

Case Study #2: Narcolepsy: 11 SNPs in 180 bp – 6 non-synonymous !

DRB1\*1501, DQA1\*0102, DQB1\*0602 haplotype.

DQB1\*0602 **HLA class II** Chromosome 6p21.3 : 32,735,225

30 to 33Mbp A C B DRB DQA DQB

167 D -> G (in allele DQB1\*0202).

Siebold 2004: closely related *DQB1\*06011* allele, differs at only nine residues: G13A L26Y Y37D A38V (4 narcolepsy underlined below)

Narcolepsy Frequency = 0.05 to 0.067% dominant allele

DQB1\*0602 frequency = 24% in general 90% in Narcoleptics.

<http://www.ncbi.nlm.nih.gov/entrez/dispmim.cgi?id=161400>

<http://www.rcsb.org/pdb/explore.do?structureId=1UVQ>

>1UVQ:B HLA-DQB1\*0602 12345678901234567890123456789012345678901234567890

RDSPEDFV**F**QFK**G**MCYFTNGTERVRL**V**TRYIYNREE**Y**ARFDS DVG VYRAVTPQGRPD AEYWNSQKEVLEGT RAELDT  
VCRHNYEVAFRGILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVRWFRNDQEETAGVVSTPLIRNGDWT  
FQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKVD

>06011

RD**P**PEDFV**L**QFK**A**MCYFTNGTERV**R****Y**VTRYIYNREED**D**VRFDS DVG VYRAVT  
PQGRPD AEYWNSQKDILERTRAELDTVCRHNYEVAFRGILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIK  
VRWFRNDQEETAGVVSTPLIRNGDWT**F**QILVMLEMTP**Q**HGDVYTCHVEHPSLQSPITVEWRAQSESAQNKMLSGIGG  
FVLGLIFLGLGLIIRQRSQKGLLH

>[L34105](#) HLA-DQB1\*0602

```
aga gac tct ccc gag gat ttc gtg ttc cag ttt aag ggc atg tgc tac
ttc acc aac ggg acg gag cgc gtg cgt ctt gtg acc aga tac atc tat
aac cga gag gag tac gcg cgc ttc gac agc gac gtg ggg gtg tac cgc
gcg gtg acg ccg cag ggg cgg cct gat gcc gag tac
```

> [L40181](#) HLA-DQB1\*06011

```
aga gac ctt ccc gag gat ttc gtg ctc cag ttt aag gcc atg tgc tac
ttc acc aat ggg acg gag cgc gtg cgt tat gtg acc aga tac atc tat
aac cga gag gag gac gtg cgc ttc gac agc gac gtg ggg gtg tat cgg
gcg gtg acg ccg cag ggg cgg cct gac gcc gag tac
```

>[gnl|dbSNP|rs12722107](#) Query= 06011

rs=12722107|pos=1001|len=2001|taxid=9606|mol="genomic"|class=1|alleles="C/T"| build=126 Length=2001

Score = 259 bits (140), Expect = 1e-66  
Identities = 159/169 (94%), Gaps = 2/169 (1%) Strand=Plus/Plus

```
Query 13 GAGGATTTTCGTGCTCCAGTTTAAAGGCCATGTGCTACTTCACCAATGGGACGGAGCGCGTG 72
Sbjct 989 .....Y.....G.....C..... 1048
```

```
Query 73 CGTTATGTGACCAGATAACATCTATAACCGAGAGGAGGACGTGCGCTTCGACAGCGACGT 131
Sbjct 1049 ...CT.....T...C..... 1107
```

```
Query 132 GGGGGTGTATCGGGCGGTGACGCCGAGGGGCGGCCTGACGCCGAGTAC 180
Sbjct 1108 .....C..C.....T..... 1156
```

**Original before Manual redo:**

```
Query 73 CGT-TATGTGACCAGATAACATCTATAACCGAGAGGAGGACGTGCGCTTCGACAGCGACGT 131
Sbjct 1049 ...C..-.....T...C..... 1107
```

>[gnl|dbSNP|rs29029547](#) Query = 0602  
rs=29029547|pos=1001|len=2001|taxid=9606|mol="genomic"|class=1|alleles="C/G"|build=125 Length=2001

Score = 311 bits (168), Expect = 4e-82  
Identities = 168/168 (100%), Gaps = 0/168 (0%) Strand=Plus/Plus

