The Utility of Genomic Variant Databases in Genetic Counseling

Colleen Ahern  
*Sarah Lawrence College, cahern@gm.slc.edu*

Elly Brokamp  
*Sarah Lawrence College, ebrokamp@gm.slc.edu*

Follow this and additional works at: [http://digitalcommons.slc.edu/genetics_etd](http://digitalcommons.slc.edu/genetics_etd)

Part of the [Other Genetics and Genomics Commons](http://digitalcommons.slc.edu/genetics_etd)

Recommended Citation

The Utility of Genomic Variant Databases in Genetic Counseling

Colleen Ahern and Elly Brokamp

Joan H. Mark Program in Human Genetics
Sarah Lawrence College
May 2016

“Submitted in partial completion of the Master of Science Degree at Sarah Lawrence College, May 2016”
The Utility of Genomic Variant Databases in Genetic Counseling
Colleen Ahern and Elly Brokamp

Abstract

Organizations such as the American College of Medical Genetics (ACMG) and the National Society of Genetic Counselors (NSGC) are in agreement that public genomic data sharing will benefit patient care. Despite these recommendations, not all clinical laboratories share their variant data onto public databases. As the amount of genetic material being analyzed for patient care continues to increase, more variants of unknown significance (VUS) are reported as well. Genetic counselors need to properly interpret VUS results in order to aid patients in making educated health decisions. For this paper, genetic counselors were asked about genomic data sharing and how they handle VUS results for patients. While almost all genetic counselors agree that there is a need for genomic data sharing, only some took laboratories’ data sharing practices into account when deciding where to order testing. Genetic counselors do not have a standard way of processing VUS results; there is little consistency to how often genetic counselors look up variants in public databases or which databases they use.

Key Words: genetic counseling, data sharing, VUS, variant database

Introduction/Background

Many organizations, clinicians, and laboratories believe that sharing genomic data is important to research and patient care (Arias et al, 2015). On April 15th, 2015 The National Society of Genetic Counselors (NSGC) released a statement in favor of the transferring of variant, phenotypic, and interpretative data quickly into public databases. NSGC states that, “Timely data sharing in non-proprietary databases is essential to improve accuracy of variant interpretation” (NSGC Headquarters, 2015). Other organizations also support this belief. The American College of Medical Genetics (ACMG) also encourages clinical laboratories to upload their information to databases in order to help gain more information about variant classifications (Richards et al, 2015). Individual researchers and clinicians have also expressed their support of the data sharing movement. As Dr. Robert Nussbaum, Chief of the Division of Genomics at UCSF Medical Center summarizes, “it is absolutely clear that sharing information provides better medical care” (http://www.freethe-data.org/learn).

Despite public opinion, not all laboratories place their information in public databases; some prefer to use private databases for the purpose of competitive advantage or
convenience (Cook-Deegan et al, 2013). For example, Myriad Genetics ceased data sharing in 2004 and has since maintained their own private databases, the largest repository of variant information for the BRCA1 and BRCA2 genes (Nguyen et al, 2013). Through the company’s 25 years of molecular testing they have tested over 2 million individuals and created an internal variant classification system with 99.98% percent analytic sensitivity (Myriad Genetics, 2016). Myriad claims that the variant data they have compiled over the years puts their testing and variant classification method ahead of the competition (Matloff et al, 2015). Some laboratories and clinicians feel that Myriad’s privatization of their BRCA1 and BRCA2 data hampers other laboratories ability to correctly classify variant information resulting in diminished patient care (Nguyen et al, 2013).

In response to Myriad’s private database, other laboratories offering BRCA testing created the Free the Data movement, which encourages clinicians, scientists, and patients to obtain and share their genetic testing information (Nguyen et al, 2013). These laboratories choose to advertise their data sharing methods in a hope that providers and patients will show a preference toward laboratories that participate in data sharing (Matloff et al, 2015). Whether or not this transparency has brought in more business for these laboratories has not been analyzed.

