HLA-DQA1 3'UTR is an Unexplored Source of Disease Associations in the HLA Complex

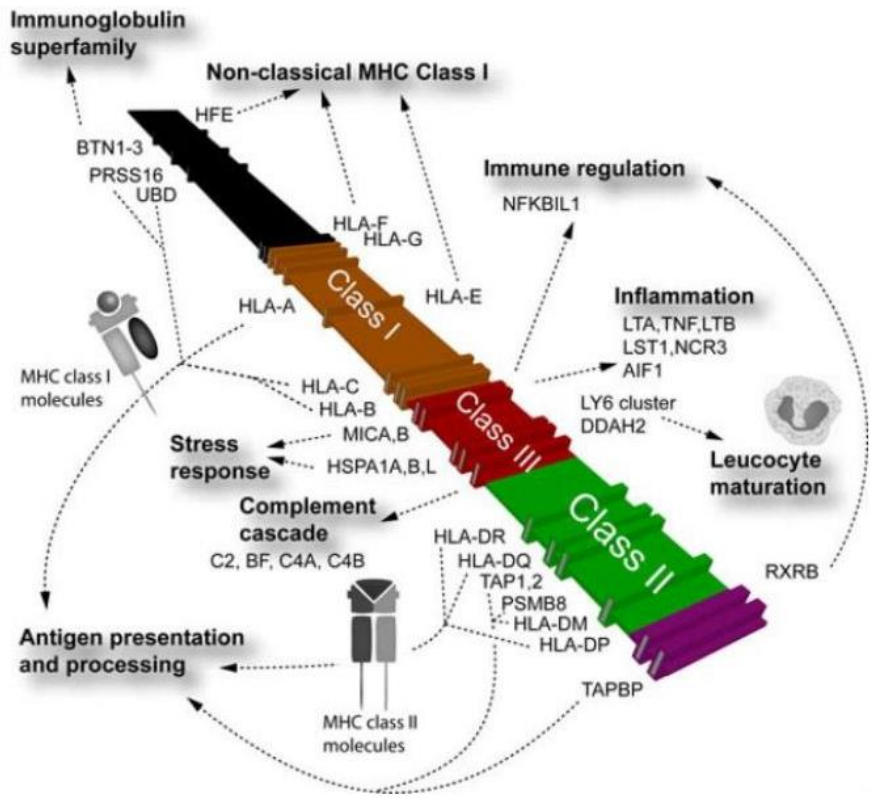
Sonay Temurhan[1], Fatma Savran Oguz[1], Mehmet Tevfik Dorak[2]

[1] Department of Medical Biology, Istanbul University Medical School, Istanbul, Turkey;
[2] School of Health Sciences, Liverpool Hope University, Liverpool, U.K.

EFI 2015, Geneva

YOUR FUTURE
STARTS WITH HOPE

BACKGROUND



The Wellcome Trust Centre for Human Genetics

HLA Region:

- **The most gene-dense and polymorphic**
- **Shows the highest number of disease associations**

Most associations are attributed to variation in antigen presentation caused by polymorphisms

Liverpool Hope University EST. 1844

BACKGROUND

Elsewhere in the genome, the most common *intermediate phenotype* between a polymorphism and disease development is alterations in gene expression regulation

