Invitae exome interpretation: leading with science and innovation to bring exomes to scale

Exome sequencing is a powerful tool for evaluating patients with suspected genetic disease. But pinpointing rare disease-causing changes within the 30 million base pairs of the human exome requires deep knowledge of the connections between genes and disease, and that knowledge needs to be constantly updated to keep pace with the rapidly developing science of medical genetics. To address these needs, Invitae's clinical exome analysis combines automated gene and variant curation with expert evaluation for fast, reproducible, and accurate exome interpretation.

Exome sequencing is one of the most comprehensive clinical genetic tests available. By targeting the coding sequence of nearly every gene in a patient's genome, exome sequencing produces a rich trove of data that can be used to identify the cause of a previously elusive genetic disease. Exome sequencing is especially useful for patients with complex or vague phenotypes that could be explained by several conditions or by changes in genes not previously suspected of being associated with disease. Studies show that clinical exome sequencing achieves positive diagnostic yields of 20%–40% in patients with neurological impairment, congenital anomalies, or syndromic clinical presentations.[1-4] The results of exome analysis can impact medical management through pharmacotherapy choices, reproductive planning, clinical trial participation, and the discontinuation of additional diagnostic studies (i.e., ending the "diagnostic odyssey").

Although exome analysis has significant benefits, it also carries significant challenges for addressing the needs of many patients in a timely manner. Exome sequence data are vast, and connections between patient phenotypes and genes are constantly evolving. New associations between genes and diseases are published in the scientific literature nearly every day. In addition, an individual patient's phenotype (e.g., signs and symptoms, biochemical data) can change over time. Clinical labs must perform reproducible and unbiased evaluations of genetic variation in patients that weigh all relevant information at the time of testing and, in some cases, revisit those evaluations as knowledge grows (Box 1).

