Supplementary Material

1. Simulation based on randomized variations.

We randomized the chromosomal positions of all the polymorphism annotations. Then we re-ran our procedure (with and without Swiss-Prot) using these randomized variants to obtain new ORFs. An ORF is said to be new if it shares no same-frame codons with the ‘old ORF’.

Below is the summary:

<table>
<thead>
<tr>
<th></th>
<th>&gt;=1 same-frame codons</th>
<th>No same-frame codons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Swiss-Prot</td>
<td>64,213</td>
<td>22,203</td>
</tr>
<tr>
<td>After Swiss-Prot</td>
<td>58,357</td>
<td>18,726</td>
</tr>
</tbody>
</table>

Number of changes in ORFs after modification with random polymorphisms

<table>
<thead>
<tr>
<th></th>
<th>&gt;=1 same-frame codons</th>
<th>No same-frame codons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Swiss-Prot</td>
<td>15,612</td>
<td>5,930</td>
</tr>
<tr>
<td>After Swiss-Prot</td>
<td>9,390</td>
<td>3,004</td>
</tr>
</tbody>
</table>

Note that the number of ORF changes that share no same-frame codons with old ORFs are smaller than those that shares with >=1 same-frame codons. Also, when we consider random polymorphism, the number of changes after Swiss-Prot is smaller than real polymorphism (3,004/5,930 versus 18,726/22,203). The latter indicates that most of the real SNPs coincide more with start/stop codon than random SNPs.

We also looked at how many of these random novel ORFs overlap with other species. The table below exhibits the comparison. Table below showed that the number of overlaps from the ORFs generated by random SNP is fewer than non-random.

<table>
<thead>
<tr>
<th></th>
<th>Overlap Non Random</th>
<th>Overlap Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>chicken (galGal3)</td>
<td>65</td>
<td>10</td>
</tr>
<tr>
<td>medaka (oryLat2)</td>
<td>68</td>
<td>10</td>
</tr>
<tr>
<td>zebrafish (danRer7)</td>
<td>71</td>
<td>9</td>
</tr>
<tr>
<td>zebrafinch (taeGut1)</td>
<td>72</td>
<td>8</td>
</tr>
<tr>
<td>panda (ailMel1)</td>
<td>76</td>
<td>10</td>
</tr>
</tbody>
</table>
2. The occurrence of new ORFs in 11 populations.

Table below shows the number of occurrences (frequency) of ORFs that appear in exactly 1 to 11 populations. For example there are 2,912 ORFs that appear in exactly 4 populations. It shows that majority (54%) of these ORFs do appear in all the 11 populations and around 23% appear in 4 populations (CEU, CHB, JPT and YRI).

<table>
<thead>
<tr>
<th>Number of Populations</th>
<th>Frequency of New ORFs</th>
<th>Percentage from total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96</td>
<td>0.80%</td>
</tr>
<tr>
<td>2</td>
<td>84</td>
<td>0.70%</td>
</tr>
<tr>
<td>3</td>
<td>214</td>
<td>1.79%</td>
</tr>
<tr>
<td>4</td>
<td>2,812</td>
<td>23.47%</td>
</tr>
<tr>
<td>5</td>
<td>292</td>
<td>2.44%</td>
</tr>
<tr>
<td>6</td>
<td>216</td>
<td>1.80%</td>
</tr>
<tr>
<td>7</td>
<td>316</td>
<td>2.64%</td>
</tr>
<tr>
<td>8</td>
<td>326</td>
<td>2.72%</td>
</tr>
<tr>
<td>9</td>
<td>346</td>
<td>2.89%</td>
</tr>
<tr>
<td>10</td>
<td>801</td>
<td>6.69%</td>
</tr>
<tr>
<td>11</td>
<td>6,479</td>
<td>54.07%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>11,962</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

We compute the frequency of ORFs that are affected by polymorphism by populations. Below is the table.
3. Number of gene change by type of polymorphism.

The polymorphisms included are deletion (DLT), indel (IND), insertion (INS), microsatellite (MIC), multiple nucleotide polymorphism (MNP) and single nucleotide polymorphism (SNP). The last row shows the initial overall frequency of the polymorphism in the annotation file (snp131.txt).

