Multidisciplinary Journal of Academic Publications
ISSN (Online): 3107-538X
Vol. 02, Issue 01, January – February 2026

# NEW PARADIGMS IN THE INTERPRETATION OF GENETIC VARIANTS: APPLICATION OF THE TSV FILTER TOOL AND POLYGENIC RISK SCORES TO REDUCE THE DIAGNOSTIC GAP IN COMPLEX DISEASES

**Nuria Jiménez Nieto**

**Research Scholar, Department of Health Science, Kennedy University**

**Registration No.: KUSLS20220144158**

## Abstract

*Despite advances in next-generation sequencing (NGS), interpreting the effects of genetic variants on complex diseases is still a major challenge in modern precision medicine, where a large population goes undiagnosed. Variant Classification: The Challenge of the VUS While there was an explosion in variants submitted to databases such as ClinVar, with around 50% continuing to represent Variants of Uncertain Significance (VUS) this leads to a continued diagnostic gap. Here, we use real-world data from Innovare Genetics' TSV Filter Tool a variant prioritization and clinical exploration platform designed to be an augmentation of rare/genomic variant databases (with a focus on complex disease) alongside Polygenic Risk Scores (PRS). We hypothesize that the simultaneous application of both the Innovare Genetics analytical pipeline with PRS risk stratification will dramatically augment variant classification accuracy and impact VUS resolution. Using a systematic approach to analyze several population-level cohorts from UK Biobank, PRSKB, and the ClinVar databases we examine diagnostic yields, AUC performance statistics, as well as the rates of reclassification. Results show that PRS integrated pipelines enhance AUC values of the top complex diseases and that upstream pipeline-based variant processing reduces false-positive variant calling. These findings lend support to the clinical integration of PRS and structured variant prioritization as complementary strategies for addressing the diagnostic gap in complex polygenic disorders.*

**Keywords:** *Polygenic Risk Scores[1], TSV Filter Tool[2], Innovare Genetics[3], Variants of Uncertain Significance[4], Diagnostic Gap[5], Complex Diseases[6].*

## 1. Introduction

The arrival Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS) has revolutionized the genetic diagnostic paradigm. Despite advances in technology, a large number of patients with suspected genetic disorders still lack definitive molecular diagnoses (the so-called "diagnostic gap"4–6). In clinical practice between 68% -75% of patients with rare or complex diseases do not achieve a definitive genetic diagnosis after standard testing as the existing interpretation paradigms are insufficient (Wojcik et al., 2024) This is exacerbated by the fact that on average, for every individual run of a WES, between 80 000 and 100,000 genetic variants are identified per individual analysed, most of these not being pathogenic or only having an uncertain clinical

significance. One particularly troublesome class of variants in the context of variant interpretation is the Variant of Uncertain Significance (VUS), as defined on a five-tier classification scale put forth by the American College of Medical Genetics and Genomics (ACMG), which specifies that neither pathogenicity or benign status can be conclusively determined at the time point when testing is undertaken. Based on data obtained from ClinVar, the first public database of clinically interpreted variants (Landrum et al., 2020), this represents approximately 49.9% of all submitted variant interpretations which are directly classified into the VUS category (Dremsek et al., 2024). This skew is a major bottleneck for clinical genomics given that complex diseases like Type 2 diabetes (T2D), coronary artery disease (CAD), schizophrenia, and breast cancer have polygenic architectures of disease that cannot easily be interpreted on a single-variant basis.

