Insights into Type 1 Diabetes and Multiple Sclerosis Pathogenesis from Common Genetic Associations

Sandeep K. Singh [1], Aziz A. Chentoufi [2], Jacob L. McCauley [3], Mehmet Tevfik Dorak [4]

[1]Florida International University, Miami, FL, USA; [2]King Fahad Medical City, Riyadh, Saudi Arabia; [3]University of Miami, John P. Hussman Institute for Human Genomics, Miami, FL, USA; [4]School of Health Sciences, Liverpool Hope University, Liverpool, U.K.

BACKGROUND

The ancestral HLA haplotype A3-B7-C7-DR15-DQ6 (AH 7.1) harbours loci that confer protection for type 1 diabetes (T1D), and risk for multiple sclerosis (MS). The mechanism of this dual effect is not known.

Two recent studies reviewed T1D GWAS results and applied functional annotations. While no specific molecular mechanism has been implicated, the overall conclusion was that the risk modification by most SNPs was likely via their effect on gene expression levels.

We reasoned that by functionally annotating common T1D/MS markers, we may gain insight into the pathogenesis of both diseases.

BACKGROUND

Hum Immunol. 2015 Oct;76(10):753-8. doi: 10.1016/j.humimm.2015.09.033. Epub 2015 Sep 30.

Functional relevance for type 1 diabetes mellitus-associated genetic variants by using integrative analyses.

Qiu YH[1], Deng FY[1], Tang ZX[1], Jiang ZH[2], Lei SF[3].

CONCLUSIONS: These findings indicated that integrative analyses can yield important functional information to link genetic variants and type 1 DM.

J Immunol. 2016 Apr 1;196(7):3043-53. doi: 10.4049/jimmunol.1502056. Epub 2016 Feb 24.

Systematic Evaluation of Genes and Genetic Variants Associated with Type 1 Diabetes Susceptibility.

Ram R[1], Mehta M[1], Nguyen QT[1], Larma I[1], Boehm BO[2], Pociot F[3], Concannon P[4], Morahan G[5].

⊕ Author information

Abstract

Genome-wide association studies have found >60 loci that confer genetic susceptibility to type 1 diabetes (T1D). Many of these are defined only by anonymous single nucleotide polymorphisms: the underlying causative genes, as well as the molecular bases by which they mediate susceptibility, are not known. Identification of how these variants affect the complex mechanisms contributing to the loss of tolerance is a challenge. In this study, we performed systematic analyses to characterize these variants. First, all known genes in strong linkage disequilibrium ($r(2) > 0.8$) with the reported single nucleotide polymorphisms for each locus were tested for commonly occurring nonsynonymous variations. We found only a total of 22 candidate genes at 16 T1D loci with common nonsynonymous alleles. Next, we performed functional studies to examine the effect of non-HLA T1D risk alleles on regulating expression levels of genes in four different cell types: EBV-transformed B cell lines (resting and 6 h PMA stimulated) and purified CD4(+) and CD8(+) T cells. We mapped cis-acting expression quantitative trait loci and found 24 non-HLA loci that affected the expression of 31 transcripts significantly in at least one cell type. Additionally, we observed 25 loci that affected 38 transcripts in trans. In summary, our systems genetics analyses defined the effect of T1D risk alleles on levels of gene expression and provide novel insights into the complex genetics of T1D, suggesting that most of the T1D risk alleles mediate their effect by influencing expression of multiple nearby genes.

METHODS

We compiled common risk markers reported by the WTCCC and International MS Genetics Consortium studies.

Using the statistical significance threshold of $P < 5x10^{-4}$, 119 SNPs that modify risk for both T1D and MS were identified and annotated using SNPnexus, PheGenI and SNiPA.

Their additional disease associations were examined on GWAS catalog and GRASP.

We also examined correlations between these SNPs and HLA haplotypes using our ImmunoChip data on 95 HLA-typed IHWG cell lines.

