

genomDE Berlin 2025

Genome and Epigenome Diagnostics with Nanopore Sequencing

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

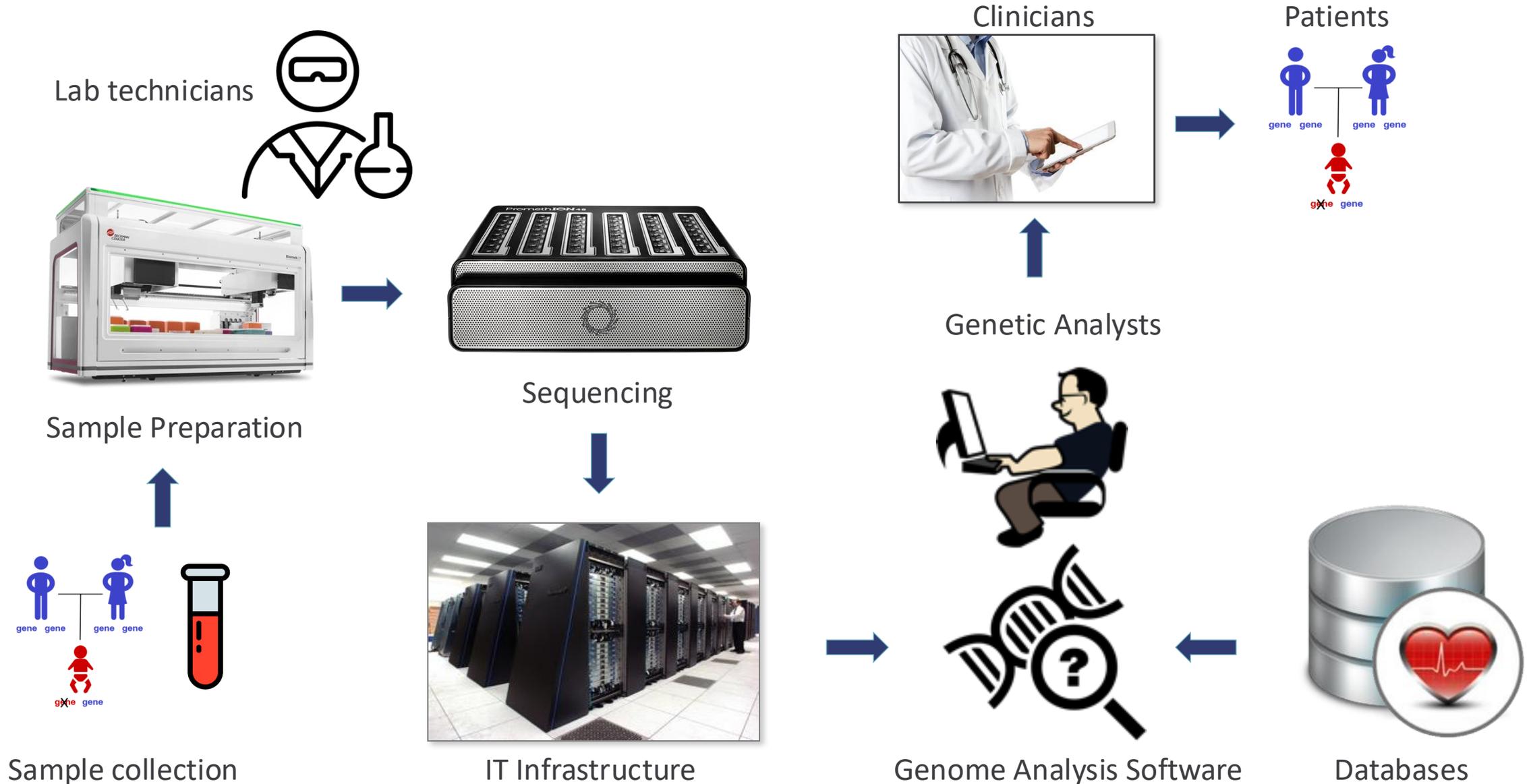


Stephan Ossowski
Institute of Medical Genetics and Applied
Genomics

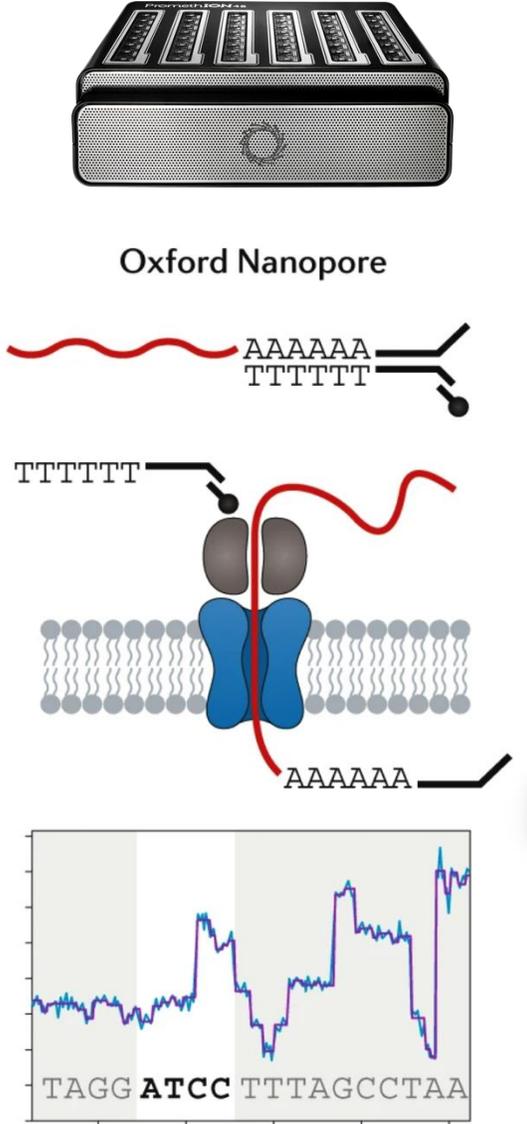


Universitätsklinikum
Tübingen

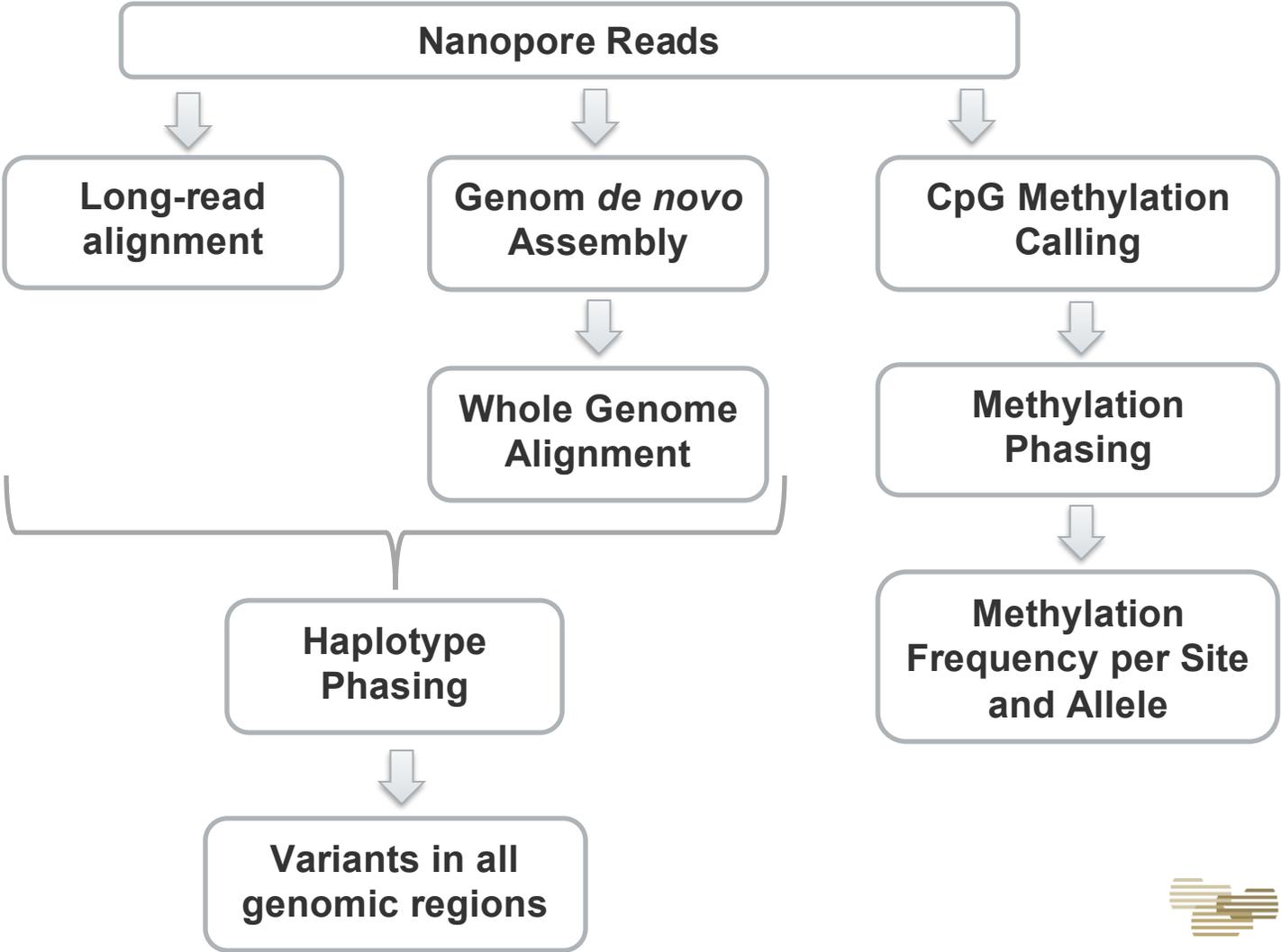
Nanopore Diagnostics Platform (Accredited)



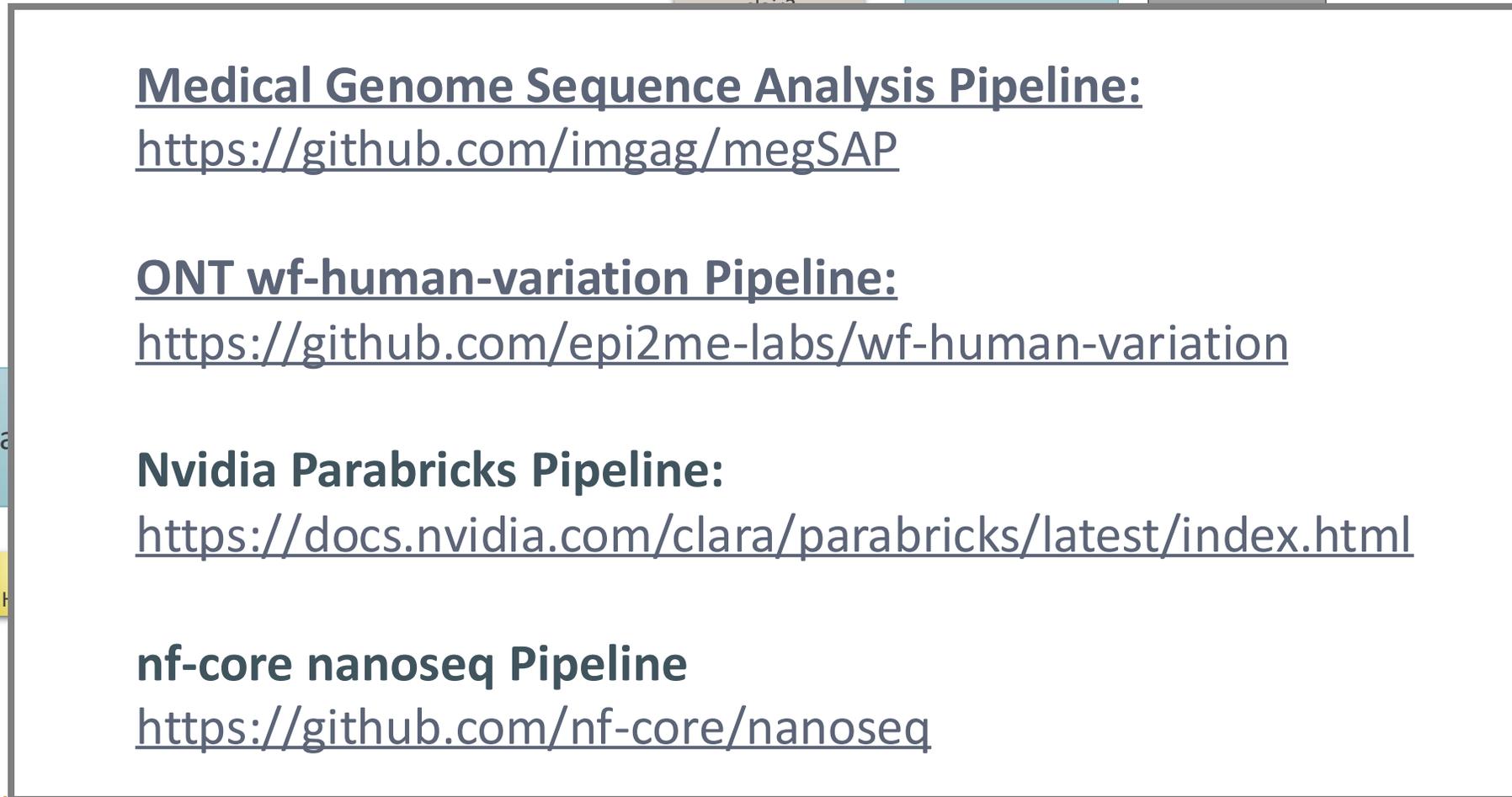
Genome & Methylome Diagnostics with Nanopore



New analysis options:



Nanopore Analysis Pipelines



 Run on GPU (either on PromethION or GPU server)

 Run on CPU (on normal CPU server)

Clinical Decision Support for Nanopore Diagnostics



GSvar

The screenshot displays the GSvar software interface with several key components:

- Main Table:** A table listing genomic variants with columns for chr, start, end, ref, obs, genotype, filter, quality, gene, variant_type, ing_and_splic, jgultc, OMIM, and ClinVar. A highlighted variant shows a G>A mutation in the CLP1 gene.
- Repeat Expansion:** A window titled 'Repeat Expansions of single-sample analysis DX197989_01' showing a table of repeat units with columns for chr, start, end, repeat_id, repeat_unit, repeats, wt_repeat, repeat_oi, filter, locus_coverage, reads_flanking, reads_in_repeat, and reads_spanning.
- Circos Plot:** A circular plot titled 'Circos Plot of single-sample analysis DX197989_01' showing genomic data across chromosomes, including SV (Structural Variants) and CNV (Copy Number Variations).
- Variant Details:** A window for 'chr11:57427367-57427367 G>A (hom)' showing gene details for CLP1, including its Ensembl ID (ENST00000302731), impact (LOW), and protein domain information.
- Filters:** A 'Filters' panel on the right with 'Main filters' and 'Target region filters'. The main filters include Allele frequency, Count NGSD, Impact, and Annotated pathogenicity.
- Phenotype Browser:** A window titled 'Phenotype browser' showing a list of HPO terms. The selected term is 'Abnormal heart valve morphology (HP:0001654)', with a definition and synonyms provided.
- Expression Data:** A window titled 'Expression Data of RNA200898AT_02' showing a table of gene expression data with columns for gene_id, gene_name, gene_type, size, tpm, cohort_mean, log2fc, zscore, pval, hpa_tissue_tpm, and hpa_t1.