Causal variant → **Gene expression**
Splicing
Protein structure → **Disease**

Liverpool Hope University EST. 1844

BACKGROUND

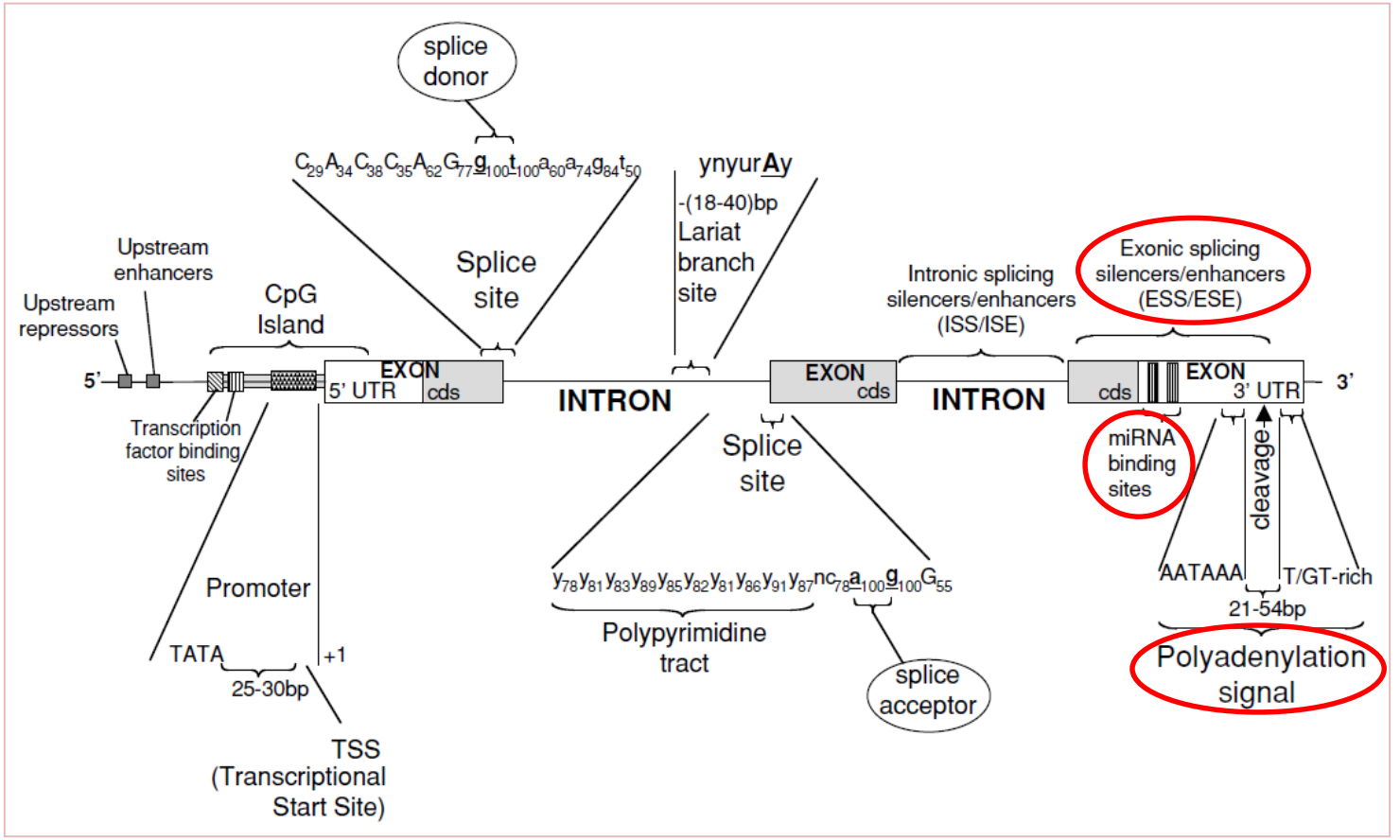


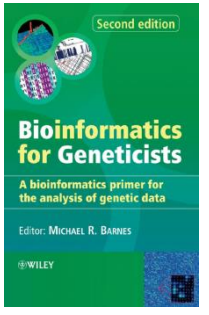
Figure 11.2 The anatomy of a gene. This figure illustrates some of the key regulatory regions that control the transcription, splicing and post-transcriptional processing of genes and transcripts. Polymorphisms in these regions should be investigated for functional effects

11

Predictive Functional Analysis of Polymorphisms: An Overview

Mary Plumpton[1] and Michael R. Barnes[2]

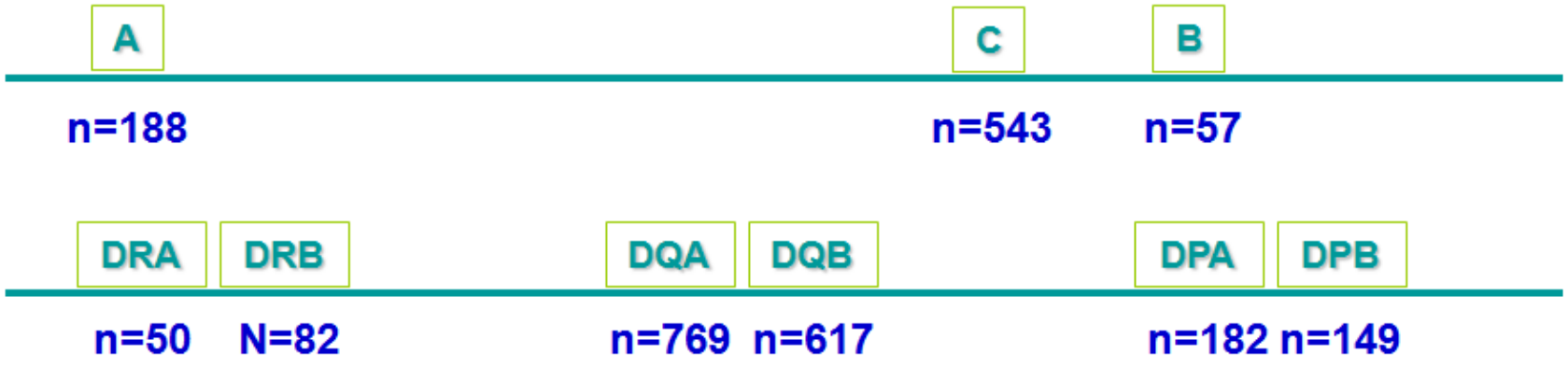
Bioinformatics, GlaxoSmithKline Pharmaceuticals, [1]Stevenage, Hertfordshire, UK, [2]Harlow, Essex, UK



Liverpool Hope University EST. 1844

BACKGROUND

Of the genes encoding six classical HLA antigens, *HLA-DQA1* 3'UTR (<6.3kb) has the highest number of variants (n=769)



Liverpool Hope University EST. 1844

AIM

To explore the polymorphism content of *HLA-DQA1* 3' UTR and potential functional consequences of these polymorphisms

METHODS

Functional annotation of *HLA-DQA1* 3'UTR variants using web-based bioinformatic tools and genome browsers