RESULTS: Overall

SNP ID	Position on chromosome 6 (Build 36.2)	*P* value for T1D association	In Gene	Nearest Gene	CADD score	DANN score	FATHMM score	GWAWA score	eQTL target(s) in lymphoblastoid cells	TF Binding Site Alteration	On AH 7.1 (ImmunoChip)
rs2517538	31121520	2.31E-09	Intergenic	C6orf205	1.37	0.321	0.0705	0.26	**HCG22**; HLA-C	.	Yes
rs2523881	31150587	0.000238518	Intergenic	C6orf15	1.463	0.57	0.07909	0.21	**HCG22**	.	No
rs1265052	31188450	1.06E-12	Intergenic	C6orf15	10.03	0.784	0.0823	0.44	**HCG22**; HLA-C	.	Yes
rs3130558	31205162	1.10E-08	PSORS1C1	PSORS1C2	7.627	0.695	0.56171	0.36	HLA-C	.	Yes
rs3131009	31206811	7.39E-10	PSORS1C1	PSORS1C2	0.72	0.295	0.03974	0.19	HLA-C	.	Yes
rs2106074	31241488	7.81E-11	POU5F1	POU5F1	1.975	0.532	0.18297	0.2	HLA-C	.	Yes
rs3130501	31244432	2.11E-06	POU5F1	POU5F1	1.322	0.513	0.04483	0.22	HLA-C	.	Yes
rs2736177	31694073	2.23E-60	Intergenic	AIF1	5.883	0.624	0.06585	0.41	HLA-C; HLA-DRB5	.	Yes
rs760293	31719756	5.81E-54	BAT3	BAT3	4.391	0.421	0.17617	0.34	HLA-C; HLA-DRB5	.	Yes
rs3130287	32158522	6.47E-91	TNXB	TNXB	5.271	0.652	0.07981	0.32	HLA-C; -DQA1/DQB1; -DRB5	.	Yes
rs3131294	32288124	1.75E-88	NOTCH4	NOTCH4	1.2	0.766	0.10702	0.27	HLA-C; -DQA1/DQB1; -DRB5	.	Yes
rs415929	32297010	6.73E-09	NOTCH4	NOTCH4	0.052	0.49	0.06559	0.15	HLA-DQB1	CREB; C/EBP alpha	No
rs910049	32423705	5.42E-27	C6orf10	C6orf10	0.537	0.401	0.03904	0.15	HLA-DRB5; HLA-DQA1	.	Yes
rs3129934	32444165	3.23E-22	C6orf10	C6orf10	6.401	0.683	0.1208	0.24	HLA-DRB5; HLA-DQB1	.	Yes
rs3806156	32481676	5.12E-12	BTNL2	BTNL2	2.858	0.745	0.18487	0.12	HLA-DQB1	.	No
rs9268645	32516505	6.56E-95	HLA-DRA	HLA-DRA	1.435	0.371	0.21164	0.27	HLA-DRB5	.	No
rs3135392	32517220	1.69E-12	HLA-DRA	HLA-DRA	4.612	0.287	0.11902	0.37	HLA-DRB5	.	No
rs9268831	32535726	2.21E-35	Intergenic	HLA-DRA	10.78	0.718	0.87492	0.33	HLA-DRB5; -DQA1/DQB1	ATF6	Yes
rs9268877	32539125	3.82E-78	Intergenic	HLA-DRA	0.662	0.566	0.02851	0.2	HLA-DQA1	.	Yes
rs9276429	32820082	2.98E-77	HLA-DQA2	HLA-DQA2	0.453	0.739	0.02996	0.29	HLA-DQA1	.	Yes
rs9276431	32820225	1.30E-77	HLA-DQA2	HLA-DQA2	0.247	0.564	0.06304	0.29	HLA-DQA1	.	Yes
rs7453920	32837990	2.18E-76	Intergenic	HLA-DQA2	0.457	0.152	0.04007	0.25	HLA-DQA1	.	Yes

Low statistical threshold ($P < 10^{-4}$) and no LD pruning → Proxy SNPs were not examined.

RESULTS: Classical HLA genes

All common risk markers were located in the classical HLA region flanked by *HLA-F* and *HLA-DPA1*.

None of them were in classical HLA genes (except two intronic SNPs in *HLA-DRA*). Only one SNP (in *CDSN*) was missense, and none were splice-site SNPs.

Thus, antigen presentation differences via HLA alleles of AH 7.1 did not appear to be involved in the differential risk to T1D and MS.