One of the largest genomic databases currently available is ClinVar, a repository for variant data from clinical laboratories, clinicians, expert groups, patients, researchers, and other databases maintained by the National Institutes of Health. As of May 4th 2015, ClinVar had 172,055 variants submitted from 314 different submitters. Out of all the variants submitted, 11% have been submitted by at least two different sources. Of variants that have been submitted more than once, 17% have conflicting classifications (Rehm et al, 2015).

Besides ClinVar, there are other broad variant databases. Online Mendelian Inheritance in Man (OMIM) contains information on all known Mendelian disorders, focusing on the relationship between genotype and phenotype. It is available to the public for free and is updated daily (omim.org). Human Gene Mutation Database (HGMD) contains published gene mutations that are associated with human inherited disease. It is a publically available, up-to-date, comprehensive source of human gene mutations (Stenson et al, 2014). Leiden Open Variation Database (LOVD) is an open source of DNA variations, even variations outside of genes; LOVD is updated once a month (www.lovd.nl/3.0/home).

Population databases include sequence information from large populations and are not disease or variant specific. They are used to find variant frequencies within a population or more broadly. The 1000 Genomes Project was the first project to sequence a large number of people’s genomes. Its goal is to find genetic variants that have at least a 1% frequency in
the general population, to provide a comprehensive resource on human genetic variation. 1000 Genome’s data is freely available online (www.1000genomes.org). Single Nucleotide Polymorphism Database (dbSNP) is run by the National Center for Biotechnology Information (NCBI). It contains short variations, including insertions/deletions and repeats, in sequences from different types of organisms (http://www.ncbi.nlm.nih.gov/snp). Exome Aggregation Consortium (ExAC) is a database containing sequence data from a variety of projects. Its purpose is to make summary data widely available. It currently holds sequence information on exomes from 60,706 individuals (exac.broadinstitute.org). The NHLBI GO Exome Sequencing Project (ESP) also known as Exome Variant Server (EVS), is made up of many collaborating groups with the goal of discovering genes that contribute to heart, lung, and blood disorders (https://esp.gs.washington.edu/drupal/).

Locus specific databases are a curated listing of variants in a specific gene or causing a specific disease. Breast Cancer Information Core (BIC) is a repository for variants causing breast cancer run by National Human Genome Research Institute (NHGRI). Only members of BIC have access to their database (http://research.nhgri.nih.gov/bic/).

Newer forms of genetic testing such as panel testing and whole exome/genome sequencing are able to examine a large amount of the human genome, bringing along challenges in interpreting results (Lerner-Ellis et al, 2015). Whole exome and whole genome sequencing is the biggest challenge to properly interpreting genomic variants (Cook-Deegan et al, 2013). Rather than the standard negative or positive test results these tests often identify genetic variants that have not been seen before. In many cases it is not possible to interpret the significance of these variants when they are first identified. These novel variants have been termed variants of uncertain significance (VUS) (Aronson et al, 2012).

The correct VUS classification is important in the application of genetic testing to patient clinical care. Guidelines for clinical practice depend greatly on the accurate classification of actionable variants and their potential pathogenicity (Richards et al, 2015). The accuracy of the interpretation and methodology varies and resulting discrepancies between labs calls for improved standardization in variant classification (Craig et al, 2011). In order to help patients make medical decisions, genetic counselors often cannot rely on laboratory results, and are required to access the relevant databases and the assess the information available on certain VUS results (Ormond et al, 2015).

The increase in the number of VUS results with which they must deal intensifies questions about the approaches used by genetic counselors. Exactly how genetic counselors are processing VUS results is still unknown. While counselors presumably know databases are available, whether or not they are accessing these resources has not been assessed.
Therefore, this study was designed to ask practicing genetic counselors how they are processing VUS results and whether or not they are using genomic databases to compare classifications.

Though the NSGC has released a statement encouraging the sharing of genetic variant data by laboratories, genetic counselors’ opinions on the relative importance of this and other laboratory attributes have not been investigated. Additionally, this study aimed to uncover whether or not genetic counselors incorporate a laboratory’s data sharing practices when choosing a testing laboratory.