1KG, Ensembl, HapMap, Illumina (NextBio) browser

AURA Atlas of UTR Regulatory Activity, JASPAR, dbSTEP

NCBI eQTL Browser, GTEx, RegulomeDB, miRBase

Examination of disease associations by *HLA-DQA1* 3'UTR variants

PheGenI, SNPnexus, GRASP

RESULTS

Summary of variation consequences in ENSG00000196735

Switch to tree view

Show All entries

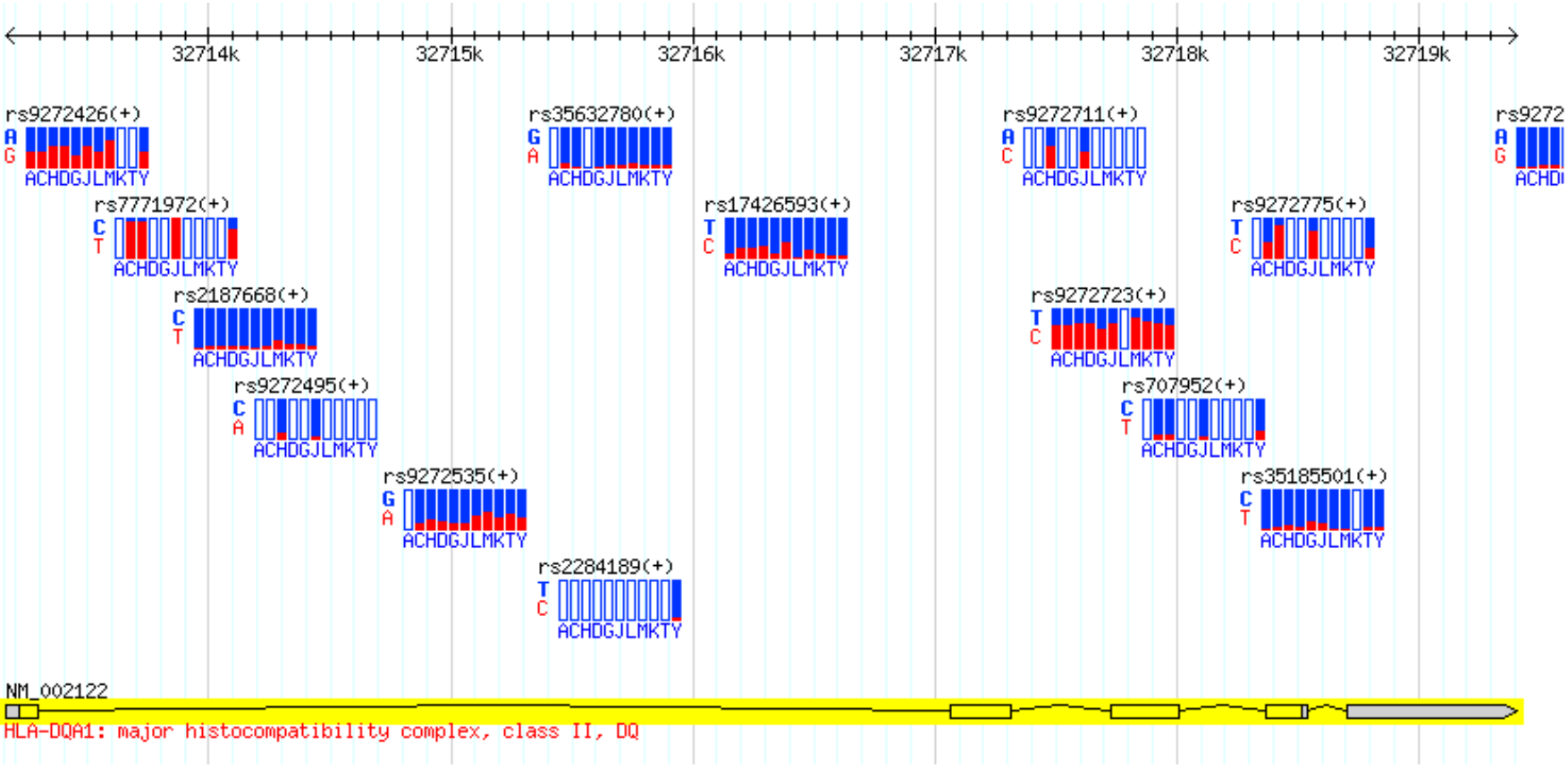
Filter

Number of variant consequences		Type	Description
0	-	Transcript ablation	A feature ablation whereby the deleted region includes a transcript feature (SO:0001893)
12	Show	Splice donor variant	A splice variant that changes the 2 base region at the 5' end of an intron (SO:0001575)
6	Show	Splice acceptor variant	A splice variant that changes the 2 base region at the 3' end of an intron (SO:0001574)
1	Show	Stop gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript (SO:0001587)
44	Show	Frameshift variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three (SO:0001589)
0	-	Stop lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript (SO:0001578)
0	-	Initiator codon variant	A codon variant that changes at least one base of the first codon of a transcript (SO:0001582)
0	-	Transcript amplification	A feature amplification of a region containing a transcript (SO:0001889)
4	Show	Inframe insertion	An inframe non synonymous variant that inserts bases into in the coding sequence (SO:0001821)
8	Show	Inframe deletion	An inframe non synonymous variant that deletes bases from the coding sequence (SO:0001822)
367	Show	Missense variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved (SO:0001583)
76	Show	Splice region variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron (SO:0001630)
0	-	Incomplete terminal codon variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed (SO:0001626)
157	Show	Synonymous variant	A sequence variant where there is no resulting change to the encoded amino acid (SO:0001819)
0	-	Stop retained variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains (SO:0001567)
4	Show	Coding sequence variant	A sequence variant that changes the coding sequence (SO:0001580)
0	-	Mature miRNA variant	A transcript variant located with the sequence of the mature miRNA (SO:0001620)
92	Show	5 prime UTR variant	A UTR variant of the 5' UTR (SO:0001623)
769	Show	3 prime UTR variant	A UTR variant of the 3' UTR (SO:0001624)
998	Show	Non coding exon variant	A sequence variant that changes non-coding exon sequence (SO:0001792)
5993	Show	Intron variant	A transcript variant occurring within an intron (SO:0001627) (WARNING: table may not load for this number of variants!)
1110	Show	NMD transcript variant	A variant in a transcript that is the target of NMD (SO:0001621)
1696	Show	NC transcript variant	A transcript variant of a non coding RNA (SO:0001619)
4839	Show	Upstream gene variant	A sequence variant located 5' of a gene (SO:0001631)
4505	Show	Downstream gene variant	A sequence variant located 3' of a gene (SO:0001632)
17787	Show	ALL	All variations (WARNING: table may not load for this number of variants!)

