Title: Broad Implications of Sustained, Proactive Clinical Validity Curation

Authors: Deepali N. Shinde, Bess Wayburn, Haley Keller, Meghan C. Towne, Anant Babtiwale, Devon L. Thrush, Kelly D. Farwell Hagman, Kelly Radtke

Introduction: As the use of next generation sequencing continues to grow, discoveries of new gene-disease associations are increasing. As a result, our understanding of the phenotypic spectrum of diseases is expanding, and emerging evidence supports that some clinical phenotypes previously considered distinct are more likely along the same disease spectrum. However, a slow adoption of clinical validity assessment standards and a lack of guidelines for disease naming contributes to an increased number of potentially duplicated disease associations. This is evident in reference databases like OMIM and HGMD which have on average 3-4 disease associations per gene, and some entries in HGMD have more than 100 distinct disease associations listed. Collaborative efforts such as ClinGen provide helpful insight into this issue; however, given the vast need and limited resources that ClinGen has, the scope is limited. While these reference databases are valuable resources when considering gene content for diagnostic panel design, genetic testing laboratories must still conduct complete standardized curation of gene-disease associations in order to ensure development of tests with clear clinical utility.

Methods: To overcome these limitations, our clinical laboratory has established standardized guidelines for curating and scoring the clinical validity of gene-disease associations which have been independently validated (Smith, ED et al. (2017) Hum Mutat 38(5): 600-608). Gene-disease associations with a score of Moderate, Strong or Definitive are categorized as characterized, and those with a score of Limited or lower are uncharacterized. A proactive curation workflow enables the maintenance of an up-to-date clinical validity database which currently includes approximately 9000 curated gene-disease relationships, out of which ~58% have a score of Moderate or higher clinical validity. Currently, there are 3272 genes with one or more disease associations classified as Moderate or higher, and this number increases as relevant literature is published. This up-to-date database containing newly discovered gene-disease relationships is useful for data reanalysis leading to reclassifications of previously negative or uncertain diagnostic exome sequencing results, and for assessing gene relevance for inclusion in comprehensive and phenotype-focused genetic testing panels.

Outcomes: In August 2020, 1406 genes associated with neurodevelopmental disorders were identified for inclusion on a comprehensive neurodevelopmental panel and further assigned to smaller, more focused panels. Only characterized gene-disease associations were included. This curated gene list was compared to the gene content of panels for similar clinical indication and scope from other clinical testing laboratories. In total, 3114 genes were included in one of the three laboratories' large neurodevelopmental panels, and 1182 (37.96%) were included on all 3. Our panel included 60 unique genes, out of which 50% were characterized in the last 2 years. Some of the genes included on other clinical panels were excluded from ours due to either limited evidence for clinical validity or because they were determined to be outside the phenotypic scope for the panel in development. Gene content for these panels are continuously reviewed by our laboratory personnel and updated quarterly as new clinical information becomes available for upgrading previously uncharacterized gene-disease associations. After the first quarterly review, we added 27 genes to our neurodevelopmental panels due

to new reports in the literature resulting in an internal clinical validity classification of Moderate or higher. As an example, the gene KCNN2 was added in the month of December 2020 following the publication of its association with an autosomal dominant neurodevelopmental movement disorder in November 2020.

Conclusion: Navigating the massive amount of genetic and clinical data to provide accurate molecular diagnoses following clinical genetic testing critically depends on the proper curation of available genetic and experimental data. Sustained and proactive clinical validity curation has wide reaching benefits in defining phenotype which in turn has a major impact on panel design, panel updates, redesign, and reclassification to increase diagnostic rates.