Introduction

Background: One key criterion used in the clinical interpretation of sequence variants is the allele frequency observed in the general population. In the recently published American College of Medical Genetics and Genomics (ACMG) guidelines (2015), an allele “frequency greater than expected for a disorder” is considered to be strong evidence for a benign classification. In principle, allele-frequency thresholds can be derived for each gene based on the disease incidence and penetrance of pathogenic mutations in that gene. In practice, however, this is difficult, since accurate estimates of incidence and penetrance are not available for most genes. Moreover, this analysis ignores the frequency distribution of reported pathogenic mutations: some genes have only a few relatively common mutations, while others have hundreds of very rare or private mutations. Finally, accurate and precise frequency measurements based on large populations have only recently become available. For all these reasons, diagnostic laboratories have typically set their thresholds for benign allele frequency conservatively high. To better understand the frequency distribution of disease-causing alleles, we aggregated pathogenic variants from the ClinVar database and determined the frequency of these variants in the recently released Exome Aggregation Consortium (ExAC) dataset.

Data source: We identified putative pathogenic variants by selecting any variant classified as Pathogenic or Likely Pathogenic by one or more submissions from a CLIA-certified laboratory in the April 2015 dataset. For CFTR, this data was supplemented with data from HGMD. ExAC data is based on version 0.3 (Jan 13, 2015) of the database.

Pathogenic-variant spectrum-based approach to variant interpretation

Example: CDH1 c.1298A>G (p.Asparagine33Gly)
- All population MAF: 0.007% (9 hets), East Asian MAF: 0.09% (8 hets).
- Not previously reported in literature.

Traditional approach:
- Per GLOBOCAN (http://globocan.iarc.fr/), the population prevalence of hereditary diffuse gastric cancer is <0.1 per 100,000 (0.0001%).
- Per literature, penetrance of CDH1 is considered high.
- 0.007% > 0.0001%; therefore, VUS or Likely Benign depending on the confidence level of prevalence and penetrance and on lab policy on MAF.

Pathogenic-variant spectrum-based approach:
- In ExAC, the cumulative sum of all known Pathogenic and Likely Pathogenic CDH1 mutations is 3.
- This variant has observed 9 times in ExAC.
- This previously never-reported variant is not likely to account for 3 times as many HDGC cases as all previously characterized mutations combined; therefore, Likely Benign.

Distribution of pathogenic variants

1. The vast majority of pathogenic mutations are observed never or only once in ExAC.
2. Most commonly observed pathogenic mutations are known founder mutations, including 7 of the 9 variants observed 3 or more times.
3. Pathogenic missense and intronic mutations are observed at frequencies similar to those of pathogenic truncating mutations.
4. Pathogenic mutations are observed at frequencies substantially lower than the industry standard for a Benign or Likely Benign classification.

1. Despite the lower penetrance of BRCA2, the overall conclusions are very similar.
2. Two common BRCA2 variants have discrepant interpretations.

Identifying common pathogenic mutations by literature search

To avoid the risk of erroneously disregarding a relatively high-frequency pathogenic mutation, we are developing an approach based on publication count to identify potentially known, pathogenic mutations. We expect common pathogenic mutations to have been observed and characterized by many researchers and clinicians.

Publication counts of common pathogenic BRCA1 variants

- c.68_69delAG (AJ Founder) Publication count: 52
- c.1687C>T (Swedish Founder) Publication count: 93
- c.23974870 (putative AJ founder) Publication count: 54
- c.3846G>A (putative AJ founder) Publication count: 51
- c.853A>T (Lithuanian Founder) Publication count: 16
- c.7480C>T (Korean Founder) Publication count: 16
- c.7865A>G (ClinVar: LP, 2x VUS, LB, Benign) Publication count: 8
- c.2132A>G (Polish Founder) Publication count: 9

Publication counts of common similar MAF

<table>
<thead>
<tr>
<th>Variant</th>
<th>DNA Change</th>
<th>ExAC allele count in EA</th>
<th>User-generated pathogenic score</th>
<th>User-generated VUS score</th>
<th>ClinVar Interpretation</th>
<th>ClinVar List</th>
<th>PMID Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.37392A&gt;G</td>
<td>p.Val1247Ile</td>
<td>0.00000013</td>
<td>Benign</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.6120C&gt;G</td>
<td>p.Lys204Met</td>
<td>0.00000013</td>
<td>VUS</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.21674G&gt;A</td>
<td>p.Leu725Ile</td>
<td>0.00000013</td>
<td>Benign &amp; VUS</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.37392C&gt;G</td>
<td>p.Val1247Leu</td>
<td>0.00000013</td>
<td>VUS</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.23974870</td>
<td>(putative AJ founder)</td>
<td>0.00000013</td>
<td>Pathogenic</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.5098A&gt;G</td>
<td>p.Arg170Glu</td>
<td>0.00000013</td>
<td>VUS</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.2132A&gt;G</td>
<td>p.Glu64Ile</td>
<td>0.00000013</td>
<td>VUS</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.37392C&gt;G</td>
<td>p.Arg1726Gly</td>
<td>0.00000013</td>
<td>VUS</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.1687C&gt;T</td>
<td>(Swedish Founder)</td>
<td>0.00000013</td>
<td>Benign &amp; LB</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.37392C&gt;G</td>
<td>p.Arg170Glu</td>
<td>0.00000013</td>
<td>VUS</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The publication count of founder mutations tends to be much higher than those of other variants within the same MAF range.

Conclusions & Discussions

The collective knowledge of the clinical and research community about disease-causing variants now allows for the analyses of the pathogenic mutation load on massive population datasets like ExAC. This combined database provides an invaluable tool to scientists for evaluating the potential clinical impact of sequence variants.

The presence of founder mutations remains a concern in any approach based on population frequency, including the method presented here. However, the risk is mitigated in part due to the well-studied nature of most of the subpopulations in the ExAC database. Furthermore, an approach based on simple automatable publication count may provide additional guidance in identifying potential founder mutations. Nevertheless, careful evaluation of the literature remains an absolutely essential part of any variant interpretation process.

Acknowledgments

The authors of this poster would like to thank all scientists, researchers, and clinicians who contributed data to the ClinVar and ExAC databases. Furthermore, we would like to acknowledge the work of the Clinical Genomics Team and the BioMed Team at Invitae, which contributed tools and data necessary for this work.