A Systematic Analysis of the Gene and Variation Content of the Extended HLA Region

Ertan Kanbur[1], Mustafa Dogan[2], Mehmet Tevfik Dorak[1]

[1] School of Health Sciences, Liverpool Hope University, Liverpool, UK;
[2] Baskent University, Department of Electrical-Electronics Engineering, Ankara, Turkey

YOUR FUTURE
STARTS WITH HOPE

BACKGROUND

The extended HLA (xHLA) region is already known to have the highest gene density and extreme polymorphism

It also contains the highest number of disease-associated variants, trans-eQTLs, and a high frequency of deleterious variants

We aimed to compare genomic features of the xHLA with the rest of the genome

METHODS

We explored the unique genomic features of the extended HLA (xHLA) region (chr6:25,726,131 to 33,400,601bp) in the latest genome assembly (GRCh38) to gain insight into the gene and variation content

The gene list was obtained from NCBI Map Release 108.6 (n=674)

We extracted the current SNP list (GRCh38.p7) from Ensembl (n=470,343)

Liverpool Hope University EST. 1844

RESULTS: Gene content

xHLA makes up 0.24% of the genome

674 genes

(1.1% of total genes)

453 protein-coding genes (67.2%)
(2.3% of total protein-coding genes)
(67.2% vs 32.7% (genome-wide proportion); *P* < 0.0001)

Non-protein coding genes (8.0% of all x HLA genes)
(42.6% in the rest of the genome; *P* < 0.0001)

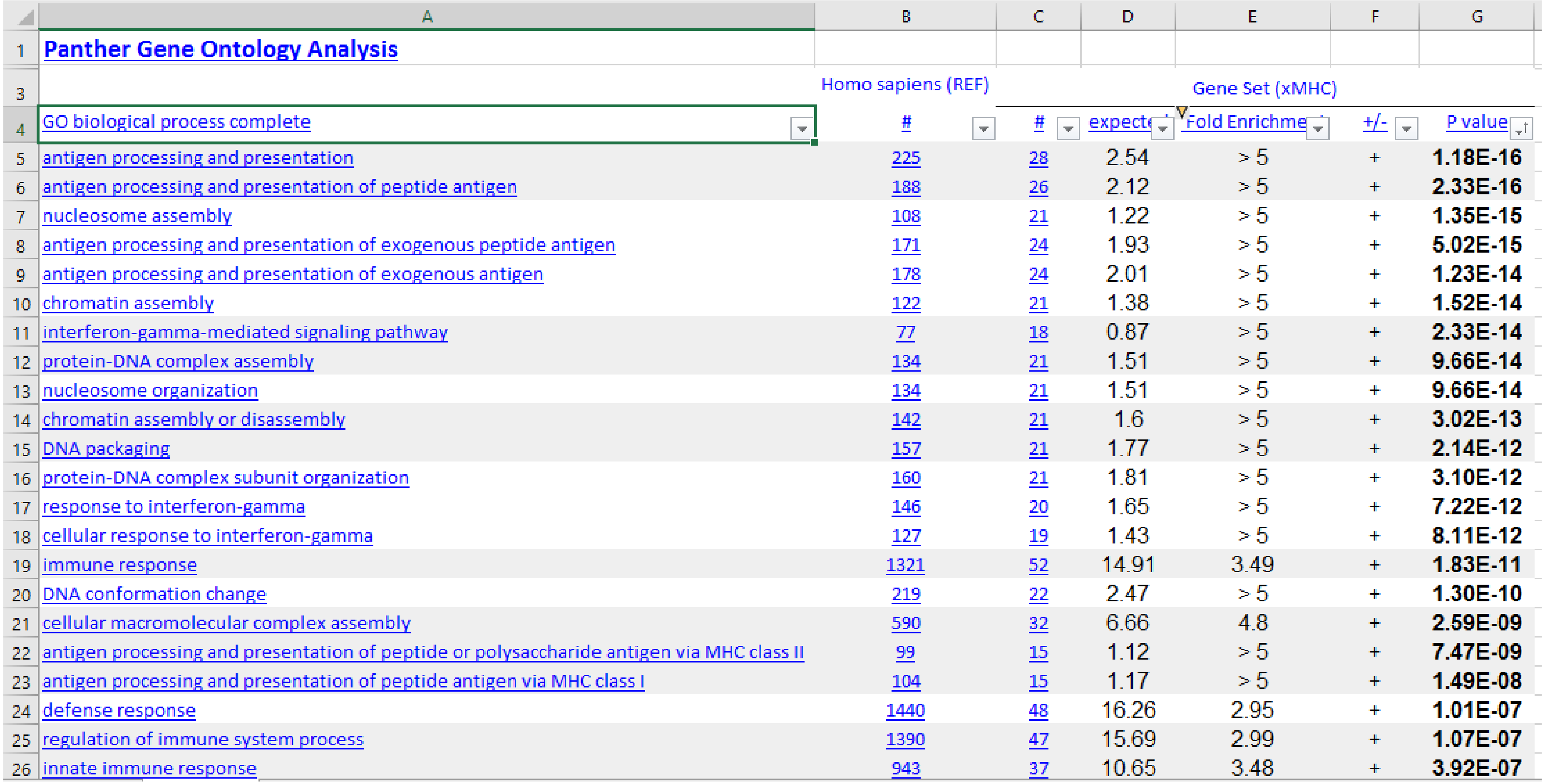
Only 13 microRNA and seven recognised
long non-coding RNA genes in the xHLA