Repeat Expansion

SV

CNV

Filter:
HPO-Terms





Benchmarking the Quality of Reads and Variant Calls

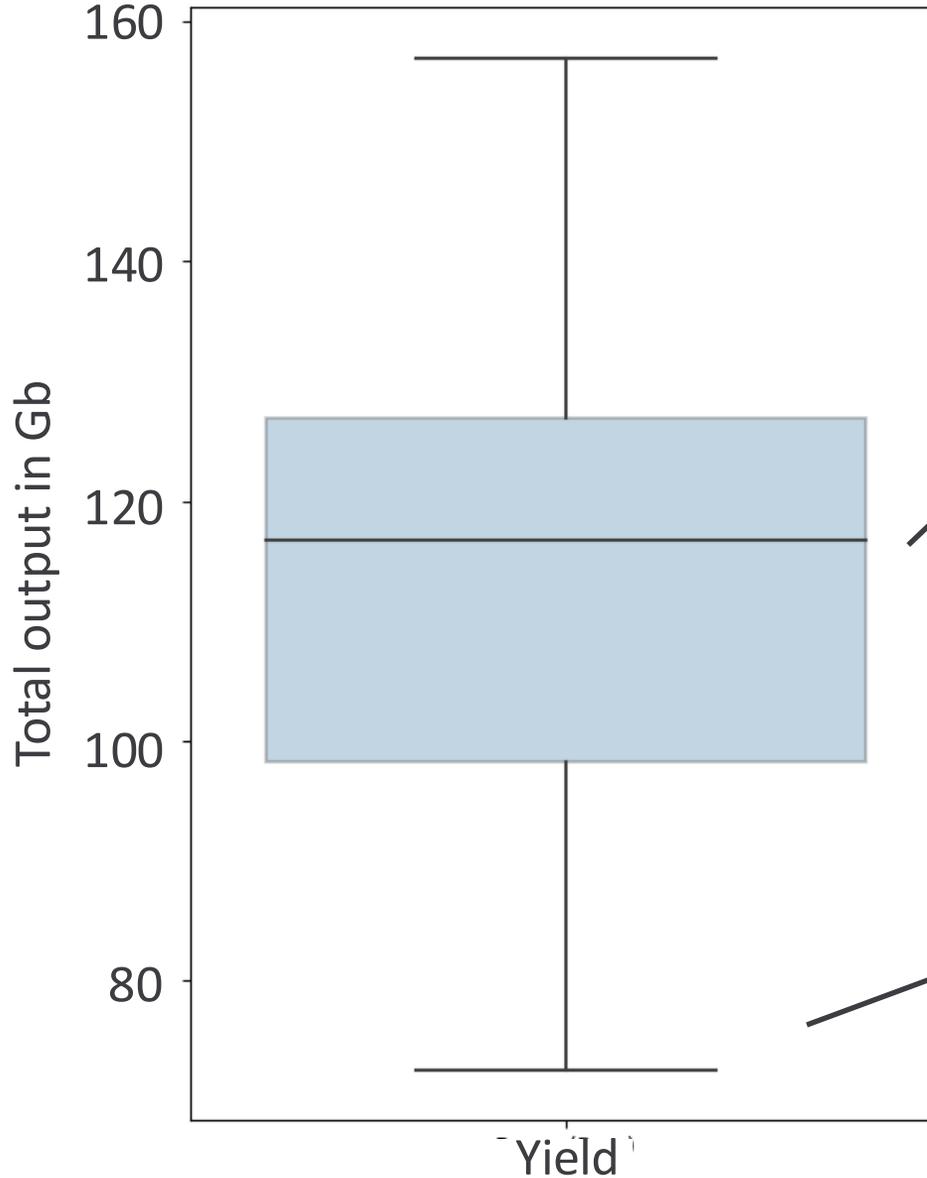


Yield per Flowcell and reaching "Diagnostic Coverage"



Diagnostic Genome:
35-40x Coverage

30x Coverage

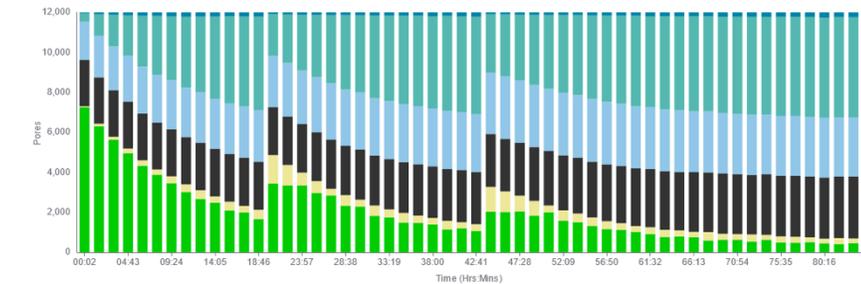
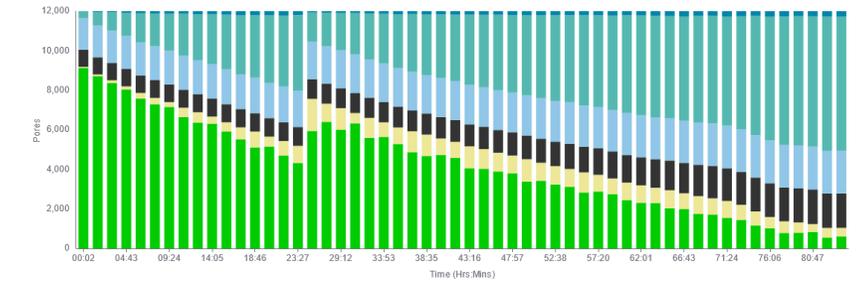


PORE SCAN

A Pore scan is performed at configurable time intervals to determine the current status of pores within channels on a Flow Cell. For this run a Pore scan is performed every 1.5 hrs.

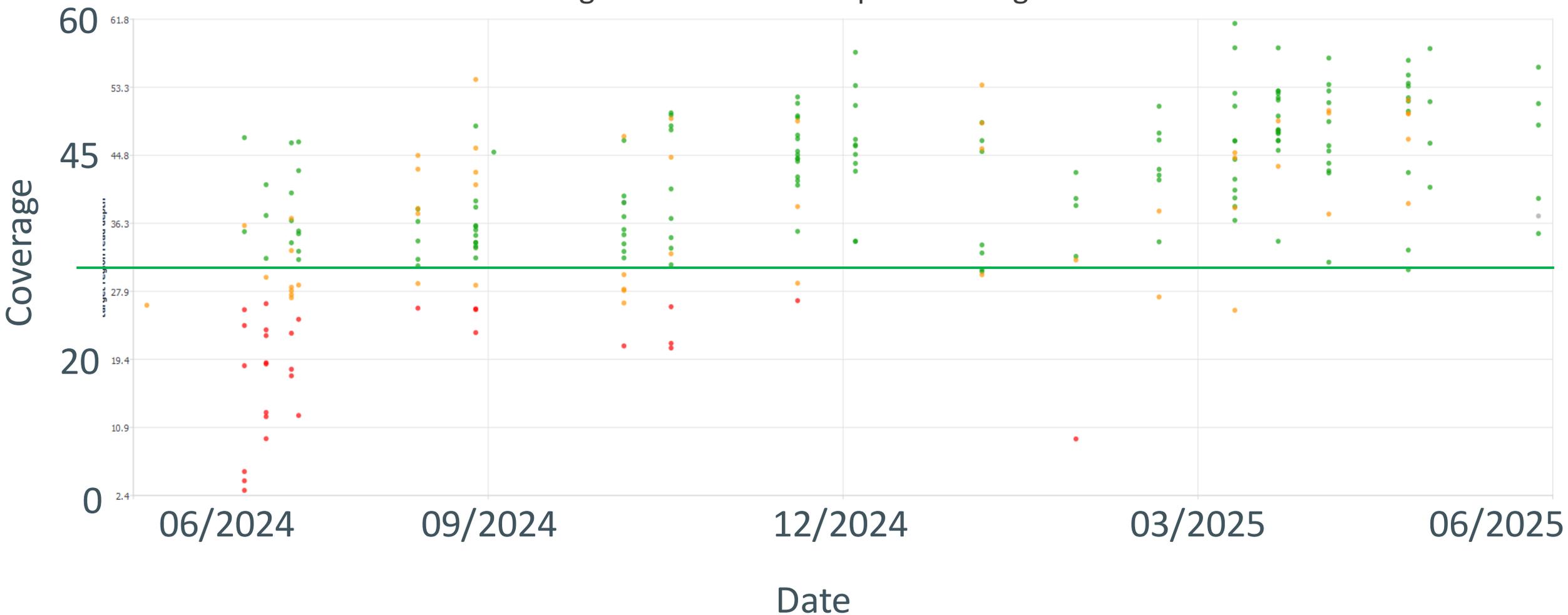
Legend

- Single Pore
Pore in channel available for sequencing
- Reserved Pore
Pore in reserve, will return to available when required
- Saturated
Possible contamination in the sample
- Zero
No current is passing through this pore, possibly due to bubbles on the membrane
- Unavailable
Pore inhibited from sequencing
- Inactive
Pore no longer suitable for further sequencing



One Human Genome per Flowcell: Coverage Yield

Coverage with one flowcell per human genome



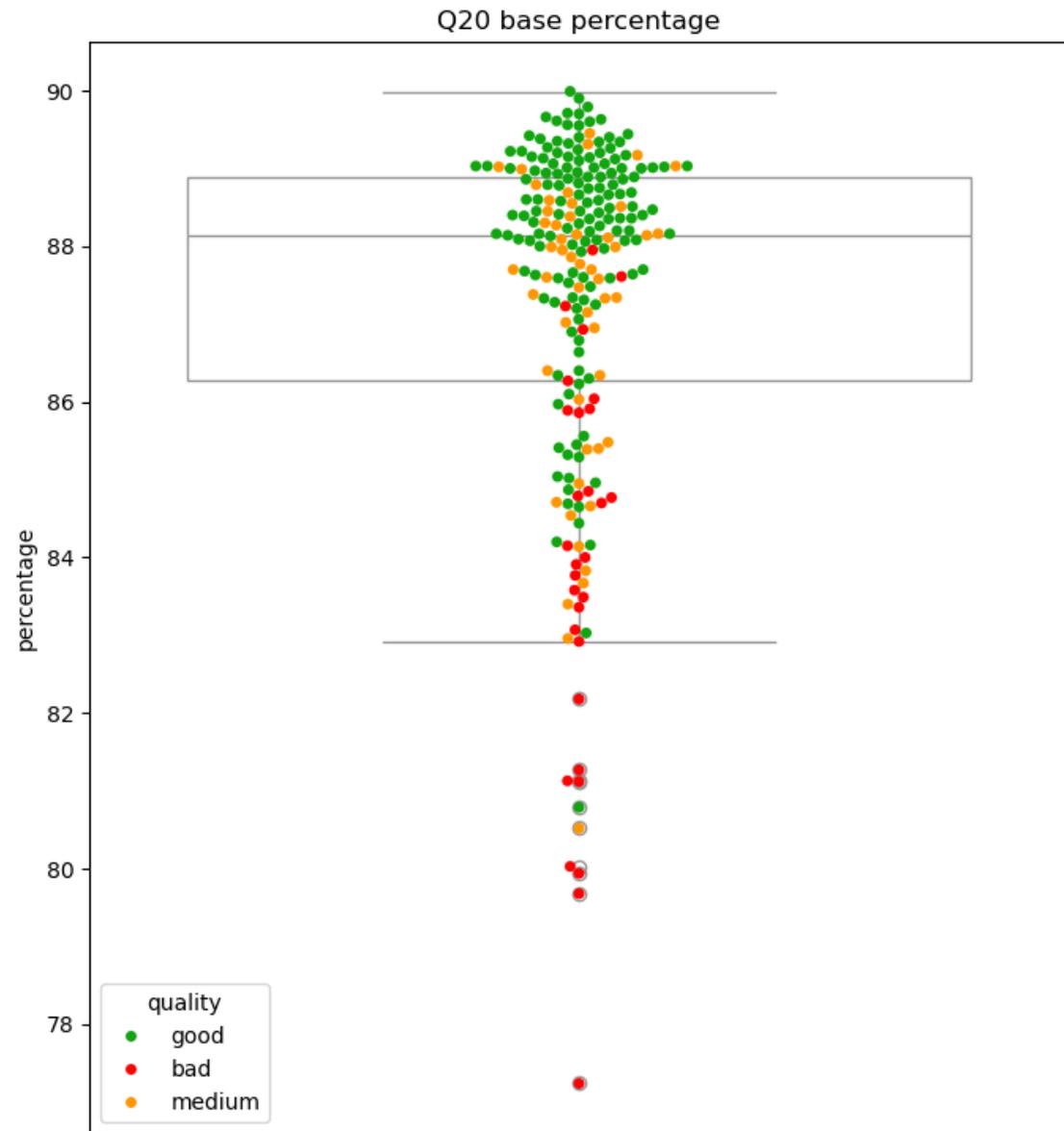
Base Quality



QC-Analysis on 233 diagnostic
Nanopore-GS:

Good quality: 80% of bases > Q20

Note: we remove reads with Read-
Q < 9 before base quality analysis!



Summary: Essential QC Parameter



1. Coverage: $\geq 30x$
2. Saturation: $\geq 20x$ coverage in $> 95\%$ of the genome
3. N50 read length: $>10\text{kb}$ (optimal: 20-25kb)
4. Base quality: 80% of all bases have $Q > 20$
5. Remove reads with Read-Quality < 9



Benchmarking with Genome in a Bottle Reference Data



NIST

PROJECTS/PROGRAMS

Genome in a Bottle

Summary

Consortium hosted by NIST dedicated to authoritative characterization of benchmark human genomes. Sign up for [General GIAB](#) and [Analysis Team](#) email lists. [Public workshops](#) held annually - next workshop will be rescheduled after COVID-19. **Interested in job opportunities with us? Contact Justin Zook at the email in the right panel.**

[Click here for the GIAB FAQ](#)

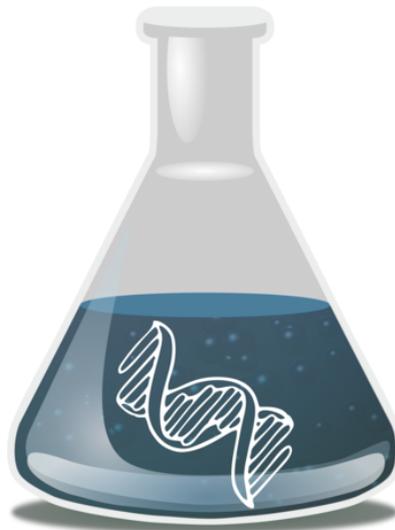
DESCRIPTION

Consortium goals:

The Genome in a Bottle Consortium is a public-private-academic consortium hosted by NIST to develop the technical infrastructure (reference standards, reference methods, and reference data) to enable translation of whole human genome sequencing to clinical practice and innovations in technologies. The priority of GIAB is authoritative characterization of human genomes for use in benchmarking, including analytical validation and technology development, optimization, and demonstration.

