Table 1. Trans-membrane helix analysis within loci

Predictions for transmembrane helices made for sequences from loci with multiple variants. In the figure TM is transmembrane proteins, TMH is transmembrane helices, SP-PRO is signal peptide predicted with SignalP and transmembrane helices predicted with PRODIV, SP-ENS is signal peptide predicted with SignalP, transmembrane helices predicted with ENSEMBLE1.0 and PHOBIUS is transmembrane helices predicted with Phobius.

Table 2. 30 loci identified with altered MDA

Classes of behavior assigned to each of the 30 loci identified as coding for isoforms with altered MDAs. The way MDAs were altered in splice isoforms were characterized into five different behaviors. *Truncation* indicates isoforms lost N or C termini causing an alteration to the MDA; the truncation is not expected to substantially alter a remaining domain. *Deletion* indicates that an isoform has lost a domain from within the MDA; again the alteration is not expected to alter other domains. *Substitution* indicates that a domain has been swapped in place of another. *Damaged* indicates that the one (or more) of the isoforms is expected to have a partial, probably non-folding, domain. *Trivial* is a catch-all category for MDA variation that is an artifact of the domain prediction process, unreliable underlying sequence data or other non-biologically meaningful process. Blue highlighted rows indicate loci that are further discussed in the report. Loci may belong to more than one category. Truncation was the most common reason for an MDA to be altered.

Table 3. ENCODE peptides with abnormal domain size

67 GENCODE peptides with domains of suspicious length. 53 of the ENCODE transcripts encoding the peptides are partial sequences. Many of these annotations have since been amended.

Table 4. Comparison of experimental and predicted sets of interactions with ENCODE genes

The columns ‘Genes’ and ‘Interactions’ give the total numbers of genes and interactions in the respective data sets. The columns ‘ENCODEn genes’ and ‘ENCODEn interactions’ indicate the number of ENCODE genes with interaction partners and the number of interactions with ENCODE genes, respectively. ‘Non-ENCODEn genes’ is the number of interacting genes without counting ENCODE genes. The columns ‘ENCODEn–ENCODEn’ and ‘non-ENCODEn – non-ENCODEn’ refer to interactions solely between ENCODE genes and between non-ENCODEn genes, respectively. ‘Domain-Domain’ lists the number of Pfam domain-domain interaction predicted to be responsible for specific ENCODE interactions of gene products. The column ‘Different domains’ contains the number of distinct Pfam domains involved in all domain-domain interactions. The column ‘MIM disorders’ shows the number of different diseases associated with non-ENCODEn genes.

Table 5. ENCODE Genes and Disease

ENCODE genes associated with diseases contained in OMIM. Only genes coding for multiple splice isoforms are shown for clarity.

Table 6. ENCODE Gene Interaction Partners and Disease

All ENCODE genes and interaction partners associated with the same OMIM diseases.
<table>
<thead>
<tr>
<th>Loci encoding for variants identically predicted as globular</th>
<th>SP-PRO</th>
<th>SP-ENS</th>
<th>PHOBIUS</th>
<th>CONSENSUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>184</td>
<td>163</td>
<td>166</td>
<td></td>
<td>185</td>
</tr>
<tr>
<td>Loci encoding for TM variants predicted with the same TM structure</td>
<td>26</td>
<td>26</td>
<td>34</td>
<td>28</td>
</tr>
<tr>
<td>Loci encoding for both globular and TM(^1) variants</td>
<td>19</td>
<td>37</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>Loci encoding for TM variants with different number of TMH(^2)</td>
<td>17</td>
<td>22</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>Loci encoding for TM variants with inverted orientation</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1
Table 1. Trans-membrane helix analysis within loci

Predictions for transmembrane helices made for sequences from loci with multiple variants. In the figure TM is transmembrane proteins, TMH is transmembrane helices, SP-PRO is signal peptide predicted with SignalP and transmembrane helices predicted with PRODIV, SP-ENS is signal peptide predicted with SignalP, transmembrane helices predicted with ENSEMBLE1.0 and PHOBIUS is transmembrane helices predicted with Phobius.

Table 2. 30 loci identified with altered MDA

Classes of behavior assigned to each of the 30 loci identified as coding for isoforms with altered MDAs. The way MDAs were altered in splice isoforms were characterized into five different behaviors. Truncation indicates isoforms lost N or C termini causing an alteration to the MDA; the truncation is not expected to substantially alter a remaining domain. Deletion indicates that an isoform has lost a domain from within the MDA; again the alteration is not expected to alter other domains. Substitution indicates that a domain has been swapped in place of another. Damaged indicates that the one (or more) of the isoforms is expected to have a partial, probably non-folding, domain. Trivial is a catch-all category for MDA variation that is an artifact of the domain prediction process, unreliable underlying sequence data or other non-biologically meaningful process. Blue highlighted rows indicate loci that are further discussed in the report. Loci may belong to more than one category. Truncation was the most common reason for an MDA to be altered.

Table 3. ENCODE peptides with abnormal domain size

67 GENCODE peptides with domains of suspicious length. 53 of the ENCODE transcripts encoding the peptides are partial sequences. Many of these annotations have since been amended.

Table 4. Comparison of experimental and predicted sets of interactions with ENCODE genes

The columns ‘Genes’ and ‘Interactions’ give the total numbers of genes and interactions in the respective data sets. The columns ‘ENCODE genes’ and ‘ENCODE interactions’ indicate the number of ENCODE genes with interaction partners and the number of interactions with ENCODE genes, respectively. ‘Non-ENCODE genes’ is the number of interacting genes without counting ENCODE genes. The columns ‘ENCODE-ENCODE’ and ‘non-ENCODE – non-ENCODE’ refer to interactions solely between ENCODE genes and between non-ENCODE genes, respectively. ‘Domain-Domain’ lists the number of Pfam domain-domain interaction predicted to be responsible for specific ENCODE interactions of gene products. The column ‘Different domains’ contains the number of distinct Pfam domains involved in all domain-domain interactions. The column ‘MIM disorders’ shows the number of different diseases associated with non-ENCODE genes.

Table 5. ENCODE Genes and Disease

ENCODE genes associated with diseases contained in OMIM. Only genes coding for multiple splice isoforms are shown for clarity.

Table 6. ENCODE Gene Interaction Partners and Disease

All ENCODE genes and interaction partners associated with the same OMIM diseases.