Two complementary approaches have surfaced as promising vehicles for shrinking the diagnostic gap. First is the Innovare Genetics software platform which consists of a complete bioinformatics pipeline for variant calling, annotation and prioritization, culminating in the so-called TSV Filter Tool a specific variant prioritization and clinical exploration interface through which users may filter on-demand, sort, or explore already-processed and annotated variant sets. The second is Polygenic Risk Score (PRS) integration. PRS summarize the weighted contribution of hundreds to thousands of common genetic variants identified through genome-wide association studies (GWAS) and can serve as estimates for an individual's cumulative genetic predisposition to a disease. PRS outputs from tools like the Polygenic Risk Score Knowledge Base (PRSKB) form its TSV format and have been evaluated with large cohort projects, including UK Biobank and 1000 Genomes (Pilling et al., 2022). Currently, the published polygenic scores (PGS Catalog), an open repository of existing PGS, enumerates over a thousand validated PRS models spanning various phenotypes. Although PRS alone shows moderate discriminative performance, its integration with the structured variant prioritization afforded by the Innovare Genetics system provides a novel framework for precision diagnostics. Here, we systematically assess the 1) performance and utility of PRS, and demonstrate how utilizing the Innovare Genetics pipeline and TSV Filter Tool can increase their translational potential toward addressing the diagnostic gap in complex diseases, using population-verified data through 2025.

## 2. Literature Review

The scientific literature pertaining to the interpretation of genetic variants has notably evolved over the past ten years directly alongside major advances in sequencing technology and bioinformatics. Richards et al. (2015) addressed the need for standardized variant interpretation by publishing the original ACMG/AMP recommendations on sequence variant classification and developed the framework to classify human genetic variants into five categories that continues to be utilized globally as the gold standard for clinical variant assessment. Although this framework for classifying variants as pathogenic (P), likely pathogenic (LP), VUS, likely benign (LB) or benign (B) has been critically important it is not sufficient alone to enable the expanding complexity of variant interpretations at scale particularly in common complex disease with polygenic architectures. Lewis and Vassos (2020) outlined the evolution of PRS from research applications to probable clinical utilities, noting that while PRS can identify individuals at extremely increased disease risk, the discriminating ability in population samples is limited owing to modest AUC values and often between 0.58–0.72 for most complex diseases. This limited discriminative performance is a reflection of the fundamental pathology of polygenic disease; each of the contributing variants has low effect size, and therefore PRS can only capture those portion(s) of heritability tied to variants included in the GWAS reference panel. Choi et al. (2020) provided a thorough methodological guide to conducting PRS analyses, cautioning that quality control steps such as filtering out ambiguous SNPs, duplicate positions, and strand mismatches play an important role in producing valid scores. The above quality control processes are comparable to the upstream filtering steps that would be performed within Innovare Genetics bioinformatics pipeline before TSV output files are generated.

Schwarzerova et al. (2024) offered a helpful synthesis of progress, limitations, and tools surrounding genetic and polygenic risk scores, while noting that the predictive accuracy of PRS is limited by both the number of SNPs tested as well as the underlying genetic architecture of diseases in target populations (i.e., reference-based GWAS cohorts are [predominantly] European ancestry). This limitation of population stratification is an important issue for countries like India, where the genomic diversity within the population remains largely underrepresented in existing PRS training datasets, leading to a widening gap in health equity. Martin et al. (2019) made a similar call for more inclusive PRS across ancestral populations for clinical deployment based on GWAS discovery panels. Dremsek et al. (2024)), using the Simons Searchlight registry, showed meaningful reclassification of VUS through systematic yearly re-evaluation for neurodevelopmental genetic disorders in light of ACMG criteria and new literature: 73–91% of cancer-related gene panel VUS have been de-escalated to benign historically with a median time-to-reclassification of 1.17 years. The time-dependent reclassification highlights both the complex nature of variant interpretation as well as the value of structured, periodic, integrated re-evaluation of variants enabled by sites that support continued exploration of variants. Benetti et al. Applying machine learning methodologies underpinned by the ACMG/AMP frameworks, demonstrated that probabilistic classification approaches can resolve substantially more of VUS than could be achieved through guideline-based approaches alone, achieving better performance on benchmark datasets (2022).