Methods

Study Design

This study received exemption from Sarah Lawrence College’s Institutional Review Board in November of 2015. An email was sent in January 2016 by NSGC to all of its members who had previously agreed to be contacted by students for research, inviting them to participate. This email included a general description of the study, the informed consent form, and the link to the survey. The survey, which was created using Survey Monkey, was made up of 14 questions including four free response prompts.

Participants

Survey participants included laboratory and clinical genetic counselors who are members of the NSGC. No direct contact occurred between the researchers and the participants. The data from the survey was compiled at the end of February 2016 and a combined 216 responses were collected from both laboratory and clinical genetic counselors. In order to fill out the survey, participants were asked to identify as either working in a clinical or laboratory setting. If the participant worked in a clinical setting they had to have received a VUS result for a patient at some point in their career in order to complete the remainder of the survey.

Measures

Of the 216 total responses, 178 identified as clinical genetic counselors and 38 identified as laboratory genetic counselors. Seven questions were asked only of clinical genetic counselors and did not apply to counselors working for a laboratory. For example, only clinical counselors were asked “Does whether or not a laboratory participates in data
sharing influence your decision to order tests from that company?" There were no questions asked only of laboratory genetic counselors.

**Data Analysis**

Three types of questions were analyzed from the survey: multiple choice, Likert scale, and free response.

Two of the survey questions with the multiple choice options of “yes” or “no” were asked to both clinical and laboratory counselors. The answers to these questions were analyzed by calculating the percentage of each response for both professional settings separately.

Two survey questions had three choice response options including: yes, no, or I do not have a choice (referring to their role in choosing a laboratory for testing). These questions were only asked of clinical genetic counselors.

Two survey questions used a Likert scale. Likert scale questions were used to assess genetic counselor’s behavior toward contacting laboratories and searching databases.

The first free response survey question “Please list the databases you use to compare VUS results?” was asked to both clinical and laboratory genetic counselors. The answers to this question were analyzed for the number of times each specific database was mentioned and for the total number of databases mentioned. The results were sorted by databases mentioned more than or less than five times by clinical and laboratory counselors combined.

The second free response question “Please use this space to include any additional comments you may have about publicly shared databases or VUS classification.” was asked to both clinical and laboratory genetic counselors. The responses to this question were analyzed separately by two researchers to assess for recurrent themes.

Results

Only clinical counselors were asked “are you aware if the laboratories you order genetic tests from participate in data sharing?” 91.01% (162) of all clinical counselors responded. Results listed in Graph 1.
Both clinical and laboratory counselors were asked if they support the sharing of data by clinical laboratories. 88.89% (192) of all participants responded to this question. 97.92% (188) of both clinical and laboratory counselors combined said that they support the sharing of data by clinical laboratories. Results listed in Graph 2.

Both clinical and laboratory counselors were asked if genetic counselors should encourage clinical laboratories to share their variant interpretations. 88.89% (192) of all participants responded to this question. 96.88% (186) of both clinical and laboratory counselors combined said that genetic counselors should encourage clinical laboratories to share their variant interpretation. Results listed in Graph 3.
Graph 3

*Should genetic counselors encourage clinical laboratories to share their variant interpretation?*

Clinical counselors only were asked whether or not a laboratory’s VUS classification method comes into consideration when they are choosing a laboratory to use for testing. Results listed in Graph 4.

Graph 4

*Does the laboratories VUS classification method come into consideration when you are choosing from which laboratory to order testing from?*

Clinical counselors were asked whether a laboratory’s data sharing practices influences their decision to order testing from that company. Of those who have control over choosing a laboratory, 66.14% said that they take the laboratory’s data sharing practices into consideration. Results listed in Graph 5.
Clinical counselors only were asked to rank on a Likert scale from 1 to 5 (1 being “almost never” and 5 being “almost always”) how often they speak to the testing laboratory about a specific VUS results. A total of 161 clinical genetic counselors responded to this question and the mean likert score was 3.16. Results listed in Table 1.

