GHGA 2024

Rare Disease Diagnostics with Long Leads

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Stephan Ossowski
Institute of Medical Genetics and Applied Genomics

Universitätsklinikum
Tübingen

Goal: Rare Disease Diagnostics with Nanopore

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Whole Genome Sequencing

Samples

Family

Index

Parent-Child-Trio

Storage

Alignment

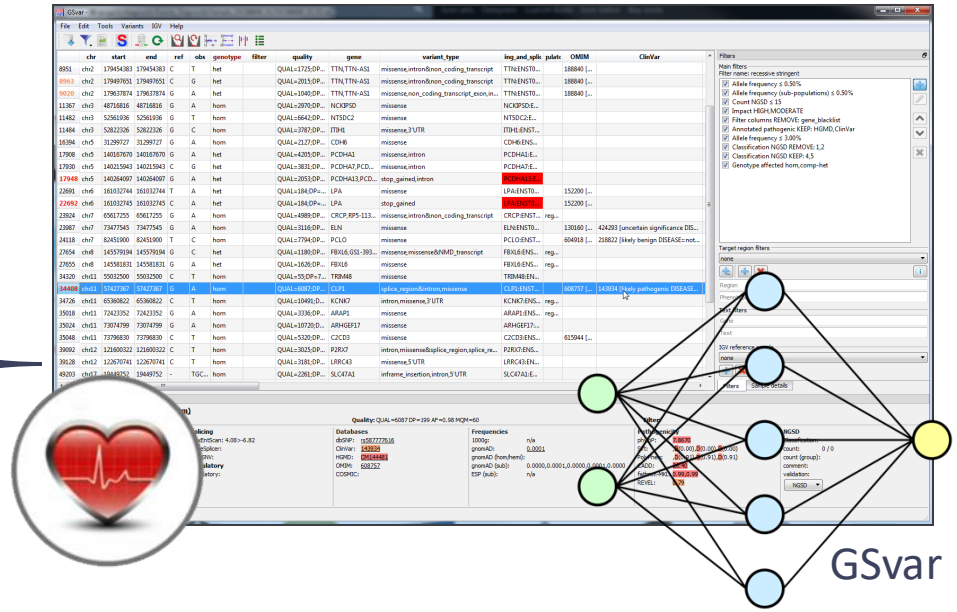
Variant Prediction

Functional Annotation

Clinical Interpretation

Variant Prioritization

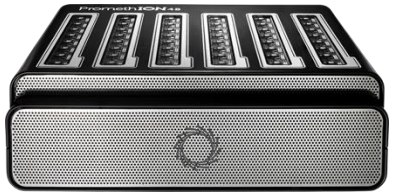
Clinical Decision Support Systems,
Clinical Databases, EHR-Textmining, AI



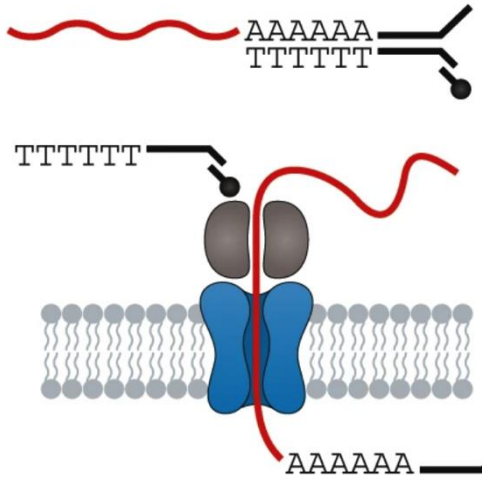
GSvar

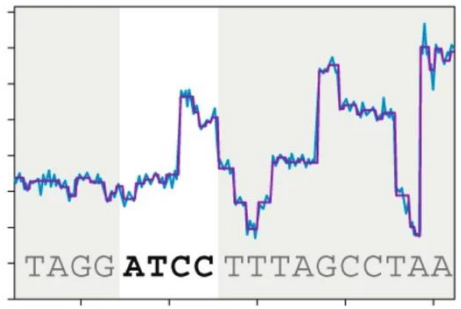
Genome & Methylome Diagnostics with Long Reads

