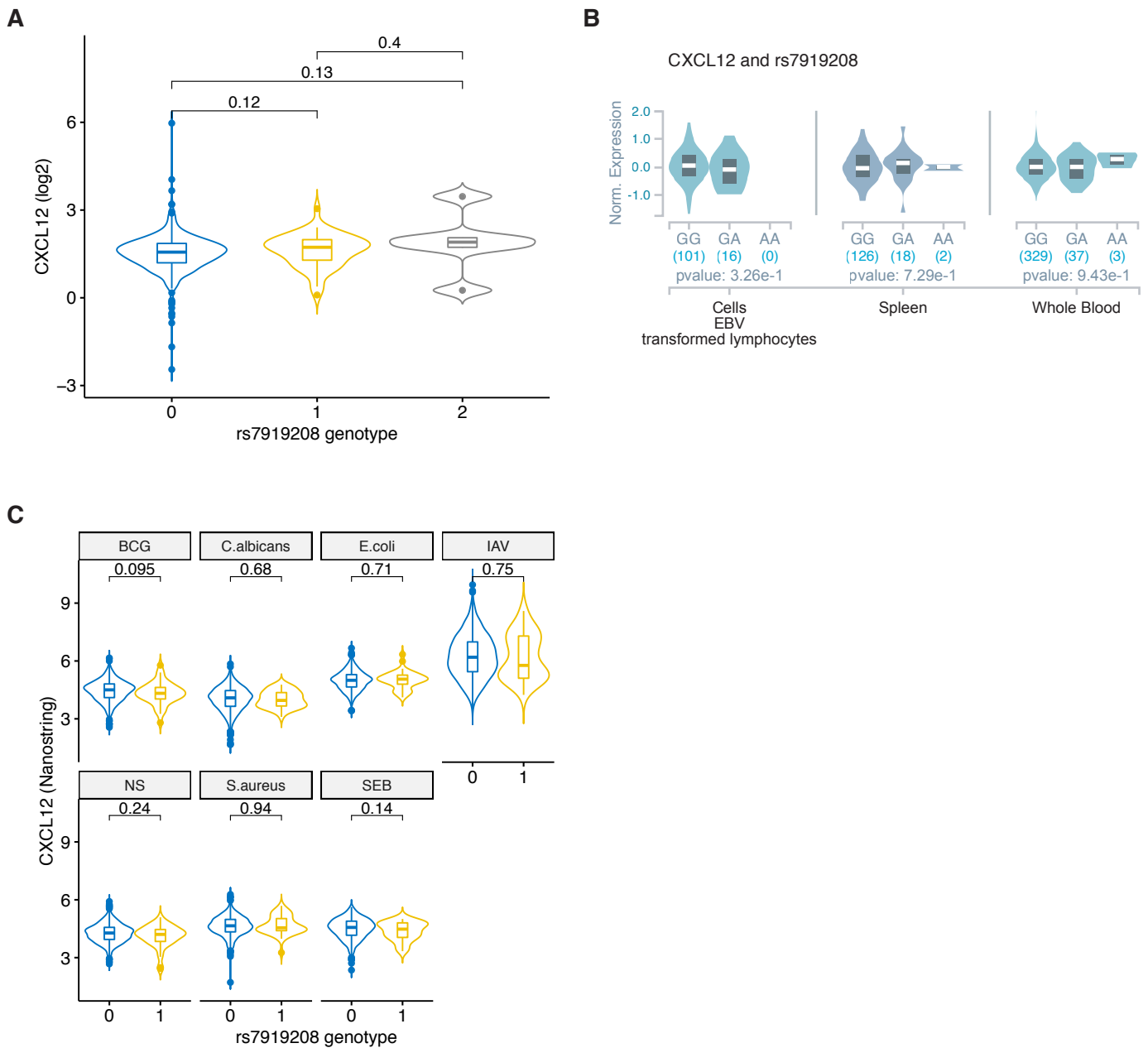
Supplementary Figures



Supplemental Figure 1. Principal component analyses (PCA) with the HapMap project. The black crosses represent individuals genotyped and included in this study. Individuals of European ancestry colocalizes with the HapMap reference samples from CEU (Northern Europeans from Utah) and TSI (Tuscans from Italy). (A) The SHCS cohort. (B) The ANRS cohort. (C) The MACS cohort.



Supplemental Figure 2. Quantile-quantile plots for the initial cohort level GWAS. Lambda indicates the genome-wide inflation factor. Values ~ 1 denotes the lack of genomic inflation due to confounding factors. (A) Plot for the GWAS in the SHCS cohort. (B) Plot or the GWAS in the ANRS cohort. (C) Plot for the GWAS in the ANRS cohort.



Supplemental Figure 3. eQTL information on rs7919208. (A) Expression levels of CXCL12 in EBV transformed lymphocytes from the GEUVADIS consortium according to the rs7919208 genotype. Differences between genotype groups were tested using Wilcoxon rank-sum tests with the obtained p-values shown on the figures. (B) Relationship between CXCL12 and the rs7919208 genotype in GTEx across EBV transformed lymphocytes, Spleen and Whole Blood. (C) Relationship between CXCL12 expression and rs1919208 using Nanostring in the Milieu Interieur Consortium for stimulated and non-stimulated (NS) PBMCs. Stimulants used were Mycobacterium bovis (BCG), Candida albicans, Escherichia coli, Influenza A virus (IAV), Staphylococcus aureus and Staphylococcal enterotoxin B (SEB).