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>New ORFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASW</td>
<td>8,236</td>
<td></td>
</tr>
<tr>
<td>CEU</td>
<td>11,620</td>
<td></td>
</tr>
<tr>
<td>CHB</td>
<td>11,594</td>
<td></td>
</tr>
<tr>
<td>CHD</td>
<td>7,409</td>
<td></td>
</tr>
<tr>
<td>GIH</td>
<td>7,707</td>
<td></td>
</tr>
<tr>
<td>JPT</td>
<td>11,574</td>
<td></td>
</tr>
<tr>
<td>LWK</td>
<td>8,288</td>
<td></td>
</tr>
<tr>
<td>MEX</td>
<td>7,968</td>
<td></td>
</tr>
<tr>
<td>MKK</td>
<td>8,146</td>
<td></td>
</tr>
<tr>
<td>TSI</td>
<td>7,907</td>
<td></td>
</tr>
<tr>
<td>YRI</td>
<td>11,674</td>
<td></td>
</tr>
</tbody>
</table>

It must be noted that not all the polymorphism data in dbSNP131 are population specific. Out of 25,877,929 polymorphism data in only 4,164,537 are population specific.
4. Cumulative count of the E-value from the alignment of new ORFs to Swiss-Prot protein

5. Effect of second SNP remodification of new mRNA.

In order to verify whether the new ORF could be significantly affected by other polymorphisms, we perform a further experiment. From the new mRNA after initial modification we reapply polymorphisms onto it (2\textsuperscript{nd} modification). The table below shows the number of ORFs derived from the initial mRNA versus ORFs derived from second modification.

<table>
<thead>
<tr>
<th></th>
<th>mRNA (1\textsuperscript{st} modification)</th>
<th>Newer mRNA (2\textsuperscript{nd} modification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unchanged</td>
<td>202,232</td>
<td>202,172</td>
</tr>
<tr>
<td>Changed</td>
<td>18,726</td>
<td>18,786</td>
</tr>
</tbody>
</table>

Although the second modification does change the counts of the disrupted ORFs, the difference is not significant.
6. Comparison of GO terms of ORFs before and after modification.

In this experiment the genes from the human proteome in Uniprot are used as the background set. Secondly we apply new ORFs and old ORFs as target set separately. The results showed that polymorphism does indeed change the GO assignment.

We find it is surprising that the GO terms of ORFs before modification are significant. Presumably, this is because the gene names are filtered through Swiss-prot, which contain primarily human curated genes. The P-value of GO terms for ORFs after modification are lower (more significant) than those before modification. Furthermore, the enrichment of the GO terms is higher for all for those after modification.

The total number of GO terms from all 3 categories is 11,109.