The pseudogene content of xHLA is similar to the rest of
the genome (25.5% vs 24.0%)

RESULTS: Gene content

Genome (3.2Gb)	xHLA Region (25.7 to 33.4Mb)	Comparison
Total No of Genes 60155	Total No of Genes 674	…
Protein-coding genes 19881	Protein-coding genes 453	32.66 vs 67.21% *P*<0.0001
Non-coding RNA Genes 25411	Non-coding RNA Genes 54	42.63 vs 8.01% *P*<0.0001
Long non-coding RNA genes 15877	Long non-coding RNA genes 13	1.93 vs 26.39% *P*<0.0001
Small non-coding RNA genes 9534	Small non-coding RNA genes 7	1.04 vs 15.85% *P*<0.0001
Pseudogenes 14467	Pseudogenes 172	24.03 vs 25.52% *P* = 0.37

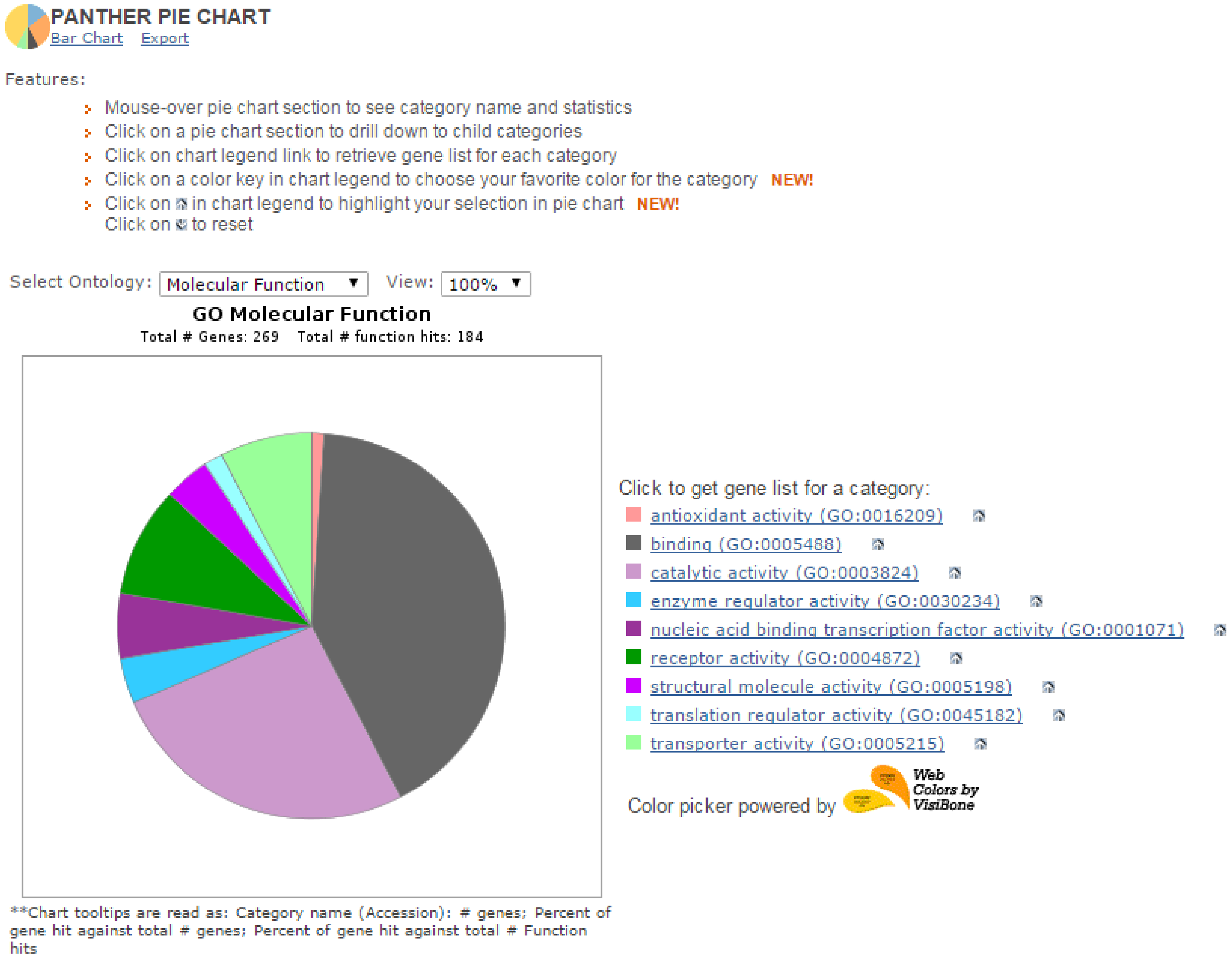
RESULTS: Gene ontology



	A	B	C	D	E	F	G
1	Panther Gene Ontology Analysis						
3		Homo sapiens (REF)	Gene Set (xMHC)				
4	GO biological process complete	#	#	expecte	Fold Enrichme	+/-	P value
5	antigen processing and presentation	225	28	2.54	> 5	+	1.18E-16
6	antigen processing and presentation of peptide antigen	188	26	2.12	> 5	+	2.33E-16
7	nucleosome assembly	108	21	1.22	> 5	+	1.35E-15
8	antigen processing and presentation of exogenous peptide antigen	171	24	1.93	> 5	+	5.02E-15