### Table 1

*How often do you speak with the testing laboratory about a specific VUS result?*

<table>
<thead>
<tr>
<th>Likert score</th>
<th>Number of responses</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 “almost never”</td>
<td>12</td>
<td>7.45%</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>21.12%</td>
</tr>
<tr>
<td>3 “sometimes”</td>
<td>53</td>
<td>32.92%</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>24.84%</td>
</tr>
<tr>
<td>5 “almost always”</td>
<td>22</td>
<td>13.66%</td>
</tr>
</tbody>
</table>

Both clinical and laboratory genetic counselors were asked to rank on a Likert scale from 1 to 5 (1 being “almost never” and 5 being “almost always”) how often they search any databases for conflicting or agreeing classification information for a VUS result. A total of
160 clinical genetic counselors responded to this question and the mean response score was 3.68. Twenty-four laboratory genetic counselors responded to this question and the mean response score was 3.83. Full results listed in Table 2.

Table 2
How often do you search any databases for conflicting or agreeing classification information for a VUS result you have received?

<table>
<thead>
<tr>
<th>Likert score</th>
<th>Number of responses</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical counselors:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 “almost never”</td>
<td>23</td>
<td>14.38%</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>7.5%</td>
</tr>
<tr>
<td>3 “sometimes”</td>
<td>18</td>
<td>11.25%</td>
</tr>
<tr>
<td>4</td>
<td>47</td>
<td>29.38%</td>
</tr>
<tr>
<td>5 “almost always”</td>
<td>60</td>
<td>37.50%</td>
</tr>
<tr>
<td><strong>Laboratory counselors:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 “almost never”</td>
<td>5</td>
<td>20.83%</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4.17%</td>
</tr>
<tr>
<td>3 “sometimes”</td>
<td>1</td>
<td>4.17%</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>12.50%</td>
</tr>
<tr>
<td>5 “almost always”</td>
<td>14</td>
<td>58.33%</td>
</tr>
</tbody>
</table>

Both clinical and laboratory counselors were asked in a free response question to list the databases they used to compare VUS results. A total of 138 (71.88%) clinical and laboratory genetic counselors answered this question. Out of the 23 different databases mentioned, nine databases were mentioned more than five times by all respondents combined. ClinVar was mentioned the most amount of times by far with a total of 119, the second most mentioned database was ClinVitae with 25. All databases mentioned five times or more are listed in Table 3.
<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Curator</th>
<th>Clinical</th>
<th>Laboratory</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClinVar</td>
<td>Contains variants submitted by various laboratories/groups (~300). Variants are classified on their pathogenicity.</td>
<td>NCBI</td>
<td>105</td>
<td>14</td>
<td>119</td>
</tr>
<tr>
<td>ClinVitae</td>
<td>Contains variants collected from six different database sources, including Invitae and ClinVar.</td>
<td>Invitae</td>
<td>24</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>HGMD (Human Gene Mutation Database)</td>
<td>Data from variants that are associated with diseases reported in the literature and links to the publication(s)</td>
<td>BIOBASE</td>
<td>14</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>ExAC (The Exome Aggregation Consortium)</td>
<td>Data on ~60,000 people from population studies and disease specific groups</td>
<td>Broad Institute</td>
<td>16</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>LOVD (Leiden Open Variant Database)</td>
<td>A gene centered collection of variants</td>
<td>Leiden University Medical Center</td>
<td>8</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>EVS (Exome Variant Server)</td>
<td>Includes variants on ~6,500 people. Data collected from both unaffected people and people from specific disease groups.</td>
<td>NHLBI</td>
<td>9</td>
<td>2</td>
<td>11</td>
</tr>
</tbody>
</table>
Ethnicity is broken down into European Americans & African Americans.

