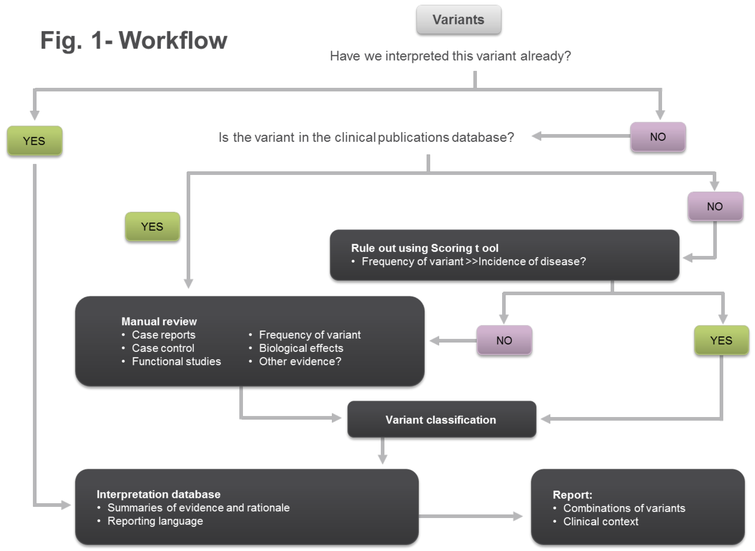
Strength in Numbers: Efficiency and Quality Improvements in Clinical Whole Genome Interpretation.

E.R. Ramos, C.L. Mead, M.L.A. Hogue, J.M. Sardina, A. Khouzam, D.A. Mote, S.J.K. Pond, A.G. Crawford, S.S. Ajay, V. Rajan, J. Silhavy, S. Chowdhury, T.M. Hambuch
Illumina Clinical Services Laboratory, San Diego, CA

As the cost of whole genome sequencing rapidly decreases and technical limitations are surmounted, one of the most significant remaining challenges of clinical human whole genome sequencing (WGS) is the interpretation of rare variants.

- Since October 2012, the Illumina Clinical Services Laboratory has sequenced and interpreted the genomes of approximately 500 primarily healthy adults; with interpretation focused on 1,600 genes associated with 1,221 of the most commonly tested monogenic conditions.
- All single nucleotide variants were assessed to determine clinical significance by a team of trained geneticists and genetic counselors, in accordance with the American College of Medical Genetics & Genomics guidelines.
  - This process included a manual review of the literature for variants detected within the coding regions and flanking intronic regions (+/- 15 base pairs) of the 1,600 interpreted genes as outlined in Fig. 1.

Fig. 1- Workflow

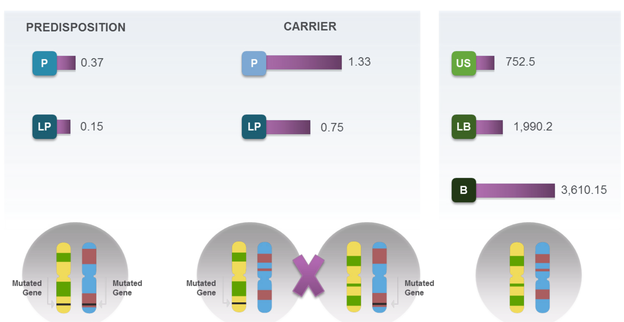


- As the Illumina Variant Database has grown to greater than 60,000 variants, including more than 650 variants with potential clinical significance, the average number of variants per genome that are novel to our laboratory has decreased significantly.
- Currently, of the approximately 5,000 single nucleotide variants detected within the subset of 1,600 interpreted genes, an average of 65 novel variants are found per genome.
  - This is a six-fold decrease from the 360 novel variants identified per genome in May 2013, resulting in a substantial concomitant reduction in manual effort.
- The required manual effort per genome has been further reduced by the development of custom software to facilitate the clinical process of classifying, interpreting and reporting on variant disease relationships.
  - The Variant Interpretation & Reporting Tool (VIT).
- Growth of the variant database and implementation of custom software have contributed to an overall decrease in time required to analyze variants in the 1,600 genes for a single WGS sample from 35 hours in May 2013 to less than ten hours today.

Continued growth in the total number of variants within the database, as well as improvements to the software that aids in the annotation and interpretation process, is anticipated to result in a continued diminution of interpretation effort, which will improve the efficiency and quality of ongoing interpretation, as well as mitigate the cost of offering WGS in the clinic.

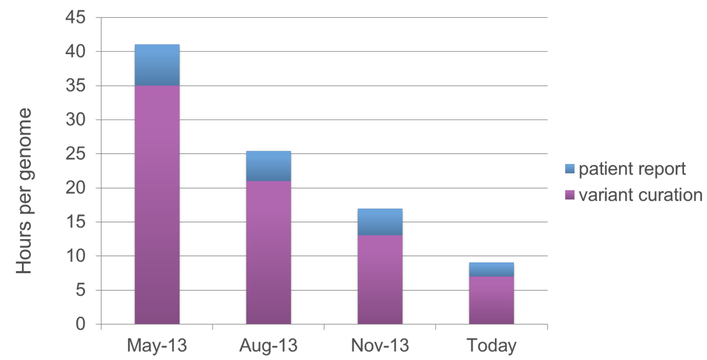
Although challenges involving the clinical interpretation of whole genome sequence data will persist, these advances are anticipated to facilitate the implementation of WGS in the clinical laboratory testing arena.

Fig. 2 Results: Average SNVs Per Individual



P- Pathogenic **LP**- Likely Pathogenic **US**- Unknown Significance **LB**- Likely Benign **B**- Benign

Fig. 3 Efficiency changes due to scale and tool augmentation



illumina®