9	antigen processing and presentation of exogenous antigen	178	24	2.01	> 5	+	1.23E-14
10	chromatin assembly	122	21	1.38	> 5	+	1.52E-14
11	interferon-gamma-mediated signaling pathway	77	18	0.87	> 5	+	2.33E-14
12	protein-DNA complex assembly	134	21	1.51	> 5	+	9.66E-14
13	nucleosome organization	134	21	1.51	> 5	+	9.66E-14
14	chromatin assembly or disassembly	142	21	1.6	> 5	+	3.02E-13
15	DNA packaging	157	21	1.77	> 5	+	2.14E-12
16	protein-DNA complex subunit organization	160	21	1.81	> 5	+	3.10E-12
17	response to interferon-gamma	146	20	1.65	> 5	+	7.22E-12
18	cellular response to interferon-gamma	127	19	1.43	> 5	+	8.11E-12
19	immune response	1321	52	14.91	3.49	+	1.83E-11
20	DNA conformation change	219	22	2.47	> 5	+	1.30E-10
21	cellular macromolecular complex assembly	590	32	6.66	4.8	+	2.59E-09
22	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	99	15	1.12	> 5	+	7.47E-09
23	antigen processing and presentation of peptide antigen via MHC class I	104	15	1.17	> 5	+	1.49E-08
24	defense response	1440	48	16.26	2.95	+	1.01E-07
25	regulation of immune system process	1390	47	15.69	2.99	+	1.07E-07
26	innate immune response	943	37	10.65	3.48	+	3.92E-07