<table>
<thead>
<tr>
<th>Description</th>
<th>P-value</th>
<th>FDR q-value</th>
<th>Enrichment (N,B,n,b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>membrane-bounded organelle</td>
<td>5.64E-13</td>
<td>7.24E-10</td>
<td>(17184,6973,16844,6897)</td>
</tr>
<tr>
<td>intracellular membrane-bounded organelle</td>
<td>6.09E-13</td>
<td>3.91E-10</td>
<td>(17184,6968,16844,6892)</td>
</tr>
<tr>
<td>organelle</td>
<td>1.61E-12</td>
<td>6.89E-10</td>
<td>(17184,7913,16844,7819)</td>
</tr>
<tr>
<td>intracellular organelle</td>
<td>2.08E-12</td>
<td>6.68E-10</td>
<td>(17184,7896,16844,7802)</td>
</tr>
<tr>
<td>nucleus</td>
<td>2.84E-11</td>
<td>7.29E-09</td>
<td>(17184,4640,16844,4598)</td>
</tr>
<tr>
<td>plasma membrane</td>
<td>7.25E-10</td>
<td>1.55E-07</td>
<td>(17184,3269,16844,3244)</td>
</tr>
<tr>
<td>cytoplasm</td>
<td>7.67E-07</td>
<td>1.41E-04</td>
<td>(17184,3892,16844,3850)</td>
</tr>
<tr>
<td>membrane</td>
<td>2.60E-06</td>
<td>4.18E-04</td>
<td>(17184,5393,16844,5324)</td>
</tr>
<tr>
<td>intrinsic to plasma membrane</td>
<td>4.07E-06</td>
<td>5.82E-04</td>
<td>(17184,1062,16844,1058)</td>
</tr>
<tr>
<td>integral to plasma membrane</td>
<td>8.31E-06</td>
<td>1.07E-03</td>
<td>(17184,1021,16844,1017)</td>
</tr>
<tr>
<td>cytosol</td>
<td>1.56E-05</td>
<td>1.82E-03</td>
<td>(17184,2282,16844,2268)</td>
</tr>
<tr>
<td>nuclear part</td>
<td>2.82E-05</td>
<td>3.02E-03</td>
<td>(17184,2436,16844,2412)</td>
</tr>
<tr>
<td>plasma membrane part</td>
<td>2.17E-04</td>
<td>2.15E-02</td>
<td>(17184,1865,16844,1847)</td>
</tr>
<tr>
<td>Golgi apparatus</td>
<td>2.66E-04</td>
<td>2.44E-02</td>
<td>(17184,631,16844,629)</td>
</tr>
<tr>
<td>cell part</td>
<td>2.67E-04</td>
<td>2.29E-02</td>
<td>(17184,12362,16844,12147)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>P-value</th>
<th>FDR q-value</th>
<th>Enrichment (N,B,n,b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell part</td>
<td>1.11E-33</td>
<td>1.43E-30</td>
<td>(17184,12362,5262,4108)</td>
</tr>
<tr>
<td>intracellular part</td>
<td>5.77E-29</td>
<td>3.71E-26</td>
<td>(17184,11801,5262,3922)</td>
</tr>
<tr>
<td>organelle</td>
<td>9.28E-21</td>
<td>3.98E-18</td>
<td>(17184,7913,5262,2703)</td>
</tr>
<tr>
<td>intracellular organelle</td>
<td>1.60E-20</td>
<td>5.15E-18</td>
<td>(17184,7896,5262,2696)</td>
</tr>
<tr>
<td>cytoplasmic part</td>
<td>4.34E-18</td>
<td>1.11E-15</td>
<td>(17184,6578,5262,2268)</td>
</tr>
<tr>
<td>cytosol</td>
<td>5.23E-18</td>
<td>1.12E-15</td>
<td>(17184,2282,5262,878)</td>
</tr>
<tr>
<td>membrane-bounded organelle</td>
<td>2.53E-16</td>
<td>4.65E-14</td>
<td>(17184,6973,5262,2377)</td>
</tr>
<tr>
<td>intracellular membrane-bounded organelle</td>
<td>2.86E-16</td>
<td>4.60E-14</td>
<td>(17184,6968,5262,2375)</td>
</tr>
<tr>
<td>Biological Process (Before Modification vs. All Genes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td><strong>P-value</strong></td>
<td><strong>FDR q-value</strong></td>
<td><strong>Enrichment (N,B,n,b)</strong></td>
</tr>
<tr>
<td>cellular process</td>
<td>2.58E-18</td>
<td>2.87E-14</td>
<td>1.01 (17184,11815,16844,11658)</td>
</tr>
<tr>
<td>regulation of cellular process</td>
<td>8.70E-18</td>
<td>4.83E-14</td>
<td>1.01 (17184,7402,16844,7330)</td>
</tr>
<tr>
<td>regulation of biological process</td>
<td>1.52E-16</td>