Liverpool Hope University EST. 1844

1000 Genomes
A Deep Catalog of Human Genetic Variation

RESULTS



Liverpool Hope University EST. 1844

International HapMap Project

RESULTS

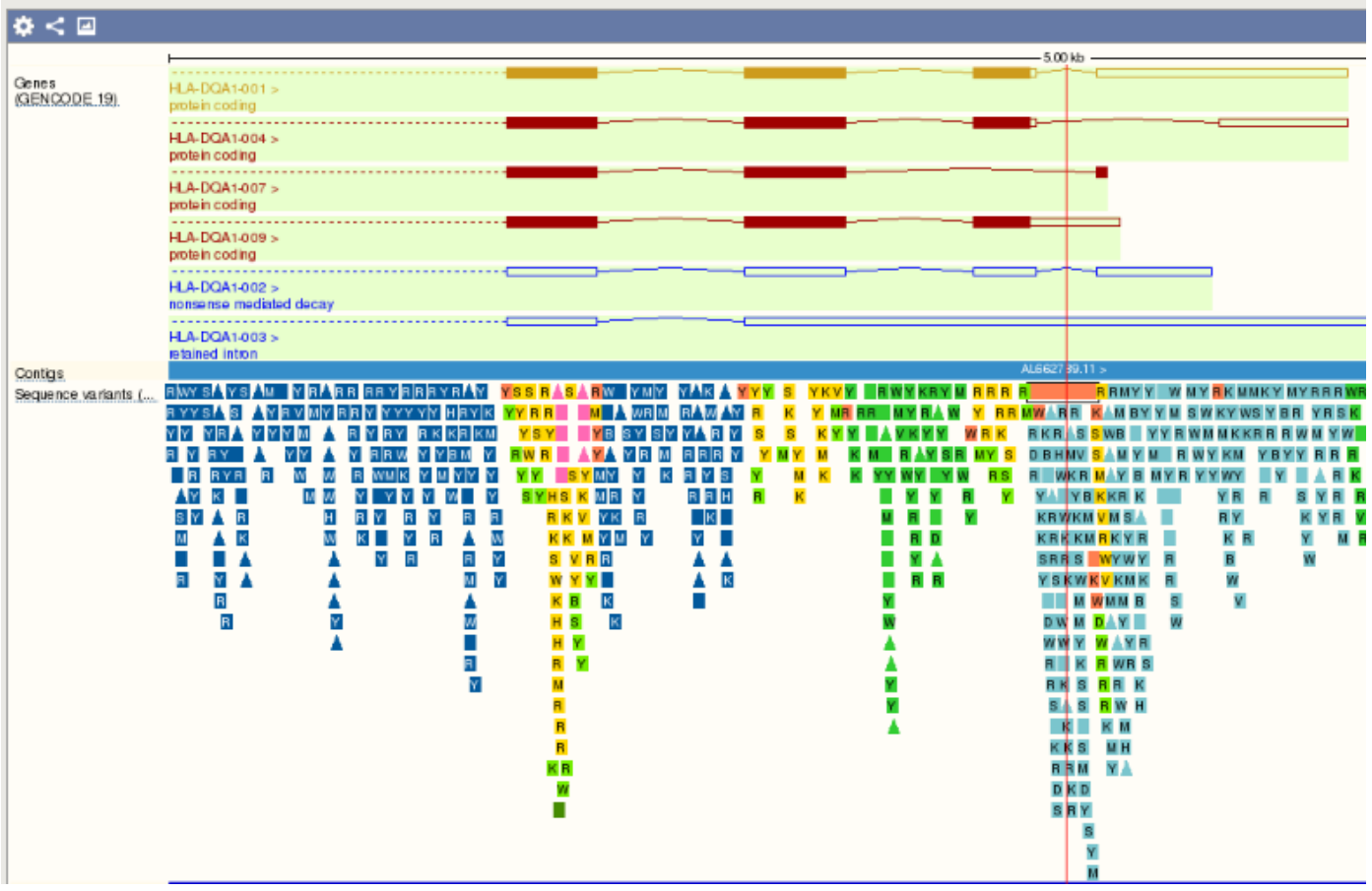
AURA Atlas of UTR Regulatory Activity:

- 534 to 6246bp
- Transcript half life:
  - 291.2 to 395.1 minutes (Array probe ID: 211654_X_AT)
  - 435.8 to 463.9 minutes (Array probe ID: 212671_S_AT)

Descriptive statistics:

- 668 SNPs
- 76 indel polymorphisms (one of which 167bp)
- 6 splice site SNPs (out of 49 in the whole of MHC in dbSTEP)
- 467 listed in NCBI dbSNP
- Only 2 in GWAS chips
- 14 observed in cancer genomes as somatic mutations

RESULTS



Splice donor variant
Inframe deletion
Synonymous variant
Non coding exon variant
Splice acceptor variant
Missense variant
Coding sequence variant
Intron variant
Frameshift variant
Splice region variant
3 prime UTR variant

Liverpool Hope University EST. 1844

1000 Genomes
A Deep Catalog of Human Genetic Variation

RESULTS

SNP	TFBS_id	chromStart_hg19	chromEnd_hg19	TFBS_Accession	TFBS_Species	TFBS_name	SwissProt_Accession
rs9272591	V$GATA_C	32607534	32607545	M00203	human	GATA-1	P15976
rs1129737	V$MYCMAX_01	32609087	32609101	M00118	human	c-Myc	P01106
rs1048134	V$AP1FJ_Q2	32609788	32609799	M00172	human	c-Fos	P01100
rs1048134	V$AP1FJ_Q2	32609788	32609799	M00172	human	c-Jun	P05412
rs41555012	V$AP1FJ_Q2	32609788	32609799	M00172	human	c-Fos	P01100
rs41555012	V$AP1FJ_Q2	32609788	32609799	M00172	human	c-Jun	P05412
rs1048173	V$MYCMAX_02	32609848	32609860	M00123	human	c-Myc	P01106

SNPnexus

Barts and The London School of Medicine and Dentistry
Barts Cancer Institute

Liverpool Hope University EST. 1844

RESULTS

Human | Location: 6:32,595,956-32,614,839 | Gene: HLA-DQA1 | Variation: rs71680105

Variation displays
- Explore this variation
- Genomic context
  - Genes and regulation (12)
  - Flanking sequence
- Population genetics
- Individual genotypes
- Linkage disequilibrium
- Phenotype Data
- Phylogenetic Context
- Citations
- External Data
  - LOVD

Configure this page
Add your data
Export data
Get VCF data
Bookmark this page
Share this page
View in Ensembl
Download view as CSV