Liverpool Hope University EST. 1844

RESULTS: Gene ontology



PANTHER PIE CHART

Bar Chart Export

Features:

- Mouse-over pie chart section to see category name and statistics
- Click on a pie chart section to drill down to child categories
- Click on chart legend link to retrieve gene list for each category
- Click on a color key in chart legend to choose your favorite color for the category **NEW!**
- Click on ⇱ in chart legend to highlight your selection in pie chart **NEW!**
  Click on ↺ to reset

Select Ontology: Molecular Function ▼ View: 100% ▼

GO Molecular Function

Total # Genes: 269 Total # function hits: 184

Click to get gene list for a category:

- 🟥 antioxidant activity (GO:0016209)
- ⬛ binding (GO:0005488)
- 🟪 catalytic activity (GO:0003824)
- 🟦 enzyme regulator activity (GO:0030234)
- 🟪 nucleic acid binding transcription factor activity (GO:0001071)
- 🟩 receptor activity (GO:0004872)
- 🟪 structural molecule activity (GO:0005198)
- 🟦 translation regulator activity (GO:0045182)
- 🟩 transporter activity (GO:0005215)

Color picker powered by Web Colors by VisiBone

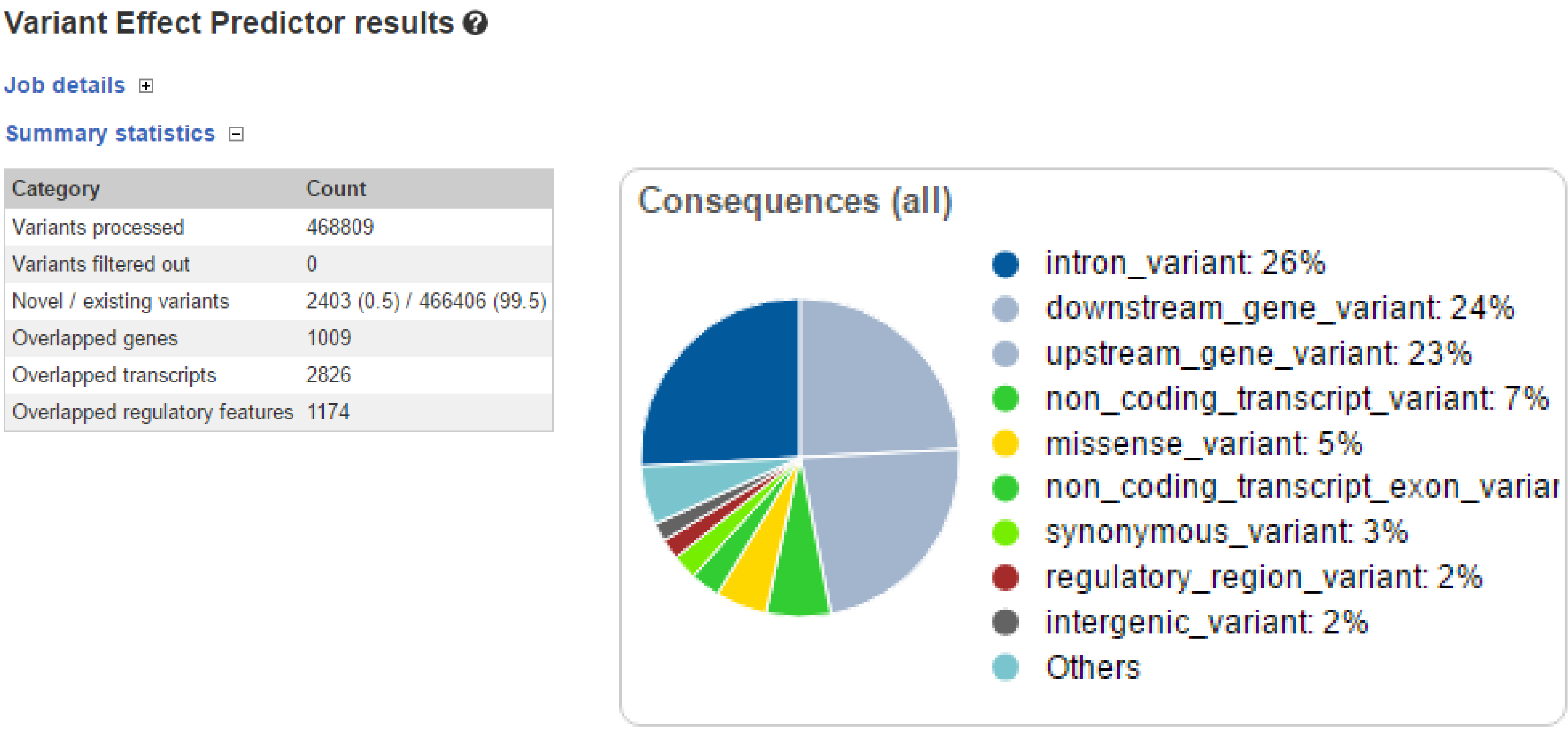
**Chart tooltips are read as: Category name (Accession): # genes; Percent of gene hit against total # genes; Percent of gene hit against total # Function hits