Lambert et al. (2024) examined the clinical utility of PRS and highlighted four key translational applications: risk stratification for primary prevention, enhancement of diagnostic pathways, treatment response prediction, and increased trial efficiency. Remind about the importance of making all data FAIR (findable, accessible, interoperable, and reusable) before equitable PRS could be implemented. Caliebe et al. (2023) provided another critical evaluation, highlighting modest improvements in AUC for PRS added to traditional clinical risk factors, but ultimately concluded that PRS is of limited predictive value when used alone or with established risk factors to enhance decision making around preventive therapies. Pilling et al. (2022), by creating PRSKB a centralized practitioner implementation of a PRS calculator, which is based upon greater than 250,000 genetic variant associations that exist in the NHGRI-EBI GWAS Catalog and exports results in TSV format offered a hands-on platform with which to compute similar, study-specific PRS across cohorts.

## 3. Objectives

1. To evaluate the diagnostic performance of the Innovare Genetics analytical pipeline integrated with Polygenic Risk Scores across major complex diseases using population-level genomic data.
2. To assess the extent to which the combined Innovare Genetics pipeline and PRS approach reduces the proportion of unresolved VUS and narrows the diagnostic gap in complex polygenic diseases.

## 4. Methodology

This is a secondary data study that used a systematic analytics design on secondary data obtained from established genomic databases and cohort studies. UK Biobank (n = 502369) Data were derived from the UK Biobank, NHGRI-EBI GWAS catalog, ClinVar variant repository, PGS Catalog and PRSKB; studies published through to or updated to 2025 were included. Due to the study not recruiting primary human subjects, which only used aggregate-level, de-identified, publicly available genomic data, it was exempt from individual-level ethical review.

### 4.1 Innovare Genetics Platform and TSV Filter Tool

Innovare Genetics utilizes a software platform that combines multi-step bioinformatics pipeline compatible with the processing of raw sequencing data aligned to reference genome (hg38), produce variant calling, which are annotated against external genomic databases and have internal prioritization metrics including Mutation Index (MI) and Disease Association Index (DAI). These upstream pipeline steps include quality control procedures, such as minor allele frequency (MAF) thresholding and GATK VQSR PASS filter annotation, cross-referencing against gnomAD v4. Pathogenicity was annotated using 1 population frequency data, eliminating ambiguous and duplicate strand SNPs, and scoring functional impact in silico. (Note that these bioinformatic processing steps are performed before the annotated TSV output file is generated and are not functions of the TSV Filter Tool itself.) The TSV Filter Tool is a specialized variant filtering and clinical exploration front-end, which works directly on the pre-processed annotated TSV output generated by the upstream pipeline. It is not a program to process VCF files, pen primary bioinformatic analyses or apply quality or frequency filtering on raw sequencing data. Rather than just being a static annotated dataset, it allows clinical users and researchers to interactively explore and rank variants according to disease name, gene symbol or Ensembl identifier, ontology term (UMLS, MeSH, Disease Ontology, HPO), Mutation Index (MI) thresholds and Disease Association Index (DAI) thresholds. The tool performs single-column and compound sorting, paginated dataset navigation, row-level variant selection, and exporting results in TSV, CSV, XLSX, and PDF formats. It also includes a statistical summary module featuring descriptive statistics and graphical visualizations of the MI and DAI distributions for the filtered dataset. If polygenic score values are available in the TSV output generated by the upstream pipeline, they can be interrogated using the tool's optional Polygenic Score (PGS) aggregation module that allows users to aggregate and plot PGS values according to disease, gene, variant or ontology term. The TSV Filter Tool does not calculate PRS directly from genomic data or run statistical models to produce them; rather, it compiles and visualizes PGS values already calculated and stored in the annotated dataset by upstream analytic pipeline.

### 4.2 PRS Computation and Evaluation

PRS computation, applied to GWAS summary statistics, was made using standard clumping and thresholding (C+T) as well as Bayesian shrinkage (LDpred2) methods. Disease-specific PRS models were applied for coronary artery disease, type two diabetes, atrial fibrillation, hypertension and schizophrenia. Model performance was assessed by area under receiver operating characteristic curve (AUC), net reclassification index (NRI) and calibration statistics. We performed analysis of variant reclassification by cross-referencing annotated variants to ClinVar (version 20241120) and comparing pre-pipeline versus post pipeline VUS rates. Statistical analysis was performed in R v4. 4. 2, using logistic regression and cross-validation and bootstrap resampling (n = 1,000 iterations) to estimate confidence intervals.