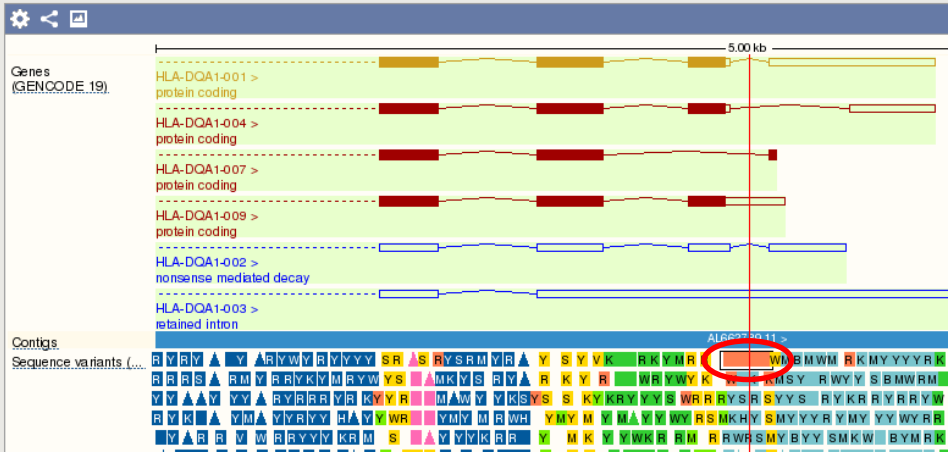
rs71680105 DELETION

Original source	Variants (including SNPs and indels) imported from dbSNP (release 138) | View in dbSNP
Alleles	(LARGEDELETION)/-
Location	Chromosome 6:32610560-32610726 (forward strand) | View in location tab
HGVS names	This variation has 6 HGVS names - click the plus to show

Context

Structural variants display

There are 430 structural variants overlapping this region. Some of them might not be displayed as we limit the size of each structural v



1000 Genomes
A Deep Catalog of Human Genetic Variation

NCBI
dbSNP
Short Genetic Variations

```
SubSNP Detail:
NCBI Assay ID:        ss99476824
Submitter SNP ID:     INDEL_GP36_2007a_208672_0
Synonyms:             .
LOCUSID:              Not submitted
Submitter STS ID:     Not submitted
STS Accession:        Not submitted
GenBankK Accession:   NT_007592
Comment:
   largedeletion=
   aggtaagattgagactggttacagttgaagcggcagtatgaaaggaagga
   aagtgggagggcgttgtggacatgaatgtggttgaaagttgtaggggaat
   tgggaagtggcatgatgatgacacaggagccccctcggacccattgatct
   catgtctgccctgttgc
```

Liverpool Hope University EST. 1844

RESULTS

2 putative sites were predicted with these settings (95%) in sequence named DQA1

Model ID	Model name	Score	Relative score	Start	End	Strand	predicted site sequence
MA0130.1	ZNF354C	7.916	0.956240771454112	66	71	-1	gtccac
MA0133.1	BRCA1	7.329	0.951002797117118	161	167	-1	gcaacag

Comment: This type of analysis has a high sensitivity but abysmal selectivity. In other words: while true functional will be detected in most cases, most predictions will correspond to sites bound in vitro but with no function in vivo. A number of additional contraints of the analysis can improve the prediction; phylogenetic footprinting is the most common. We recommend using the ConSite service, which uses the JASPAR datasets.

The review Nat Rev Genet. 2004 Apr;5(4):276-87 gives a comprehensive overview of transcription binding site prediction

Liverpool Hope University EST. 1844

The high-quality transcription factor binding profile database

RESULTS

dbSTEP

Database of Splice Translational Efficiency Polymorphisms

49 results for region chr6: 29 to 34MB

Variant	External ID	Intron	Alleles	Motif Position	Disease Impact	Minor Allele Freq	Allelic Score Difference	BLAST EST Evidence		Splice Prediction Difference		
								Unspliced	Spliced	HSF	NNSplice	NetUTR
rs2295665	BAT4	6:31740664..31741329	G/A	3' Splice Site	8.0225	0.164	1.0000	NO	NO	YES	YES	NO
rs2857692	BAT2	6:31696615..31698445	G/T	5' Splice Site	NULL	NULL	1.0000	NO	NO	YES	YES	YES
rs2857692	BAT2	6:31696615..31698482	G/T	5' Splice Site	NULL	NULL	1.0000	NO	NO	YES	YES	YES
rs9271906	HLA-DQA1	6:32704026..32707696	G/A	5' Splice Site	NULL	NULL	1.0000	NO	NO	YES	YES	NO
rs41544522	HLA-DPA1	6:33149466..33156434	G/T	5' Splice Site	NULL	NULL	1.0000	NO	NO	YES	YES	YES
rs41544522	HLA-DPA1	6:33149528..33156434	G/T	5' Splice Site	NULL	NULL	1.0000	NO	NO	YES	YES	YES
rs41544522	HLA-DPA1	6:33149425..33156434	G/T	5' Splice Site	NULL	NULL	1.0000	NO	NO	YES	YES	YES
rs56260633	TNXB	6:32157429..32157790	T/C	5' Splice Site	NULL	NULL	1.0000	NO	NO	YES	YES	YES
rs36224777	HLA-DQA1	6:32708665..32713175	G/A	3' Splice Site	NULL	NULL	1.0000	NO	YES	YES	YES	NO
rs3207966	HLA-DQA1	6:32708665..32713175	G/A	3' Splice Site	NULL	NULL	1.0000	NO	YES	YES	YES	NO

