SNP Data Integration Using Genome Browsers

Belinda M. Giardine, Riemer, C., Hardison, R.C. Pennsylvania State University, Center for Comparative Genomics and Bioinformatics, University Park, PA 16802, USA

giardine@bx.psu.edu
What can we learn about these SNPs using a Genome Browser?

- **rs538089**
  - Known coding SNP

- **Novel SNP found in !Gubi (KB1)**
  - Shuster et al. Nature 2010

- **rs6983267**
  - Loci at 8q24 have been implicated in
    - Prostate cancer
    - Breast cancer
    - Other cancers
  - 1.2 Mb region with no known protein-coding genes
  - Adjacent to MYC
Tools I will use (others can be used)

- BX Genome Browser
  - Partial mirror of UCSC Genome Browser specializing in Personal Genomes
  - main.genome-browser.bx.psu.edu

- PhenCode (via Locus Variants track)
  - Compilation of variants from LSDBs
    - Suggestion for curators not yet participating: More data makes this more useful and brings in more users.
  - phencode.bx.psu.edu
rs538089: a known coding SNP
Basic

Read frequency and/or scores

SIFT and/or PolyPhen predictions

Coding sequence effects
Compilation of Human Variants from LSDBs (NM_170707(LMNA):c.861T>C)

- **HGVS name:** NM_170707(LMNA):c.861T>C
- **Position:** chr1.154371652-154371652
- **Band:** 1q22
- **Genomic Size:** 1
- **Strand:** +

*Note the DNA retrieved by the above link is the genomic sequence.

- **source:** Human Intermediate Filament Database
- **location:** exon
- **type:** substitution

- **Common name:** c.861T>C
- **External links:**
  - [HIFD - c.861T>C](#)

- **Effect on Protein:** p.=

- **Phenotype association:** not phenotype-associated
| Type | Gene Name | Protein Name | cDNA RefSeq ID | Protein RefSeq ID | ID | DNA Changes | cDNA Variant | Protein Changes | Protein variant | Affected Domain | Associated Disease | PubMed | Title                                                                 | Authors                                                                 | Journal   | Year |
|------|-----------|--------------|----------------|------------------|----|-------------|--------------|----------------|----------------|----------------|----------------|------------------|--------|----------------------------------------------------------------------|------------------------------------------------------------------------|----------|------|
Summary of what we learned

- In the LMNA gene, codon 287
- Synonymous
- Not phenotype associated
- Average allele frequency
  - $T = .27$
  - $C = .73$
- Caucasian allele frequency
  - $T = .11$
  - $C = .89$
- Found in 4 YRI, and 4 Bushmen
Novel coding SNP found in KB1
**KB1 Genome Variants, 454 (T)**

Position: [chr11:63895434-63895434](https://example.com/)
Band: 11q13.1
Genomic Size: 1
Strand: +

View DNA for this feature

Alleles are relative to forward strand of reference genome:

<table>
<thead>
<tr>
<th>Allele</th>
<th>Frequency</th>
<th>Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>2</td>
<td>not available</td>
</tr>
</tbody>
</table>

Coding sequence changes are relative to strand of transcript:

u001oa 1: starting positions codon 736 cds 2206
CGC > CTC
R > L

<table>
<thead>
<tr>
<th>Amino acid properties</th>
<th>R</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>polarity</td>
<td>polar</td>
<td>nonpolar</td>
</tr>
<tr>
<td>acidity</td>
<td>basic (strongly)</td>
<td>neutral</td>
</tr>
<tr>
<td>hydropathy</td>
<td>-4.5</td>
<td>3.8</td>
</tr>
</tbody>
</table>
**Gene detail page**

**Comments and Description Text from UniProtKB**

**ID:** ESSA4 HUMAN  
**DESCRIPTION:** RecName: Full=Ribosomal protein S6 kinase alpha-4; EC=2.7.11.1; AltName: Full=Serum- and stress-activated protein kinase 2; AltName: Full=90 kDa ribosomal protein S6 kinase 4; AltName: Full=Ribosomal protein kinase B; Short=SERK2.  
**FUNCTION:** Serine/threonine kinase that may play a role in mediating the growth-factor and stress-induced activation of the transcription factor CREB. Essential role in the control of RELA transcriptional activity in response to TNF. Phosphorylates Ser-10 of histone H3 in response to mitogenes, stress stimuli and epidermal growth factor (EGF) and result in the transcriptional activation of several immediate early genes, including proto-oncogenes FOS and JUN (by similarity). Mediates the mitogen- and stress-induced phosphorylation of high mobility group protein 14 (HMGB14).  
**CATALYTIC ACTIVITY:** ATP + a protein = ADP + a phosphoprotein.  
**COFACTOR:** Magnesium.  
**ENZYME REGULATION:** Appears to be activated by multiple phosphorylations on threonine and serine residues. ERK1/2 and p38 kinases may play a role in this process.  
**SUBUNIT:** Forms a complex with either ERK1 or ERK2 in quiescent cells which transiently dissociates following mitogenic stimulation. Also associates with MAPK14/p38-alpha.  
**INTERACTION:** Q65385 MAPK14, NupExp=1, IntAct=BPI-75933, BPI-75946, FJ7361 MAPK3, NupExp=1; IntAct=BPI-75933, BPI-75998.  
**SUBCELLULAR LOCATION:** Nuclear.  
**MISCELLANEOUS:** Enzyme activity requires the presence of both kinase domains (by similarity).  
**SIMILARITY:** Belongs to the protein kinase superfamily. AGC Ser/Thr protein kinase family. S6 kinase subfamily.  
**SIMILARITY:** Contains 1 AGC-kinase C-terminal domain.  
**SIMILARITY:** Contains 2 protein kinase domains.

