SUPPLEMENTARY MATERIAL

A genome-wide association study of IgM antibody against phosphorylcholine: shared genetics and phenotypic relation to chronic lymphocytic leukemia

Xu Chen, Stefan Gustafsson, Thomas Whitington, Yan Borné, Erik Lorentzen, Jitong Sun, Peter Almgren, Jun Su, Robert Karlsson, Jie Song, Yi Lu, Yiqiang Zhan, Sara Hägg, Per Svensson, Karin E. Smedby, Susan L. Slager, Erik Ingelsson, Cecilia M. Lindgren, Andrew P. Morris, Olle Melander, Thomas Karlsson, Ulf de Faire, Kenneth Caidahl, Gunnar Engström, Lars Lind, Mikael C. I. Karlsson, Nancy L. Pedersen, Johan Frostegård, and Patrik K. E. Magnusson

Figure S1. Individual genome-wide association studies of IgM anti-PC in the discovery phase

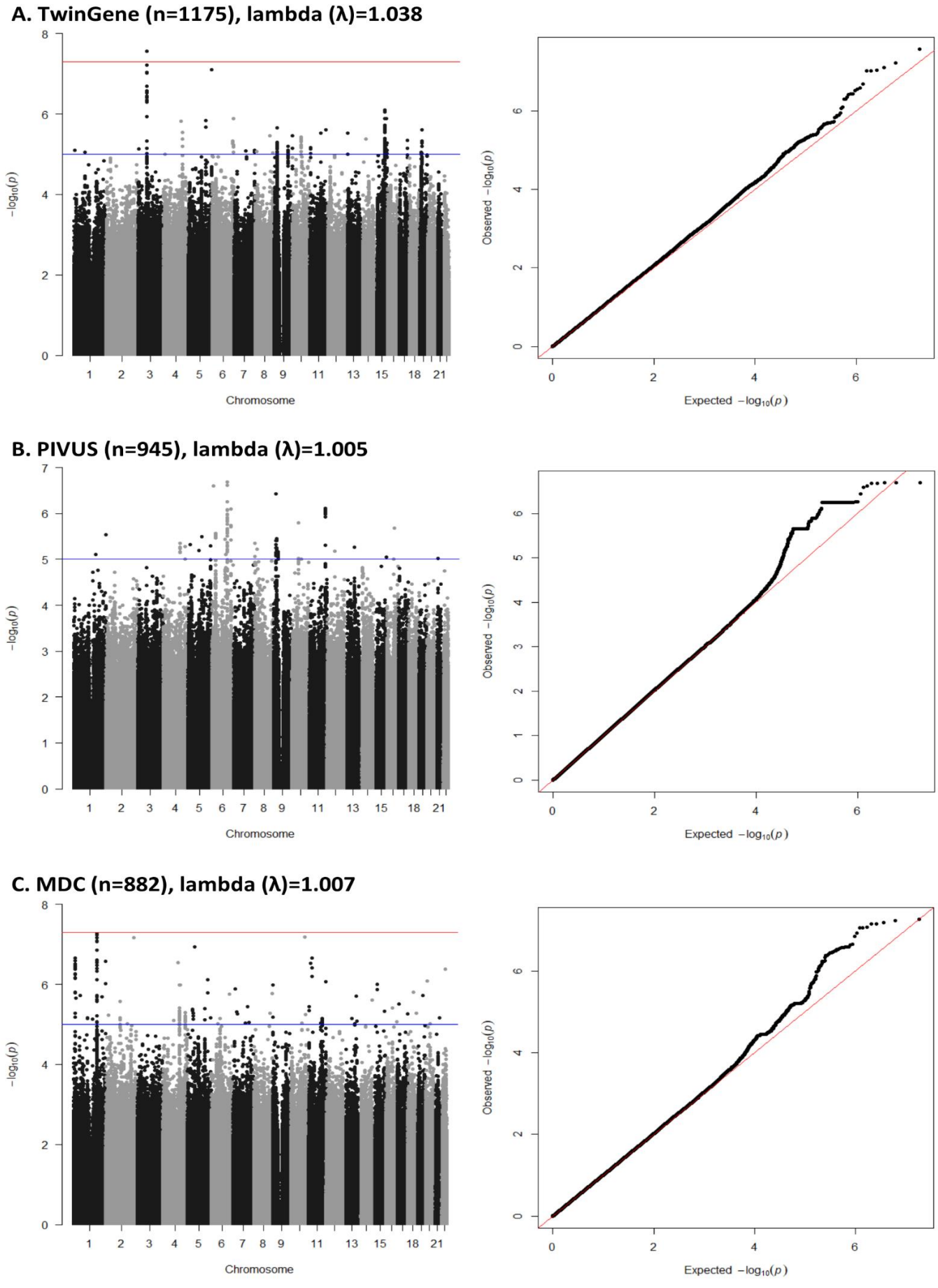


Figure S2. Marks of regulation in the LD block of 11q24.1 from UCSC Genome Browser

