

Science Spotlight

← Science Spotlight

From Data to Diagnosis: GREGoR aims to demystify rare diseases

From the Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium, including the Fred Hutch/University of Washington/Seattle Children's Cancer Consortium

JANUARY 5, 2026 • BY K MITCHELL

Rare diseases, though individually uncommon, collectively affect millions worldwide. Despite major advances in genome sequencing and interpretation of genomic data on affected individuals, for many families the search for a genetic diagnosis remains inconclusive. This diagnostic gap stems from multiple obstacles: genetic variants that fall outside protein coding regions, limitations in analysis pipelines, missing reference genome annotations, limited functional data, and the sheer diversity of rare conditions.

To address this important need, five research centers—including University of Washington—came together in 2021 to form the Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium. GREGoR—supported by the National Human Genome Research Institute (NHGRI)—aims to systematically evaluate the usefulness of emerging sequencing technologies for different disease types, build multi-omic datasets, develop new computational approaches, and rapidly share data with the global scientific community. A new Perspective, published in *Nature* in November 2025, lays out progress made by the GREGoR Consortium and outlines remaining challenges that lie ahead.

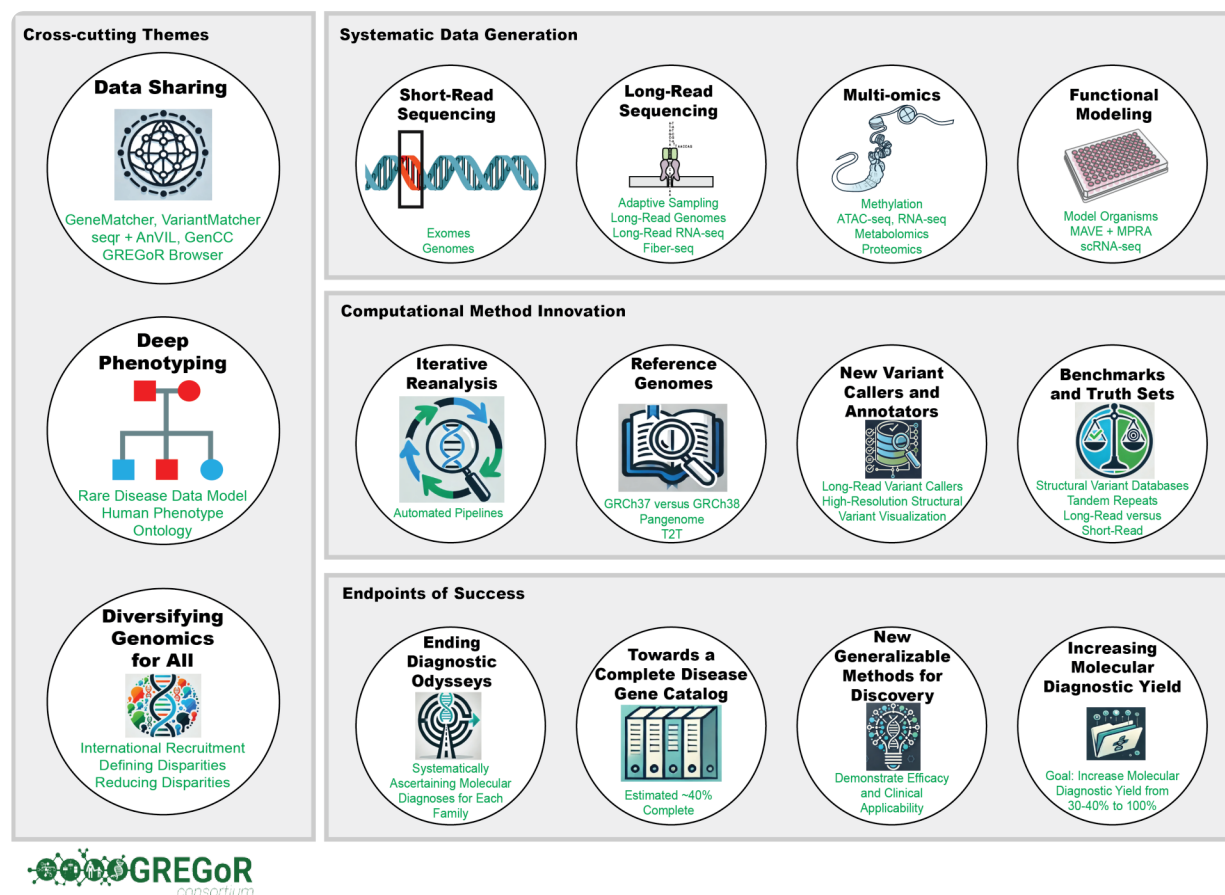
What has the consortium accomplished so far? “**GREGoR** collects and shares data from over 7,500 individuals across ~3,000 families — many of which had previously undergone clinical testing (e.g., exome sequencing) and remained undiagnosed. It makes all generated data publicly available via a shared, globally accessible platform (NIH’s AnVIL), lowering barriers for researchers worldwide to reanalyze, reuse, and build on the data.” lead author Moez Dawood shares. “The work essentially shifts rare-disease research from isolated case studies to a scalable, collaborative, global infrastructure — a “foundation” for future discovery. The consortium model helps tackle the problem that more than half of rare-disease cases remain unsolved even after clinical sequencing.”