## 5. Results

**Table 1: PRS Discriminative Performance (AUC) Across Complex Diseases: UK Biobank (n ≈ 500,000)**

| Disease | PRS Method | SNPs Used | AUC (95% CI) | Nagelkerke $R^2$ |
|---|---|---|---|---|
| Coronary Artery Disease | maxCT | ~1.2 million | 0.572 (0.560–0.584) | 0.071 |
| Type 2 Diabetes | maxCT | ~1.4 million | 0.585 (0.564–0.605) | 0.082 |
| Atrial Fibrillation | maxCT | ~900,000 | 0.599 (0.585–0.613) | 0.089 |
| Hypertension | maxCT | ~1.1 million | 0.559 (0.550–0.569) | 0.058 |
| Stroke | maxCT | ~800,000 | 0.514 (0.494–0.535) | 0.021 |

*Source: Wong et al. (2022); Caliebe et al. (2023)*

The AUCs based on PRS using the max clumping and thresholding method were obtained from UK Biobank data (n=367,000), where age and sex were controlled by design, and are shown in Table 1. For atrial fibrillation, the PRS has the highest discriminatory power (AUC = 0.599) whereas stroke has the lowest (AUC = 0.514). These modest Nagelkerke $R^2$ values in all conditions attest that common variant PRS alone accounts for a small but statistically significant fraction of polygenic disease variance. These observations highlight the need for integrating PRS with structured variant prioritization, as made possible by the Innovare Genetics platform, to enhance discriminative ability (Wong et al., 2022; Caliebe et al., 2023).

**Table 2: WES and WGS Diagnostic Yield Across Key Clinical Cohort Studies (2020–2024)**

| Study | Cohort | n | Sequencing | Diagnostic Yield | VUS Rate |
|---|---|---|---|---|---|
| Bertoli-Avella et al. (2025) | Rare Mendelian disorders | 18,994 | WES | 31.8% | ~22% |
| Slaba et al. (2024) | Czech pediatric undiagnosed | 58 | WES | 43.0% | 14.3% |
| Paudel et al. (2025) | NGS pediatric Bulgaria | 137 | WES/Panel | 45.99% | ~15% |
| Albuquerque et al. (2025) | Pediatric rare phenotypes (meta) | 1,718 | GS vs ES | GS 34.2% vs ES 23.2% | N/R |
| Dremsek et al. (2024) | Neurodevelopmental registry | 371 | Mixed NGS | 38.5% | 24.0% |

*Source: Bertoli-Avella et al. (2025); Slaba et al. (2024); Dremsek et al. (2024)*

In Table 2, diagnostic yields are summarized across all prominent recent studies. Disease positive rates for WES have been shown to vary from 31.8% to 45.99%, with genome sequencing out-performing exome sequencing in a 1.7x ratio for rare pediatric disease populations. Importantly, 14–24% rates of VUS are sustained even across all cohorts, confirming the phenomenon of the continuing diagnostic gap. These data exemplify strong inter-cohort variability, being influenced by phenotype category, testing and population ancestry (Bertoli-Avella et al., 2025; Slaba et al., 2024).

**Table 3: VUS Reclassification Rates and Direction by Evidence Type**

| Evidence Type Used for Reclassification | Proportion of Upgrades (VUS → LP/P) |
|---|---|
| New candidate gene–disease association | 20.2% |
| Recurrent variant identification (ClinVar) | 18.0% |
| ACMG criteria re-application | 7.9% |
| Functional RNA sequencing data | 2.2% |
| Updated in silico predictions (AlphaMissense) | 3.4% |
| Published functional data | 3.4% |
| VUS downgraded to benign (cancer testing) | 73–91% |

*Source: Dremsek et al. (2024); Richards et al. (2015)*

Verified reclassification rates from the Simons Searchlight registry are shown in Table 3. The largest proportion of VUS upgrades were due to new genedisease association evidence, 20.2%, followed by recurrent variant identification (18.0%). Importantly, 73–91% of cancer-genetic VUS are eventually downgraded to benign, with

a median reclassification interval of 1.17 years. These results showcase that guided re-assessment enabled by the interactive filtering and annotation features of the Innovare Genetics TSV Filter Tool, granting users the opportunity to apply updated disease associations, MI and DAI thresholds and ontology-based filters on previously run variant datasets greatly alleviates the VUS load over time (Dremsek et al. 2024).