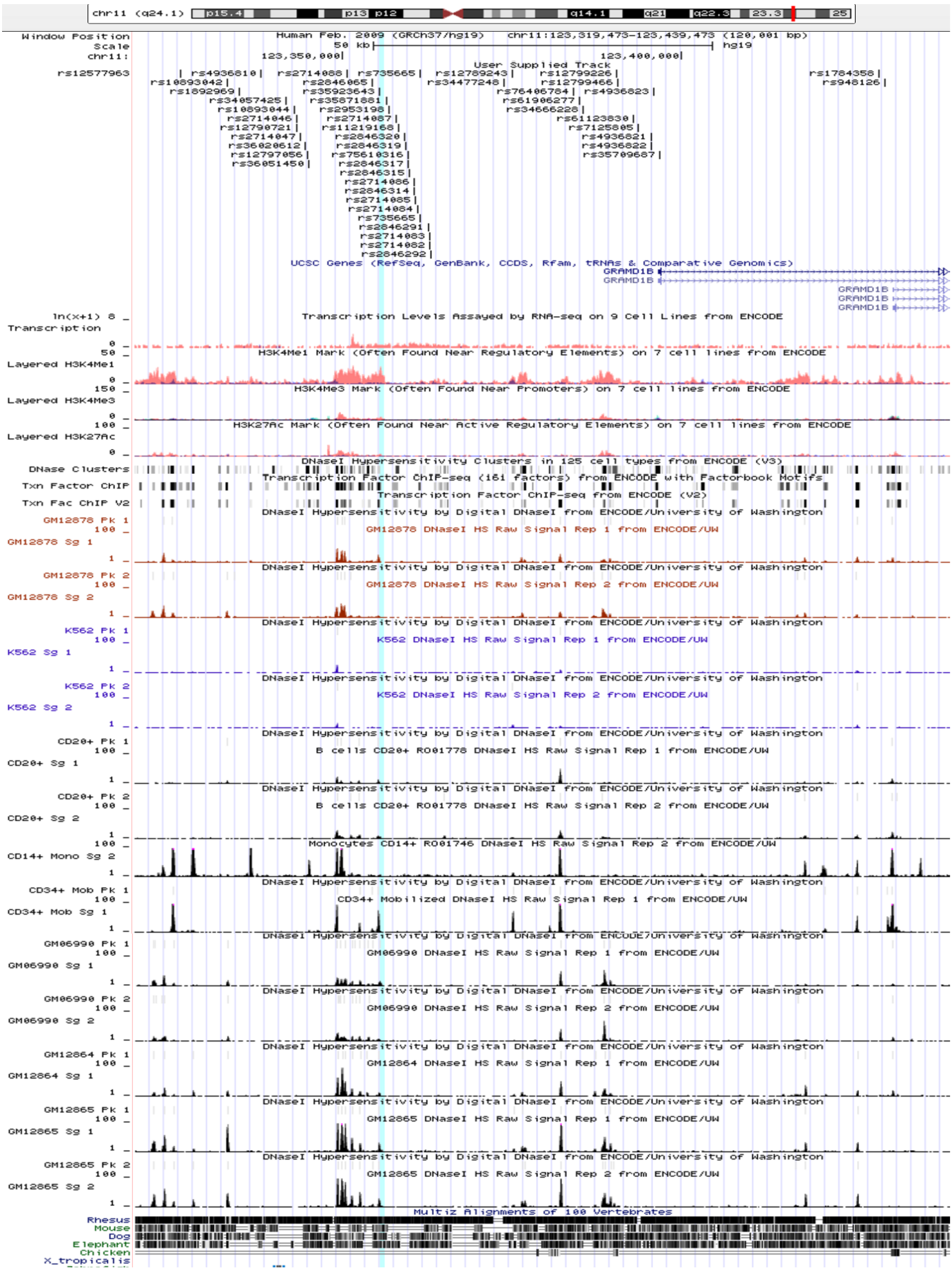


Figure S3. Functional prediction of rs35923643 in leukemia- or