RESULTS: ImmunoChip

Most of the 69 SNPs present in the ImmunoChip showed non-exclusive correlation with AH 7.1.

None were the known proxy markers for the HLA alleles of AH 7.1.

Twenty common risk markers for MS and T1D were on non-AH 7.1 haplotypes. Of those, a subset within the class I region appeared to correlate with *HLA-B*4402*, which has not been implicated in either disease.

RESULTS: eQTL effects

Overall, the common risk markers were eQTLs for HLA genes (*HLA-C*, -*DRB5*, -*DQA1*, -*DQB1*) and an HLA class I region ncRNA gene *HCG22* in lymphoblastoid cells. Strikingly, all six eQTLs for *HCG22* were among the top ten statistically most significant eQTLs ($P \leq 5x10^{-14}$). Of these, rs1265052 was also an eQTL in adipose tissue (and several other tissues) in GTEx ($P = 4x10^{-27}$; effect size = -0.74).

The target gene for the strongest eQTL effects *HCG22* is one of four novel genes implicated in MS development in a recent integrative study of GWAS and transcriptomics.

RESULTS: eQTL effects (*HCG22*)

SNP ID	Position on chromosome 6 (Bld38)	Gene	P-Value	Probe GI	Probe Chr	Probe Location
rs2523880	31042768	HCG22	8.72E-30	hmm31752-S	6	31024074
rs1265052	31080470	HCG22	1.36E-26	hmm31752-S	6	31024074
rs2523881	31042607	HCG22	1.61E-26	hmm31752-S	6	31024074
rs2517538	31013540	HCG22	2.86E-22	hmm31752-S	6	31024074
rs3135392	32409241	HLA-DRB5	1.72E-15	GI_26665892-S	6	32485222
rs9268831	32427747	HLA-DQA1	8.06E-15	GI_18426974-S	6	32610832
rs9268877	32431146	HLA-DQA1	2.82E-14	GI_18426974-S	6	32610832
rs7382297	31247066	HCG22	3.78E-14	hmm31752-S	6	31024074
rs3130532	31208452	HCG22	4.80E-14	hmm31752-S	6	31024074
rs2516460	31418699	.	8.36E-12	GI_27483210-S	6	31362547
rs7382297	31247066	HLA-C	1.46E-11	ENSE00001469831	6	31324465
rs3130952	31177914	HLA-C	1.46E-11	ENSE00001		
rs3997982	31344293	HLA-C	3.09E-11	ENSE00001		
rs2523485	31351034	HLA-C	5.98E-11	ENSE00001		
rs3129934	32336186	HLA-DRB5	2.84E-10	ENSE00001		

Susceptibility Genes for Multiple Sclerosis Identified in a Gene-Based Genome-Wide Association Study

Xiang Lin[a,b,c]
Fei-Yan Deng[a,b]
Xin Lu[a,b]
Shu-Feng Lei[a,b]

[a]Center for Genetic Epidemiology and Genomics,
[b]Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, School of Public Health, Soochow University, Suzhou, Jiangsu, People's Republic of China
[c]Department of Tuberculosis Control, Ningbo Municipal Center for Disease Control & Prevention, Ningbo, Zhejiang, People's Republic of China

Background and Purpose Multiple sclerosis (MS) is a demyelinating and inflammatory disease of the central nervous system. The aim of this study was to identify more genes associated with MS.

Methods Based on the publicly available data of the single-nucleotide polymorphism-based genome-wide association study (GWAS) from the database of Genotypes and Phenotypes, we conducted a powerful gene-based GWAS in an initial sample with 931 family trios, and a replication study sample with 978 cases and 883 controls. For interesting genes, gene expression in MS-related cells between MS cases and controls was examined by using publicly available datasets.

Results A total of 58 genes was identified, including 20 "novel" genes significantly associated with MS ($p<1.40\times10^{-4}$). In the replication study, 44 of the 58 identified genes had been genotyped and 35 replicated the association. In the gene-expression study, 21 of the 58 identified genes exhibited differential expressions in MS-related cells. Thus, 15 novel genes were supported by replicated association and/or differential expression. In particular, four of the novel genes, those encoding myelin oligodendrocyte glycoprotein (*MOG*), coiled-coil alpha-helical rod protein 1 (*CCHCR1*), human leukocyte antigen complex group 22 (*HCG22*) and major histocompatibility complex, class II, DM alpha (*HLA-DMA*), were supported by the evidence of both.