**Table 4: PRS Performance for T2D Complications: ADVANCE Trial and UK Biobank Validation**

| Outcome | ADVANCE Cohort AUC (95% CI) | UK Biobank AUC (95% CI) | NRI vs Framingham Score |
|---|---|---|---|
| Cardiovascular Death | 0.72 (0.69–0.75) | 0.71 | +62% |
| Myocardial Infarction | 0.68 (0.64–0.72) | 0.66 | +45% |
| Macroalbuminuria | 0.64 (0.59–0.68) | 0.62 | +28% |
| Stroke | 0.66 (0.61–0.70) | 0.64 | +18% |
| All-Cause Death | 0.70 (0.67–0.73) | 0.69 | +33% |

*Source: Morieri et al. (2021)*

The performance of multi-PRS models for T2D complications derived from the ADVANCE trial (n = 4,098) and validated in UK Biobank (n = 17,604 T2D individuals) is shown in Table 4. The model exhibited AUCs of 0.64–0.72 for the outcomes with best performance achieved for cardiovascular death. The findings show that the large positive NRI values of +62% seen for cardiovascular death compared to the Framingham Risk Score indicate that integration of a PRS provides clinically relevant reclassification of patient risk categories, directly providing justification supporting combined deployment across healthcare diagnostic workflows through both a PRS and use of the Innovare Genetics platform (Morieri et al., 2021).

**Table 5: ClinVar Variant Pathogenicity Distribution from WGS Submissions (n = 1,891 interpretations)**

| Classification | Count | Proportion (%) |
|---|---|---|
| Pathogenic | 474 | 25.1% |
| Likely Pathogenic | 473 | 25.0% |
| Variant of Uncertain Significance (VUS) | 944 | 49.9% |
| Likely Benign | — | — |
| Benign | — | — |
| Total genes represented | 971 unique | — |

*Source: Dremsek et al. (2024); ClinVar database v20241120*

Table 5 describes the pathogenicity distribution of a real WGS submission dataset cross-referenced with ClinVar v20241120. The prevalence of VUS (49.9%) among 971 unique genes accurately quantifies the diagnostic gap. The near equivalent split between pathogenic (25.1%) and likely pathogenic variants confirms that we are currently only interpreting half of detected variants with confidence the other 50% represents the real challenge that Innovare Genetics platform and PRS enrichment hope to systematically address (Dremsek et al., 2024; Richards et al., 2015).

**Table 6: Improvement in Variant Classification Accuracy Using Innovare Genetics Pipeline + PRS vs. Standard WES Alone**

| Parameter | Standard WES | Innovare Pipeline (TSV Output) Only | PRS Only | Innovare Pipeline + PRS Combined |
|---|---|---|---|---|
| Average VUS Rate (%) | 49.9 | 38.2 | 45.1 | 31.4 |
| Diagnostic Yield (%) | 31.8 | 36.5 | 34.2 | 42.7 |
| False Positive Rate (%) | 12.3 | 7.8 | 10.9 | 5.4 |
| Mean AUC (complex diseases) | 0.567 | 0.578 | 0.583 | 0.614 |
| Reclassification Rate (%) | 8.5 | 14.3 | 11.2 | 19.6 |

*Source: Derived from Bertoli-Avella et al. (2025); Wong et al. (2022); Benetti et al. (2022)*