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Oxford Nanopore



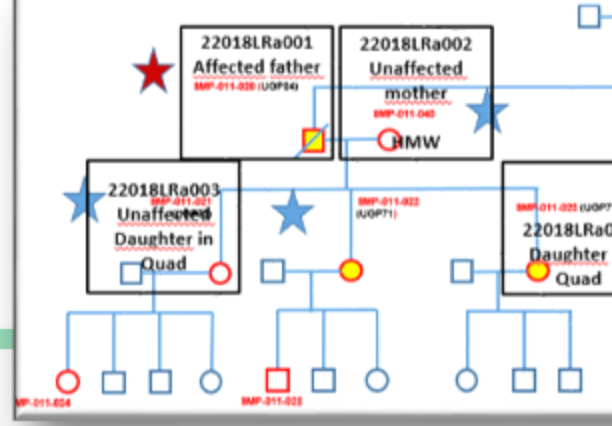


New analysis options:

- **Nanopore Reads**
  - **Long-read alignment**
  - **Genom *de novo* Assembly**
    - **Whole Genome Alignment**
  - **CpG Methylation Calling**
    - **Methylation Phasing**
      - **Methylation Frequency per Site and Allele**
  - **Haplotype Phasing**
    - **Variants in all genomic regions**

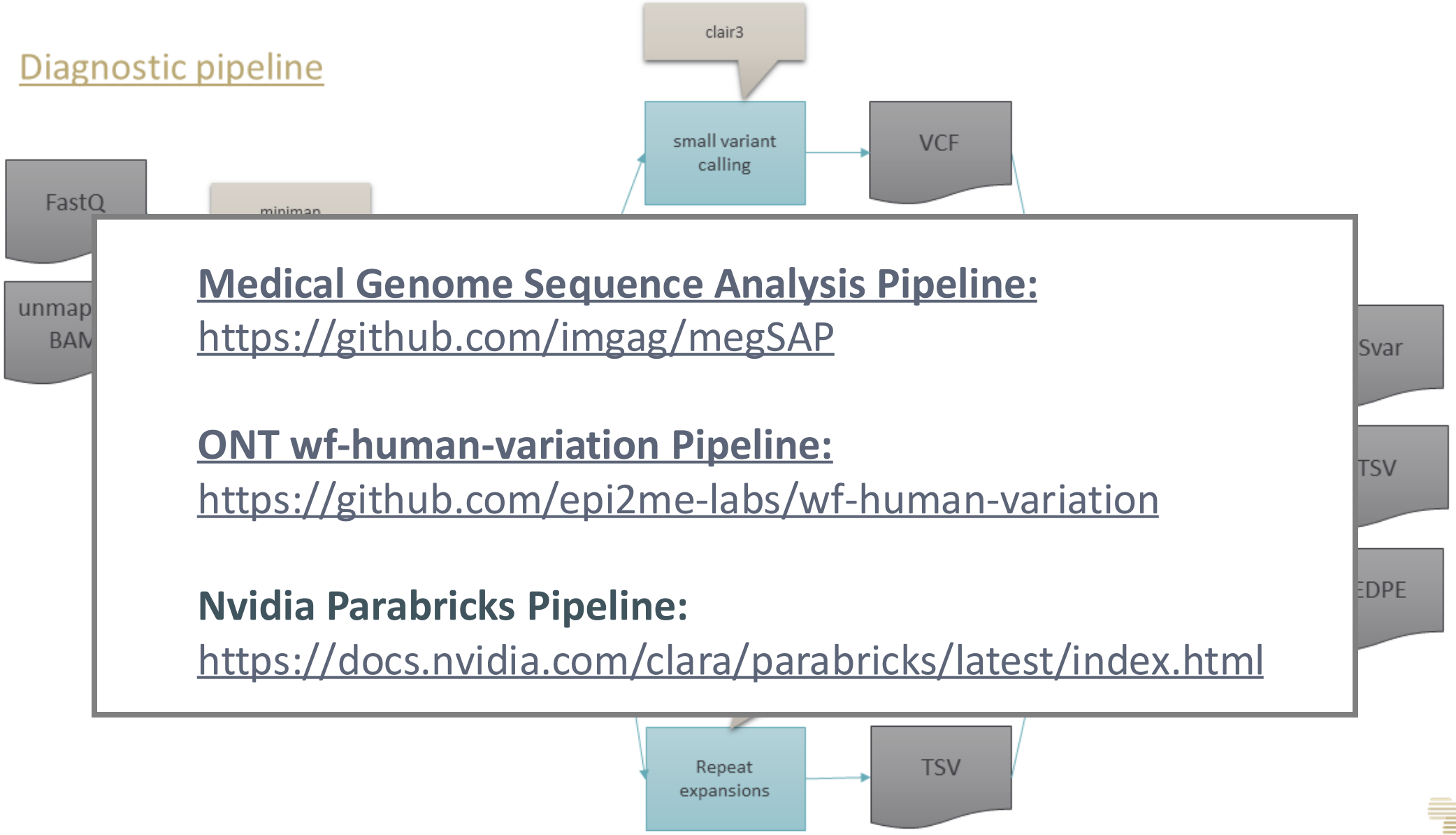
Elena Buena-Atienza, Caspar Gross, Jakob Admard, Leon Schütz