Four of them in *HLA-DQA*1 3' UTR

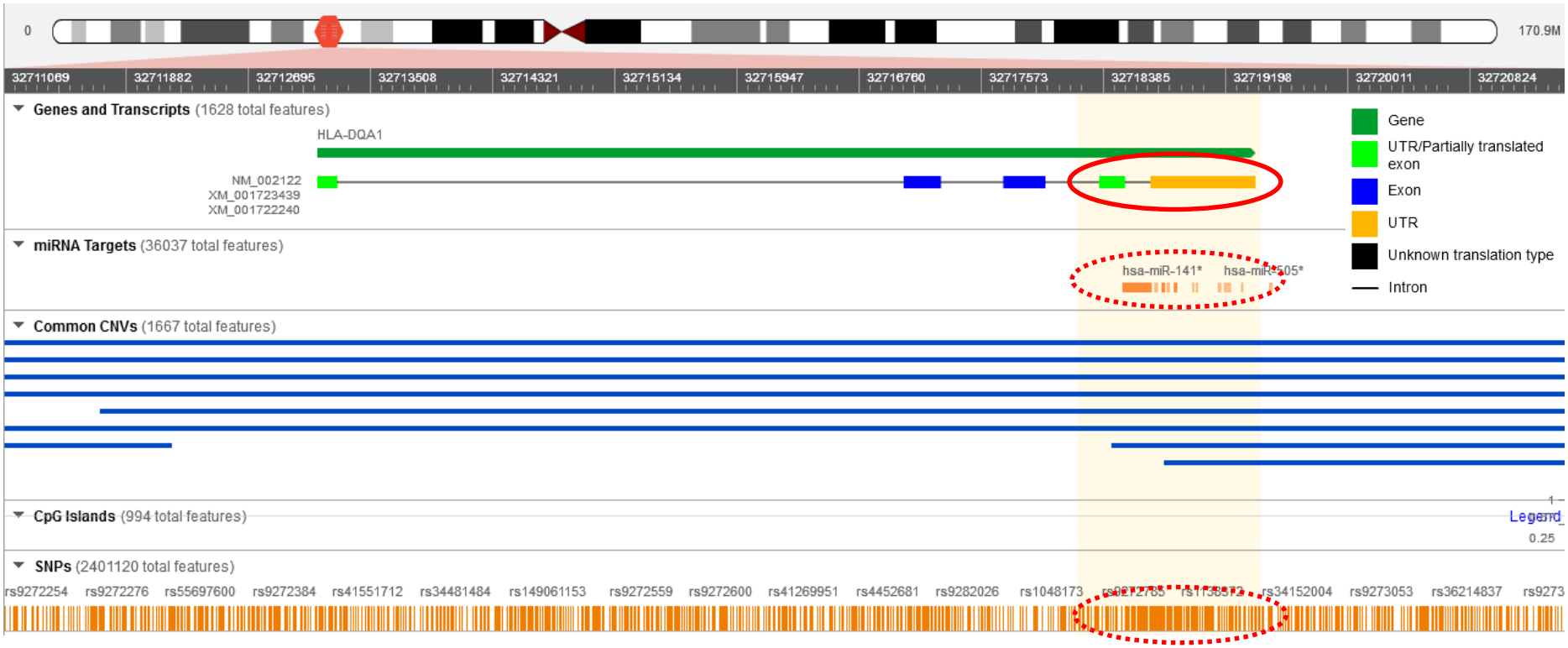
Variant	External ID	Intron	Alleles	Motif Position	Disease Impact	Minor Allele Freq	Allelic Score Difference	BLAST EST Evidence		Splice Prediction Difference		
								Unspliced	Spliced	HSF	NNSplice	NetUTR
rs36224777	HLA-DQA1	6:32708665..32713175	G/A	3' Splice Site	NULL	NULL	1.0000	NO	YES	YES	YES	NO
rs3207966	HLA-DQA1	6:32708665..32713175	G/A	3' Splice Site	NULL	NULL	1.0000	NO	YES	YES	YES	NO
rs9272426	HLA-DQA1	6:32708665..32713175	A/G	Polypyrimidine Tract	14.9176	0.422	0.0100	YES	YES	YES	YES	NO
rs36224778	HLA-DQA1	6:32708665..32713175	A/G	Polypyrimidine Tract	NULL	NULL	0.0100	YES	YES	YES	YES	NO

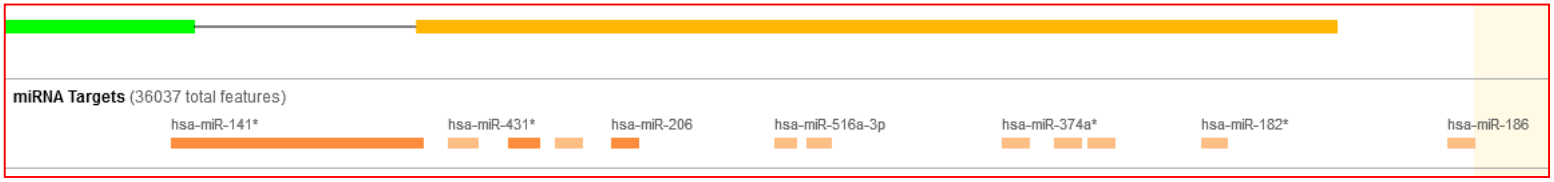
Liverpool Hope University EST. 1844

Genome-Wide Data-Mining of Candidate Human Splice Translational Efficiency Polymorphisms (STEPs) and an Online Database