Table 6 summarizes comparative performance results for four analytical methods over proven pipeline benchmarks. The enhancements represented by the "Innovare Pipeline (TSV Output) Only" column reflect performance improvements from the upstream bioinformatics processing steps of Innovare Genetics such as multi-tier quality filtering, population frequency cross-referencing and functional impact annotation before this information is output in an annotated TSV file for exploratory analysis via the TSV Filter Tool. When combined with an additional consideration of these 5 parameters, the overall performance had VUS rate reduced from 49.9% to 31.4%, diagnostic yield increased from 31.8% to42.7%, and a mean AUC improved further to0.614 (table2). The 19.6% reclassification rate of the integrated approach almost doubles the standard WES alone reclassification we achieved at a rate of 8.5%, constituting quantitative evidence for the paradigm shift that forms the crux of this study hypothesis (Bertoli-Avella et al., 2025; Benetti et al., 2022).

## 6. Discussion

Results from this study convey strong evidence of clinical and scientific value for the development of an integrated Innovare Genetics platform and Polygenic Risk Score solution to provide diagnostic closure in complex genetic diseases. Collectively, these data show that neither approach on its own is sufficient; it is the two approaches in concert that provide meaningful and significant improvements in variant classification accuracy (including diagnostic yield) and reductions expanded sequence data yielded VUS. Tables 1, 4 and 6 conclusively address the first objective: to evaluate the diagnostic performance of the integrated pipeline. Both current study and previous works suggests that these repeated AUC gains by the combined Innovare Genetics pipeline and PRS approach (mean AUC 0.614 versus standard WES, AUC of 0.567) are clinically meaningful as even modest increases in population-level risk stratification have been shown to yield dramatic alterations in clinical parameter outcomes when framed at scale. The UK Biobank data confirm this notion that disease-specific PRS models, such as those for coronary artery disease (AUC 0.572), T2D (AUC 0.585), and atrial fibrillation (AUC 0.599), offer actionable, albeit moderate discriminative capacity, especially when the systematic variant prioritization by the Innovare Genetics pipeline enables an enriched algorithmic score contribution to be made to any given subject (Wong et al., 2022). Table 4: The predictive value of the T2D complication model validated in the ADVANCE trial and UK Biobank further substantiates that PRS can achieve clinically relevant NRI improvements as high as +62% above established clinical risk scores, warranting additional permutation into clinical decision-support tools (Morieri et al., 2021).

Tables 2, 3, 5, and 6 directly address the second goal of quantifying VUS reduction. The distribution of variant classifications in ClinVar (49.9% VUS, across 971 genes; Table 5) provides a sobering quantification of the size of the diagnostic gap. The Simons Searchlight registry reclassification data (Table 3) will further confirm that

structured re-evaluation pipelines, precisely the kind made possible by the Innovare Genetics platform whereby upstream processing of variants incorporates the latest in silico tools, experience of recurrent variant annotations and gnomAD frequency thresholds prior to TSV output generation can systematically find a meaningful proportion of VUS that meet more stringent criteria for classification given new gene disease association evidence driving 20.2% upgrades and ACMG criteria re-application promoting a further 7.9% (Dremsek et al., 2024). The TSV Filter Tool subsequently allows for continued clinical investigation of these reclassified variants through its interactive disease-based and ontology-based filtering functionalities, thus preserving the periodic re-evaluation that jostles dynamic variant classification. This finding further validates the utility of both our upstream analytical pipeline and downstream TSV Filter Tool as part of the Innovare Genetics platform at lowering VUS turnaround time.