Box 1

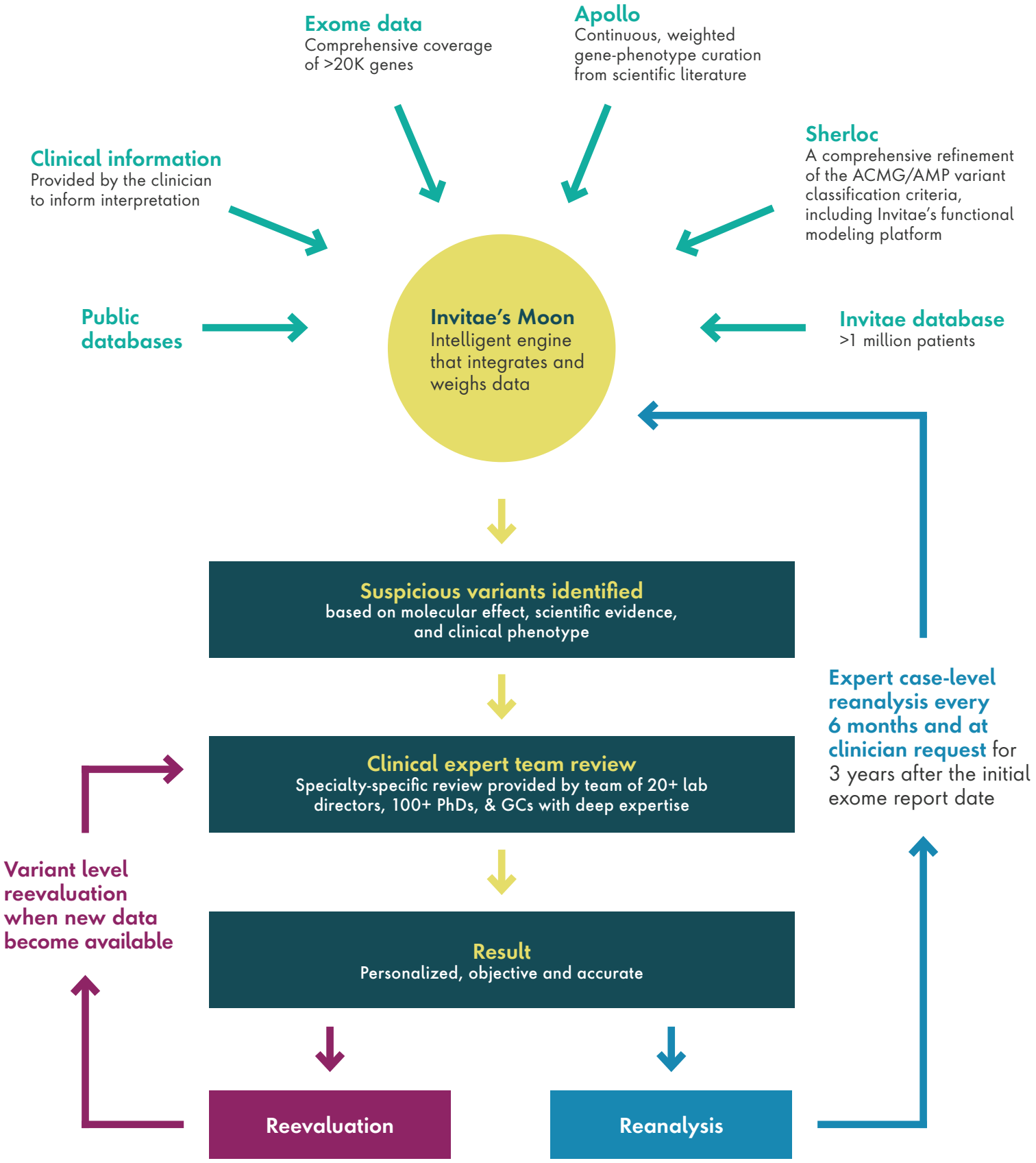
Reanalysis vs reevalution

At times, clinical labs may need to reconsider previous genetic test results. Per the American College of Medical Genetics and Genomics, the two types of review are:

- **Variant-level reevaluation** of previously reported variants
- **Case-level reanalysis** of previously generated sequencing data

Variant-level reevalution is limited to the reclassification of individual variants, most often changing a variant of uncertain significance to a more definitive classification. Conversely, case-level reanalysis is more robust and involves complete review of all variants in an exome or genome, both reported and unreported. Often this identifies new potentially causative variants after additional patient information or gene-disease associations are revealed. Invitae's exome analysis includes biannual case-level reanalysis, facilitated by continual updates to our gene-disease association databases.

Figure 1


Exome data
Comprehensive coverage of >20K genes
Apollo
Continuous, weighted gene-phenotype curation from scientific literature
Clinical information
Provided by the clinician to inform interpretation
Sherloc
A comprehensive refinement of the ACMG/AMP variant classification criteria, including Invitae's functional modeling platform
Public databases
Invitae's Moon
Intelligent engine that integrates and weighs data
Invitae database
>1 million patients
Suspicious variants identified
based on molecular effect, scientific evidence, and clinical phenotype
Clinical expert team review
Specialty-specific review provided by team of 20+ lab directors, 100+ PhDs, & GCs with deep expertise
Result
Personalized, objective and accurate
Reevaluation
Reanalysis
Variant level reevaluation when new data become available
Expert case-level reanalysis every 6 months and at clinician request for 3 years after the initial exome report date

Invitae addresses the challenges of exome analysis with expert-driven methods that can scale to patient demand and provide objective, accurate results that are personalized to each individual tested. Critical to providing fast, reproducible, and unbiased exome interpretation is an automated exome analysis tool called Moon (Figure 1). This artificial intelligence-powered software weighs clinical and genetic information such as age of onset, family information, human phenotype ontology terms from the patient's clinical records, and known gene-phenotype associations to identify the variants most relevant to each patient's case. Invitae's Moon then ranks potential causative variants based on weighted parameters of clinical and genetic information, including each variant's predicted molecular effects. Moon's output is a short list of potential positive variants sifted from the comprehensive exome sequence. These variants are reviewed by scientists, board-certified genetic counselors, and laboratory directors and then clinically evaluated within our proven variant-interpretation framework.[5]

Moon is trusted by genomic medicine pioneers at the University of Liège, Leiden University, Uppsala University, and Rady Children's Hospital, who have used the software to power their clinical exome analyses.[6-8] A study of genetic testing in 96 children at Rady Children's found 99% agreement with expert manual curation.[7] In addition to being precise, Moon enhances the speed of exome interpretation. In fact, Moon's automation helped earn the world record for fastest genetic diagnosis in 2018. In an internal Invitae study of 150 previously solved exome cases, Moon alone correctly identified more than 97% of causative variants in less than two minutes per exome. Moon can even identify disease-causing variants in challenging cases, like those with dual diagnoses,[7] or when additional post-testing data (e.g., de novo status, functional data) confirm pathogenicity. It can also reduce incidental findings by focusing only on the condition of clinical concern.