RESULTS: SNPs

xHLA makes up 0.24% of the genome, but contains 0.40% of all SNPs in the human genome

The most SNP-dense regions:
HLA-DR region (18,071 in 32.5 to 32.6Mb)
HLA-DQ region (12,189 in 32.6 to 32.7Mb)

RESULTS: Ensembl Variant Effect Predictor



Variant Effect Predictor results ❓

Job details ⊞

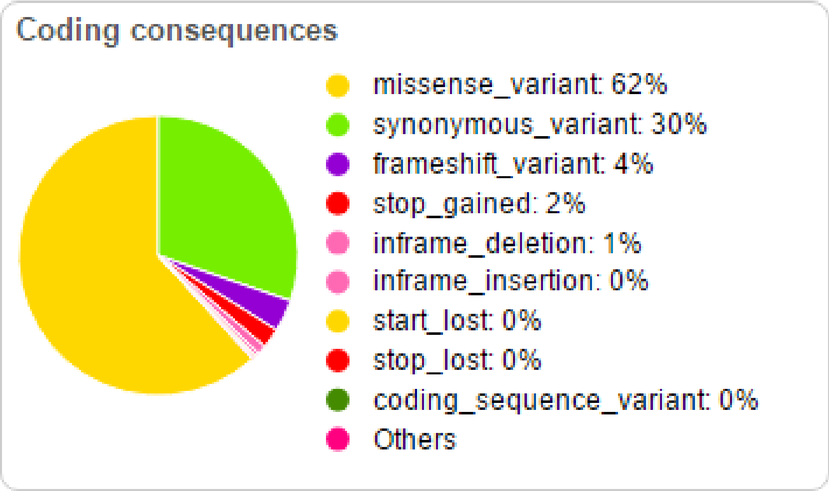
Summary statistics ⊟

Category	Count
Variants processed	468809
Variants filtered out	0
Novel / existing variants	2403 (0.5) / 466406 (99.5)
Overlapped genes	1009
Overlapped transcripts	2826
Overlapped regulatory features	1174

Consequences (all)
● intron_variant: 26%
● downstream_gene_variant: 24%
● upstream_gene_variant: 23%
● non_coding_transcript_variant: 7%
● missense_variant: 5%
● non_coding_transcript_exon_variar
● synonymous_variant: 3%
● regulatory_region_variant: 2%
● intergenic_variant: 2%
● Others