Christopher A. Raistrick, Ian N. M. Day, Tom R. Gaunt

RESULTS





Liverpool Hope University EST. 1844

illumina GENOME BROWSER

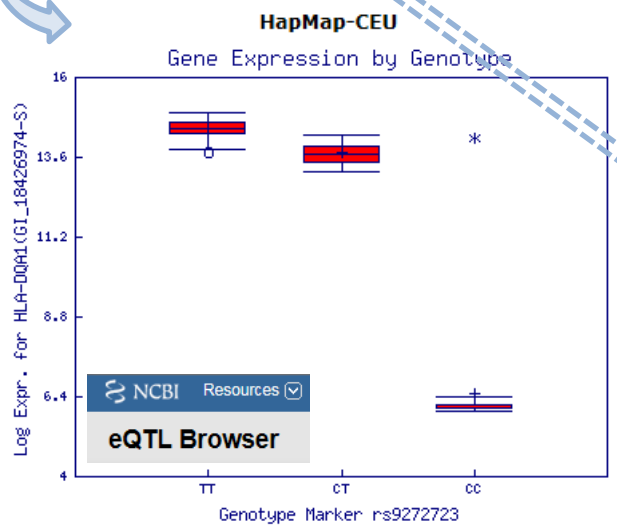
RESULTS

eQTL Data

PheGenI
Phenotype-Genotype Integrator

1 - 2 of 2 Download Modify Search

#	Tissue (Analysis ID)		rs#	SNP Location	Probe ID	Probe Location	Gene	P-value	R^2
1	Lymphoblastoid (7)		rs9272723	6: 32,609,426	GI_18426974-S	6: 32,610,832	HLA-DQA1	4.183×10^{-47}	0.6379
2	Liver (2)		rs9272723	6: 32,609,426	HLA-DRB1	6: 32,546,549	HLA-DRB1	2.430×10^{-39}	



***P* value for *HLA-DQA1* eQTLs effect in GTEx is 10E-36.**

Snp Id	Pvalue	PMID	Location	Phenotype	Phenotype Category	chr	pos	InGene	Total Samples	Platforr
rs9272723	2.2E-75	17554300	Webdata	Type 1 diabetes, Rheumatoid arthritis, combined case analysis, gender differentiated	Cardiovascular disease (CVD),Myocardial infarction (MI),Neuro,Behavioral, disorder,Blood pressure,CVD risk factor (CVD RF),Blood-related,Type 1 diabetes (T1D),Type 2 diabetes (T2D),Developmental,A arthritis,Crohn's disease	6	32641650	(HLA-DQA1)	4806	Affymetrix [469557]
rs9272723	1.2E-31	20453842	FullScan	Rheumatoid arthritis	Inflammation,Arthriti... arthritis	6	32641650	(HLA-DQA1)	41282	Affymetrix & Illumina [~271625 (imputed)
rs9272723	7.6E-17	21151127	Text	Primary sclerosing cholangitis	Hepatic,Inflammation	6	32641650	(HLA-DQA1)	6539	Affymetrix [2466182] (imputed)
rs9272723	1.1E-11	20228798	Table S3	Ulcerative colitis	Gastrointestinal,Infla... colitis	6	32641650	(HLA-DQA1)	10713	Affymetrix [1897764] (imputed)
rs9272723	1.7E-09	21156761	Table S3	Rheumatoid arthritis (ACPA-positive)	Inflammation,Arthriti... arthritis	6	32641650	(HLA-DQA1)	9129	Illumina [1723056]
rs9272723	4.8E-09	20622879	Table S2	Behçet's disease	Inflammation	6	32641650	(HLA-DQA1)	1607	Affymetrix [320438]

NIH National Heart, Lung, and Blood Institute

GRASP Search - v2.0.0.0

Liverpool Hope University EST. 1844

RESULTS

Complete Sequencing of *HLA-DQA1* 3'UTR and Correlations with Ancestral Lineages

Brittany A Morrison & M Tevfik Dorak
Genomic Immunoepidemiology Laboratory, HUMIGEN LLC, The Institute for Genetic Immunology, Hamilton, NJ 08690, USA

ASHI 2008

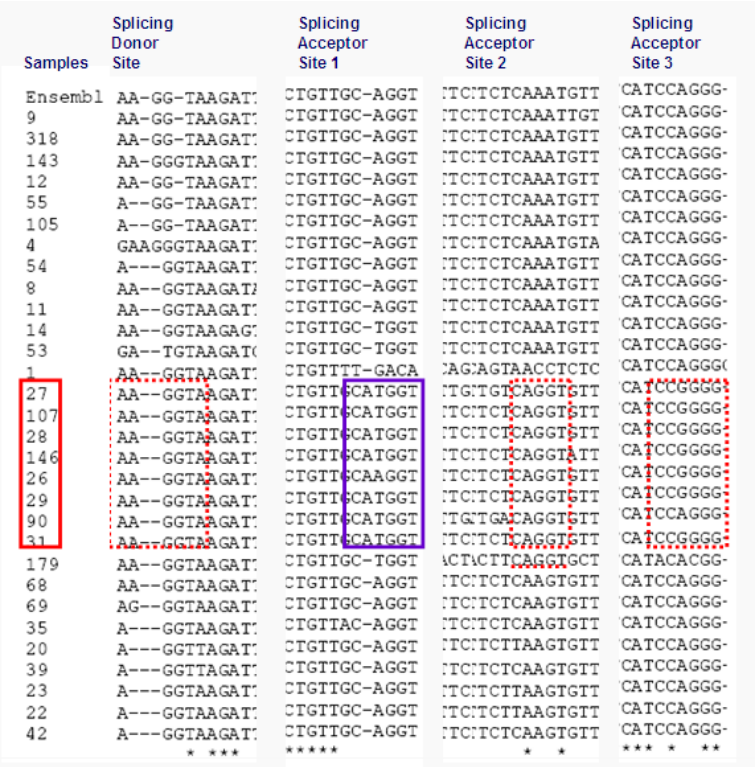


Figure 4. HLA-DQA1 3'UTR sequences at functionally relevant parts of the HLA-DQA1 3'UTR. The cell lines belonging to the DRB4 lineage are marked at the splicing sites and they have a unique sequence in the splicing acceptor site 1.

Figure 3: HLA-DQA1 3' UTR sequencing results in a reference cell line panel representing ancestral lineages, The samples in the box are all of the DRB4 lineage samples in the panel. Indel polymorphisms exclusive to the DRB4 lineage are indicated in the sequence.

