Contact: jeanie.tirch@personalis.com

User-Friendly Genomic Results: Leveraging a Novel Approach that has the Potential to Decrease Turn-Around Time and Preserve Opportunities for Novel Discoveries

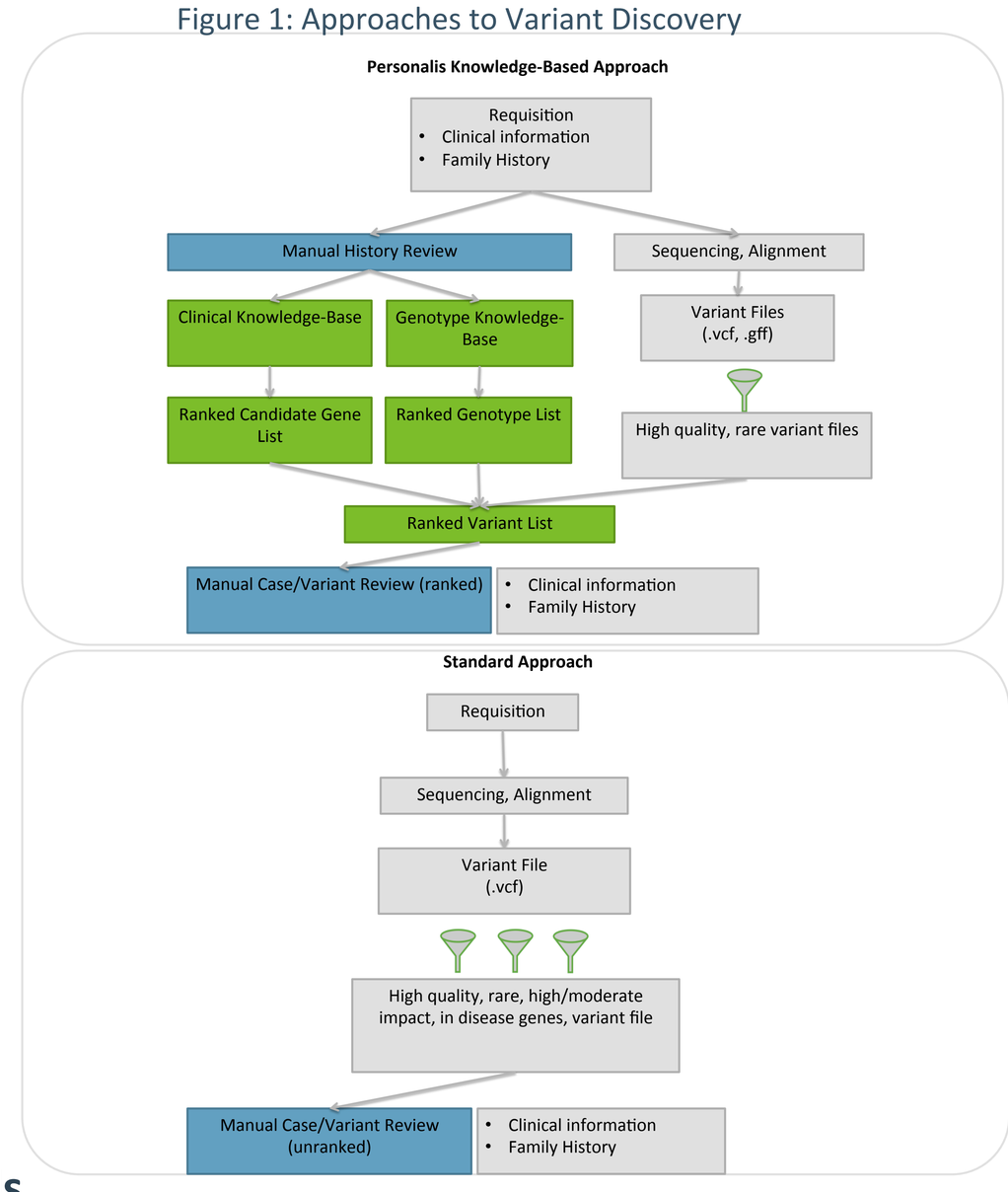
Personalis, Inc., Menlo Park, CA

Jeanie Tirch, MS, CGC; Sarah Garcia, PhD, MS, CGC; Gemma Chandratillake PhD, MS, CGC; Michael Clark, PhD; Anil Patwardhan, PhD; Jason Harris, PhD; Stephen Chervitz, PhD; Richard Chen, MD, MS; John West, MS

Introduction

The identification of causal variants in exome testing continues to rely on appropriate filtering of the tens of thousands of variants identified in an affected individual. Protocols typically apply hard filters to exclude variants least likely to be disease-causing. Sensitivity suffers when criteria are too strict, potentially missing novel candidate genes or cases that expand the phenotypic spectrum of known diseases. Conversely, relaxation of filtering criteria may result in overwhelming numbers of candidates, impeding diagnosis. We designed a knowledge-based ranking system that combines clinical and curated phenotype information as inputs into genotype-based analytics (Figure 1). This system increases the likelihood that causal findings are highly ranked while preserving the ability to detect novel genes/variants.

Figure 1: Approaches to Variant Discovery



Methods

Known samples

28 samples with clinical features and described variants representing a broad range of conditions and variant types (including 21 samples with deletions and duplications of various sizes) were purchased from the Coriell Cell Repositories (http://ccr.coriell.org). Sequencing (exome, or Personalis ACE Exome™) and pipeline analysis was performed in-house.

Personalis Knowledge-based Approach: Variants were ranked by genotype, phenotype, effect impact, and allele frequency.