immune- cells

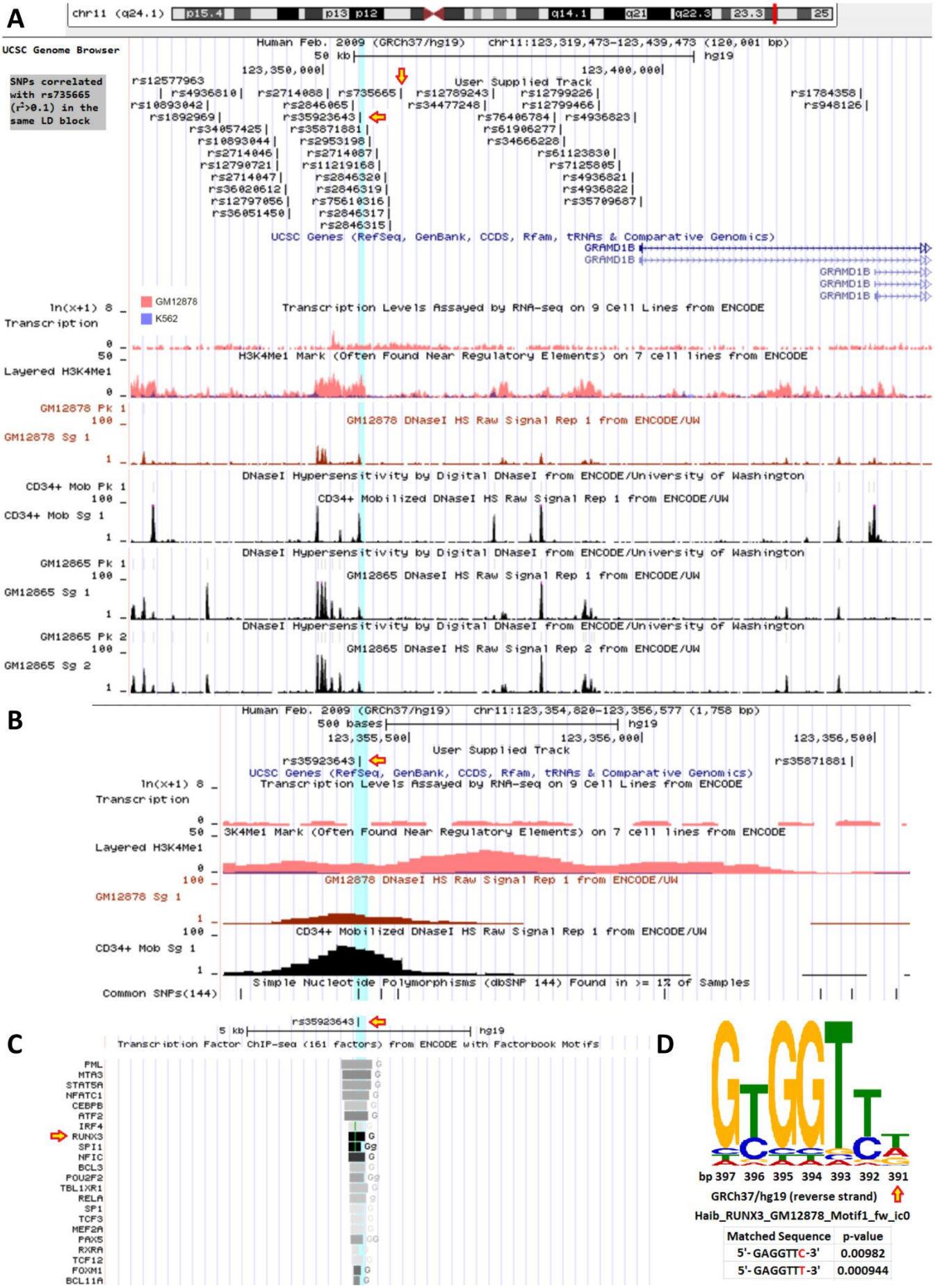
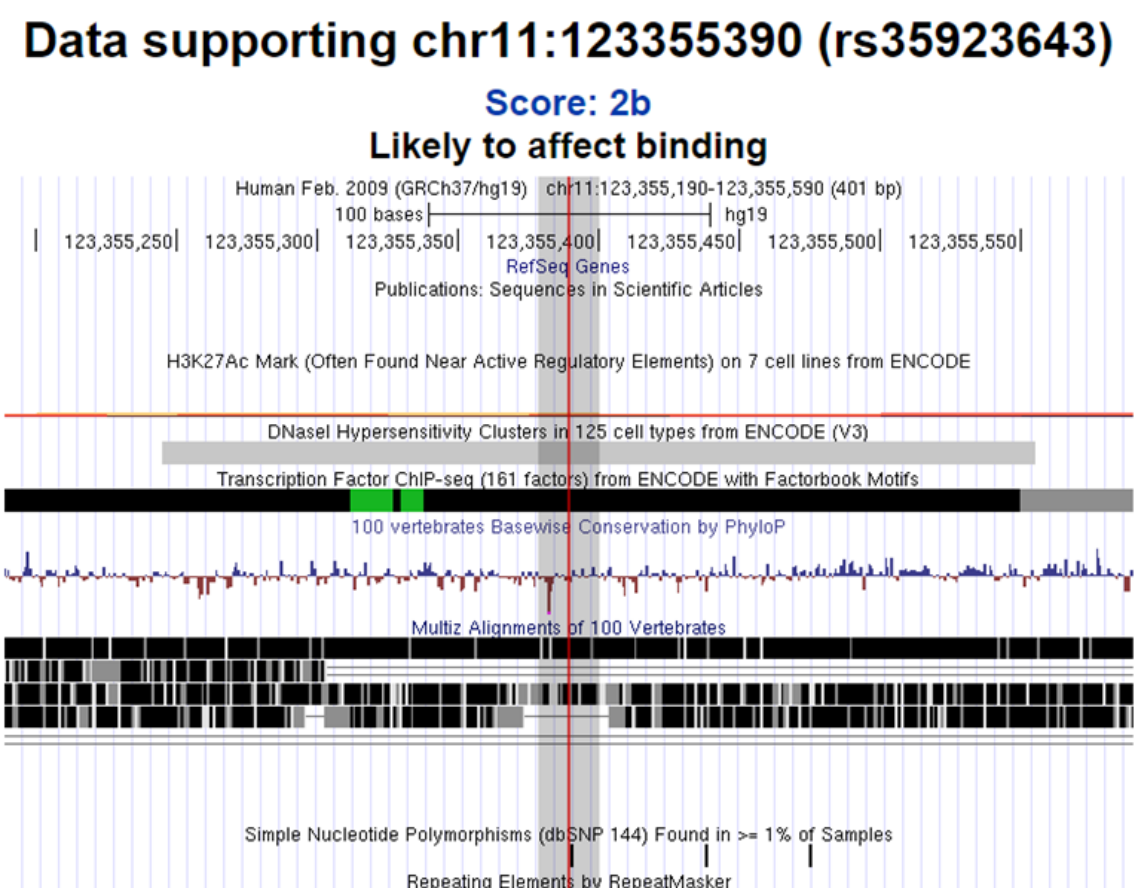


