Development of a novel score based system for germline variant interpretation of clinical next-generation sequencing data

MP Powers, M Anderson, J Garcia, K Nykamp, Y Kobayashi, F Monzon, S Topper
Invitae Corporation, San Francisco, CA
www.invitae.com

Abstract

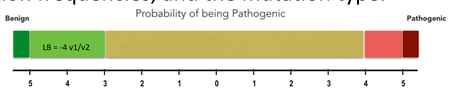
Introduction: With the advent of next-generation sequencing (NGS), the volume of sequencing data generated for each individual patient has increased exponentially. Additionally, sequence data can be generated on patients without strong histories of the clinical conditions associated with the genes tested. Because of this, a system is needed to classify both novel and previously reported genetic variants with rapidity and ease. **Methods**: Based on the ACMG guidelines, we developed a score based variant interpretation system to assign points in favor of a "pathogenic" or "benign" classification to each variant based on available evidence (segregation/prior association with disease, population frequency, functional characterization, computational, etc). Evidence is weighted by the absolute value of points, and different variant types have specific rules and cut-offs. A second version of the scoring system was implemented recently based on the updated version of the proposed ACMG/CAP/AMP guidelines, and a third version incorporated additional evidence that was absent from the proposed guidelines. The system has now been implemented within our clinical reporting system where variant specific information is archived can be easily scored, stored, and updated. **Results**: The implementation of these guidelines allowed for the rapid interpretation of all variants in a 29 gene cancer panel that often had more than one low frequency or novel genetic variants. The amount of time scientists spent interpreting variants decreased with each iteration of the scoring systems, and also led to a decrease in the rate of variants of uncertain significance (VUS). **Conclusions**: The adoption of a score- based system for variant interpretation based on ACMG/CAP/AMP guidelines allowed for a systematic way to classify all variants in our clinical laboratory. This system increases the efficiency and consistency of variant interpretation and allows for a standardized way to capture information about variants for later updates and curation. Additionally, an update after 6 months of use reduced the rate of uncertain variants and we have now integrated this system within a clinical reporting software application.

Background

- Genetic testing is more widely available with the decreased costs of NGS.
  - There are more genes sequenced per patient and there are more patients genotyped overall.
    - Including patients who don't fit the classic presentation of the disease.
  - More variants are identified per patient
    - There is a pressing need for accurate and rapid classification and to avoid the prior pitfalls of false positives and negatives (Bell et al, 2011; Norton et al, 2012; Xue et al, 2012).
- Typical variant classification systems rely on a 5 tier discrete system of classification, with every variant being classified as either Pathogenic, Likely Pathogenic, Uncertain, Likely Benign, Benign (Richards et al, 2007; Plon et al, 2008).
  - There have been several systematic approaches proposed to help guide the assessment of the available evidence, including including a recent ACMG/ISV/AMP update that is expected to be published soon (Duzkale et al, 2013; Eggington et al, 2014; Richards et al, unpublished).
- Here we propose point system toward pathogenicity or neutrality as a way to efficiently classify and store variant information.

Methods

- We have developed three versions of our scoring system where different pieces of evidence are given differing weight towards informing the classification of a variant.
  - Our system quantifies the relative importance of experimental evidence, segregation data, computational predictions, phenotype concordance, population frequencies, and the mutation type.



  - Version 1: From Oct 14 2013 to May 14 2014
  - Version 2: May 15 2014 to August 15 2014
  - Version 3: August 15 to Nov 3 2014
- 29 gene cancer panel: APC, ATM, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDK4, CDKN2A, CHEK2, EPCAM, MEN1, MET, MLH1, MSH2, MSH6, MUTYH, NBN, PALB2, PALLD, PMS2, PTCH1, PTEN, RAD51C, RET, SMAD4, STK11, TP53, VHL
- For the purposes of this experiment, we calculated the number of VUSes per gene ordered, the percentage of negative reports, and the overall average time spent in the variant interpretation phase of report writing.

Version 1: example criteria	Points (+ = pathogenic; - = benign)
Truncating variant	+5
Consensus splice junction	+4
Moderate segregation	+2.5
Very High/High MAF (defined by inheritance)	-5/-4
De novo	+2.5
Matching phenotype	+2
Version 2: example additional criteria	**Points (+ = pathogenic; - = benign)**
De novo and matching phenotype	+4
Addition of splicing predictors	+-2/+-1/+-0.5
Increased use of co-occurrence (trans/cis)	-4 / -2.5 / -1 (depending on disease and inheritance)
Somewhat high MAF	-1
Version 3: example additional criteria	**Points (+ = pathogenic; - = benign)**
Independent case reports	+1/+2/+3 (depending on disease and number of independent case reports)

Results

	Version 1	Version 2	Version 3
VUS rate per gene	3.07%	3.04%	2.81%
Negative report rate	37.68%	37.25%	41.54%
Avg. time (days) spent in variant interp.	1.00	1.03	0.86

Conclusions

- We decreased our VUS rate
- We increased the number of negative reports.
- Scientists spent less time, on average, during the variant interpretation phase of report writing.
- It should be noted that this analysis was based only on a well-characterized 29 gene cancer panel and it's applicability to other panels, such as inherited arrhythmias or Noonan spectrum disorders, has not been established yet.
- We expect that increasing the amount and quality of information in publically available databases such as ClinVar and ExAC will help improve classifications further.

Future directions

- We will perform a comparison of multiple standardized variants between our variant classification system, the ACMG guidelines when published, and other classification schemes for both accuracy and ease of implementation.
- Our guidelines did not take into account data from ExAC (http://exac.broadinstitute.org/), which will be included in subsequent versions of our system
- We will create gene specific models incorporating prevalence, penetrance, mutation spectrum, molecular mechanism, benign variation spectrum, etc., as well as statistical models to determine how "likely" each classifier increases or decreases pathogenicity (Macarthur et al, 2014)

References

1. Bell CJ et al. Sci Transl Med. 2011 Jan 12;3(65):65ra4.
2. Duzkale H et al. Clin Genet. 2013 Nov;84(5):453-63
3. Eggington JM et al. Clin Genet. 2014 Sep;86(3):229-37
4. EXAC website: http://exac.broadinstitute.org/
5. MacArthur DG et al. Nature. 2014 Apr 24;508(7497):469-76
6. Norton N et al. Circ Cardiovasc Genet. 2012 Apr 1;5(2):167-74
7. Plon SE et al. Hum Mutat. 2008 Nov;29(11):1282-91
8. Richards CS et al. Genet Med. 2008 Apr;10(4):294-300
9. Richards CS et al. (unpublished)
10. Xue Y et al. Am J Hum Genet. 2012 Dec 7;91(6):1022-32