1000 Genomes
- Data on ~2,500 genomes without any medical/phenotypic information. Can utilize genomes based on ethnicity (American, East Asian, South Asian, African, and European)
- NCBI

dbSNP (Database of Single Nucleotide Polymorphism)
- A collection of simple genetic variations from any type of organism, including humans
- NCBI

BIC (Breast Cancer Information Core)
- Variation database for breast cancer susceptibility genes
- NIH

Both clinical and laboratory genetic counselors were given the option to include any additional comments about publicly shared databases or VUS classification in a free response question. A total of 65 (30.09%) participants chose to give a free response. This response rate makes up a small section of our sample, therefore it may not be representative of the whole sample. The themes found within these responses are further analyzed in our discussion.
Discussion

Most genetic counselors believe that laboratories should share variant data in public databases

When asked directly, all but one clinical genetic counselor (99.38%) and the majority of laboratory genetic counselors (90.00%) said that they support the sharing of data by clinical laboratories. When asked open ended questions about data sharing, 69.23% genetic counselors who responded made positive comments about the use of public databases by clinical laboratories in order to aide variant classification.

Eighteen counselors (27.69%) stated that all laboratories should contribute their data to public databases. These counselors felt that databases are the solution to learning more about variants and streamlining the classification process. As one respondent said:

“I think they should be utilized and I think all labs should contribute to them. Data should not be considered private or corporate data- this is information that can affect lives and pooling data may be a quick way to learn more about these variants.”

While the majority of counselors had strong beliefs that the sharing of variant data will improve variant classification methods and patient care, there was one counselor that strongly opposed data sharing:

“Data sharing will not make results more accurate or improve the quality of other lab’s test results. If there is no reason to improve, competition, then innovation will be hampered. Sharing data is just unrealistic.”

The overall consensus of these responses reflect the NSGC statement supporting public data sharing. Genetic counselors surveyed were in agreement that commercial laboratories should contribute their variant data and classification methods to public databases in order to improve understanding and patient care.

Data sharing practices have some impact on genetic counselors’ choice of testing laboratory

Genetic counselors overwhelmingly support the idea of laboratories sharing variant data, but they are not uniformly supporting laboratories who share their data. Most (88.19%) genetic counselors are in control of choosing which laboratory their patients receive testing
from. Genetic counselors have the ability to support laboratories who publically share their variant data by choosing to order their patients’ testing from those laboratories. The majority of clinical genetic counselors responding (79.01%) reported being aware of whether or not the laboratories they order testing from participate in data sharing. Even if data sharing is something genetic counselors are aware of, some (11.72%) are restricted in their ability to personally select a laboratory for their patients’ testing. Of those who have the ability to choose, most but not all (66.14%) actually take data sharing practices into consideration when deciding a laboratory from which to order testing. While the majority of clinical counselors do consider data sharing, it is “one of many factors” that counselors are also taking into account when selecting a laboratory. Turnaround time, cost, continuity in ordering from a laboratory previously used by a family member, and insurance requirements are other factors that counselors take into account when choosing where to order testing from.

Even if a genetic counselor firmly believes that laboratories should be sharing their variant data, the counselor will only choose a laboratory that shares data over a laboratory that does not if the tests are otherwise equal. One respondent added:

“I completely believe in data sharing, but I will admit that I will order from labs that don’t if they have the best test for my specific patient ... but if I have a choice between two equivalent tests for a patient, I will chose a lab that data shares.”

Some counselors felt more strongly that sharing variant data shows transparency in how variant classification is performed and demonstrates overall more trustworthy, accurate variant classification:

“The withholding of variant data so it can have a (perceived or real) competitive advantage over other labs is unethical. Until internal laboratory variant classification algorithms/software is externally validated, I have no proof that claims of superior variant classification are true. All other things being equal, I send samples to labs that share data, and will continue to do so.”