RESULTS: Missense SNPs

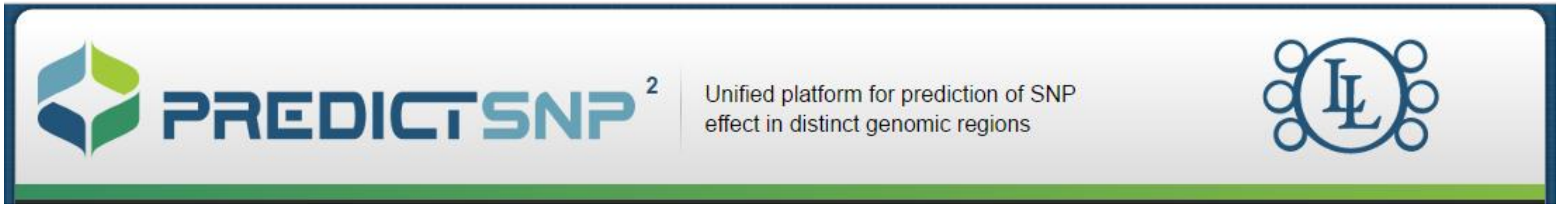
xHLA contains a higher proportion of missense SNPs (7.4%) than the rest of the genome (2.7%) as reported by NCBI ENTREZ SNP

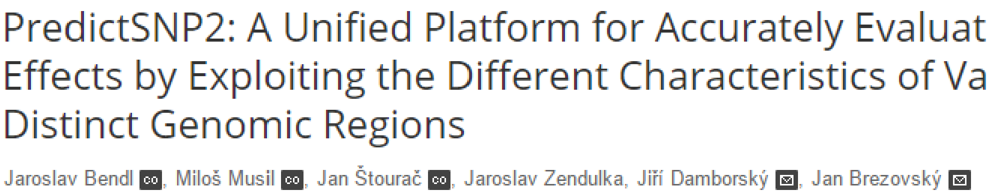


Ensembl VEP

RESULTS: Deleterious SNPs

We used PredictSNP2 algorithm to assess functionality of xHLA SNPs, and found that 45,302 (11.2%) of them were deleterious. The majority of deleterious SNPs were intergenic (18,610 or 41.1%). Rare nonsense mutations consisted of 2.7% (n=1,240) of the deleterious SNPs within xHLA.


PREDICTSNP 2
Unified platform for prediction of SNP
effect in distinct genomic regions



PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions

Jaroslav Bendl, Miloš Musil, Jan Štourač, Jaroslav Zendulka, Jiří Damborský, Jan Brezovský