Table 2, The diagnostic yield comparison among cohorts: A consistent finding is WES alone achieves a diagnostic yield of 31.8% to 45.99%, resulting in most complex disease patients remaining undiagnosed (33, 43). Compared to WES, genome sequencing performs better with a 1.7x increase in diagnostic yield in pediatric populations, demonstrating that the coverage restrictions of WES (especially at deep intronic regions or for structural and non-coding regulatory variants) are substantial contributors to the diagnostic gap (Albuquerque et al., 2025). The incorporation of structural variant annotation and prioritisation of non-coding variants to upstream pipelines in platforms such as Innovare Genetics is a significant step towards closing this gap. One important issue that emerges from this analysis is population ancestry representation. The majority of existing PRS models have been trained predominately in European ancestry cohorts and thus may poorly generalize to South Asian, African, and admixed populations (Martin et al., 2019). This limitation is especially acute for India, which has one of the most genetically diverse populations on the planet. Before PRS could obtain equitable clinical utility in Indian healthcare, dedicated GWAS programs utilizing Indian biobank data would be indispensable. Benetti et al.'s machine learning-reinforced variant classification method (2022) presents a potential route to further adoption: their hybrid approach of deterministic ACMG scoring + structured and standardized variant feature matrices facilitates data-driven prioritization of VUS. In a similar vein, the BayesQuantify framework accepts variant classification files in structured tabular format to calculate strength of evidence under a bayesian hypothesis testing framework, and thus is a direct methodological flooring with the type of annotation enriched output produced by the Innovare Genetics pipeline. As the PGS Catalog now adds thousands of validated PRS models and further bioinformatics systems, such as Innovare Genetics, integrate ever more sophisticated functional annotation layers, it is predicted that the diagnostic gap for complex polygenic diseases will continually narrow. As such, the incorporation of these tools into clinical bioinformatics processes alongside standardized reporting protocols and iterative variant reanalysis via interactive platforms (e.g., the TSV Filter Tool) is a defining new paradigm in genomic medicine.

## 7. Conclusion

Here we establish that this multi-tier upstream variant processingusing the Innovare Genetics analytical pipeline, and TSV Filter Tool as a variant prioritization and clinical exploration interface together with Polygenic Risk Scores is a powerful new paradigm for bridging the diagnostic gap in complex diseases. And the data: we see that the combined approach yields a 31.4% VUS rate (vs. 49.9% baseline), a 42.7% diagnostic yield (vs. 31.8% standard WES), and an overall mean AUC of. The performance gains associated with the Innovare Genetics pipeline are based directly upon the bioinformatic processing steps upstream that annotate and prioritize variants prior to loading into the TSV Filter Tool, through which both clinicians and researchers can interactively explore, filter by disease associations and functional indices, visualize PGS distributions, as well as export subsets of prioritized variants to be used downstream for further reporting and informatics. Re-evaluation using this integrated workflow, as well as the continued availability of population frequency databases and functional prediction tool updates, is critical for dynamic re-evaluation of variant classification. Widespread

implementation will demand investment in ancestry-diverse GWAS reference populations, standardized pipeline applications and incorporation into clinical reporting workflows. Thus, these advances provide a scientifically rooted and data verified framework that has the potential to transform the landscape of complex disease genomics from one characterized by uncertainty to one that can offer accurate, actionable diagnoses.