<td>5.61E-13</td>
<td>1.01 (17184,7815,16844,7733)</td>
</tr>
<tr>
<td>primary metabolic process</td>
<td>6.78E-16</td>
<td>1.88E-12</td>
<td>1.01 (17184,7213,16844,7140)</td>
</tr>
<tr>
<td>biological regulation</td>
<td>9.90E-16</td>
<td>2.20E-12</td>
<td>1.01 (17184,8262,16844,8170)</td>
</tr>
<tr>
<td>regulation of metabolic process</td>
<td>9.15E-13</td>
<td>1.69E-09</td>
<td>1.01 (17184,4595,16844,4557)</td>
</tr>
<tr>
<td>regulation of primary metabolic process</td>
<td>1.29E-12</td>
<td>2.05E-09</td>
<td>1.01 (17184,4093,16844,4062)</td>
</tr>
<tr>
<td>regulation of macromolecule metab. proc.</td>
<td>1.32E-12</td>
<td>1.83E-09</td>
<td>1.01 (17184,3950,16844,3921)</td>
</tr>
<tr>
<td>macromolecule metabolic process</td>
<td>6.77E-12</td>
<td>8.36E-09</td>
<td>1.01 (17184,5804,16844,5745)</td>
</tr>
<tr>
<td>regulation of nitrogen comp. metabolic proc.</td>
<td>8.07E-12</td>
<td>8.97E-09</td>
<td>1.01 (17184,3253,16844,3232)</td>
</tr>
<tr>
<td>biological_process</td>
<td>8.39E-12</td>
<td>8.48E-09</td>
<td>1.00 (17184,14122,16844,13894)</td>
</tr>
<tr>
<td>regulation of nucleobase-containing compound metabolic process</td>
<td>8.72E-12</td>
<td>8.07E-09</td>
<td>1.01 (17184,3172,16844,3152)</td>
</tr>
<tr>
<td>regulation of cellular metabolic process</td>
<td>9.96E-12</td>
<td>8.51E-09</td>
<td>1.01 (17184,4174,16844,4140)</td>
</tr>
<tr>
<td>regulation of gene expression</td>
<td>4.95E-11</td>
<td>3.93E-08</td>
<td>1.01 (17184,3144,16844,3123)</td>
</tr>
<tr>
<td>positive regulation of biological process</td>
<td>3.39E-10</td>
<td>2.51E-07</td>
<td>1.01 (17184,3318,16844,3293)</td>
</tr>
<tr>
<td>regulation of macromolecule biosynth. Proc.</td>
<td>4.47E-10</td>
<td>3.10E-07</td>
<td>1.01 (17184,3008,16844,2987)</td>
</tr>
<tr>
<td>regulation of cellular biosynthetic process</td>
<td>4.95E-10</td>
<td>3.23E-07</td>
<td>1.01 (17184,3149,16844,3126)</td>
</tr>
<tr>
<td>regulation of biosynthetic process</td>
<td>9.74E-10</td>
<td>6.01E-07</td>
<td>1.01 (17184,3178,16844,3154)</td>
</tr>
<tr>
<td>regulation of cellular macromolecule biosynthetic process</td>
<td>1.21E-09</td>
<td>7.07E-07</td>
<td>1.01 (17184,2945,16844,2924)</td>
</tr>
<tr>
<td>regulation of RNA metabolic process</td>
<td>1.32E-09</td>
<td>7.32E-07</td>
<td>1.01 (17184,2790,16844,2771)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biological Process (After Modification vs. All Genes)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>biological_process</td>
</tr>
<tr>
<td>primary metabolic process</td>
</tr>
<tr>
<td>cellular process</td>
</tr>
<tr>
<td>metabolic process</td>
</tr>
<tr>
<td>macromolecule metabolic process</td>
</tr>
<tr>
<td>Molecular Function (Before Modification vs. All Genes)</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>molecular_function</td>
</tr>
<tr>
<td>binding</td>
</tr>
<tr>
<td>small molecule binding</td>
</tr>
<tr>
<td>nucleoside phosphate binding</td>
</tr>
<tr>
<td>nucleotide binding</td>
</tr>
<tr>
<td>organic cyclic compound binding</td>
</tr>
<tr>
<td>heterocyclic compound binding</td>
</tr>
<tr>
<td>protein binding</td>
</tr>
<tr>
<td>purine nucleotide binding</td>
</tr>
<tr>
<td>purine ribonucleotide binding</td>
</tr>
<tr>
<td>ribonucleotide binding</td>
</tr>
<tr>
<td>purine ribonucleoside triphosphate binding</td>