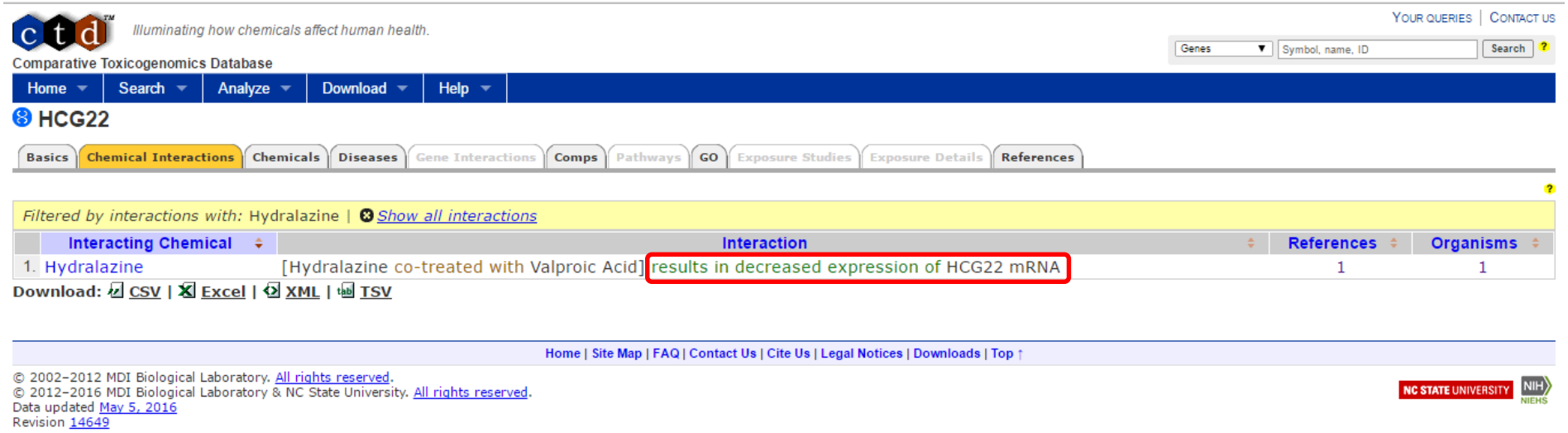
Conclusions The results of this study emphasize the high power of gene-based GWAS in detecting the susceptibility genes of MS. The novel genes identified herein may provide new insights into the molecular genetic mechanisms underlying MS.

Key Words multiple sclerosis, gene-based GWAS, gene expression, human leukocyte antigen.

Liverpool Hope University EST. 1844

RESULTS: eQTL effects (*HCG22*)

Possible translational value?