Standard approach: Defined as: minor allele frequency <1% in all populations and minimum predicted functional effect impact of "moderate" as defined by SNPeff . To simulate filtering based on "clinically interpretable genes" variants were filtered by their presence in genes in the Human Gene Mutation Database (HGMD). Simple genotype predictions were then used to filter results to a final manual review set.

Unknown Samples

30 cases (probands, trios, and families) with a suspected genetic condition of unknown etiology were obtained from multiple collaboration sites. All participants provided written informed consent as part of their enrollment in existing research studies at the participating collaborator sites with Institutional Review Board approval. Personalis, Inc. also employed an external Institutional Review Board to approve our involvement in the data analysis. Personalis ACE Exome™ sequencing was performed in-house, variants were called and annotated, and our knowledge-based ranking system approach was utilized to rank variants.

Results

Known Samples

36/41 described variants in the sample set were identified by the Personalis knowledge-based approach. The mean rank of the causative variant(s) was 2 with the Personalis approach. The number of variants and average relative rank of the causative variant(s) is detailed in Table 1.

Unknown Samples

Of the 30 initial cases with diseases of unknown etiology, our analysis identified variants in genes of interest in **17** cases. Genes with candidate variants included *PNPT1, GCNT2, COL4A1, STRA6, SKI, GDF6, NMNAT1, CNGB3, CRX, CACNA1F, PRPF31,* and *LEPREL1*. The average time spent per case was under an hour as the causative variant was the top ranked variant in **41**% of cases and was within the top **8** variants in all cases except **one**. Previously reported array findings were available for five cases. Personalis ACE Exome confirmed all five and classified one as likely pathogenic. Table 2 summarizes the clinical features and exome findings for the 17 solved unknown cases.

Table 1: Comparison of Approach Performance on Known Samples

	Mean (Range) Number of High Ranked/Filtered Variants	Mean (Mode) Rank of Causative Variant
Personalis Approach	**89 (3-238)**	**2 (1)**
Standard Approach	**174 (32-245)**	**N/A**

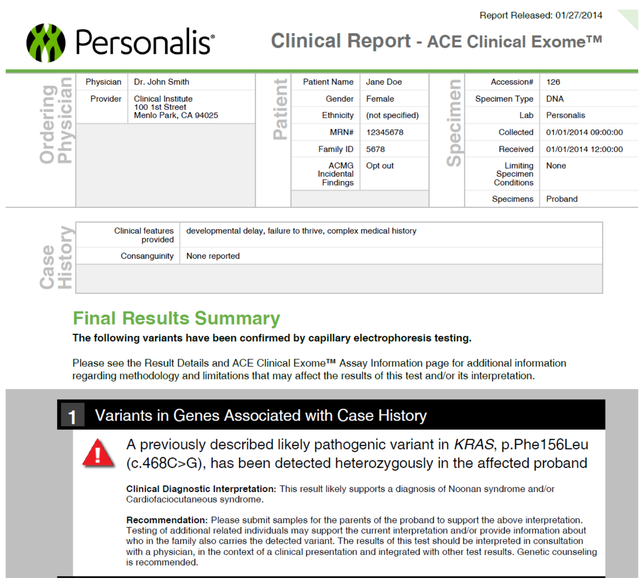


Table 2: Reported Variants for 30 Unknown Cases

Case Summary	Causative Variant(s) Description	Rank
Trio. Proband with IUGR, microcephaly, seizures, hearing loss, hypotonia, developmental delay	Compound heterozygosity for *PNPT1*	3
Trio. Congenital cataracts in two siblings	Homozygosity for *GCNT2*	2
Trio. Suspected Senger syndrome. Bilateral cataracts, cardiomyopathy	Heterozygosity for *COL4A1*	2
Trio. Anophthalmia, hearing loss, dysmorphic features	Compound heterozygosity for *STRA6*	1
Trio. Anophthalmia and facial anomalies	*De novo* heterozygous variant in *SKI*	2
Duo. Microphthalmia, heart defect, developmental delay. Mother- microphthalmia	Heterozygosity for *STRA6*	2
Proband. Anophthalmia	Heterozygosity for *GDF6*	2
Duo. Microphthalmia, coloboma	16p11.2 microdeletion	4
Family. Proband, mother, sibling all with retinitis pigmentosa	Heterozygous variant in a novel gene. Interacting partner of known RP gene.	8
Trio. Leber congenital amaurosis (LCA), panel negative	Compound heterozygosity for *NMNAT1*	1
Trio. Achromatopsia, panel negative	Homozygosity for *CNGB3*	1
Trio. Cone-rod dystrophy (CRD).	*De novo* heterozygous variant in *CRX*	1
Family. Proband, mother, and sister all with retinitis pigmentosa	Heterozygosity for *CRX*	2
Duo. Proband with retinitis pigmentosa, son with CRD	Heterozygosity for *CRX*	1
Proband. Cone-rod dysrophy, high myopia. Similarly affected brother.	Hemizygosity for *CACNA1F*	21
Family. Proband, father, and daughter all with retinitis pigmentosa	Heterozygosity for *PRPF31*	1
Duo. Brothers with "atypical Stickler"	Homozygosity for *LEPREL1*	1

*Family originally reported with an affected sister with RP. Further clinical validation revealed sister was not affected.

Conclusions

The use of our novel, knowledge-based ranking successfully prioritizes the most likely causative variants in genomic data, reducing manual review time. Given that per-variant review estimates currently span 20-60 minutes, this approach has the potential to dramatically improve turnaround time for exome/genome sequencing without sacrificing the potential for making novel discoveries.