Figure S4. Evidence for affecting binding from RegulomeDB



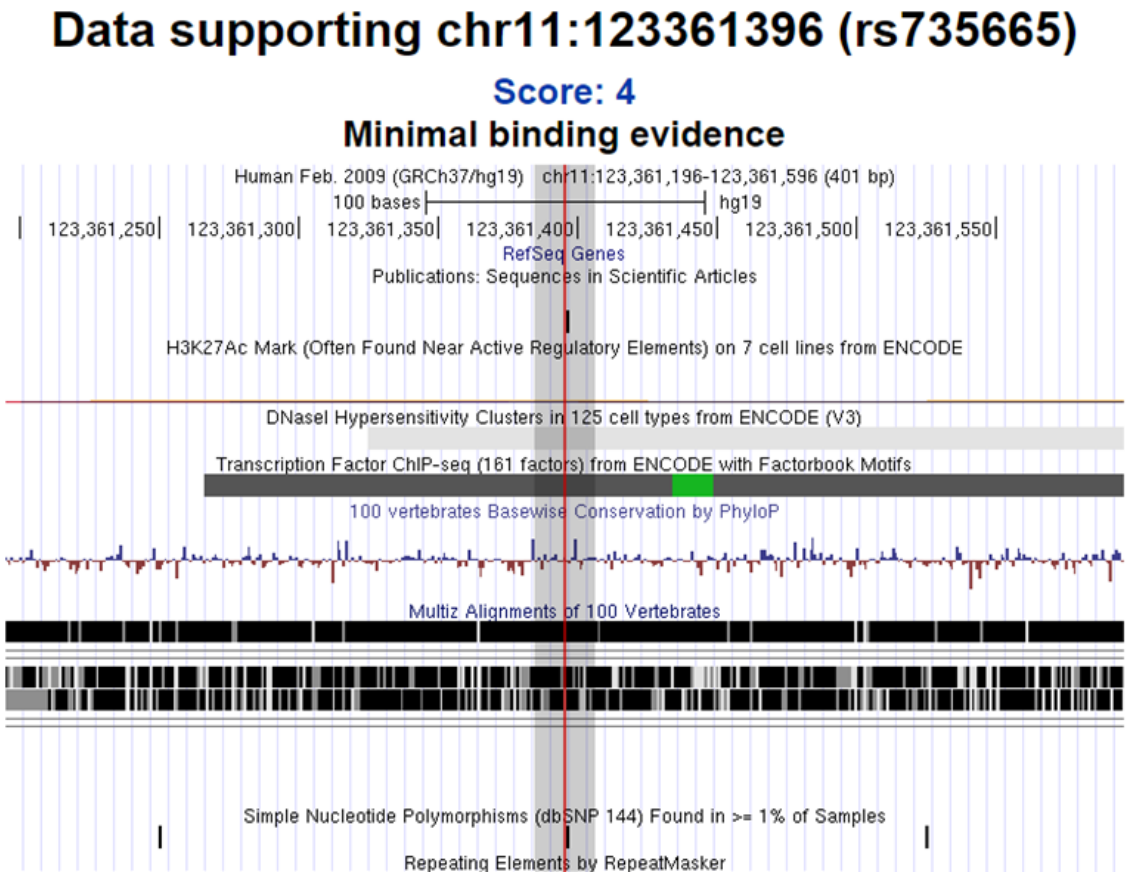


Figure S5. IgM anti-PC values before and after the first CLL diagnosis

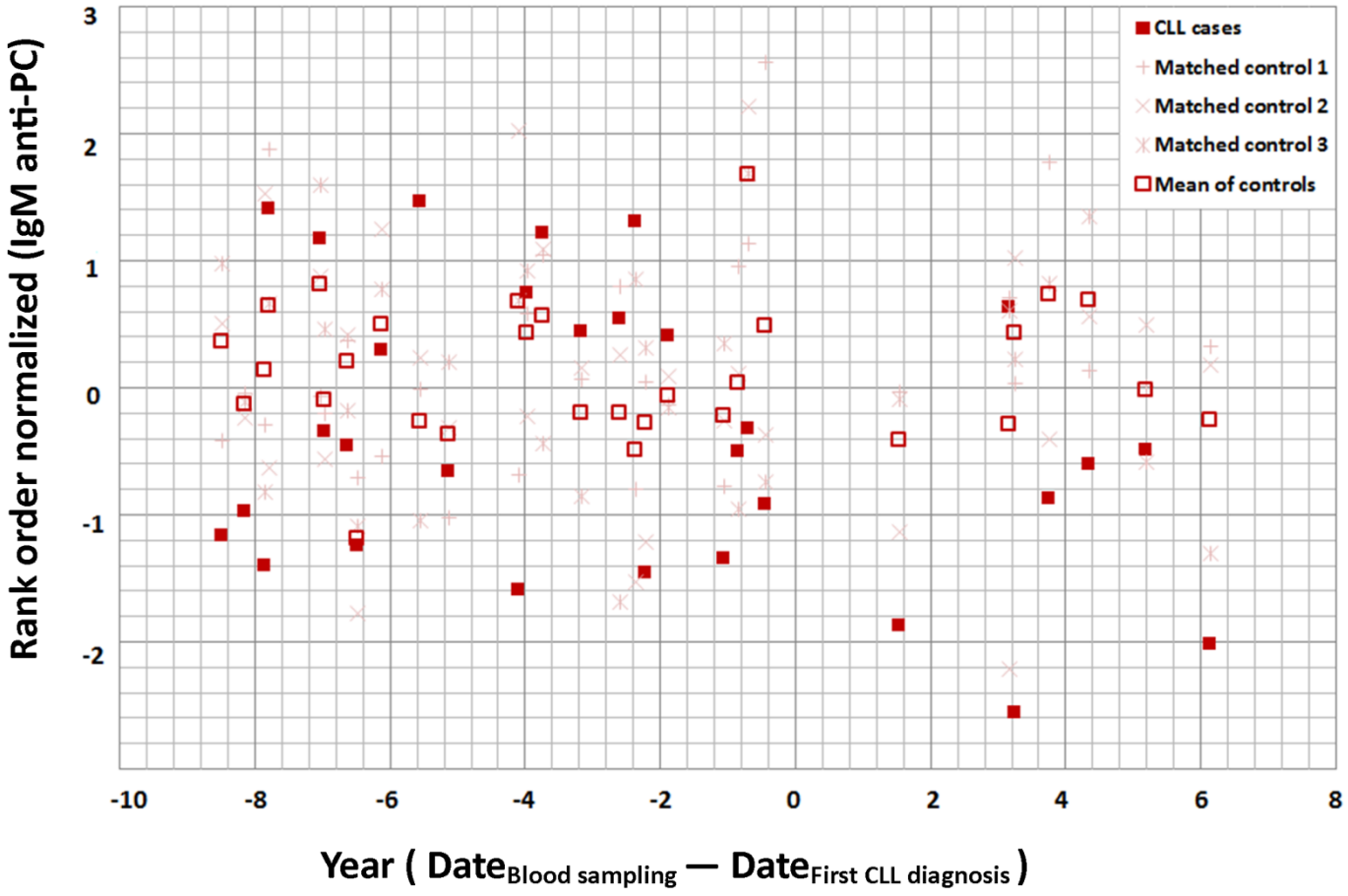


Figure S6. Global allele frequency of SNP rs735665 in the 1000 Genomes Project Phase 3 from Ensembl (GRCh37/hg19)

