What do public databases really tell us about classifications of variants in *BRCA1* and *BRCA2*?

Stephen E. Lincoln[1], Shan Yang[1], Melissa S. Cline[2], Can Zhang[2], Yuya Kobayashi[1], Scott Topper[1], David Haussler[2], Benedict Paten[2], and Robert L. Nussbaum[1,3]

[1]Invitae, San Francisco, CA; [2]University of California, Santa Cruz; [3]University of California, San Francisco

Methods

Classifying which genetic variants observed in a patient are pathogenic (and thus potentially actionable) and which are not can be complex and requires involves expert judgment by laboratory directors who must weigh various lines of evidence in determining classifications. For this reason variant classifications sometimes disagree among laboratories, even when contemporaneous. Our study focused on the commonly tested *BRCA1* and *BRCA2* genes, in which germline pathogenic variants substantially increase the risk of breast, ovarian, and other cancers, and where finding a pathogenic variant in a patient may change clinical management decisions significantly in patients. Our study aimed to measure reproducibility among independent *BRCA1/2* variant classifications.

The NIH's ClinVar database now affords the opportunity to address this in a rigorous manner and to use these results to inform clinical practice standards. Many commercial and academic laboratories voluntarily submit data to ClinVar, which has accumulated thousands of variants in *BRCA1/2*. Unfortunately, Myriad Genetics, the largest *BRCA1/2* testing lab in the world, declines to participate as do two other major reference laboratories. Fortunately, a sizable and representative set of Myriad variant classifications hasve been submitted to ClinVar by ordering clinicians.

Discordant classifications have been suggested as a reason that public databases "should be precluded from clinical use."[1] Although disagreements are real, this proposition is inconsistent with our experience. In a prior study[2] of 975 patients, we observed 99.8% concordance between *BRCA1/2* reports, (a) that were produced using publicly available data, the literature, and the most recent guidelines[3], versus (b) classifications from Myriad, who had used their proprietary databaseand slightly different criteria. Moreover, to the extent that discordances do exist in public databases, they enable

- interlaboratory quality control,
- building consensus among experts for specific variants, and
- Improving procedures for variant classification

Methods

1. *BRCA1/2* data were extracted from the May 2016 ClinVar XML release.
2. Analysis was limited to submissions that met objective criteria, namely
   - they came from established clinical laboratories,
   - who had at least 200 *BRCA1/2* classifications in ClinVar,
   - that were mostly recent (>50% were <5 years old)
3. Data integration was improved by standardizing variant nomenclature. Data were subjected to quality control, and clearly erroneous records removed.
4. Comparisons considered only differences that would significantly change management decisions under current guidelines.
5. We also collected a sequential series of 30,000 patients tested for BRCA1/2 to measure the prevalence of variants by type.

ClinVar submitter	Classified variants	Comparable variants	Most recent classification	Evidence provided	Note
Ambry	2792	1613	**Feb 2015**	**No**	
SCRP/Myriad	2327	1351	Dec 2015	**No**	Likely benign variants generally unreported.
Invitae	1998	1367	Mar 2016	Yes	
GeneDx	1216	957	Oct 2015	Yes	
Counsyl	272	256	**Feb 2015**	**No**	No variants of uncertain significance submitted
CHEO	257	220	N/A	**No**	
Emory	203	183	June 2015	**No**	
Total	**5124**	**2006**			

Legend: Comparable variants are those with two or more submitters. There are known biases in the data (noted here): Some submitters had not provided updates for more than one year (bold), and some submitters excluded certain classes of variants (see notes). Most submitters unfortunately do not provide classification evidence for specific variants. Copy number variants were also only reported by only a few submitters (not shown).