Liverpool Hope University EST. 1844

CONSEQUENCES

Research

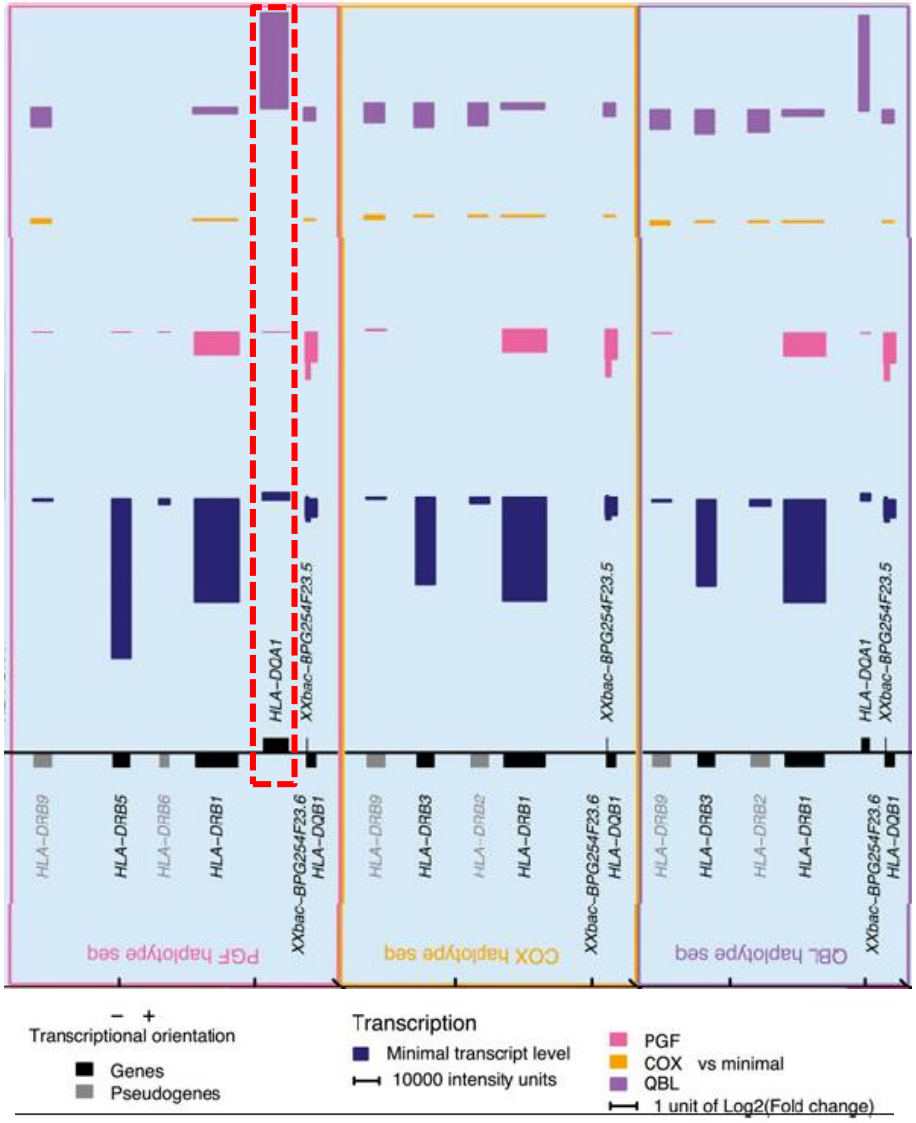
Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex

Claire Vandiedonck,[1,2,3,5] Martin S. Taylor,[1,4] Helen E. Lockstone,[1] Katharine Plant,[1] Jennifer M. Taylor,[1] Caroline Durrant,[1] John Broxholme,[1] Benjamin P. Fairfax,[1] and Julian C. Knight[1,5]

The human major histocompatibility complex (MHC) on chromosome 6p21 is a paradigm for genomics, showing remarkable polymorphism and striking association with immune and non-immune diseases. The complex genomic landscape of the MHC, notably strong linkage disequilibrium, has made resolving causal variants very challenging. A promising approach is to investigate gene expression levels considered as tractable intermediate phenotypes in mapping complex diseases. However, how transcription varies across the MHC, notably relative to specific haplotypes, remains unknown. Here, using an original hybrid tiling and splice junction microarray that includes alternate allele probes, we draw the first high-resolution strand-specific transcription map for three common MHC haplotypes (*HLA–A1–B8–Cw7–DR3*, *HLA–A3–B7–Cw7–DR15*, and *HLA–A26–B18–Cw5–DR3–DQ2*) strongly associated with autoimmune diseases including type 1 diabetes, systemic lupus erythematosus, and multiple sclerosis. We find that haplotype-specific differences in gene expression are common across the MHC, affecting 96 genes (46.4%), most significantly the zing finger protein gene *ZFP57*. Differentially expressed probes are correlated with polymorphisms between haplotypes, consistent with *cis* effects that we directly demonstrate for *ZFP57* in a cohort of healthy volunteers ($P = 1.2 \times 10^{-14}$). We establish that alternative splicing is significantly more frequent in the MHC than genome-wide (72.5% vs. 62.1% of genes, $P \leq 1 \times 10^{-4}$) and shows marked haplotypic differences. We also unmask novel and abundant intergenic transcription involving 31% of transcribed blocks identified. Our study reveals that the renowned MHC polymorphism also manifests as transcript diversity, and our novel haplotype-based approach marks a new step toward identification of regulatory variants involved in the control of MHC-associated phenotypes and diseases.

Genome Res. 2011 21: 1042-1054

Figure 1. The first transcriptional map of the human MHC



Liverpool Hope University EST. 1844

CONCLUSIONS

HLA-DQA1 **3'UTR is extremely polymorphic**

There is evidence for the functionality of 3'UTR polymorphisms

There is almost no data on the phenotypical associations of these polymorphisms

The insufficiency of GWAS platforms to assess the roles played by these polymorphisms in disease susceptibility is obvious

Specially designed sequencing-based studies to unravel HLA region disease associations is desirable

YOUR FUTURE
STARTS WITH HOPE

Liverpool Hope University EST. 1844