Moon's speed, accuracy, and focus is powered by a comprehensive and up-to-date gene-disease database called Apollo, which uses natural language processing algorithms to perform daily, automated literature searches for new gene-disease relationships. All publications are reviewed by a team of expert genetic scientists, and relevant phenotypic information is curated and stored in Invitae's Apollo database. We use automation to enhance the expertise of our large team of genetic counselors, PhD scientists, and laboratory directors, and to build on the substantial genetic knowledge housed in our databases. Importantly, phenotypic features are weighted for each gene based on the number of individuals observed in the literature with each feature. This provides the information needed to accurately rank variants for each patient tested. As of July 2020, the Apollo database has high-quality phenotypic information for more than 6,000 gene-disease relationships. This continuously growing knowledge base integrates new scientific insights, guaranteeing a comprehensive and up-to-date analysis for every patient.

Critically, Moon and Apollo's automated capabilities enable Invitae to frequently reanalyze exome data (Figure 1). When new gene-disease relationships are established in the literature or when a patient develops a new clinical sign, reanalysis of exome sequencing data may reveal new genetic insights. Given the growing advances in medical genetics and the potential for patient phenotypes to change over time, the American College of Medical Genetics and Genomics (ACMG) considers ongoing reanalysis of sequence data to be critical.[9] Further, reanalysis of exome sequence results has been shown to increase yields by 1%–4%, with the highest yield demonstrated with Moon-incorporated exome analysis.[10,11] Every six months, Invitae provides complimentary case-level reanalysis during which the full exome is reconsidered in light of new public or patient information (Box 1). Providers can also request reanalysis at any other time so we can integrate new information that may impact diagnoses. In addition to providing full-exome reanalysis, Invitae remains committed to providing variant-level reevaluation when new data become available. There is no burden on the patient or provider to remember to request this type of reanalysis or reevaluation or to decide on the best time for it to happen.

The automated processing provided by Moon and Apollo is built upon a bedrock of internal genetics knowledge that Invitae has developed from sequencing the DNA of more than 1 million individuals. Both tools fit seamlessly with our validated and quantitative variant classification system, Sherloc.[5] Based on ACMG/Association for Molecular Pathology guidelines, Sherloc ensures consistent and reproducible evaluation of variant pathogenicity. Through Sherloc, our exome analysis also draws on the advanced capabilities of our functional modeling platform, which uses machine learning techniques to create gene-specific biochemical, cellular, and computational data models to predict variant pathogenicity (see white paper on integrating functional modeling to enhance clinical variant interpretation).

Invitae is committed to providing an exome sequencing test that is accessible to all patients with rare genetic disorders and their providers. Our reproducible and unbiased artificial intelligence-enhanced approach harnesses automation to ensure that our analyses use the most current, consistent, and comprehensive gene-phenotype information available. In fact, automation is highlighted by the ACMG as a way to bolster the efforts required to provide accurate and consistent variant classification.[9] By integrating Moon and Apollo into the interpretation of our sequence data, we can bring the benefits of comprehensive clinical sequencing to many patients while maintaining exceptional accuracy and reproducibility.

References

1. Yang Y, Muzny DM, Xia F, *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312(18):1870-1879.
2. Lee H, Deignan JL, Dorrani N, *et al.* Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA*. 2014;312(18):1880-1887.
3. Prasad A, Sdano MA, Vanzo RJ, *et al.* Clinical utility of exome sequencing in individuals with large homozygous regions detected by chromosomal microarray analysis. *BMC Med Genet*. 2018;19(1):46.
4. Clark MM, Stark Z, Farnaes L, *et al.* Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med*. 2018;3:16.
5. Nykamp K, Anderson M, Powers M, *et al.* Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet Med*. 2017;19(10):1105-1117.
6. Lumaka A, Race V, Peeters H, *et al.* A comprehensive clinical and genetic study in 127 patients with ID in Kinshasa, DR Congo. *Am J Med Genet A*. 2018;176(9):1897-1909.
7. Clark MM, Hildreth A, Batalov S, *et al.* Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci Transl Med*. 2019;11(489). doi:10.1126/scitranslmed.aat6177
8. Clewemar P, Hailer NP, Hailer Y, *et al.* Expanding the phenotypic spectrum of osteogenesis imperfecta type V including heterotopic ossification of muscle origins and attachments. *Mol Genet Genomic Med*. 2019;7(7):e00723.
9. Deignan JL, Chung WK, Kearney HM, *et al.* Points to consider in the reevaluation and reanalysis of genomic test results: a statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2019;21(6):1267-1270.
10. Liu P, Meng L, Normand EA, *et al.* Reanalysis of Clinical Exome Sequencing Data. *N Engl J Med*. 2019;380(25):2478-2480.
11. James KN, Clark MM, Camp B, *et al.* Partially automated whole-genome sequencing reanalysis of previously undiagnosed pediatric patients can efficiently yield new diagnoses. *npj Genomic Medicine*. 2020;5(1):33.