Reference samples:

GIAB has currently characterized a pilot genome (NA12878/HG001) from the [HapMap project](#), and two son/father/mother trios of Ashkenazi Jewish and Han Chinese ancestry from the [Personal Genome Project](#) (selected because, unlike the pilot genome, they are consented for commercial redistribution). These samples and their IDs from [NIST](#), [Coriell](#), and [PGP](#) are in [this table](#).



Reference samples:

GIAB has currently characterized a pilot genome (NA12878/HG001) from the [HapMap project](#), and two son/father/mother trios of Ashkenazi Jewish and Han Chinese ancestry from the [Personal Genome Project](#) (selected because, unlike the pilot genome, they are consented for commercial redistribution). These samples are available from [NIST](#) and [Coriell](#), and their IDs from [NIST](#), [Coriell](#), and [PGP](#) are in [this table](#) (see [FAQ](#) for differences between NIST and Coriell).



GiaB Benchmark for Coding Regions

March 2025 – Nanopore long-read GS



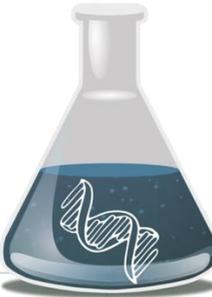
	recall/sensitivity	precision	genotyping accuracy
Exome SNVs	97.25	99.91	99.91
Exome INDELs	91.7	98.18	99.54
Nanopore SNVs	1.000	0.999	1.000
Nanopore INDELs	0.966	0.983	0.998



Structural Variant Calling srGS vs lrGS

All structural variant benchmarks are done on the GIAB reference sample NA24385/HG002 using the [draft SV benchmark v1.1](#). The analyses were performed with the short-read and long-read single sample pipelines.

Sensitivity and positive predictive value (PPV) were measured using [Hap-Eval](#).



Test	coverage	sensitivity	PPV
short-read WGS - Manta 1.6.0	39.5x	36.30%	96.53%
short-read WGS - DRAGEN 4.2.4	39.5x	50.26%	97.64%
long-read WGS (high accuracy) - Sniffles 2.4	40.5x	90.59%	98.03%
long-read WGS (super accuracy) - Sniffles 2.4	40.5x	91.02%	98.05%



The Advantage of Long Reads for Complex Clinical Use Cases



ELRIN* - European Long-Read Initiative for RD



- Form a European network of experts developing Nanopore standards for clinical testing
- Explore the potential of Nanopore sequencing in Rare Diseases and Familial Cancer
- Gather deep control data Structural Variants and Methylation analyses
- Standardize reporting of repeat expansions, duplicate genes, mobile elements, SVs and methylation disorders
- Provide long-read next-generation sequencing data to the European +1Million-Genomes initiative.



Coordinators: Olaf Rieß, Beate Kristmann, Tübingen

**ELRIN Study is supported by ONT*

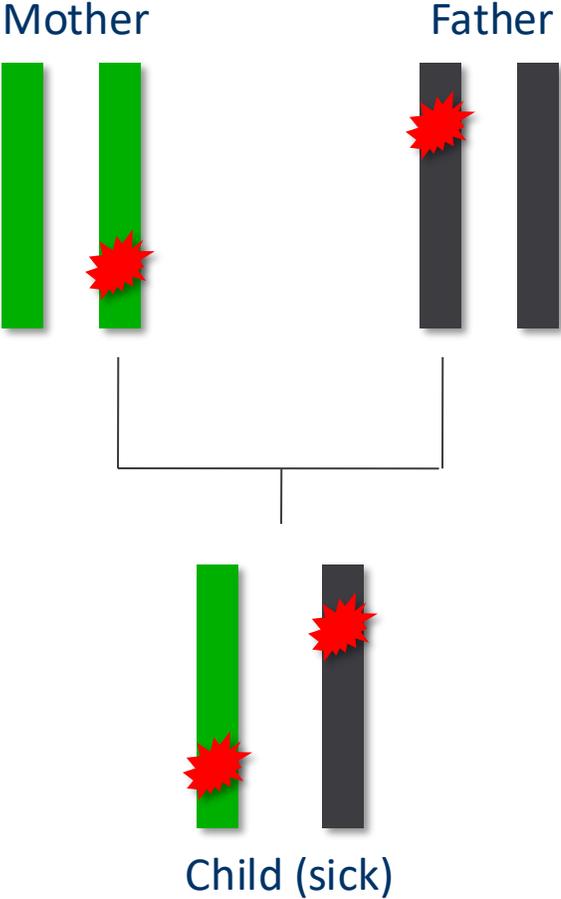


IonGER Consortium Pilot Study

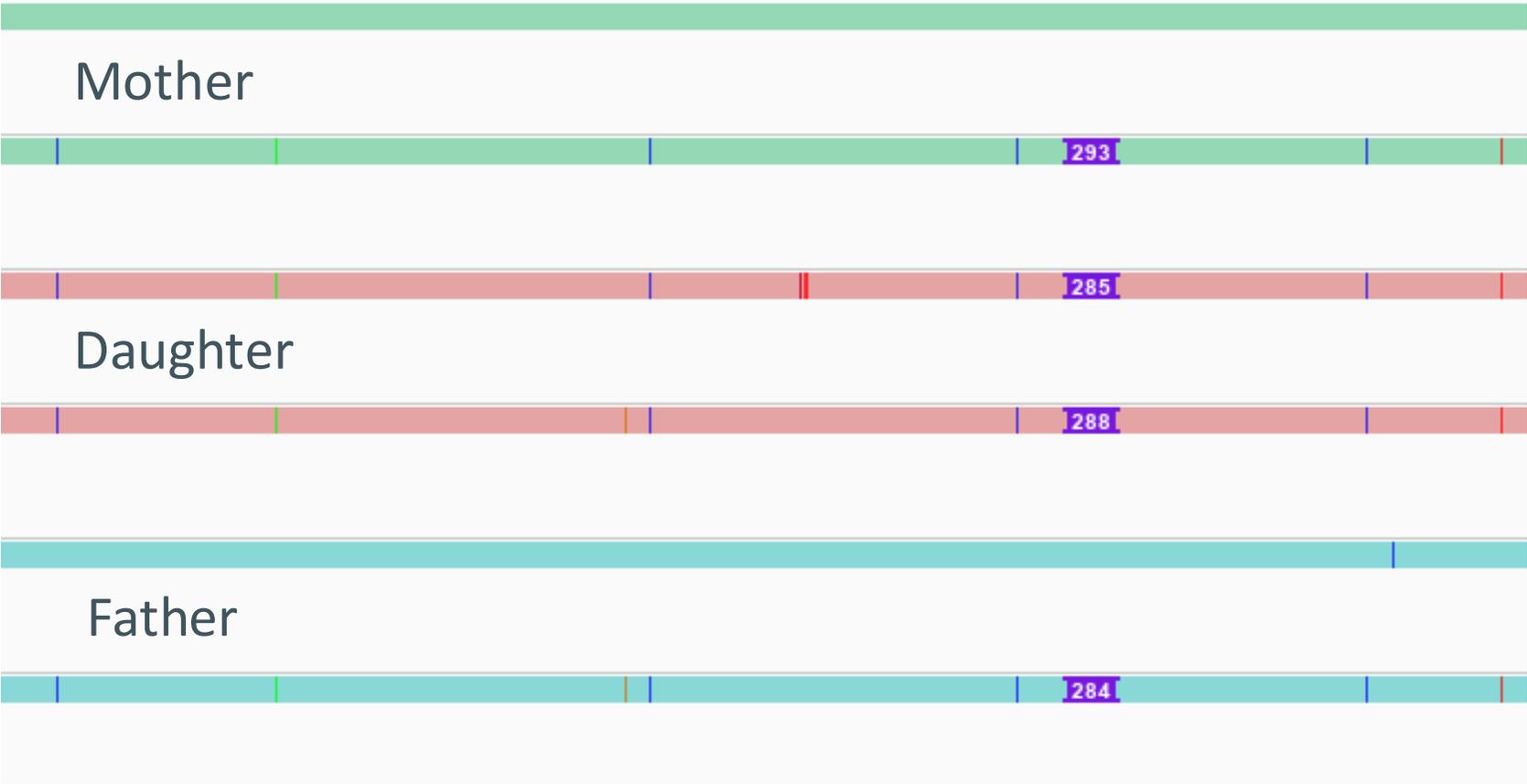
- Benchmarking **2 Genome in a Bottle** cell lines sequenced at each site
- **1000 RD patient** samples (1 FC per genome)
- **Clinical use cases**
 - Compound heterozygotes
 - Complex structural variants and mobile elements
 - Repeat expansions
 - Duplicated genes (genes with pseudogene copy)
 - Pathogenic haplotypes (e.g. OPN1-Cluster)
 - Methylation (imprinting disorders)



Haplotype-Phased Genomes – Compound Hets



Compound Heterozygotes can be resolved sequencing only the index case

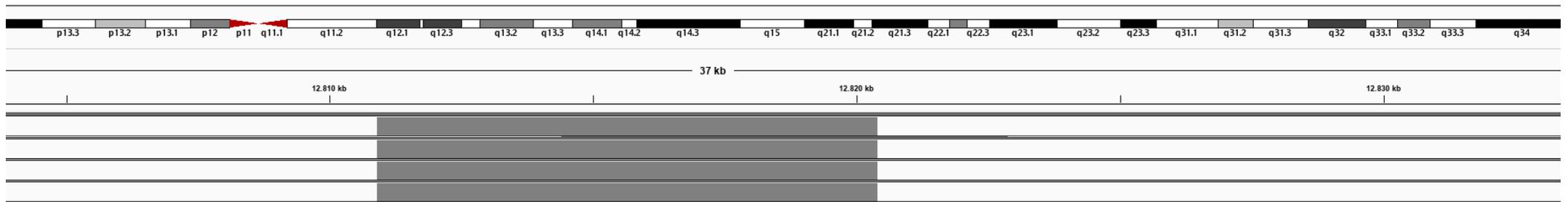


Much Higher Sensitivity for SV Detection

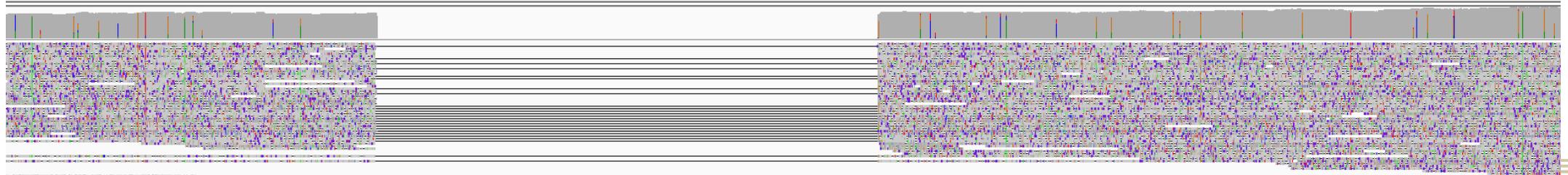


- Much higher sensitivity
- Accurate breakpoints
- True positive SVs per case:
 - Short reads: 10,000
 - Long reads: **23,000**

BioNano*



Nanopore

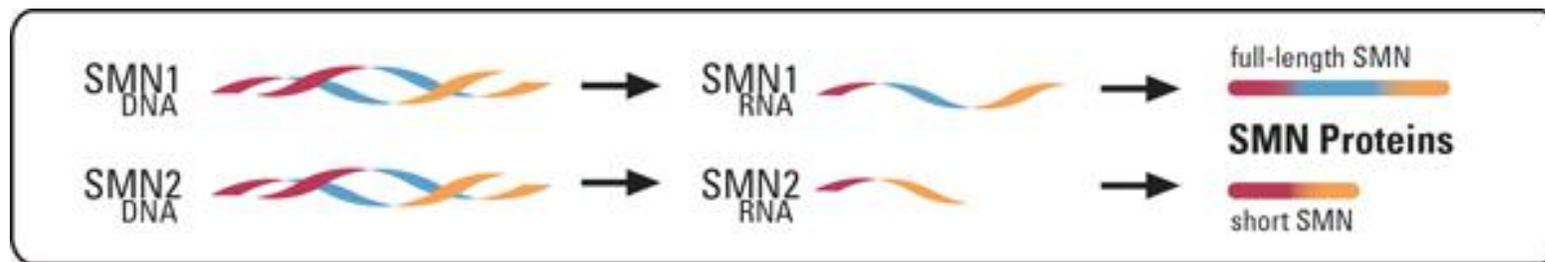
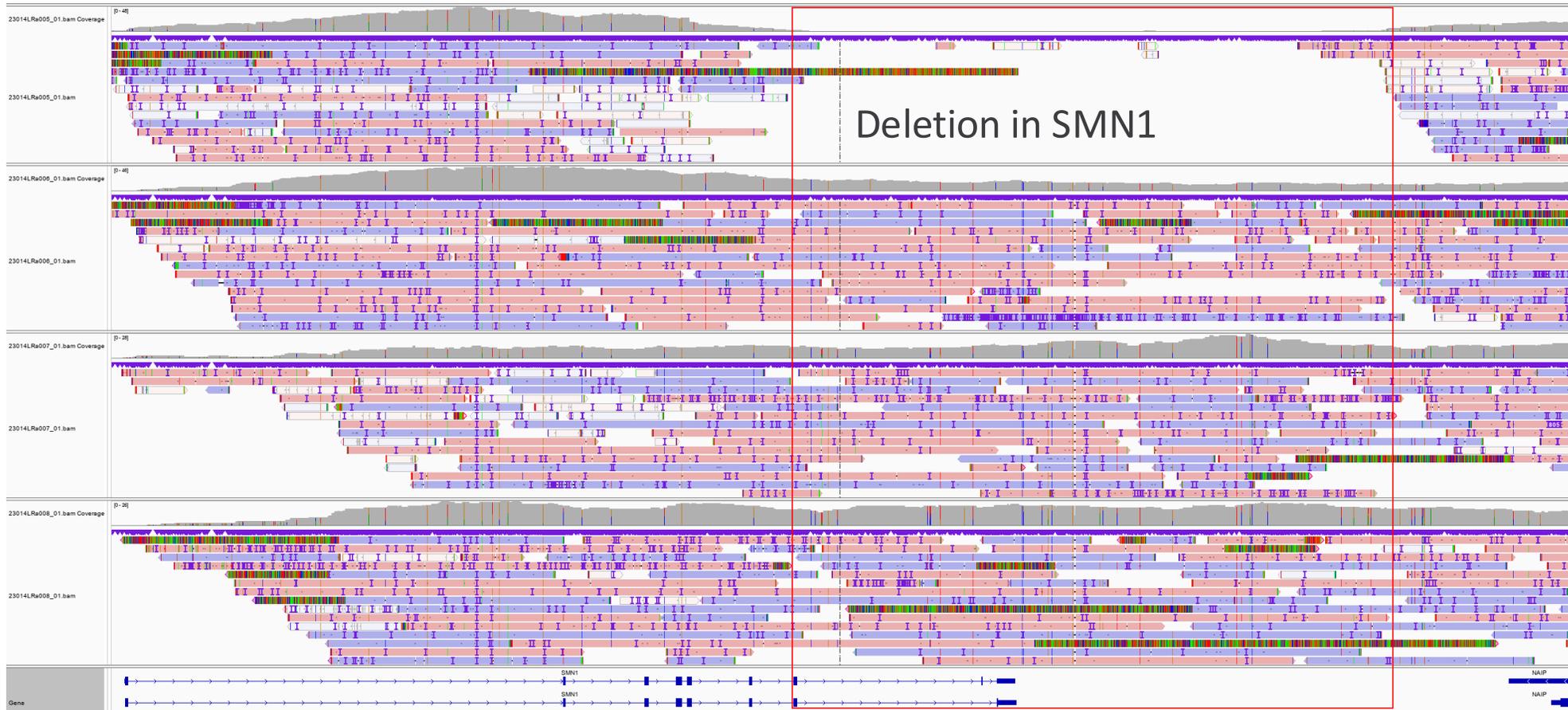