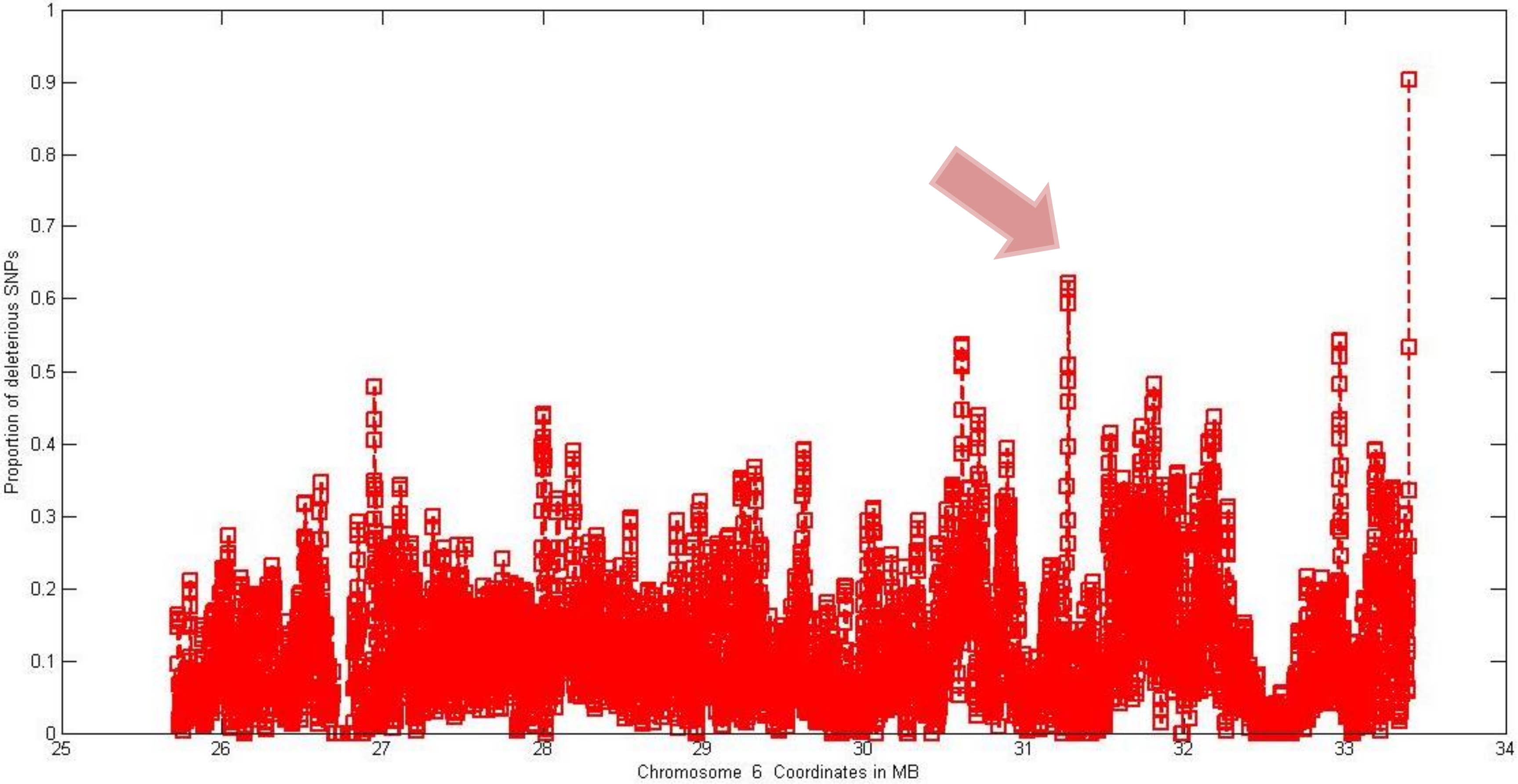
RESULTS: Deleterious SNPs

Plotting the density of deleterious SNPs across xHLA and sliding window analysis identified a hotspot (305/477 = 63.9%) for deleterious SNPs between 31,274kb and 31,281kb centromeric to *HLA-C* and containing two pseudogenes (*USP8P1*, *RPL3P2*).

The deleterious SNPs of this region included risk markers for type 1 diabetes (rs2524067), multiple sclerosis (rs7382297) and psoriasis (rs3132486) as well as strong eQTLs for *HCG22* (rs7382307, rs9264731, rs3930575, rs7382297).

Only three of the 305 deleterious SNPs in this region were also cancer somatic mutations.