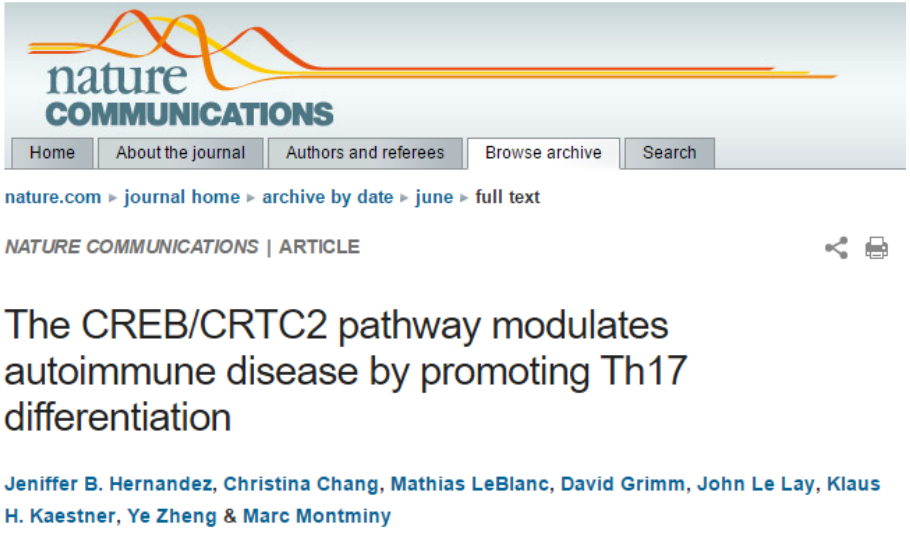


Liverpool Hope University EST. 1844

RESULTS: TF binding sites

Three SNPs were found to alter transcription factor (TF) binding sites: *SNORA38/PRRC2A* rs2736172 (XBP-1); *NOTCH4* rs415929 (CREB, delta-CREB and CEBP-alpha); *BRD2* rs17840186 (YYP-1).

These TFs have already been implicated in MS and T1D pathogenesis via their involvement in endoplasmic reticulum stress response and Th17 pathways.



Liverpool Hope University EST. 1844

RESULTS: rs3130564

The statistically most significant association by the 119 SNPs currently listed in GRASP was with the intergenic SNP rs3130564 (nearest gene: *PSORS1C1*). This SNP is highly significantly associated with T1D ($P = 1.2 \times 10^{-34}$) and MS ($P = 2.1 \times 10^{-7}$) as well as myasthenia gravis, idiopathic membranous nephropathy, rheumatoid arthritis, and lung cancer with ($P \leq 3.0 \times 10^{-8}$).

The compiled eQTL results from multiple sources in SNiPA revealed that rs3130564 is an eQTL in the pancreas for the lincRNA gene *XXbac-BPG248L24.13* and the pseudogene *WASF5P* in the HLA class I region (GTEx result). Besides, rs3130564 is also an eQTL for *HCG22* in the blood (MuTHER) and in tibial nerve (GTEx).

In our ImmunoChip data, 10 of the 95 cell lines were homozygote for the rare allele of this SNP. Two of those cell lines were biological duplicates representing AH 8.1 (HLA-A1- B8-DR3 haplotype), and five independent cell lines homozygous for *HLA-B*4402*. The rare allele of rs3130564 was not on AH 7.1.

RESULTS: rs3130564

***SNiPA*card – rs3130564**

SNP properties – Genome Assembly: grch37, Variant set: 1kgpp3v5, Population: EUR

rs3130564 (alias rs60971204, rs386579957, rs17456304, rs150646344, rs116738927)

position / outlink		allele info	
physical position	chr6: 31,101,674	alleles	C/T
genetic position [cM]	50.97	frequencies	0.840/0.160
outlink	e!	non-reference allele	C

Basic features

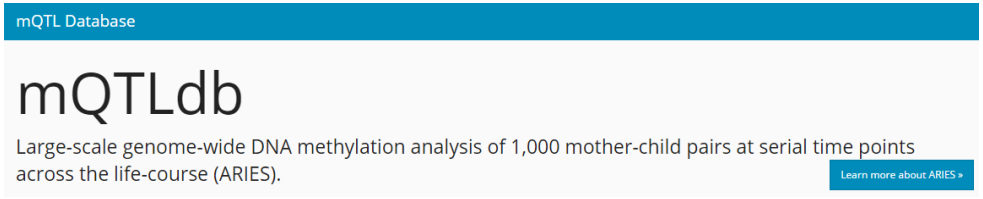
Conservation/deleteriousness		Linked genes	
phyloP	-0.243	gene(s) hit or close-by	PSORS1C1 e!, PSORS1C2 e!
phastCons	0	eQTL gene(s)	C4A e!, CCHCR1 e!, CYP21A1P e!, CYP21A2 e!, HCG17 e!, HCG22 e!, HCG27 e!, HCP5 e!, HLA-C e!, LINC00243 e!, MICB e!, PSORS1C1 e!, PSORS1C2 e!, STK19B e!, WASF5P e!, XXbac-BPG248L24.13 e!, ZNRD1 e!, ZNRD1-AS1_3 e!
GERP++	-0.266	potentially regulated gene(s)	–
CADD score	14.34	disease gene(s)	C4A e!, CYP21A2 e!
SnpEff effect impact	modifier		