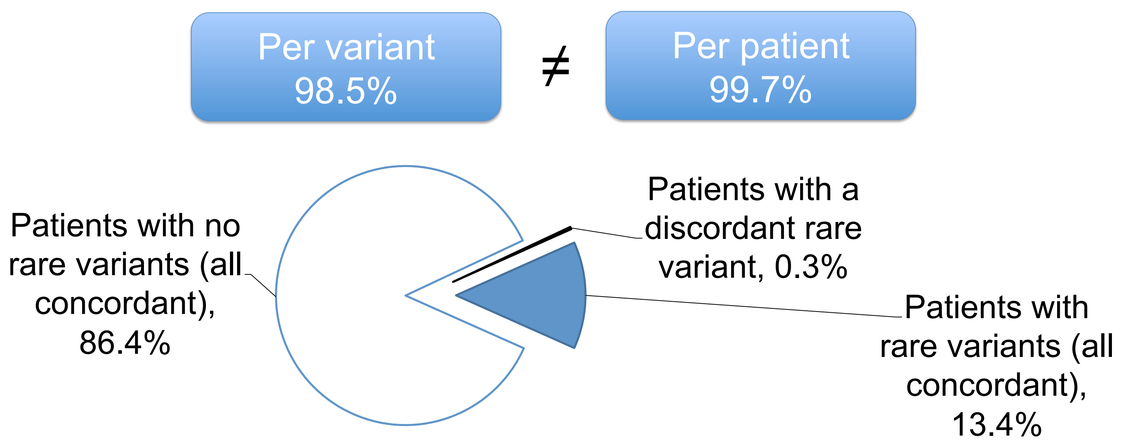
Results

We compared classifications on the basis of whether they would change management—i.e., pathogenic and likely pathogenic were considered discordant from benign, likely benign, or VUS (variant of uncertain significance). On a per-variant basis, we found high concordance: only 30 variants (1.5%) showed discordance between any two submitters. The largest class of discordant variants were rare missense changes (18/30), which composed almost half (45.5%) of our data set, although the vast majority of such variants (98.0%) had concordant classifications among all submitters. Although numerous, rare missense variants are, as a class, infrequently observed in patients (6.4% prevalence). Other discordances were in canonical RNA splice sites (5/30) or an intron (2/30) or were in-frame deletions (2/30). We contacted the directors of submitting laboratories to help resolve discordances.

	All variants	Comparable variants	Clinical prevalence	Concordance	Note
Full data set	100% n = 5124	100% n = 2006	100%	**98.5%** 1976/2006	
Of the full data set					
Common variants	1.2%	2.7%	100%	100% 55/55	Defined as having >1% population AF
Intermediate variants	3.9%	19.1%	18.4%	100% 182/182	Having a >0.05% population AF or >0.1% AF in clinical cases
Rare variants	94.9%	88.2%	12.7%	**98.3%** 1739/1769	Otherwise
Of the rare variants					
Missense	45.5%	45.5%	6.4%	**98.0%** 895/913	Most of the few discordances (18/30) were of this type
Truncating	23.6%	21.9%	2.7%	**99.8%** 438/439	The one discordance was a submission error.
Silent	10.6%	8.6%	1.9%	100.0% 173/173	
Intronic	4.9%	3.6%	1.5%	**97.3%** 71/73	
Last exon	4.1%	3.5%	0.4%	**97.2%** 69/71	Both discordant classifications predated newer key evidence
Canonical splice site	3.7%	3.0%	0.3%	**91.7%** 55/60	
In-frame indel	1.6%	1.5%	0.3%	**93.5%** 29/31	
Copy number variant	0.86%	0.40%	0.2%	100% 8/8	Underrepresented category in ClinVar
Alu insertion	0.04%	0.05%	0.01%	100% 1/1	Underrepresented category in ClinVar

Clinical prevalence is the fraction of patients (out of 30,000) with any such variants. AF is allele frequency.

Given that the variant types with discordance are rare and low prevalence, the per-patient discordance is estimated to be substantially higher than the observed per-variant discordance.



Conclusion

Although our analysis shows that clinically significant disagreements in BRCA1/2 variant classification are infrequent, discordances are of course important to patients and clinicians. We believe it is essential for the genetics community to resolve these differences collaboratively, as is standard practice in other areas of medicine, in order to deliver the best possible patient care. The open sharing of data through ClinVar affords us the opportunity to do this on a global scale.

References: 1. Vail *et al.*, *J Community Genet.* 2015; 2. Lincoln, S, *et al., J Mol Diag.* 2015; 3. Richards *et al. Genet Med.* 2015