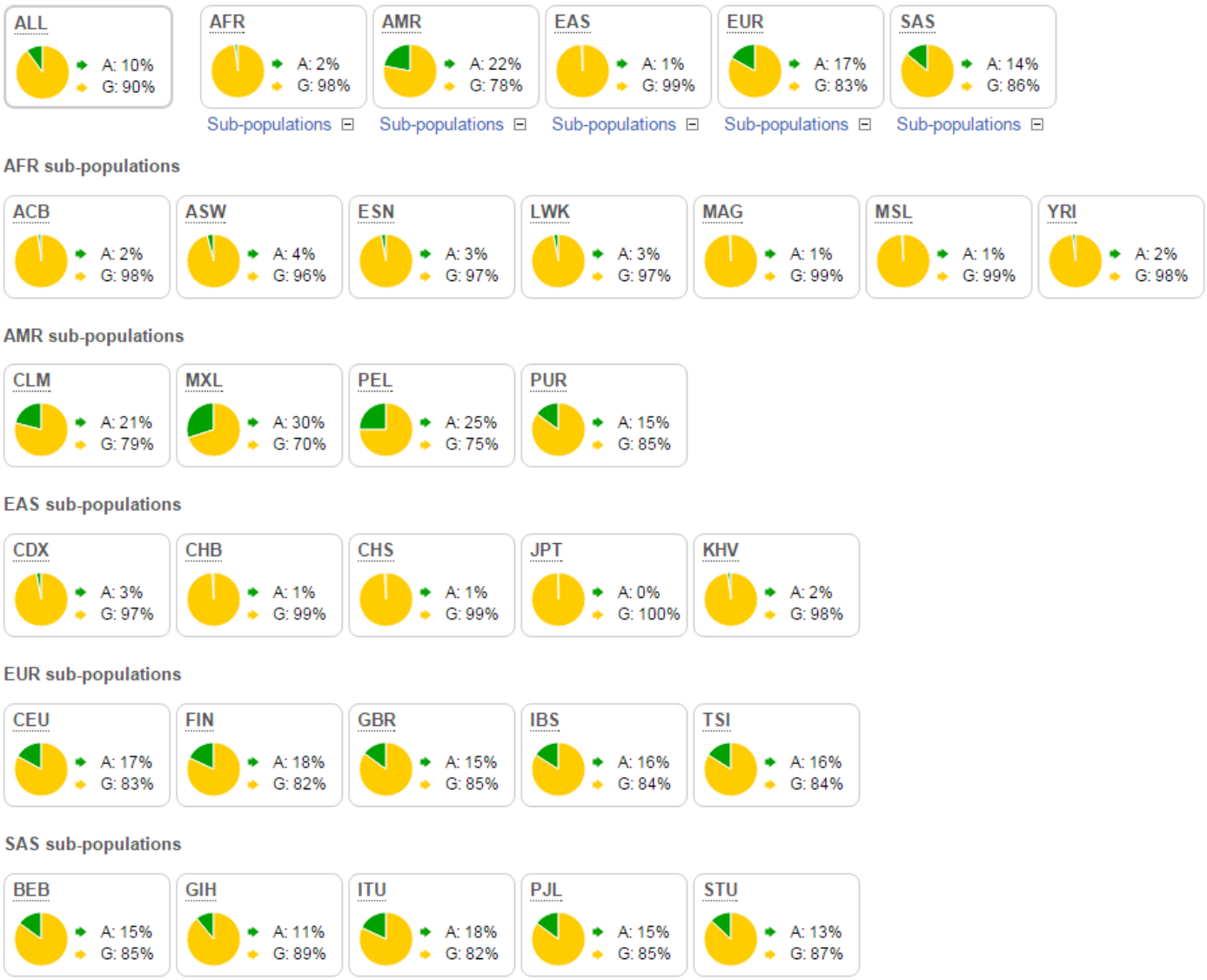
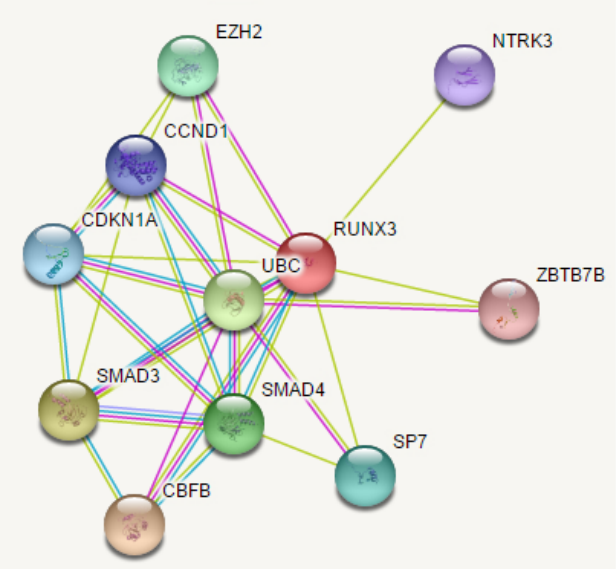


Figure S7. Functional partners with RUNX3 predicted from STRING database (version 10)



runt-related transcription factor 3; CBF binds to the core site, 5'-PYGPYGGT-3', of a number of enhancers and promoters, including murine leukemia virus, polyomavirus enhancer, T-cell receptor enhancers, lck, IL-3 and GM-CSF promoters

Identifier: ENSP00000343477, RUNX3
Organism: Homo sapiens

- show protein sequence
- homologs among STRING organisms
- Pathways, Functions, Resources (GeneCards)

Nodes:

Network nodes represent proteins

splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Size

small nodes: protein of unknown 3D structure

large nodes: some 3D structure is known or predicted

Node Color

colored nodes: query proteins and first shell of interactors

white nodes: second shell of interactors

Edges:

Edges represent protein-protein associations

associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.

Known Interactions

from curated databases

experimentally determined

Predicted Interactions

gene neighborhood

gene fusions

gene co-occurrence

Others

textmining

co-expression

protein homology

Your Input:

		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
RUNX3	*runt-related transcription factor 3; CBF binds to the core site, 5'-PYGPYGGT-3', of a number of enhancers and promoters, including murine leukemia virus, polyomavirus enhancer, T-cell receptor enhancers, lck, IL-3 and GM-CSF promoters (429 aa)*									

Predicted Functional Partners:

CBFB	*core-binding factor, beta subunit; CBF binds to the core site, 5'-PYGPYGGT-3', of a number of enhancers and promoters, includ...*					●	●	●		0.996
SMAD3	*SMAD family member 3 (425 aa)*					●	●	●		0.996
UBC	*ubiquitin C (685 aa)*					●				0.994
SMAD4	*SMAD family member 4; Common SMAD (co-SMAD) is the coactivator and mediator of signal transduction by TGF-beta (trans...*						●	●		0.987
EZH2	*enhancer of zeste homolog 2 (Drosophila) (751 aa)*					●		●		0.951
SP7	*Sp7 transcription factor; Transcriptional activator essential for osteoblast differentiation. Binds to SP1 and EKLF consensus s...*							●		0.947
CDKN1A	*cyclin-dependent kinase inhibitor 1A (p21, Cip1); May be the important intermediate by which p53/TP53 mediates its role as a...*							●		0.944
CCND1	*cyclin D1; Regulatory component of the cyclin D1-CDK4 (DC) complex that phosphorylates and inhibits members of the retino...*					●		●		0.939
NTRK3	*neurotrophic tyrosine kinase, receptor, type 3; Receptor for neurotrophin-3 (NT-3). This is a tyrosine- protein kinase receptor. K...*							●		0.933
ZBTB7B	*zinc finger and BTB domain containing 7B; Transcription regulator that acts as a key regulator of lineage commitment of imm...*							●		0.928

Figure S8. Potential functional pathways RUNX3 involved from STRING database

Network Stats

number of nodes: 11
number of edges: 29
average node degree: 5.27
clustering coefficient: 0.842

expected number of edges: 19
PPI enrichment p-value: 0.0195
your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

Note: some enrichments may be expected here (why?)