RESULTS: Deleterious SNPs



Liverpool Hope University EST. 1844

RESULTS: Deleterious SNPs

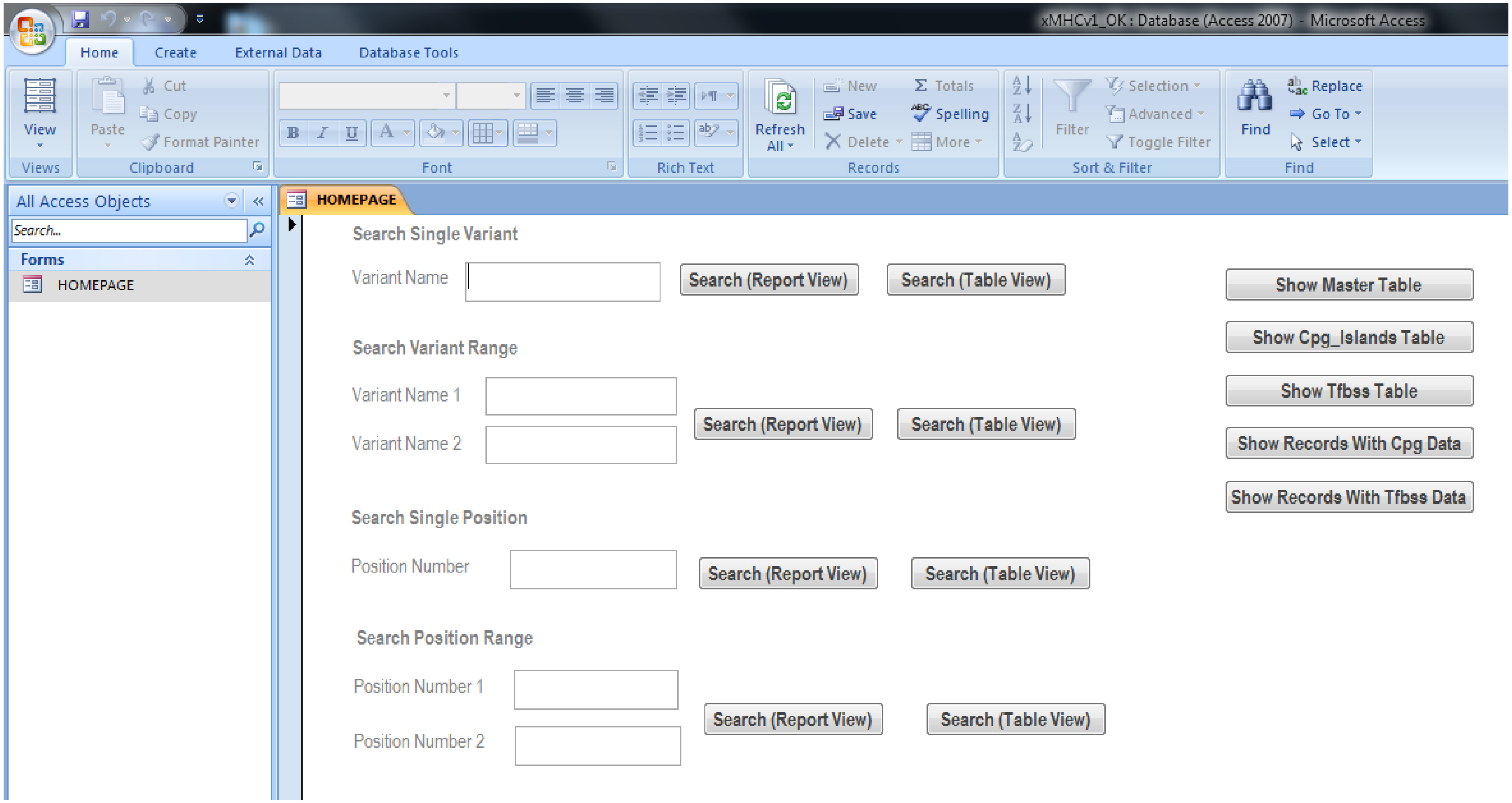
Only three of the 305 deleterious SNPs in the hotspot region were also cancer somatic mutations.

Of all xHLA SNPs, 8,139 were present in the COSMIC database as somatic cancer mutations. The proportion of COSMIC SNPs among the deleterious SNPs was higher (2.5 vs 1.9%, *P* < 0.0001).

CONCLUSIONS

In summary, xHLA makes up 0.24% of the genome, but contains 2.3% of protein-coding genes (but only 0.2% of non-coding genes) and 0.4% of all SNPs with a high missense SNP proportion. We also show that deleterious SNP distribution is not homogeneous across xHLA.

DATABASE



Search Single Variant

Variant Name [ ] Search (Report View) Search (Table View)

Search Variant Range

Variant Name 1 [ ]

Variant Name 2 [ ]

Search (Report View) Search (Table View)

Search Single Position

Position Number [ ] Search (Report View) Search (Table View)

Search Position Range

Position Number 1 [ ]

Position Number 2 [ ]

Search (Report View) Search (Table View)

Show Master Table

Show Cpg_Islands Table

Show Tfbss Table

Show Records With Cpg Data

Show Records With Tfbss Data

Available on request as an Access file. The full version will be released in the summer 2017 both as an Access file and as an online searchable database.

YOUR FUTURE
STARTS WITH HOPE

Liverpool Hope University EST. 1844