Through choosing to order their patients’ testing from laboratories who participate in data sharing, genetic counselors are assisting in increasing the amount of publically available data.
Genetic counselors’ current practice after receiving a VUS result is not uniform

Genetic counselors react in different ways after receiving a VUS result for a patient. Clinical genetic counselors are spread across the spectrum of “almost always” (37.50%) to “almost never” (14.38%) searching databases for conflicting or agreeing classification information for a VUS result. This same variation in practice is seen in how often clinical genetic counselors follow up with a laboratory directly after receiving a VUS result, where 32.92% indicate that they “sometimes” follow up. Laboratory and clinical genetic counselors collectively reported using 32 different specific databases. Only nine of these databases were named by more than five of the responding genetic counselors (Table 3). The 23 other databases were only mentioned one to two times; these responses consisted largely of disease or gene specific databases and may be specific to the counselor’s area of practice. For example, POLG Mutation Database and SCN1a Variant Database were each mentioned once. These results highlight the inconsistencies in practice among genetic counselors when are providing care to patients who receive a VUS result.

While there is variability in how often counselors are following up with laboratories and databases after receiving a VUS result, the majority of counselors are using both of these resources. 71% of clinical genetic counselors “sometimes” to “always” follow up with a laboratory about the classification of a VUS result they received from the laboratory. 78% of clinical and 75% of laboratory genetic counselors “sometimes” to “always” search databases for conflicting or agreeing classification information on a VUS result. More clinical genetic counselors (125) report searching public databases than the number (115) that report calling the laboratory in regards to a VUS result. The frequent use of public databases by clinical genetic counselors highlights the importance of these databases for patient care.

There were some explanations given for why genetic counselors who support data sharing chose not to use public databases for patient care. When asked to comment, a small number of clinical counselors stated that they trust the testing laboratories classification implicitly and do not refer to any databases:

“As a clinical genetic counselor, I largely rely on the testing laboratory’s established methods for classifying variants and often don’t look too much into it or question their reasoning when I receive results.”

Additionally, 7 out of 51 clinical counselors (13.73%) stated that they do not know how to use databases. Many of these counselors also expressed a wish for some kind of education or formal training on how to access and use databases:
“I found that as a student, my training did not prepare me to do my own variant interpretation using databases. On the job, I have also not had adequate training about how to use these databases. I think this is a barrier to using these databases and attempting to interpret VUS classifications.”

Responses like these highlight the need for educational opportunities for genetic counselors reviewing the use of variant databases to allow them to incorporate them into their practice. The inconsistencies in databases and resources used among counselors further highlight the unfamiliarity of practicing clinicians with this topic. Continuing education for clinical genetic counselors on how to incorporate public variant databases into their practice may help create a uniform standard of care for patients with a VUS result.

**Study Limitations**

This study was limited by the small overall sample size. Between January 4 and February 29, 2016 a total of 216 genetic counselors responded to the survey. Some of these counselors answered only select questions resulting in a 91.52% completion rate by clinical counselors and at 87.37% completion rate by laboratory counselors. Since each question of the survey was designed to stand alone, no surveys were discarded due to incomplete responses. This study was also limited by the small uptake of the survey by laboratory counselors. Thirty-eight out of 216 respondents identified as laboratory genetic counselors. According the 2014 professional status survey by NSGC 76% of genetic counselors work in a clinical position and 14% work as non-clinical genetic counselors (NSGC Headquarters, 2015). While, skewed due to increased response from clinical genetic counselors the responses received accurately reflect the distribution of genetic counselors in the field.

**Conclusion**

Our survey shows both clinical and laboratory genetic counselors stand behind the NSGC statement supporting the sharing of variant data, especially by clinical testing laboratories. The majority of clinical counselors are putting this support into action by allowing a laboratory’s data sharing practices to influence their decision on which laboratory to order testing from. At this time, data sharing practices by laboratories is only one of many considerations that impact genetic counselor's decisions regarding choice of testing laboratory.

Genetic counselors have been called upon to incorporate the use of public databases into clinical care in order to help patients make medical decisions based on an uncertain result
This study shows that most genetic counselors are using variant databases after receiving a VUS result for a patient. But the use of these databases by genetic counselors is far from uniform. Clinical genetic counselors are using a wide variety of genomic databases inconsistently. This inconsistency highlights that counselors need more education on how to use public genomic databases for the benefit of patient care. Adopting a standardized process regarding the handling of a VUS by genetic counselors would ultimately benefit patient care.
References