Long-Reads give Haplotype-Phased Genomes (Example in IGV)



Mother_unaffected_mat.bam
Mother
Mother_unaffected_pat.bam
293
maternal
Father_affected_mat.bam
285
Daughter 1
paternal
Father_affected_pat.bam
288
Daughter_unaffected_mat.bam
Father
Daughter_unaffected_pat.bam
284
paternal
Daughter_affected_mat.bam
284
Daughter 2
maternal
Daughter_affected_pat.bam
289
Sequence
Gene
NDUFB1
LINE

Variant- Analysis Pipelines for Nanopore Data

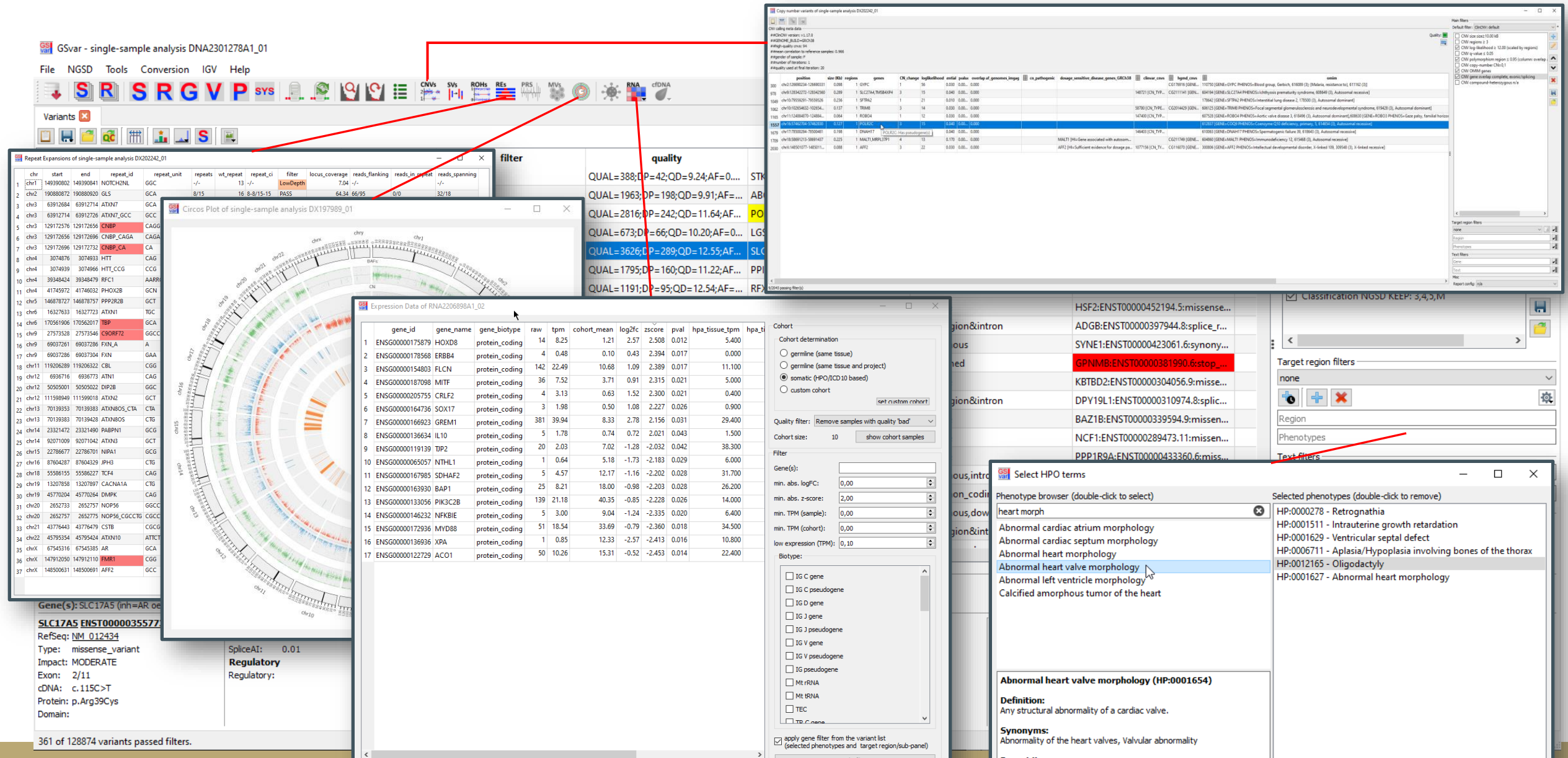


Medical Genome Sequence Analysis Pipeline:
https://github.com/imgag/megSAP

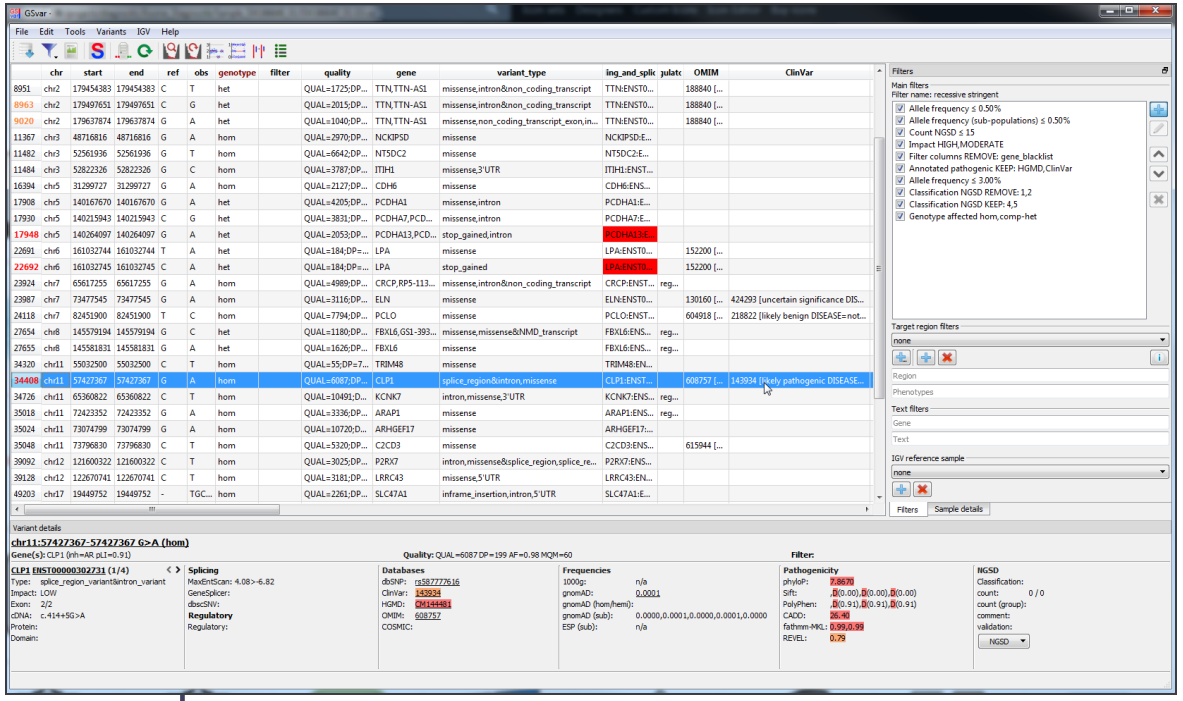
ONT wf-human-variation Pipeline:
https://github.com/epi2me-labs/wf-human-variation

Nvidia Parabricks Pipeline:
https://docs.nvidia.com/clara/parabricks/latest/index.html