</tr>
<tr>
<td>catalytic activity</td>
</tr>
<tr>
<td>metal ion binding</td>
</tr>
<tr>
<td>cation binding</td>
</tr>
<tr>
<td>ion binding</td>
</tr>
<tr>
<td>adenyl nucleotide binding</td>
</tr>
<tr>
<td>adenyl ribonucleotide binding</td>
</tr>
<tr>
<td>ATP binding</td>
</tr>
<tr>
<td>receptor activity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecular Function (After Modification vs. All Genes)</th>
<th>Description</th>
<th>P-value</th>
<th>FDR q-value</th>
<th>Enrichment (N,B,n,b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>biological_process</td>
<td>3.56E-25</td>
<td>3.96E-21</td>
<td>1.05 (17184,14122,5262,4558)</td>
<td></td>
</tr>
<tr>
<td>primary metabolic process</td>
<td>3.09E-22</td>
<td>1.72E-18</td>
<td>1.13 (17184,7213,5262,2497)</td>
<td></td>
</tr>
<tr>
<td>cellular process</td>
<td>3.68E-22</td>
<td>1.36E-18</td>
<td>1.07 (17184,11815,5262,3885)</td>
<td></td>
</tr>
<tr>
<td>metabolic process</td>
<td>2.74E-21</td>
<td>7.6E-18</td>
<td>1.12 (17184,7829,5262,2681)</td>
<td></td>
</tr>
<tr>
<td>macromolecule metabolic process</td>
<td>3.83E-18</td>
<td>8.5E-15</td>
<td>1.14 (17184,5804,5262,2025)</td>
<td></td>
</tr>
<tr>
<td>biological regulation</td>
<td>1.29E-17</td>
<td>2.38E-14</td>
<td>1.10 (17184,8262,5262,2786)</td>
<td></td>
</tr>
<tr>
<td>Process</td>
<td>FDR</td>
<td>p-value</td>
<td>q-value</td>
<td>Gene Symbol</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>cellular metabolic process</td>
<td>1.75E-15</td>
<td>2.78E-12</td>
<td>1.11</td>
<td>17184,7102,5262,2410</td>
</tr>
<tr>
<td>cellular macromolecule metabolic process</td>
<td>2.22E-15</td>
<td>3.08E-12</td>
<td>1.14</td>
<td>17184,5317,5262,1849</td>
</tr>
<tr>
<td>macromolecule modification</td>
<td>1E-14</td>
<td>1.24E-11</td>
<td>1.25</td>
<td>17184,1995,5262,762</td>
</tr>
<tr>
<td>regulation of biological process</td>
<td>9.71E-14</td>
<td>1.08E-10</td>
<td>1.09</td>
<td>17184,7815,5262,2615</td>
</tr>
<tr>
<td>phosphorylation</td>
<td>1.77E-13</td>
<td>1.78E-10</td>
<td>1.44</td>
<td>17184,634,5262,280</td>
</tr>
<tr>
<td>phosphate-containing comp. metab. proc.</td>
<td>2.19E-13</td>
<td>2.03E-10</td>
<td>1.39</td>
<td>17184,805,5262,342</td>
</tr>
<tr>
<td>phosphorus metabolic process</td>
<td>2.19E-13</td>
<td>1.87E-10</td>
<td>1.39</td>
<td>17184,805,5262,342</td>
</tr>
<tr>
<td>protein phosphorylation</td>
<td>3.3E-13</td>
<td>2.62E-10</td>
<td>1.46</td>
<td>17184,580,5262,259</td>
</tr>
<tr>
<td>cellular protein modification process</td>
<td>3.43E-13</td>
<td>2.54E-10</td>
<td>1.24</td>
<td>17184,1901,5262,721</td>
</tr>
<tr>
<td>protein modification process</td>
<td>3.43E-13</td>
<td>2.38E-10</td>
<td>1.24</td>
<td>17184,1901,5262,721</td>
</tr>
<tr>
<td>regulation of cellular process</td>
<td>6.41E-12</td>
<td>4.19E-09</td>
<td>1.09</td>
<td>17184,7402,5262,2470</td>
</tr>
<tr>
<td>protein metabolic process</td>
<td>9.83E-12</td>
<td>6.07E-09</td>
<td>1.17</td>
<td>17184,2958,5262,1061</td>
</tr>
<tr>
<td>transmembrane receptor protein tyrosine kinase signaling pathway</td>
<td>9.11E-11</td>
<td>5.33E-08</td>
<td>1.44</td>
<td>17184,491,5262,217</td>
</tr>
<tr>
<td>response to stimulus</td>
<td>1.53E-10</td>
<td>8.47E-08</td>
<td>1.10</td>
<td>17184,5621,5262,1901</td>
</tr>
</tbody>
</table>