Duplicate Genes and Pathogenic Haplotypes

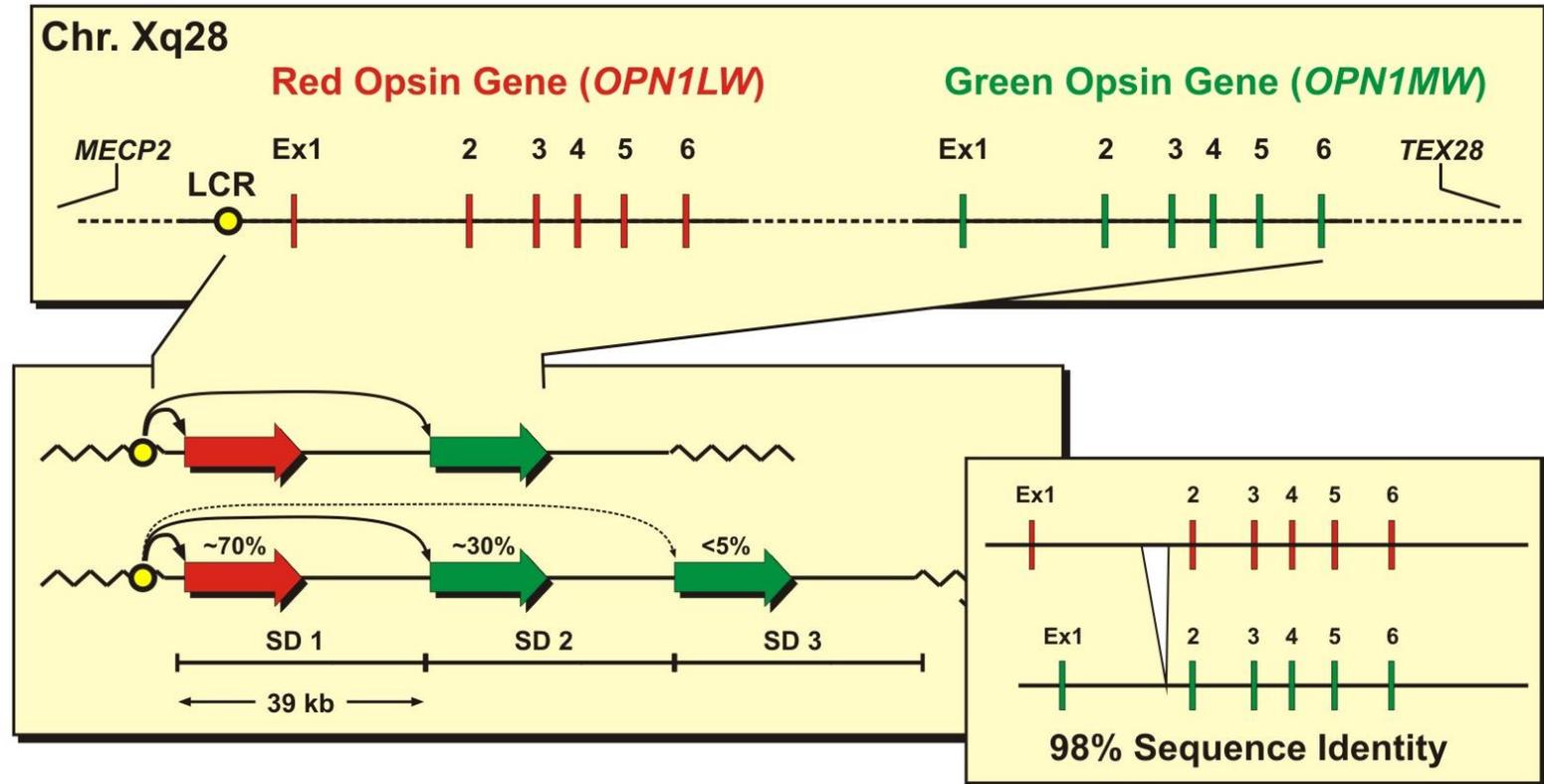


Duplicate Genes: SMN1 and SMN2 in Spinal Muscular Atrophy



Human *Opsin* Gene Cluster (Color Blindness)

- Genes OPN1LW and OPN1MW
- Copy no. variability (n=2-8)
- Expression gradient (only first two copies are relevant)
- High sequence conservation
- 5 pathogenic variants
- Very difficult for diagnostics



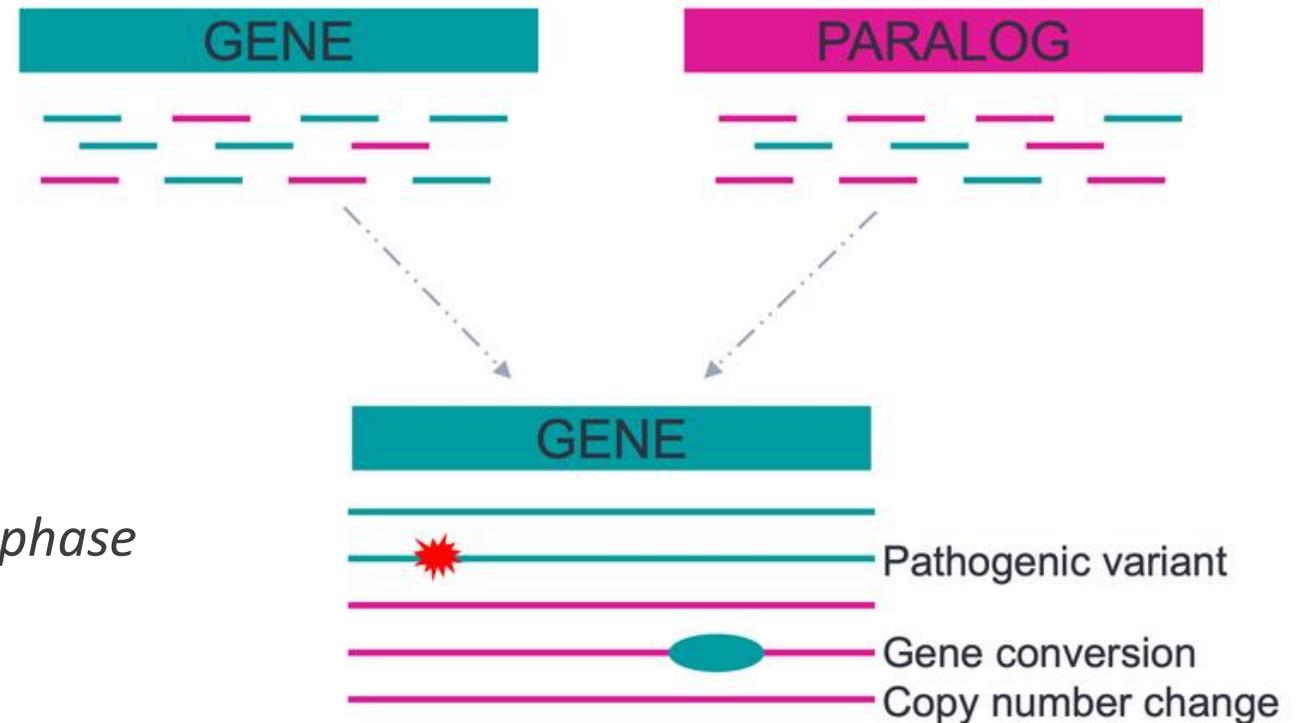
Human *Opsin* Gene Cluster (Color Blindness)

- Characterization of LW/MW gene variants not possible with srWGS
- Nanopore: evaluate copy number, order of copies, hetero-/hemizygous variants, haplotypes

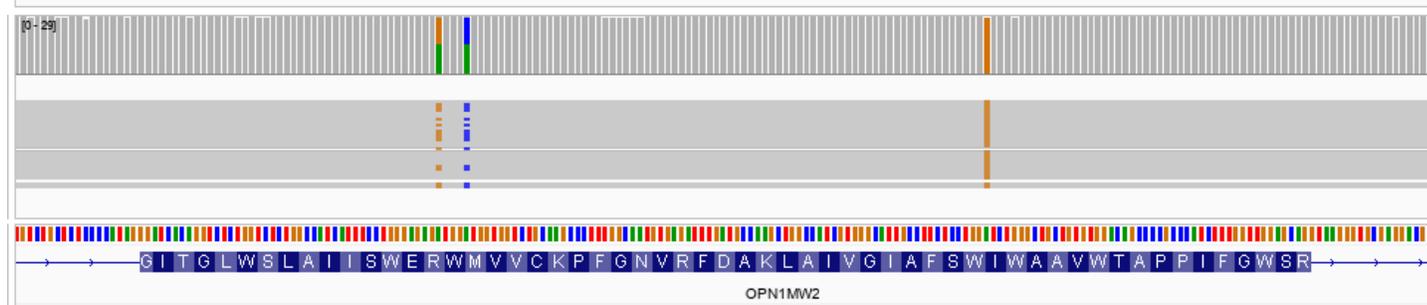
- Diseases: e.g.
 - Blue Cone Monochromacy,
 - X-linked cone dysfunction disorders
 - TODO

- Paraphrase tool for long reads:

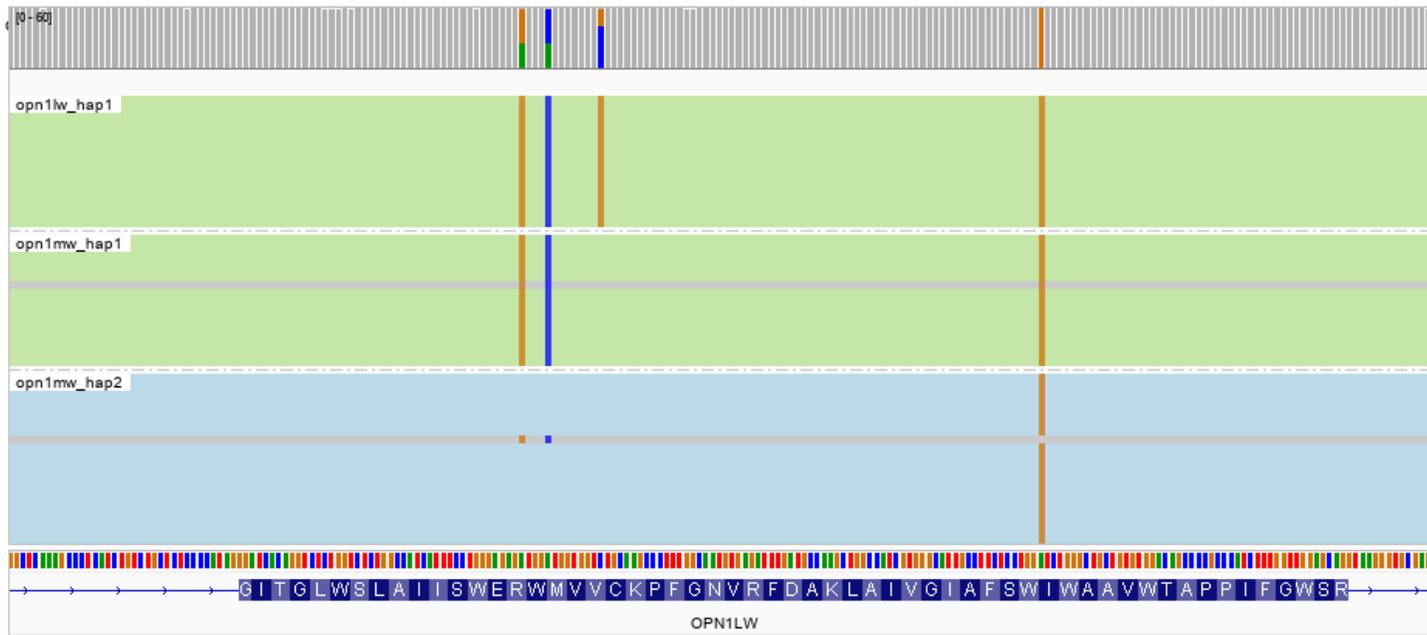
<https://github.com/PacificBiosciences/paraphrase>