Biological Process (GO)

pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
GO:0010604	positive regulation of macromolecule metabolic process	10	6e-05
GO:2000738	positive regulation of stem cell differentiation	4	6e-05
GO:0045786	negative regulation of cell cycle	6	0.000151
GO:0045930	negative regulation of mitotic cell cycle	5	0.000187
GO:0001932	regulation of protein phosphorylation	7	0.000273

(more ...)

Molecular Function (GO)

pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
GO:0005072	transforming growth factor beta receptor, cytoplasmic mediator activity	2	0.0121
GO:0044212	transcription regulatory region DNA binding	5	0.0121
GO:0043565	sequence-specific DNA binding	5	0.0274
GO:0070412	R-SMAD binding	2	0.0274
GO:0016538	cyclin-dependent protein serine/threonine kinase regulator activity	2	0.0323

Cellular Component (GO)

pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
GO:0071141	SMAD protein complex	2	0.00258
GO:0000790	nuclear chromatin	4	0.00452
GO:0000228	nuclear chromosome	4	0.00902
GO:0000307	cyclin-dependent protein kinase holoenzyme complex	2	0.00902
GO:0031981	nuclear lumen	8	0.00902

KEGG Pathways

pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
05220	Chronic myeloid leukemia	3	0.00171
04068	FoxO signaling pathway	3	0.00304
04110	Cell cycle	3	0.00304
05161	Hepatitis B	3	0.00304
05206	MicroRNAs in cancer	3	0.00304

(more ...)

PFAM Protein Domains

pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
PF03166	MH2 domain	2	0.00471
PF03165	MH1 domain	2	0.00589

INTERPRO Protein Domains and Features

pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
IPR001132	SMAD domain, Dwarfin-type	2	0.00336
IPR013019	MAD homology, MH1	2	0.00336
IPR013790	Dwarfin	2	0.00336
IPR003619	MAD homology 1, Dwarfin-type	2	0.00503
IPR017855	SMAD domain-like	2	0.00503

(more ...)

Statistical background

For the above enrichment analysis, the following statistical background is assumed: Whole Genome ▼ UPDATE

Table S1. Genome-wide significant SNPs in the discovery GWAS meta-analysis of IgM anti-PC

SNP	CHR	POS	A1	A2	FREQ_ A1	INFO	Beta	SE	I^2(%)	Phet	P-value
rs35871881	**11**	**123356451**	**G**	**A**			**0.174**	**0.031**	**61.2**	**0.08**	**1.95e-08**
TwinGene			G	A	0.250	0.986	0.158	0.049			0.001
PIVUS			G	A	0.241	0.970	0.263	0.053			7.72e-07
MDC			G	A	0.215	0.969	0.086	0.059			0.145
rs12798553	**11**	**123353693**	**A**	**G**			**0.174**	**0.031**	**61.0**	**0.08**	**1.98e-08**
TwinGene			A	G	0.248	0.980	0.160	0.049			0.001
PIVUS			A	G	0.240	0.969	0.263	0.053			8.58e-07
MDC			A	G	0.215	0.967	0.085	0.059			0.151
rs35923643	**11**	**123355391**	**G**	**A**			**0.177**	**0.032**	**53.9**	**0.11**	**2.51e-08**
TwinGene			G	A	0.230	0.999	0.153	0.050			0.002
PIVUS			G	A	0.225	0.980	0.263	0.054			1.18e-06
MDC			G	A	0.188	0.987	0.101	0.061			0.097
rs141373354	**1**	**68736970**	**A**	**G**			**0.578**	**0.104**	**0.0**	**0.71**	**2.60e-08**
TwinGene			A	G	0.980	0.910	0.625	0.159			9.22e-05
PIVUS			A	G	0.986	0.901	0.658	0.203			0.001
MDC			A	G	0.980	0.913	0.457	0.181			0.011
rs116361873	**1**	**68728886**	**G**	**A**			**0.576**	**0.104**	**0.0**	**0.68**	**2.99e-08**
TwinGene			G	A	0.981	0.918	0.626	0.159			8.83e-05
PIVUS			G	A	0.986	0.910	0.660	0.202			0.001
MDC			G	A	0.981	0.935	0.446	0.181			0.014
rs735665	**11**	**123361397**	**A**	**G**			**0.174**	**0.031**	**54.3**	**0.11**	**3.12e-08**
TwinGene			A	G	0.232	NA	0.145	0.050			0.004
PIVUS			A	G	0.229	NA	0.261	0.053			1.05e-06
MDC			A	G	0.188	NA	0.103	0.060			0.089
rs36020612	**11**	**123344435**	**T**	**C**			**0.182**	**0.033**	**61.1**	**0.08**	**3.73e-08**
TwinGene			T	C	0.222	0.963	0.161	0.052			0.002
PIVUS			T	C	0.218	0.909	0.279	0.057			9.47e-07
MDC			T	C	0.186	0.900	0.090	0.064			0.156
rs36051450	**11**	**123344959**	**C**	**T**			**0.179**	**0.033**	**60.6**	**0.08**	**4.26e-08**
TwinGene			C	T	0.224	0.962	0.158	0.052			0.002
PIVUS			C	T	0.221	0.932	0.274	0.056			1.00e-06
MDC			C	T	0.187	0.924	0.089	0.063			0.154

CHR: chromosome number; POS: position of the SNP in human genome (GRCh37/hg19); A1: effect allele; A2: non-effect allele; FREQ_A1: frequency of effect allele; INFO: imputation quality info; Beta: effect size per standard deviation of rank order normalized IgM anti-PC per allele; SE: standard error; I^2: the percentage of variation across studies that is due to heterogeneity rather than chance; Phet: P-value for heterogeneity; NA: not available, because SNP was directly genotyped; PIVUS: Prospective Investigation of the Vasculature in Uppsala Seniors; MDC: Malmö Diet and Cancer study.

Table S2. Replication results for the eight genome-wide significant SNPs in PRACSIS

SNP	CHR	POS	A1	A2	FREQ_ A1	INFO	Beta	SE	P-value
rs35871881	11	123356451	G	A	0.230	0.953	0.184	0.067	0.0064
rs12798553	11	123353693	A	G	0.229	0.946	0.185	0.068	0.0064
rs35923643	11	123355391	G	A	0.205	0.975	0.247	0.069	0.0004
rs141373354	1	68736970	A	G	0.982	0.883	0.305	0.219	0.1644
rs116361873	1	68728886	G	A	0.982	0.904	0.306	0.219	0.1640
rs735665	11	123361397	A	G	0.206	0.993	0.244	0.068	0.0004
rs36020612	11	123344435	T	C	0.199	0.933	0.250	0.072	0.0005
rs36051450	11	123344959	C	T	0.201	0.947	0.250	0.071	0.0004

CHR: chromosome number; POS: position of the SNP in human genome (GRCh37/hg19); A1: effect allele; A2: non-effect allele; FREQ_A1: frequency of effect allele; INFO: imputation quality info; Beta: effect size per standard deviation of rank order normalized IgM anti-PC per allele; SE: standard error.