IHWG-ID	CELL LINE	CEH	rs1736913_A	s1610603_	1633070_	s1737060_	s2517646_	s3094123_	3095345_	2517538_	rs4713429_C_	9262615_	2523881_	s1265052_	s2233967_	s3095314_	s3130558_	s3131009_	rs3130564_A_	2073723_	s2106074_	s3130501_	s3095238_
9081	EA	7.1	2	2	2	2	2	1	2	2	0	0	0	2	0	0	2	2	0	2	2	2	2
9001	SA	7.2	0	0	0	0	0	1	2	2	0	0	0	2	0	0	2	2	0	2	2	2	2
9002	MZO70782	65.1?	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0
9003	KAS116	.	2	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0
9004	JESTHOM	.	0	0	0	0	0	0	2	0	2	2	2	0	2	2	0	0	0	0	0	0	0
9005	HOM2	.	2	2	2	2	1	0	0	0	2	2	2	0	2	2	0	0	0	0	0	0	0
9006	WT100BIS	35.2	2	0	0	0	0	0	0	0	2	2	2	0	0	0	2	2	0	0	2	0	0
9008	DO208915	18.1	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0
9009	KAS011		0	0	0	0	2	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0
9010	AMAI		2	0	0	0	2	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0
9011	E4181324	52.1?	0	0	0	0	2	0	0	0	0	0	2	0	2	2	0	0	0	2	2	2	0
9012	WJR076		0	0	0	0	0	0	0	2	0	2	0	0	0	2	0	0	2	0	0	0	0
9014	MGAR		0	0	0	0	2	0	2	0	0	0	0	0	0	2	0	0	2	0	0	0	0
9015	WT24		0	0	0	0	0	0	0	2	0	0	0	2	0	0	2	2	0	2	2	2	0
9016	RML REM		0	0	0	0	0	0	0	0	2	2	2	0	2	2	0	0	0	0	0	0	0
9017	WT8		2	2	2	2	2	2	2	2	0	0	0	2	0	0	2	2	0	2	2	2	2
9019	DUCAF	18.2	0	0	0	0	2	0	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0
9020	QBL	18.2	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0
9021	RSH, RSHD	42.1	1	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
9022	COX	8.1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0
9024	KT17	.	1	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	0	0	1	0	0
9026	YAR	38.1	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0
9029	WT51		2	2	2	2	0	0	0	0	2	2	2	0	0	2	0	0	0	0	0	0	0
9031	BOLETH BO	62.1	0	0	0	0	2	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0

RESULTS: rs3130564

Strong mQTL during pregnancy and childhood in cis and trans

Timepoint	SNP	SNP Chr	SNP Pos	CpG	CpG Chr	CpG Pos	beta	Effect Size	*P* value	Trans
Pregnancy	rs3130564	6	31101674	cg01517384	6	31323856	-0.48168	0.01846	1.14E-22	N
Adolescence	rs3130564	6	31101674	cg01517384	6	31323856	-0.42125	0.02521	2.10E-21	N
Pregnancy	rs3130564	6	31101674	cg15196058	**4**	130934318	-0.36988	0.02694	1.64E-20	**Y**
Childhood	rs3130564	6	31101674	cg01517384	6	31323856	-0.39573	0.01733	3.48E-19	N
Pregnancy	rs3130564	6	31101674	cg26036029	6	32552443	0.4207	0.05749	3.85E-18	**Y**
Childhood	rs3130564	6	31101674	cg15196058	**4**	130934318	-0.29462	0.01987	1.58E-17	**Y**
Middle Age	rs3130564	6	31101674	cg01517384	6	31323856	-0.39872	0.023	2.80E-16	N
Adolescence	rs3130564	6	31101674	cg15196058	**4**	130934318	-0.31518	0.01774	1.61E-15	**Y**
Childhood	rs3130564	6	31101674	cg26036029	**6**	32552443	0.34274	0.05583	1.05E-14	**Y**



CONCLUSION

These results raise doubt about the involvement of the classical HLA genes of AH 7.1 in differential susceptibility to T1D and MS, and point towards the non-HLA content of AH 7.1 and other haplotypes (*HLA-B*44:02*).

Our approach shows a good example of post-GWAS analysis and how genetic epidemiology can be used to probe disease biology.

YOUR FUTURE
STARTS WITH HOPE

Liverpool Hope University EST. 1844