Haplotype Phasing: Determine Variants in 3 Gene Copies



Un-phased haplotypes



Phased haplotypes

OPN1LW_hap1:LVAVA^{c.465G}

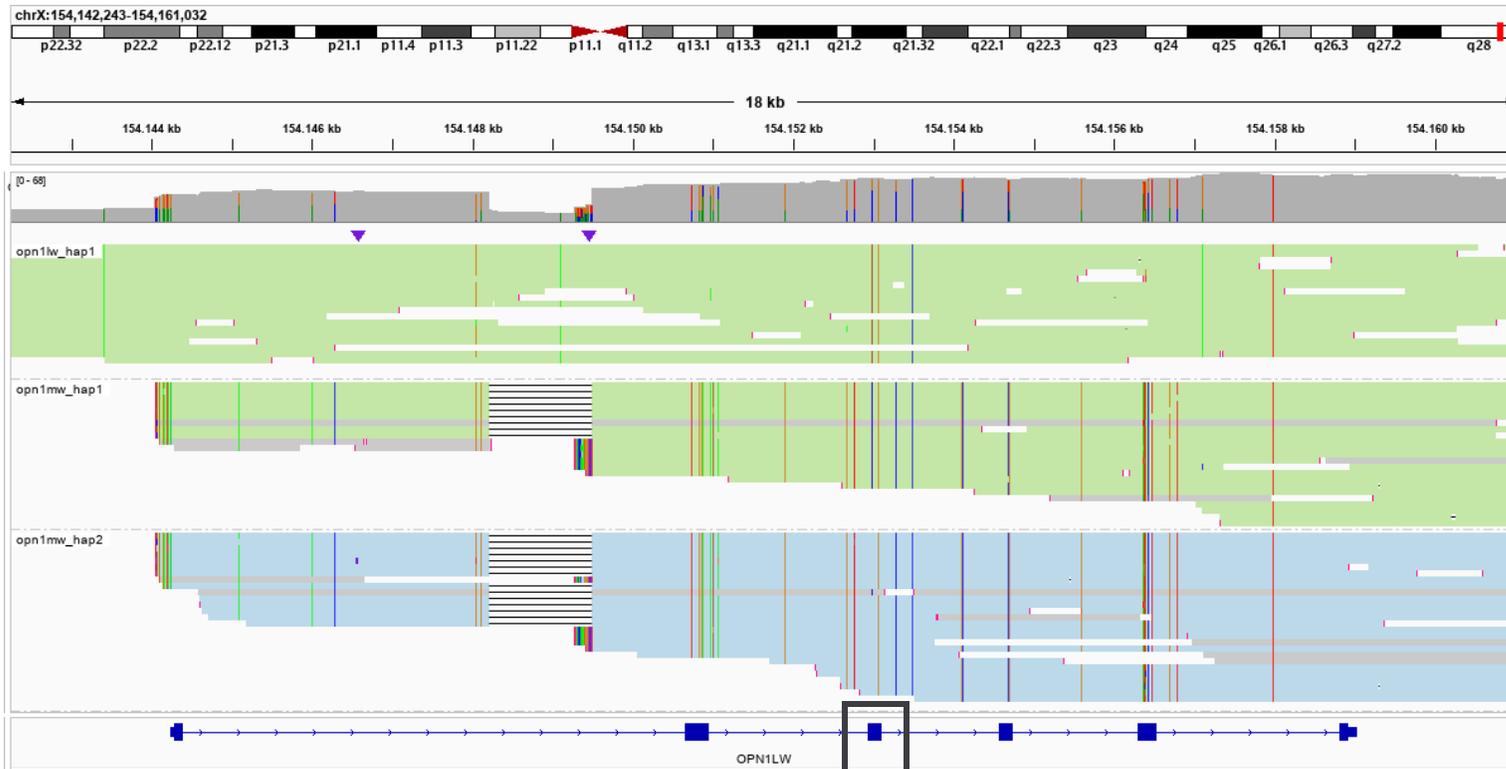
OPN1MW_hap1:LVAVA^{c.465C}

OPN1MW_hap2:MVAVA^{c.465C}



Diagnosis for Patient

Gender	N50 read length (kb)	Coverage WGS	Coverage <i>OPN1</i>	Copy Number	Haplotype 1	Haplotype 2	Haplotype 3	Phenotype	Mutation type
male	19,7	43	80	3	<i>OPN1LW_hap1</i> LVAVA	<i>OPN1MW_hap1</i> LVAVA	<i>OPN1MW_hap2</i> MVAVA	Blue Cone Monochromacy (BCM)	Exon 3 splicing-deficient Haplotypes



Exon 3

OPN1LW_hap1 LVAVA
OPN1MW_hap1 LVAVA
OPN1MW_hap2 MVAVA

Why is important?

- 2 first copies expressed
 - LVAVA and MVAVA have different pathogenicity (5% vs 50% correctly spliced transcripts)
- (Buena-Atienza et al., 2016; Neitz et al., 2021)

Collaborators: Caspar Gross, Elena Buena-Atienza, Bernd Wissinger

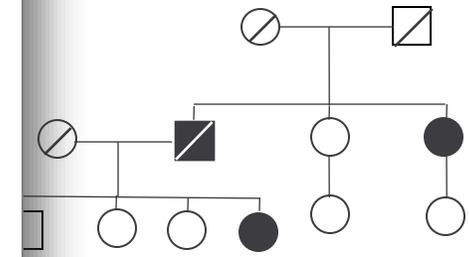
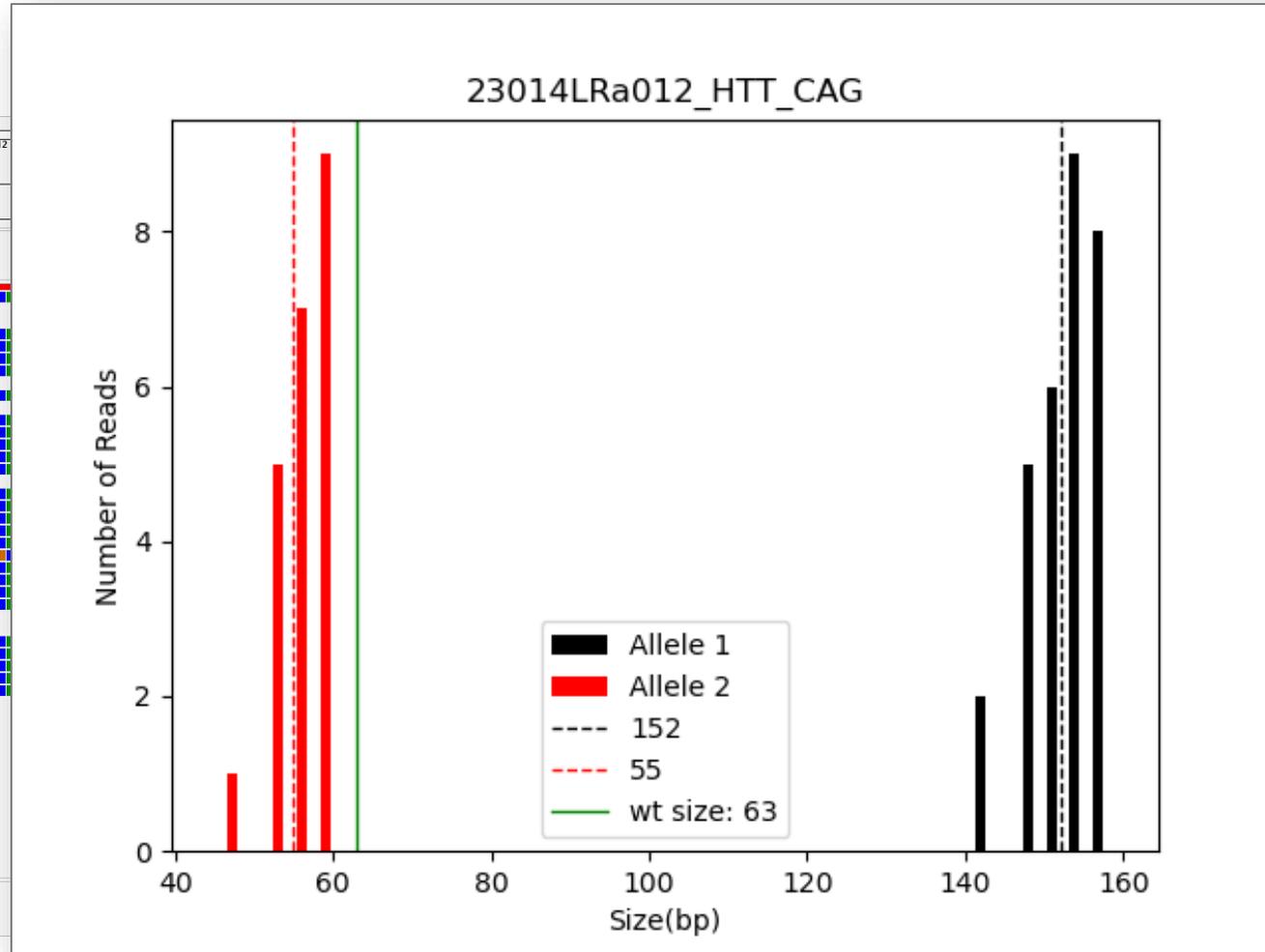




Repeat Expansion Diseases



Example: Huntington Disease (Tübingen Case 15)



n: Huntington disease

kinetic movement disorder
 disturbance
 fine motor coordination

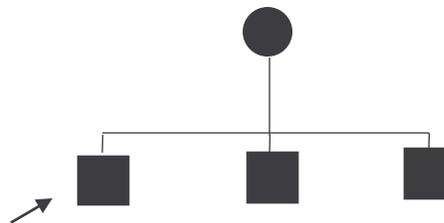
Extended allele: 53 CAG Repeats in HTT



Example: Ataxia (SCA3, Tübingen Case 16)

Indication: SCA 3

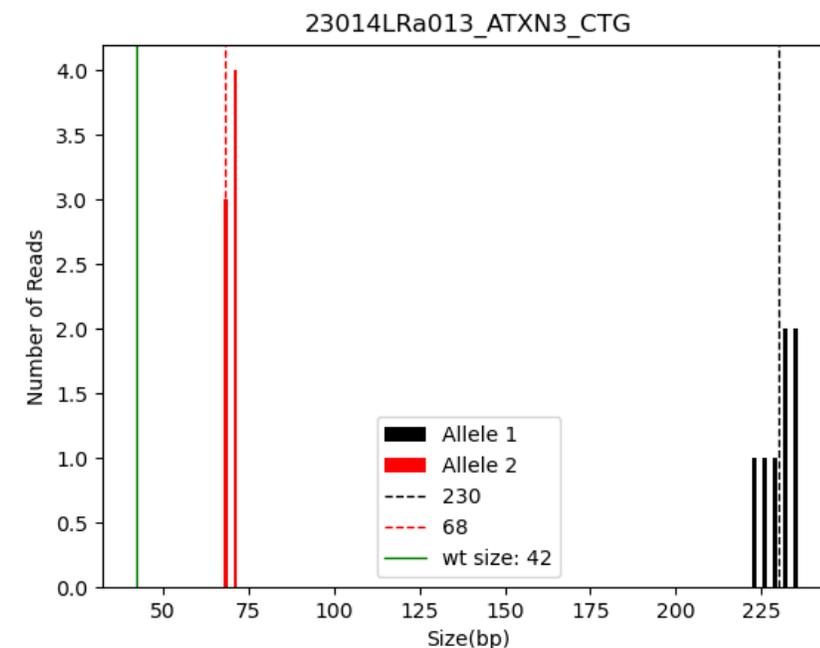
- ataxia
- dysarthria
- dysphagia
- cognitive impairment
- ophthalmoparesis
- parkinsonism



Extended allele: 77 CAG Repeats in ATXN3

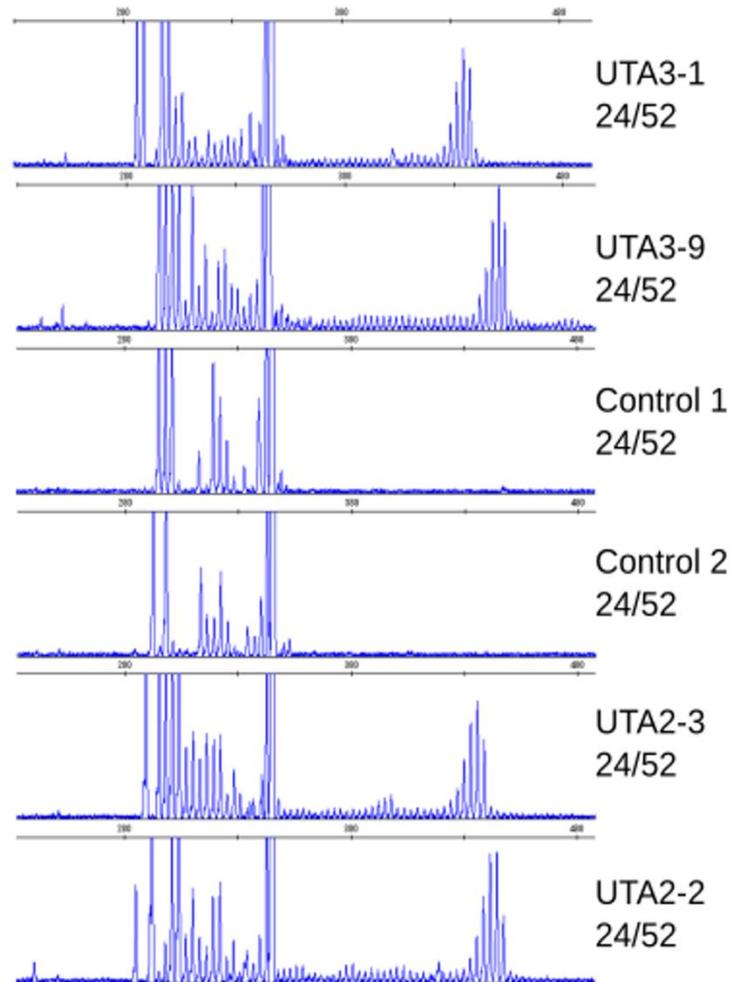
ONT Long Read Analysis

- ATNX3: (CAG)₂₃ / (CAG)₇₇

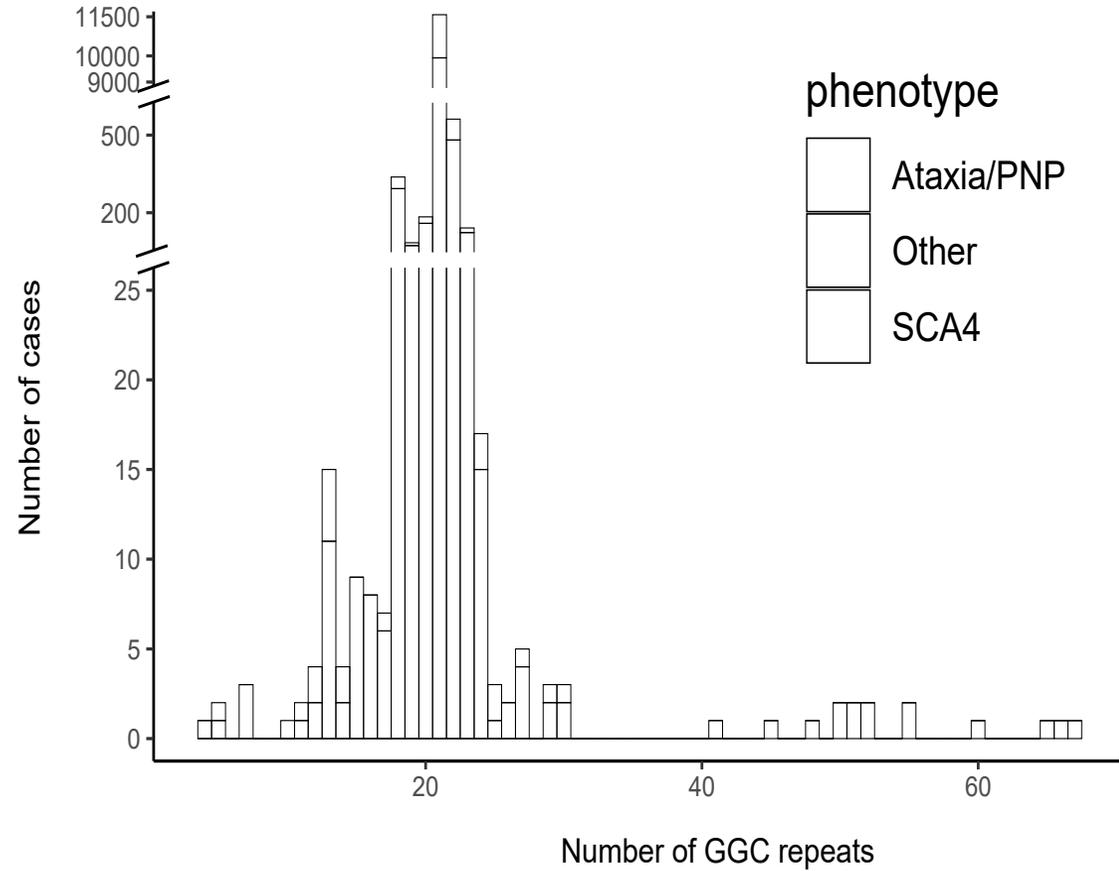


SCA4: Novel Pathogenic Repeat Expansion in ZFH3 Discovered

Validation (PCR amplification)