Table S3. Meta-analysis for the successfully replicated SNPs (sorted by association P-value)

SNP (Position)	Study	Info	Beta	SE	P-value	Phet
rs35923643 G/A	Meta		0.189	0.029	4.34×10^{-11}	0.159
(Chr11: 123355391)	*TwinGene*	0.999	0.153	0.050	0.002	
	PIVUS	0.980	0.263	0.054	1.18×10^{-6}	
	MDC	0.987	0.101	0.061	0.097	
	PRACSIS	0.975	0.247	0.069	0.0004	
rs735665 A/G	Meta		0.186	0.028	5.03×10^{-11}	0.150
(Chr11: 123361397)	*TwinGene*	NA	0.145	0.050	0.004	
	PIVUS	NA	0.261	0.053	1.05×10^{-6}	
	MDC	NA	0.103	0.060	0.089	
	PRACSIS	0.993	0.244	0.068	0.0004	
rs36020612 T/C	Meta		0.193	0.030	1.06×10^{-10}	0.118
(Chr11: 123344435)	*TwinGene*	0.963	0.161	0.052	0.002	
	PIVUS	0.909	0.279	0.057	9.47×10^{-7}	
	MDC	0.900	0.090	0.064	0.156	
	PRACSIS	0.933	0.250	0.072	0.0005	
rs36051450 C/T	Meta		0.191	0.030	1.09×10^{-10}	0.116
(Chr11: 123344959)	*TwinGene*	0.962	0.158	0.052	0.002	
	PIVUS	0.932	0.274	0.056	1.00×10^{-6}	
	MDC	0.924	0.089	0.063	0.154	
	PRACSIS	0.947	0.250	0.071	0.0004	
rs12798553 A/G	Meta		0.176	0.028	3.17×10^{-10}	0.158
(Chr11: 123353693)	*TwinGene*	0.980	0.160	0.049	0.001	
	PIVUS	0.969	0.263	0.053	8.58×10^{-7}	
	MDC	0.967	0.085	0.059	0.151	
	PRACSIS	0.946	0.185	0.068	0.0064	
rs35871881 G/A	Meta		0.176	0.028	3.25×10^{-10}	0.160
(Chr11: 123356451)	*TwinGene*	0.986	0.158	0.049	0.001	
	PIVUS	0.970	0.263	0.053	7.72×10^{-7}	
	MDC	0.969	0.086	0.059	0.145	
	PRACSIS	0.953	0.184	0.067	0.0064	

SNP is presented with effect allele/alternative allele (chromosome number and position in human genome GRCh37/ hg19). Info: imputation quality; NA: not available, because SNP is directly genotyped; Beta: effect size per standard deviation of rank order normalized IgM anti-PC per allele; SE: standard error; Phet: P-value for heterogeneity; The two SNPs (rs141373354 and rs116361873) on chromosome 1 are not successfully replicated in PRACSIS, association P-values are 0.1644 and 0.1640, respectively.

Table S4. Levels of IgM anti-PC among genotypes of rs735665 and rs35923643

		Discovery phase			Replication
		TwinGene	*PIVUS*	*MDC*	*PRACSIS*
Genotypes of rs735665	**N_{GG}**	697	568	556	407
	N_{GA}	411	322	254	212
	N_{AA}	67	55	33	27
Raw values of IgM anti-PC	**GG**	40.3 (23.2-67.2)	39.5 (24.8-68.1)	46.0 (28.6-75.1)	31.9 (19.7-56.3)
	GA	48.7 (23.8-83.2)	45.8 (28.9-75.6)	51.1 (32.8-75.5)	39.7 (24.1-71.0)
	AA	40.6 (21.1-78.6)	52.0 (34.6-110.9)	58.6 (34.4-78.1)	47.1 (28.9-59.7)
Normalized values of IgM anti-PC in GWAS	**GG**	-0.07±0.99	-0.11±0.98	-0.05±0.99	-0.11±1.01
	GA	0.14±1.03	0.13±0.99	0.05±1.00	0.18±0.98
	AA	0.01±1.08	0.43±1.09	0.22±0.86	0.23±0.69
Genotypes of rs35923643	**N_{AA}**	705	564	583	409
	N_{AG}	403	312	266	210
	N_{GG}	67	53	33	27
Raw values of IgM anti-PC	**AA**	40.0 (23.1-66.8)	39.4 (24.8-68.1)	47.2 (28.9-75.5)	31.9 (19.6-56.2)
	AG	50.7 (24.0-83.9)	46.8 (29.1-76.3)	51.7 (32.8-77.4)	39.7 (24.1-71.0)
	GG	40.6 (21.1-78.6)	49.8 (34.6-107.9)	58.6 (34.4-78.1)	47.1 (28.9-59.7)
Normalized values of IgM anti-PC in GWAS	**AA**	-0.08±0.99	-0.11±0.98	-0.04±1.00	-0.11±1.01
	AG	0.16±1.03	0.14±0.99	0.05±1.01	0.18±0.98
	GG	0.01±1.08	0.41±1.10	0.22±0.86	0.23±0.69

PIVUS: Prospective Investigation of the Vasculature in Uppsala Seniors; MDC: Malmö Diet and Cancer study; PRACSIS: Prognosis and Risk in Acute Coronary Syndromes in Sweden.