But GREGoR is going beyond just amassing genomic data— these researchers are evaluating the performance, limitations, and ideal use-cases of genomics technologies to guide clinicians and researchers on which methods to use for unsolved rare diseases. Since researchers’ understanding of the human genome is heavily weighted towards protein coding regions, if a rare disease patient does not exhibit genetic variation in the coding genome (detected by exome sequencing), it can be difficult to determine next steps. To this end, GREGoR has developed computational approaches to extract new diagnosis by reanalyzing existing exome sequencing data. For example, they have improved techniques determining whether one or both alleles for a gene are disrupted (known as “phasing variants”) and developed better tools for detecting pathogenic structural variants such as deletions, duplications, insertions, inversions, and translocations.

Aside from reanalysis of existing data, GREGoR provides a roadmap for genetic testing when exome sequencing is inconclusive, starting with short-read genomic sequencing (srGS). In fact, their data suggests that srGS could be more effective as a first line diagnostic test than exome sequencing due to its higher diagnostic yield. GREGoR is developing tools that infer structural and copy-number variants from srGS, which would replace both SNP arrays and exomes as a single, cost-effective assay.

Variation in noncoding regions could have several possible effects including gene regulation, gene expression and mRNA splicing, and to assess each of these possibilities, technologies such as short- and long-read sequencing, RNA sequencing, structural variant discovery, and epigenomic profiling can be applied. These types of data can provide critical context about when, where, and how genes are expressed, helping researchers link genetic changes to their functional consequences in disease-relevant tissues. GREGoR has made progress towards demystifying these regions by identifying noncoding regions intolerant of

mutation. GREGoR is also evaluating how short- and long-read sequencing compare in diagnostic power and cost, showing, for instance, that targeted long-read methods can uncover structural and repetitive-sequence variants that short-read sequencing often misses.



Framework, tools and goals of the GREGoR consortium.

Graphic provided by Moez Dawood.

As Dawood notes, a central hurdle will be deciphering function and proving causality for novel variants: “How do we interpret noncoding or regulatory variants at scale and with confidence? As GREGoR is likely to generate many noncoding candidate variants, a major challenge remains: distinguishing which are truly pathogenic and developing robust interpretation frameworks.” But there is certainly reason for hope. GREGoR has illuminated molecular diagnoses in 365 genes, for instance linking noncoding genes such as *RNU4-2* to neurodevelopmental disorders, *CHASERR* to developmental and epileptic encephalopathy and *STRTS* to congenital hypothyroidism. Early genetic diagnosis could enable earlier clinical management strategies aimed at mitigating symptoms—for example, by monitoring disease progression or intervening to partially ameliorate haploinsufficiency of the affected

gene(s) before irreversible pathology develops.

GREGoR emphasizes that solving rare diseases will increasingly require integrating multi-omic datasets—genome, transcriptome, methylation, chromatin accessibility, and regulatory maps—to connect sequence variation with biological function. Their platform is also designed to accommodate and integrate emerging data types as they become available. In addition, the consortium is taking strides to document phenotypic heterogeneity in rare diseases to link all possible genotypes to all possible phenotypes. Another priority is equity. Dawood brings light to an important challenge: “How to ensure equitable discovery and diagnosis across diverse ancestries and underrepresented populations [when] biases in genomic databases, reference genomes, annotation tools, variant interpretation pipelines remain?”

Together, these efforts aim to transform the diagnosis of rare diseases from a long shot into a manageable, evidence-driven process for families worldwide.

Fred Hutch/University of Washington/Seattle Children’s Cancer Consortium Members Drs. Evan Eichler, Ali Shojaie, and Chia-Lin Wei contributed to this research.

The University of Washington Center for Rare Disease Research is one of the five research centers comprising the GREGoR consortium.

The spotlighted research was funded by the National Institutes of Health NHGRI GREGoR Consortium.

Dawood M, Heavner B, Wheeler MM *et al.* 2025. GREGoR: accelerating genomics for rare diseases. *Nature*. <https://doi.org/10.1038/s41586-025-09613-8>.

Science Spotlight writer Kelly Mitchell is a postdoctoral fellow in the Paddison Lab at Fred Hutch Cancer Center. She utilizes live cell reporters and CRISPR screening to study how glioblastoma cancer cells resist chemotherapy and radiation treatment. She obtained her PhD in cellular biology from Albert Einstein College of Medicine.

Tags

[cancer consortium,](#)
[genetics and genomics,](#)
[genomics database,](#)
[University of Washington](#)

For the Media

[Contact Media Relations](#)
[News Releases](#)
[Media Coverage](#)
[About Fred Hutch](#)

Fred Hutchinson Cancer Center is an independent organization that serves as UW Medicine's cancer program.

© 2026 Fred Hutchinson Cancer Center, a 501(c)(3) nonprofit organization