Pathogenic repeat length

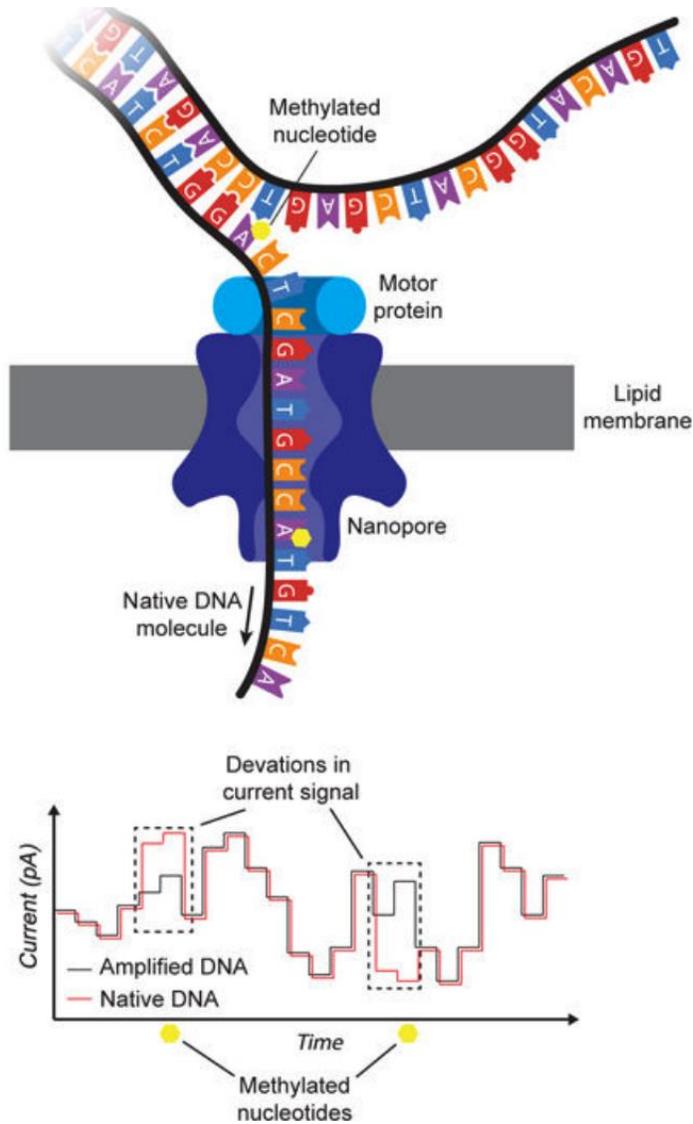




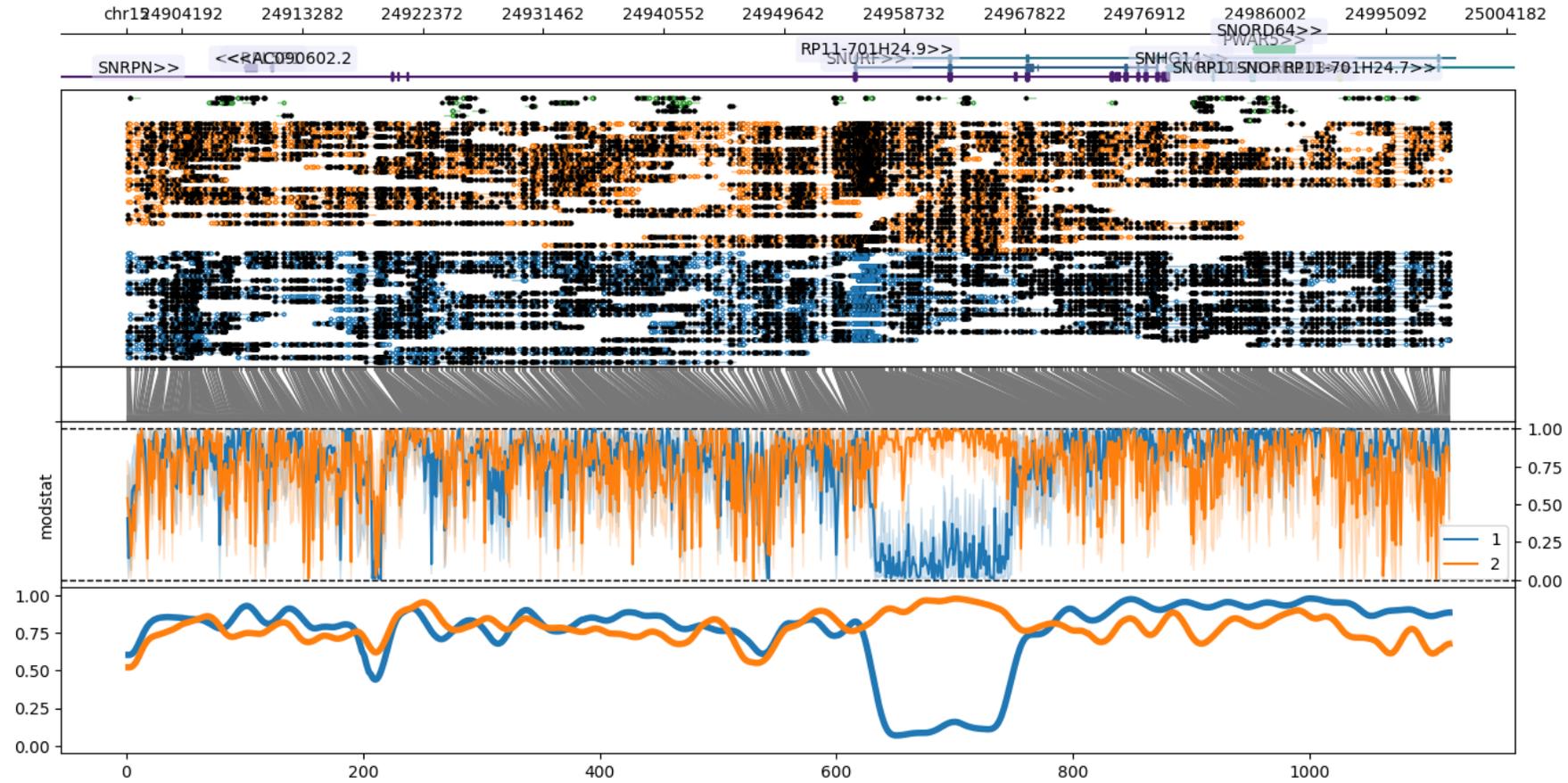
Methylation: Imprinting Diseases



Haplotype-Phased DNA Methylation Calling



Plot: MethylArtist



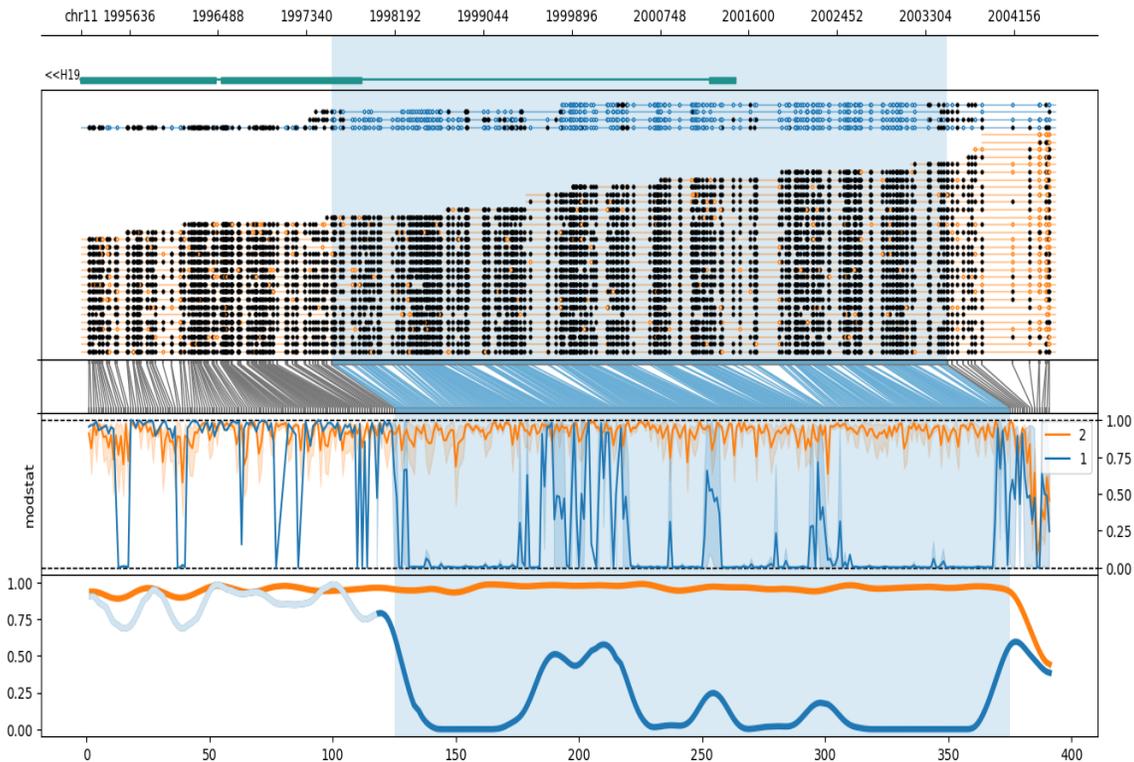
Imprinting Locus



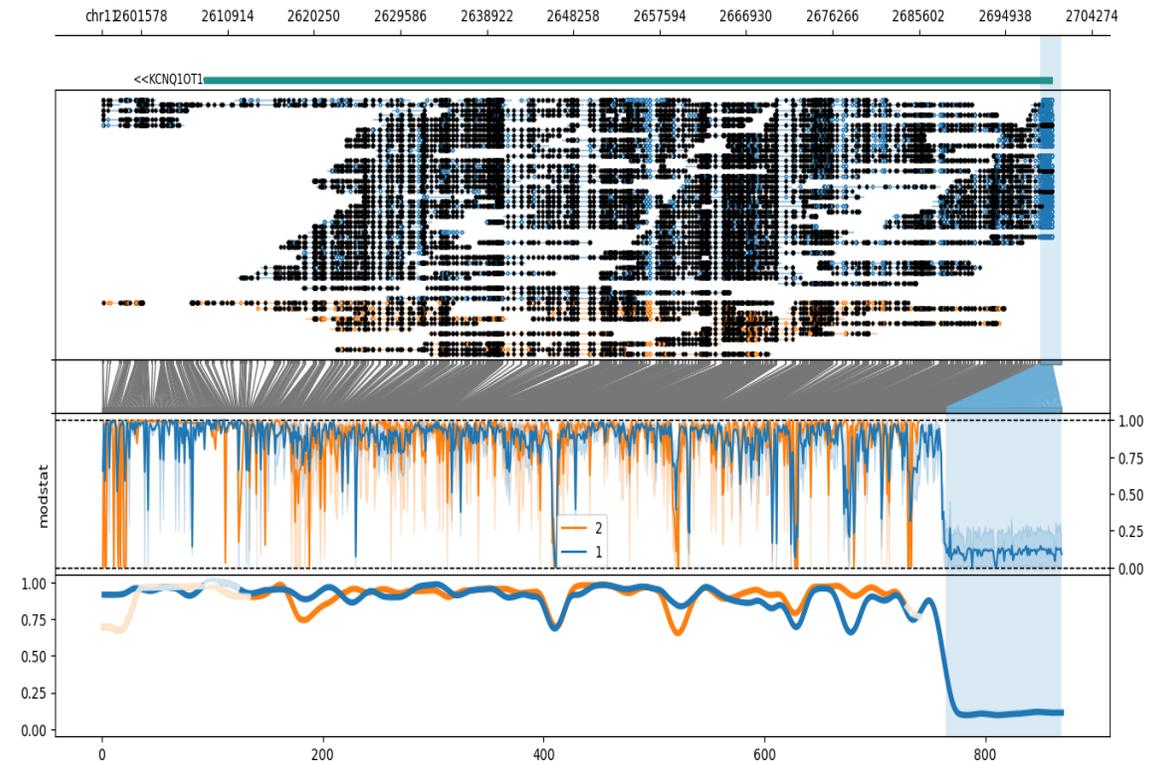
Imprinting Defects: Detecting Anomalies in Haplotype-Phased Methylation

- Beckwith-Wiedemann syndrome (BWS) caused by Uniparental Disomy 11 (UPD11)
- Genomic imprinting at genes H19 and KCNQ1OT1 – only one parental allele should be expressed