**Comparative Toxicogenomics Database (CTD)**

The following chemicals interact with this gene:
- [D000082](#) Acetaminophen
- [D014212](#) Tretinoin
- [D014415](#) Vancomycin

**Biochemical and Signaling Pathways**

- **KEGG - Kyoto Encyclopedia of Genes and Genomes**
  - [isa04010](#) - MAPK signaling pathway

- **BioCarta from NCI Cancer Genome Anatomy Project**
  - [h_mapkPathway](#) - MAPKinase Signaling Pathway
Ancestral allele is R, but KB1 is homozygous L
Coding, codon 736
Non-synonymous R>L
Changes amino acid sequence of a nuclear protein kinase that activates transcription factors in response to stress and mitogens
  - Does it change the activity?
The arginine (R) is conserved from humans to dogs
Possible interaction with Tylenol
This SNP at 8q24 has been implicated in prostate and other cancers. The closest known gene is MYC, about 335 kb away.
## Human Genome Diversity Project SNP Population Allele Frequencies

<table>
<thead>
<tr>
<th>Ancestral Allele</th>
<th>G</th>
<th>Derived Allele</th>
<th>T</th>
</tr>
</thead>
</table>

### Ancestral Allele Frequencies for rs6983267:

- Adygei: 0.6471
- Balochi: 0.3953
- BantuKenya: 1.0000
- BantuSouthAfrica: 1.0000
- Basque: 0.5000
- Bedouin: 0.6900
- Bashkortostan: 1.0000
- Brahui: 0.4200
- Burushko: 0.4200
- Cambodian: 0.3000
- Colombian: 0.5714
- Druze: 0.4500
- Dweer: 0.3829
- Dzao: 0.4524
- French: 0.5714
- Han: 0.4265
- Han-NorthChina: 0.4000
- Hazara: 0.4545
- Hazten: 0.5000
- Italian: 0.5417
- Japanese: 0.4107
- Kalash: 0.6087
- Karibiana: 0.5000
- Lahu: 0.3125
- Makrani: 0.6000
- Mandenka: 0.9545
- Maya: 0.5238

### Derived Allele Frequencies:

- MbutiPygmy: 1.0000
- MbutiPygmy2: 0.3182
- Miao: 0.4000
- Mongula: 0.3000
- Mozabite: 0.9074
- Naga: 0.9000
- Occidian: 0.3333
- Ocegen: 0.6111
- Palestinian: 0.5000
- Papuan: 0.1176
- Pathian: 0.4723
- Fiina: 0.6429
- Russian: 0.4200
- San: 1.0000
- Sardinian: 0.6643
- She: 0.6000
- Shadili: 0.5208
- Sursi: 0.6250
- Tu: 0.5500
- Tuju: 0.4000
- Tuscan: 0.5000
- Uygur: 0.8000
- Xibo: 0.4444
- Yakuti: 0.3200
- Yi: 0.3000
- Yoruba: 1.0000
Epigenetic data that illuminates a cancer-associated SNP

1. Confirm occupancy by TCF7L2 at SNP

2. Also see occupancy by RAD21, CTCF, JUND, P300

3. Also see histone modifications and DHSs

4. Gencode annotation reveals another gene, transcripts in nucleus
HGDP provides frequency of SNP in many populations

Found in 5 Caucasians and 2 Koreans; homozygous reference sequence in Africans (Bantu, San, and Yoruba)

Epigenetic support for function
Conclusion

- Genome Browsers can be used to learn about SNP consequences.
- Pooling known information into browsers makes it easier to find.
Acknowledgments:

- Our group
  - Ross Hardison
  - Webb Miller
  - Cathy Riemer

- Data Contributors
  - LSDB’s who have contributed to PhenCode
  - Individuals who contributed public genomes

- Browsers
  - UCSC Genome Browser
  - Ensembl Genome Browser

- Funding
  - Huck Institutes of the Life Sciences