Query	13	GAGGATTTTCGTGTTCCAGTTTAAGGGCATGTGCTACTTCACCAACGGGACGGAGCGCGTG	72
Sbjct	825	.....	884
Query	73	CGTCTTGTGACCAGATACATCTATAACCGAGAGGAGTACGCGCGCTTCGACAGCGACGTG	132
Sbjct	885	.....	944
Query	133	GGGGTGTACCGCGCGGTGACGCCGCAGGGGCGGCCTGATGCCGAGTAC	180
Sbjct	945	.....	992

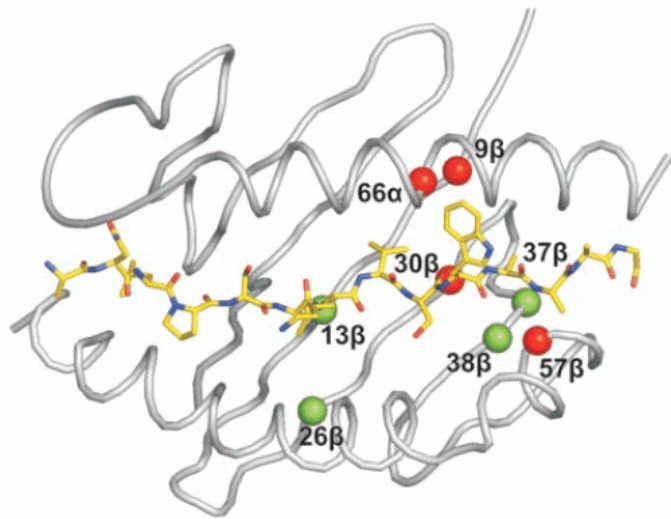



Fig. 2. Disease-associated polymorphisms. (A) Disease-associated residues as highlighted by sequence alignment of selected DQ chains. Residues associated with T1D are red, and those associated with narcolepsy are green. The binding groove pockets to which these residues contribute are indicated. (B) The DQ0602 binding groove is shown in  $C\alpha$  trace, with the hypocretin peptide depicted in the standard format (Fig. 1) and the  $C\alpha$  positions of residues highlighted in A indicated by spheres (red, T1D; green, narcolepsy).


Siebold 2004 Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?pubmedid=14769912>

<http://www.genecards.org/cgi-bin/carddisp.pl?gene=HLA-DQB1>


**Allele Frequency Data for SNP rs12722107:**

Genomic Data				Transcription Data			
Snp ID	Validation	Chr 6 pos	Sequence	Recs	AA Chg	Type	More
rs12722107 <sup>1,2</sup>	no-info	32740811(-)	GCAGAGGATTTTCGTG <b>T</b> / <b>C</b> TCCAGTTTAAGGGCA	1	--	int <sup>1</sup> ese <sup>3</sup>	

<sup>1</sup>[NCBI SNP Database](#), <sup>2</sup>[Ensembl](#), <sup>3</sup>[PupaSUITE/PupaSNP](#)

Expanded Allele Frequencies information						
Submitter	Population Class	Population Name	Sample Size	Het	frequencies chart	Allele frequencies
UWGC	MULTI-NATIONAL	HLA_PANEL	18	--		C:0.06, T:0.94

### Expression Data for SNP rs29029547:

Genomic Data				Allele Frequencies				PupaSNP information <sup>3</sup> 		
Snp ID	Validation	Chr 6 pos	Sequence	Recs	Allele freq	Pop	Total sample	ESE name	SNP effect	Transcript
rs29029547 <sup>1,2</sup>	no-info	32740647(-)	CGAGTACTGGAACAGC/GCAGAAGGAAGT CCTG	--	--	--	--	srp40	Lose	<a href="#">ENST00000374943</a>
								sf2	Lose	<a href="#">ENST00000374943</a>
								srp40	Lose	<a href="#">ENST00000334399</a>
								sf2	Lose	<a href="#">ENST00000334399</a>

<sup>1</sup>[NCBI SNP Database](#), <sup>2</sup>[Ensembl](#), <sup>3</sup>[PupaSUITE/PupaSNP](#)

Expanded Transcription Data information						
NUC Change	AA Change	Location Type	mRNA Accession	Protein Accession	Phase	Protein Position
--	--	intron	<a href="#">NM_002123</a>	<a href="#">NP_002114</a>	--	--