H19 – Gain of methylation



KCNQ1OT1 – Loss of methylation

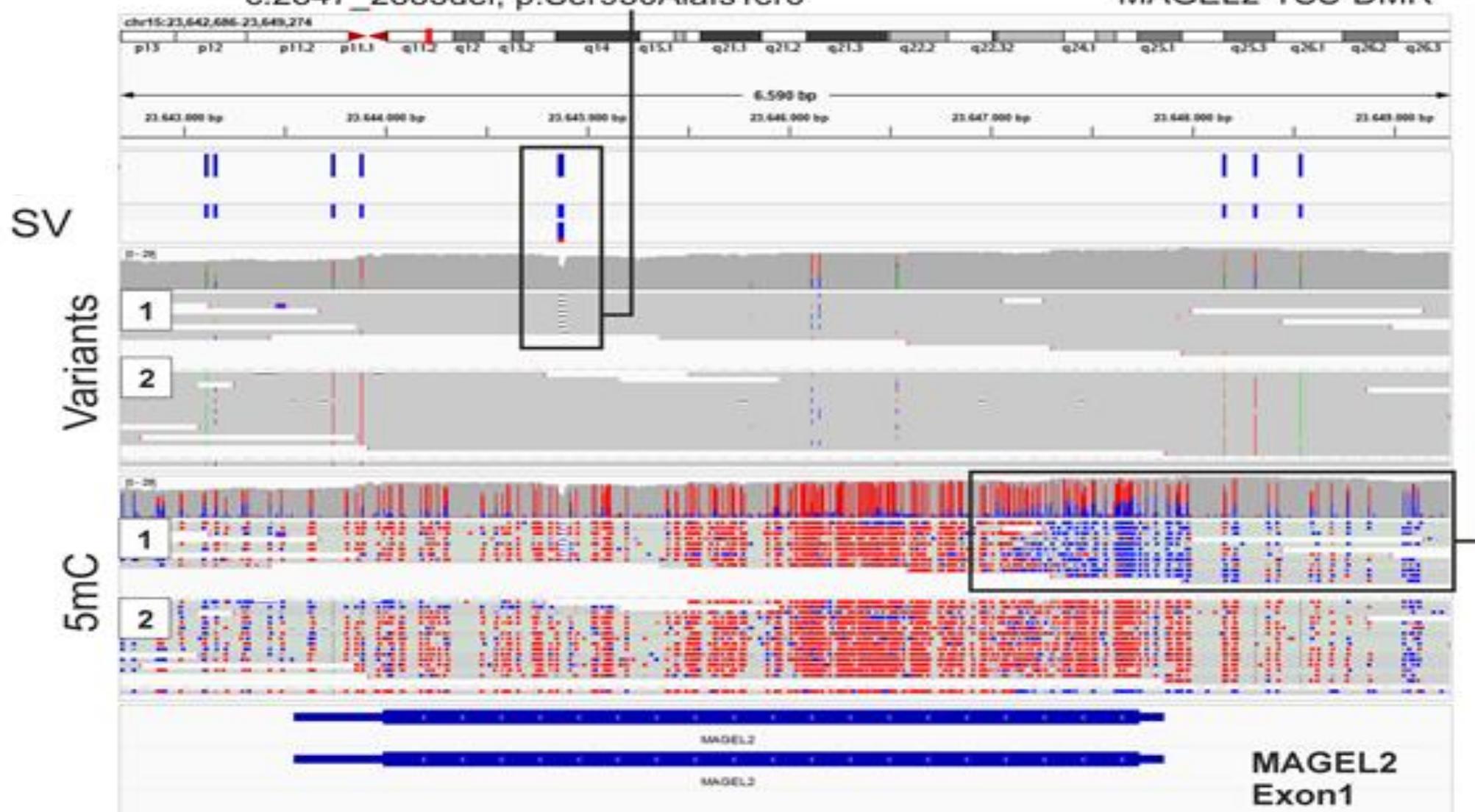


Compound Het: Methylation + Frameshift Indel

Schaaf-Yang Syndrome: **MAGEL2 Exon 1**

c.2847_2883del; p.Ser950AlafsTer6

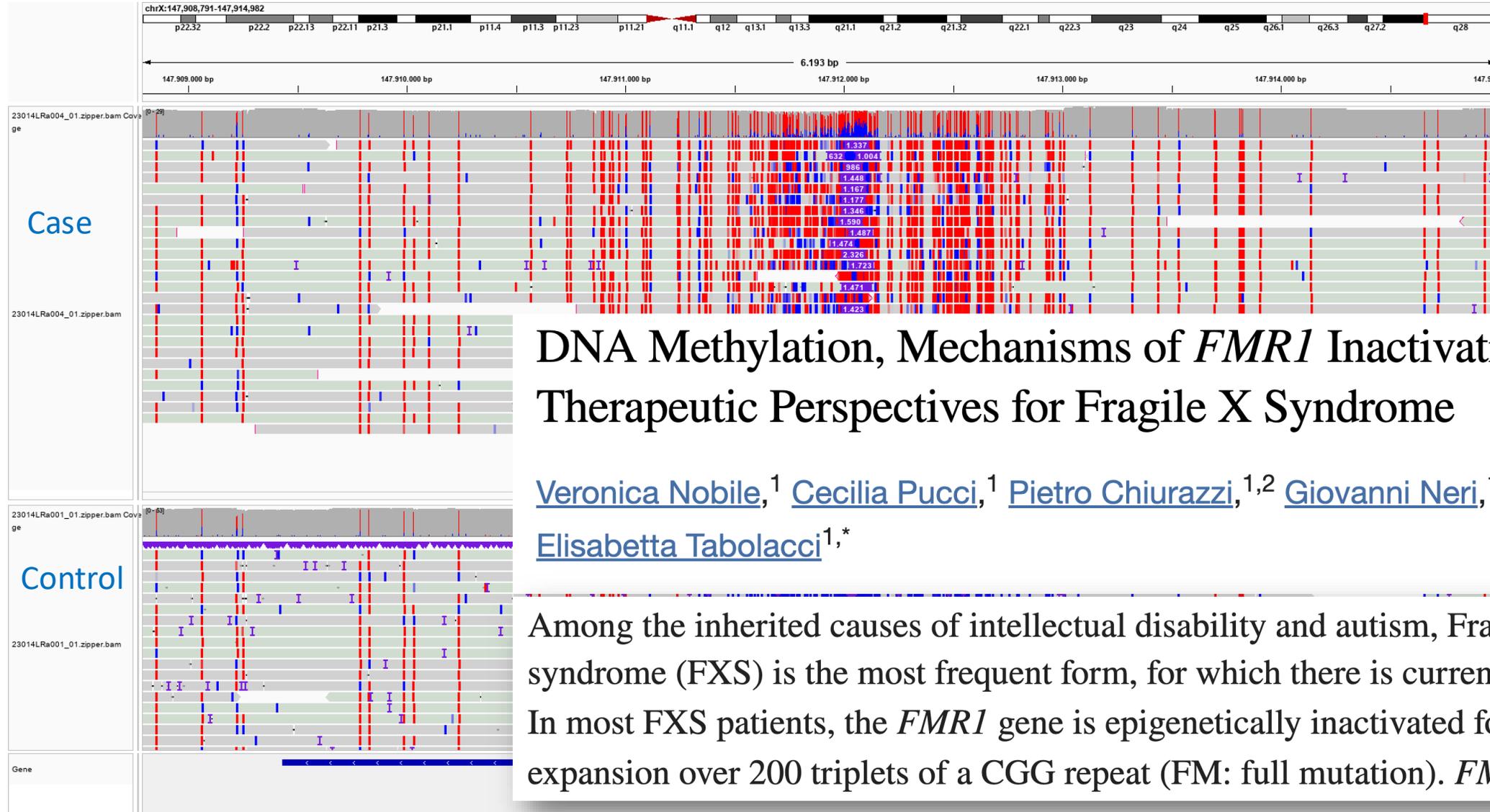
Hypomethylated allele
MAGEL2-TSS-DMR



Fragile X Syndrome: FMR1 Inactivation by Methylation



CGG expansion & Hypermethylation in FMR1 (X chromosome, male patient),

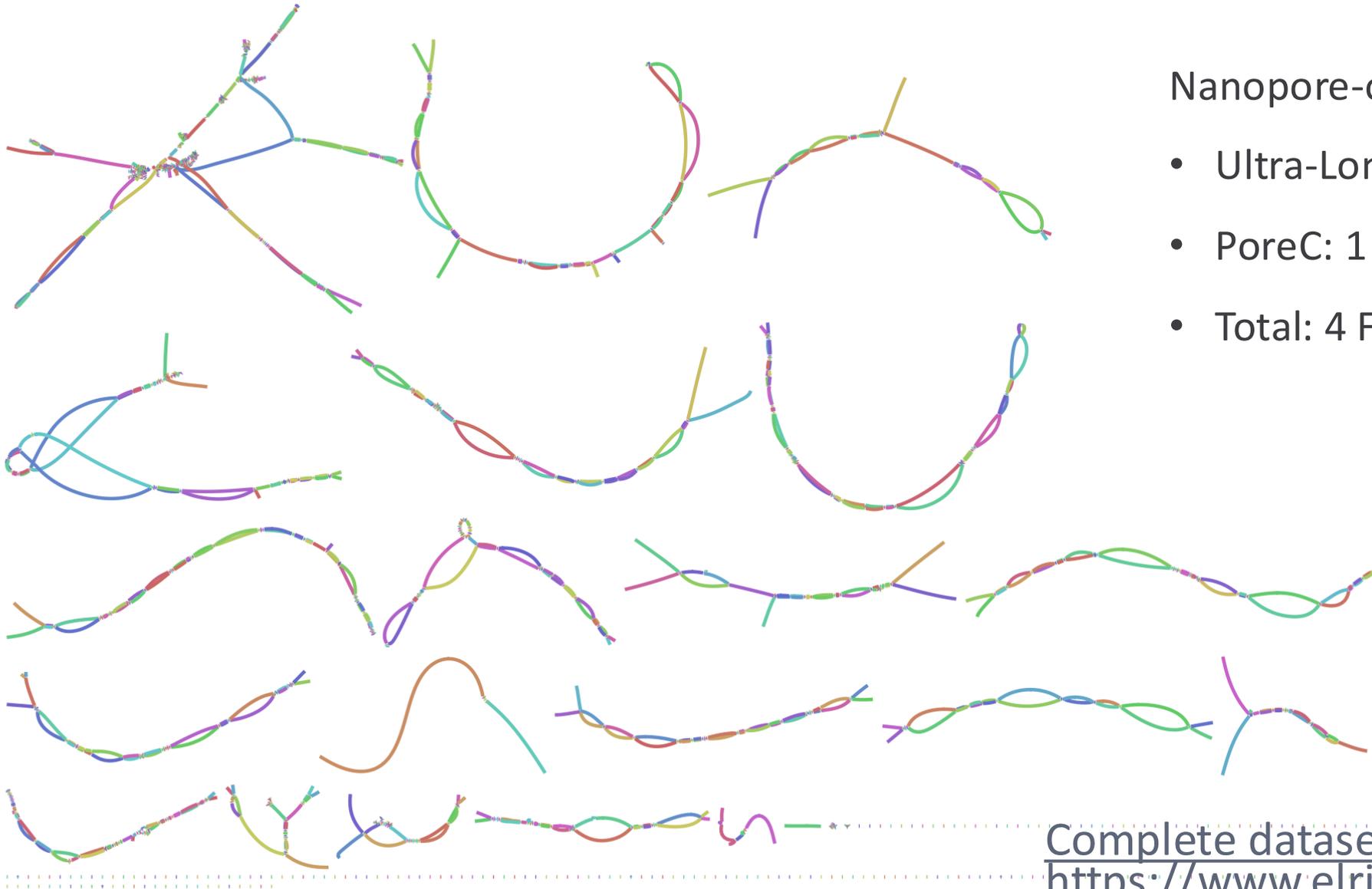


DNA Methylation, Mechanisms of *FMR1* Inactivation and Therapeutic Perspectives for Fragile X Syndrome

[Veronica Nobile](#),¹ [Cecilia Pucci](#),¹ [Pietro Chiurazzi](#),^{1,2} [Giovanni Neri](#),^{1,3} and [Elisabetta Tabolacci](#)^{1,*}

Among the inherited causes of intellectual disability and autism, Fragile X syndrome (FXS) is the most frequent form, for which there is currently no cure. In most FXS patients, the *FMR1* gene is epigenetically inactivated following the expansion over 200 triplets of a CGG repeat (FM: full mutation). *FMR1* encodes

Outlook: Telomere-Telomere Genomes



Nanopore-only T2T with:

- Ultra-Long Reads of 100kb: 3 FC
- PoreC: 1 FC
- Total: 4 Flowcells

Complete dataset:
<https://www.elrin-network.eu/>

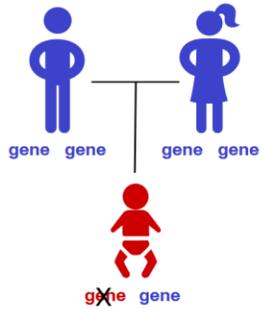




AI-assisted Diagnostics



The Problem: Finding the Causal Variant in Genetic Disease



Pathogenicity
Classification

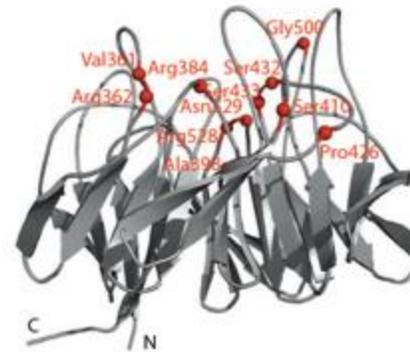
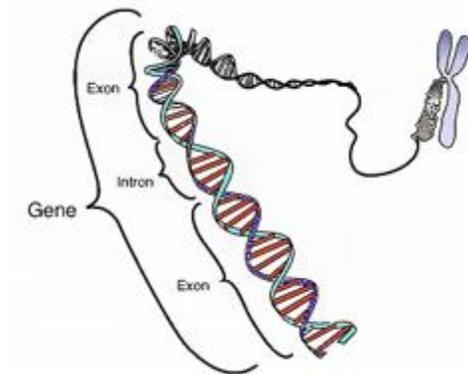
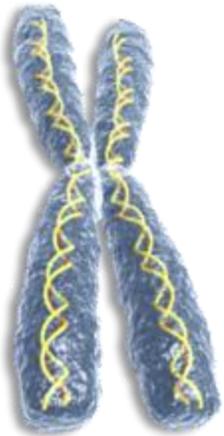
Causal Variant
Prioritization