Table S5. Polygenic risk score analyses between immunoglobulins and CLL risk

Base — Target	P_T	N_{SNPs}	OR	SE	r^2	P-value
IgA — CLL	5.0e-15	2	1.06	0.17	0.00%	0.75
	5.0e-10	10	0.89	0.11	0.01%	0.29
	5.0e-09	22	0.87	0.09	0.02%	0.12
	5.0e-08*	**31**	**0.88**	**0.08**	**0.02%**	**0.11**
	5.0e-07	41	0.93	0.07	0.00%	0.31
	1.0e-06	45	0.91	0.07	0.02%	0.20
IgG — CLL	5.0e-15	5	0.69	0.10	0.12%	0.007
	5.0e-10	10	0.68	0.10	0.14%	0.0003
	5.0e-09	12	0.67	0.10	0.15%	0.0001
	5.0e-08	17	0.73	0.09	0.12%	0.0004
	5.0e-07*	**20**	**0.68**	**0.09**	**0.18%**	**3.9e-06**
	1.0e-06	22	0.67	0.09	0.20%	1.2e-05
IgM — CLL	5.0e-15	5	0.84	0.16	0.01%	0.27
	5.0e-10	18	0.70	0.09	0.14%	0.0001
	5.0e-09	33	0.72	0.07	0.19%	5.3e-06
	5.0e-08*	**43**	**0.69**	**0.07**	**0.28%**	**4.2e-08**
	5.0e-07	59	0.79	0.06	0.14%	0.0001
	1.0e-06	66	0.75	0.06	0.23%	8.4e-07
IgM anti-PC — CLL	**5.0e-08***	**1**	**17.55**	**0.36**	**0.59%**	**1.2e-15**
	5.0e-07*	**1**	**17.55**	**0.36**	**0.59%**	**1.2e-15**
	1.0e-06*	**1**	**17.55**	**0.36**	**0.59%**	**1.2e-15**
	5.0e-06	6	1.87	0.11	0.31%	7.3e-09
	5.0e-05	39	1.11	0.04	0.05%	0.02
	5.0e-04	386	1.00	0.02	0.00%	0.87
	0.005	3237	1.00	0.01	0.00%	0.90
	0.05	22405	1.02	0.003	0.24%	3.1e-07
	0.5	137132	1.01	0.002	0.10%	0.001

Standardized polygenic risk score (PRS) of the base phenotype was used to predict the target phenotype. P_T: P-value threshold of the association between single nucleotide polymorphisms (SNPs) and the base phenotype, * means the best P_T that defined as the threshold corresponding to the largest explained variance; N_{SNPs}: number of independent SNPs included in the best P_T quantile; OR: odds ratio per standard deviation; SE: standard error; r^2: Nagelkerke r^2, the proportion of target variation explained by PRS of SNPs in the best P_T quantile.

Table S6. Signal for each transcription factor and matched sequence from FIMO

TF	Signal	Motif	Strand	Start	End	P-value	Q-value	Matched Sequence
ATF2	487	M1	-	389	399	0.00986	1	CTGAGGTTTTG
BCL11A	426	×						
BCL3	202	×						
CEBPB	264	×						
FOXM1	549	×						
IRF4	160	M2	+	381	395	0.00665	1	CCAACCCCCAAAACC
		M3	+	384	398	0.00369	1	ACCCCCAAAACCTCA
MEF2A	144	×						
MTA3	496	×						
NFATC1	331	×						
NFIC	732	M4	-	386	391	0.00862	0.528	TTGGGG
PAX5	388	M5	-	378	397	0.00555	0.499	GAGGTTTTGGGGGTTGGAGG
PML	319	×						
POU2F2	327	×						
RELA	228	×						
RUNX3	1000	M6	-	391	397	0.000944	0.216	GAGGTTT
		M7	+	391	400	0.00029	0.208	AAACCTCAGA
		M7	+	382	391	0.00409	0.602	CAACCCCCAA
		M8	+	384	399	0.000894	0.301	ACCCCCAAAACCTCAG
		M8	+	383	398	0.00279	0.335	AACCCCCAAAACCTCA
		M8	+	376	391	0.00289	0.335	CCCCTCCAACCCCCAA
		M9	+	382	399	4.5e-05	0.0607	CAACCCCCAGAACCTCAG
		M9	+	382	399	1.86e-05	0.0251	CAACCCCCAAAACCTCAG
		M9	+	391	408	0.00673	0.479	AAACCTCAGAAATTGGCT
		M10	+	391	400	0.000414	0.229	AAACCTCAGA
		M10	+	383	392	0.000972	0.341	AACCCCCAAA
		M10	+	382	391	0.00156	0.369	CAACCCCCAA
RXRA	131	M11	-	382	396	0.0012	0.423	AGGTTTTGGGGGTTG
		M11	-	381	395	0.00358	0.834	GGTTTTGGGGGTTGG
		M12	-	389	405	0.00907	0.699	CAATTTCTGAGGTTTTG
SP1	156	×						
SPI1	1000	M13	+	314	328	8.52e-06	0.0117	ATGAGGAACAGAGAG
		M14	+	314	324	0.000166	0.226	ATGAGGAACAG
STAT5A	357	M15	-	375	390	0.00373	0.367	TGGGGGTTGGAGGGGG
		M15	-	385	400	0.00495	0.367	TCTGAGGTTTTGGGGG
		M15	-	387	402	0.00503	0.367	TTTCTGAGGTTTTGGG
		M15	-	377	392	0.00982	0.431	TTTGGGGGTTGGAGGG
TBL1XR1	220	×						
TCF12	154	×						
TCF3	131	×						

The position of rs35923643 (at the 391bp of submitted sequence) is labeled in red. M1: Haib_ATF2_H1-hESC_ Motif1 _fw_ic0; M2: IRF4_full_TCAAGG20NCG_AD_NCGAAACCGAAACYN_fw_ic0; M3: Irf4_fw_ic0; M4: MA0161_NFIC_fw _ic0; M5: MA0014_Pax5_fw_ic0; M6: Haib_RUNX3_GM12878_Motif1_fw_ic0; M7: RUNX3_ DBD_TACGGA30NGGC_AI_NAACCGCAAN_fw_ic0; M8: RUNX3_DBD_TACGGA30NGGC_AI_NA ACCGC AAACCRCAN_fw_ic0; M9: RUNX3_DBD_TACGGA30NGGC_AI_WAACCRCAAWAACCRCAN_fw_ ic0; M10: RUNX3_full_TCTCCC20NGA_AE_NAACCRCAAN_fw_ic0; M11: MA0074_RXRA_VDR_fw_ic0; M12: MA0115_NR1H2_RXRA_fw_ic0; M13: Haib_SPI1_GM12878_Motif1_fw_ic0; M14: Haib_SPI1_K562_Motif1_ fw_ic0; M15: Haib_STAT5A _GM12878_Motif1_fw_ic0.