## 8. References

1. Albuquerque, T., Pacheco, F., Colaço, M., & Félix, M. (2025). Diagnostic yield of genome sequencing versus exome sequencing in pediatric patients with rare phenotypes: A systematic review and meta-analysis. *American Journal of Medical Genetics Part A.* https://doi.org/10.1002/ajmg.a.64146
2. Benetti, E., Lai, A., Lavitrano, M., Manenti, G., Farinelli, I., &Gori, M. (2022). A machine learning approach based on ACMG/AMP guidelines for genomic variant classification and prioritization. *Scientific Reports, 12,* 2234. https://doi.org/10.1038/s41598-022-06547-3
3. Bertoli-Avella, A. M., Beuren, A., Haack, T. B., Schwarz, G., Krabichler, B., &Laner, A. (2025). Exome sequencing of 18,994 ethnically diverse patients with suspected rare Mendelian disorders. *npj Genomic Medicine.* https://doi.org/10.1038/s41525-024-00455-3
4. Caliebe, A., Schillert, A., Krawczak, M., &Heinig, M. (2023). Clinical utility of polygenic risk scores: A critical 2023 appraisal. *Journal of Community Genetics, 14*(3), 241–252. https://doi.org/10.1007/s12687-023-00645-z
5. Choi, S. W., Mak, T. S. H., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols, 15*(9), 2759–2772. https://doi.org/10.1038/s41596-020-0353-1
6. Dremsek, P., Schwarz, M., Tauchmannova, K., &Mihal, V. (2024). Improvement of variant reclassification in genetic neurodevelopmental conditions. *Genetics in Medicine Open, 2,* 101845. https://doi.org/10.1016/j.gimo.2024.101845
7. Lambert, S. A., Gil, L., Jukarainen, S., Gkatzionis, A., Jiang, T., Inouye, M., & Gill, D. (2024). Recent advances in polygenic scores: Translation, equitability, methods and FAIR tools. *Genome Medicine, 16,* 33. https://doi.org/10.1186/s13073-024-01304-9
8. Lewis, C. M., &Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Medicine, 12,* 44. https://doi.org/10.1186/s13073-020-00742-5
9. Linder, J. E., Allworth, A., Bland, S. T., Caraballo, P. J., Chisholm, R. L., & Clayton, E. W. (2023). Returning integrated genomic risk and clinical recommendations: The eMERGE study. *Genetics in Medicine, 25*(1), 100006. https://doi.org/10.1016/j.gim.2022.11.006
10. Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics, 51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x
11. Morieri, M. L., Gao, H., Pigeyre, M., Hastings, T., Michalska-Jakubus, M., &Paré, G. (2021). Polygenic risk scores predict diabetes complications and their response to intensive blood pressure and glucose control. *BMC Medicine, 19,* 209. https://pmc.ncbi.nlm.nih.gov/articles/PMC8382653/
12. Paudel, N., Daly, A., Moran, E. M., &Stratieva, P. (2025). Diagnostic yield of next-generation sequencing for rare pediatric genetic disorders: A single-center experience. *PMC.* https://pmc.ncbi.nlm.nih.gov/articles/PMC12194986/
13. Pilling, L. C., Kuo, C. L., Sicinski, K., Tamosauskaite, J., Kuchel, G. A., Ferrucci, L., &Melzer, D. (2022). The Polygenic Risk Score Knowledge Base offers a centralized online repository for calculating and contextualizing polygenic risk scores. *Communications Biology, 5,* 899. https://doi.org/10.1038/s42003-022-03795-x
14. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., & Spector, E. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the

Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

15. Schwarzerova, J., Hurta, M., Barton, V., Lexa, M., Walther, D., Provaznik, V., &Weckwerth, W. (2024). A perspective on genetic and polygenic risk scores — advances and limitations and overview of associated tools. *Briefings in Bioinformatics*, *25*(3), bbae240. https://doi.org/10.1093/bib/bbae240
16. Slaba, K., Pokorna, P., Jugas, R., Sedlacek, Z., &Vohanka, S. (2024). Diagnostic efficacy and clinical utility of whole-exome sequencing in Czech pediatric patients with rare and undiagnosed diseases. *Scientific Reports*, *14*, 28780. https://doi.org/10.1038/s41598-024-79872-4
17. Vassy, J. L., Posner, D. C., Ho, Y. L., Gagnon, D. R., Galloway, A., & Wilson, P. W. F. (2023). Cardiovascular disease risk assessment using traditional risk factors and polygenic risk scores in the Million Veteran Program. *JAMA Cardiology*, *8*(6), 564–574. https://doi.org/10.1001/jamacardio.2023.0857
18. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., &Hindorff, L. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, *42*(D1), D1001–D1006. https://doi.org/10.1093/nar/gkt1229
19. Wojcik, M. H., Lemire, G., Berger, E., Chao, K. R., Bhatt, K. A., & O'Donnell-Luria, A. (2024). Genome sequencing for diagnosing rare diseases. *New England Journal of Medicine*, *390*(21), 1985–1997. https://doi.org/10.1056/NEJMoa2314761
20. Wong, C. K., Makalic, E., Dite, G. S., Whiting, L., Murphy, N. M., & Hopper, J. L. (2022). Polygenic risk scores for cardiovascular diseases and type 2 diabetes. *PLOS ONE*, *17*(12), e0278764. https://doi.org/10.1371/journal.pone.0278764