Genome:
~4M SNVs

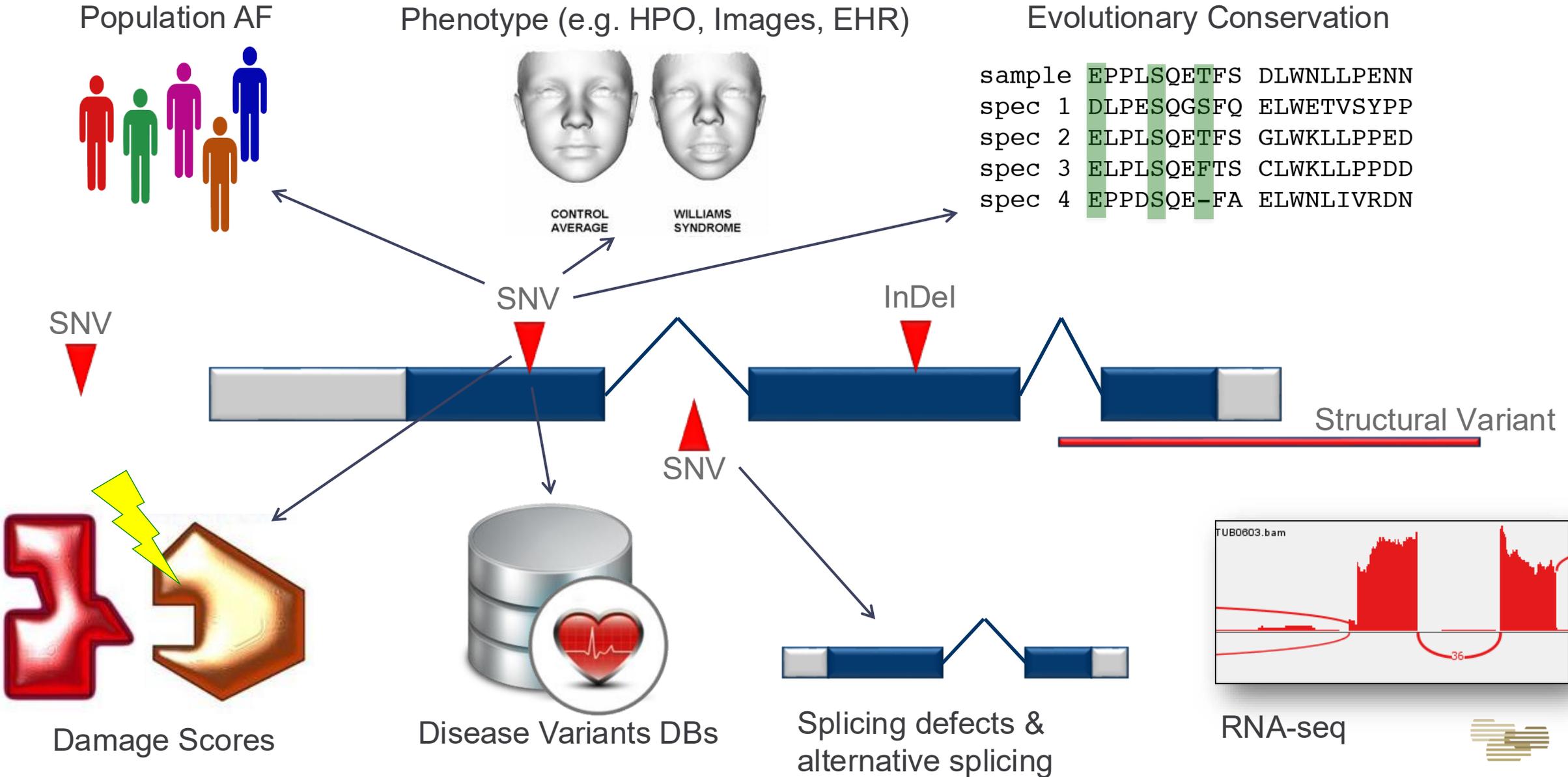
Exome:
~80k SNVs

**Alters
Protein:**
~15k SNVs

**Causal
variant:**
1 SNV

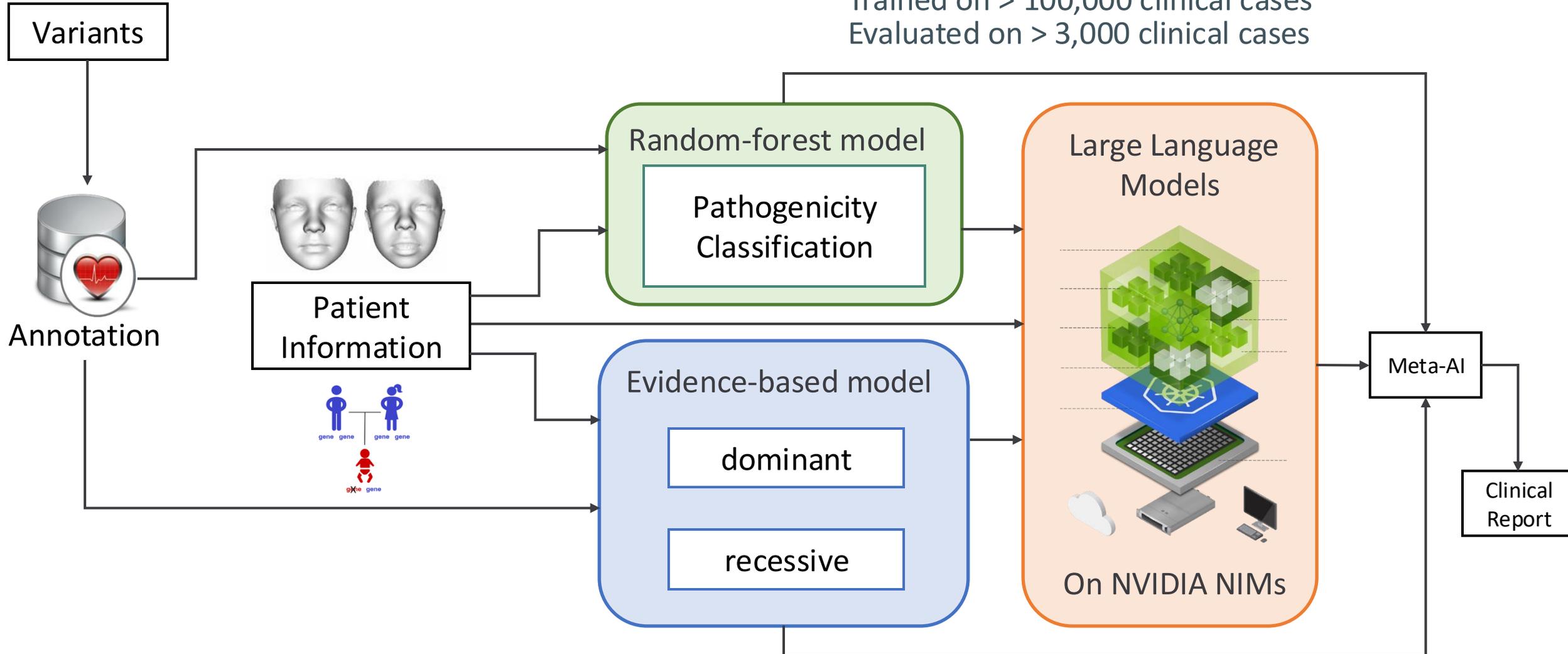


Annotation Features for AI Diagnostics

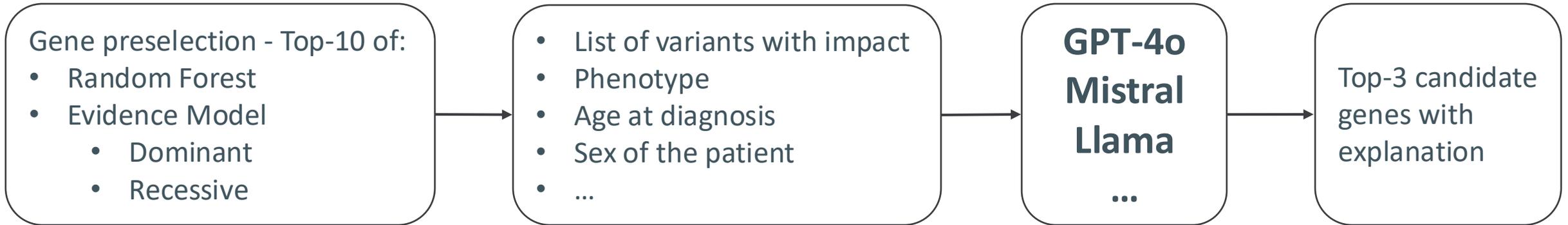


Automated Diagnostics by AI: aiDIVA ensemble-AI

Trained on > 100,000 clinical cases
Evaluated on > 3,000 clinical cases



Improving Causal Variant Prioritization with Large Language Models (LLMs)



Example Prompt:

"A female rare disease patient of age 0 has the following symptoms: Facial hemangioma, Status epilepticus. A causal variant in which of the following candidate genes would best explain these symptoms?
Candidate genes:
PIEZO1, SEC23B, EFTUD2, CDH11, AASS, KIAA0753, SCN1A, ABCC8, VPS51, NDUFA9". Please provide sources.



Example LLM “Diagnosis”



4 Websites durchsucht

For a newborn female with symptoms of facial hemangioma and status epilepticus, a mutation in the **SCN1A** gene would be the most likely causal variant.

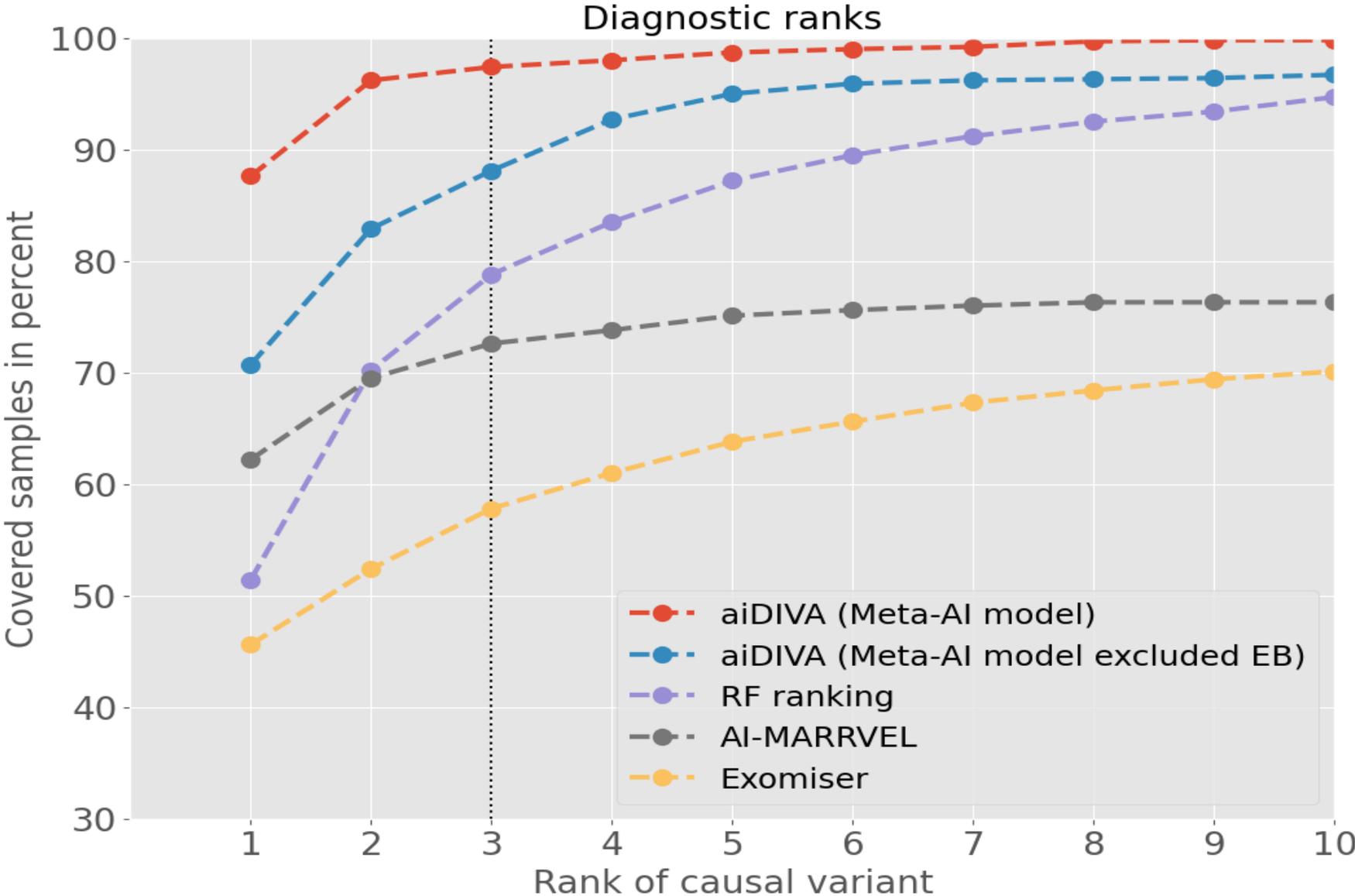
SCN1A is well known for its association with severe epilepsy syndromes, particularly Dravet syndrome, which often begins in infancy and includes prolonged seizures (status epilepticus). In addition to seizures, Dravet syndrome and other SCN1A-related epilepsies can be associated with various neurological and systemic manifestations, potentially including vascular anomalies such as facial hemangiomas ([National Organization for Rare Disorders](#)) ([Epilepsy Foundation](#)) ([Oxford Academic](#)).

The SCN1A gene encodes a subunit of the sodium channel NaV1.1, which is crucial for maintaining the proper electrical activity in neurons. Mutations in this gene can lead to disrupted neuronal function, resulting in severe epileptic conditions and other associated symptoms.

For more detailed information, you can refer to sources like the Epilepsy Foundation and academic journals on the genetic underpinnings of SCN1A-related disorders. These provide comprehensive insights into the gene's role in epilepsy and related conditions ([National Organization for Rare Disorders](#)) ([Epilepsy Foundation](#)) ([Oxford Academic](#)).



aiDIVA Finds Close to 100% Correct Diagnoses in Top-3 Ranks



Evaluated on 3,000 solved clinical cases of UKT



Method Development & Nanopore Implementation



Elena Buena-Atienza
Nanopore Sequencing



Marc Sturm
Lead Diagnostic
Bioinformatics



Dominic Bocek
AI Development



Caspar Gross
Long-read bioinformatics



Alexandr Chernov
Cloud Platform



Alexander Ott
Cancer Decision
Support System



Jakob Admard
NGS Method development



German Demidov
Structural Variants,
Statistics



**Victoria Paz Cepeda
Espinoza**
Pathogen Diagnostics



Leon Schütz
Decision Support
Systems for RD
and Nanopore



Vincent Hammer
Single Cell & Spatial
Transcriptomics



Andreas Schmidt
Association Tests



Vladislav Lysenkov
Nanopore Analysis



Ramya Potabattula
T2T, Ultra-Long Reads



Fubo Cheng
T2T, PoreC



Chia Ying Ko
Repeat Expansions



Prithivi Jung Thapa
T2T Assembly



Thanks!



Olaf Riess



Tobias Haack

Medizinische Hochschule Hannover

Bernd Auber
Gunnar Schmidt

Charité – Universitätsmedizin Berlin

Nadja Ehmke

Berlin Institute of Health

Janine Altmüller
Manuel Holtgrewe
Claudia Quedenau

University Hospital Aachen

Ingo Kurth
Florian Kraft
Sebastian Gießelmann

SCA4 Project

Malte Spielmann (U. of Lübeck and Kiel)
Karla P. Figueroa (U. of Utah)
Daniel R. Scoles (U. of Utah)
Stefan M. Pulst (U. of Utah)

University of Tübingen

Elena Buena-Atienza
Caspar Gross
Leon Schütz
Marc Sturm
Jakob Admard
Vladislav Lysenkov
German Demidov
Chia Ying Ko
Thomas Braun
Nicolas Casadei
Tobias Haack
Olaf Riess

Oxford Nanopore Technologies

Cerissa French
Tonya McSherry
Alexander Vogel
Daniel Mathow
Nicolai Stoye
Anthony Doran
Alexander Rotmann
Manuela Saathoff
Gerald Goh

NVIDIA

Uwe Samer
Harry Clifford
Lotfi Slim