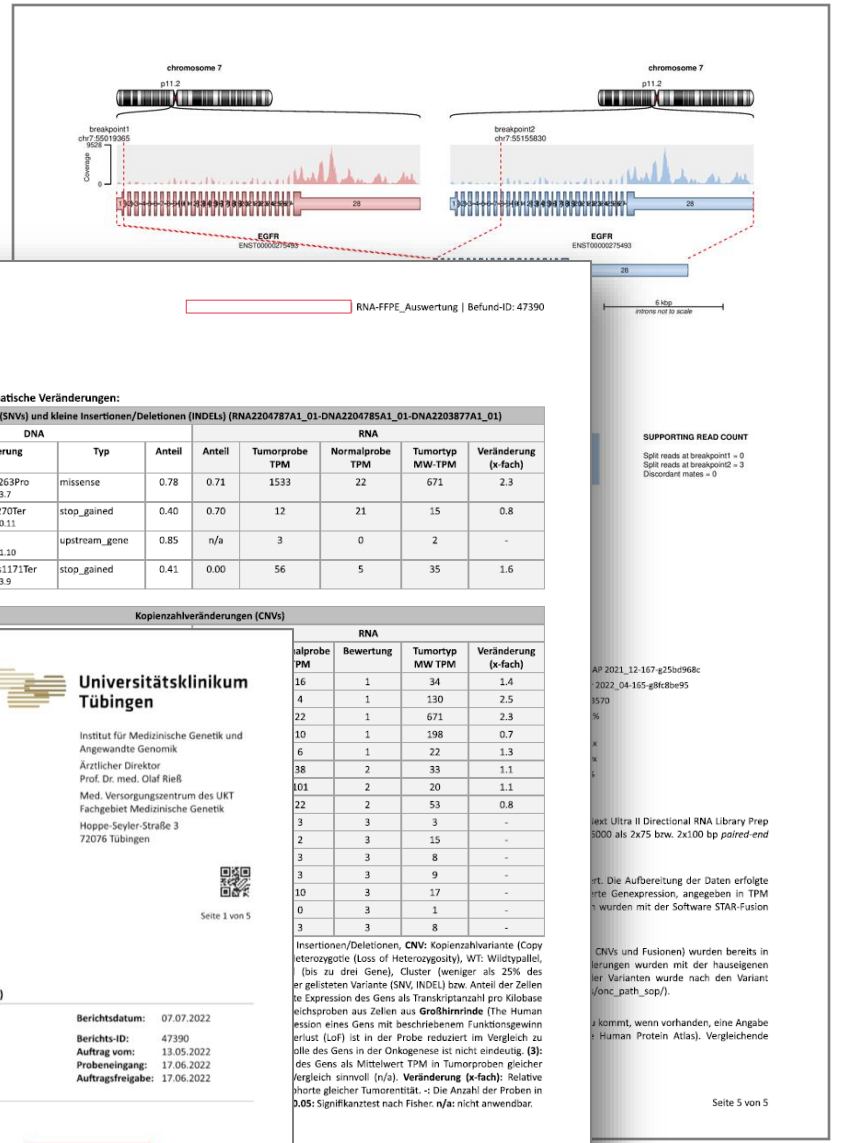
Clinical Decision Support System GSvar Fully Supports Long Read Diagnostics (DNA, RNA, Methylation)



Generating Diagnostic Reports for Long-Read Diagnostics



Tübingen: 25 genetic analysts use GSvar to identify causal variants and to generate reports

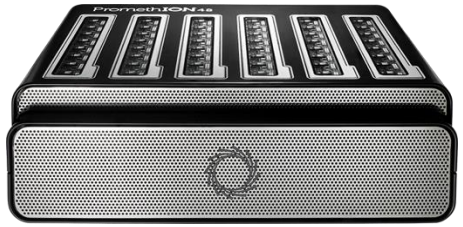


Accreditation of NGS Laboratory Developed Tests

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Lab technicians

Wet-lab/robotics



Sequencers

IT Infrastructure

Sample collection, DNA extraction, Library Prep

Clinicians

Patient

Bioinformaticians, Data Analysts, Geneticists

Quality Control

NGS Analysis Software

Databases

lonGER* - Clinical Long-read Genome Initiative

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

- National German initiative to evaluate the clinical applications of Nanopore sequencing for rare disease
- Sequencing will be done at 4 university medical centers across Germany
- Relies on a clinically well characterized multi-center cohort of rare disease patients
- 1000 samples in a 2-year period + 80 Pilot samples
- Aims:
  - Demonstration of added diagnostic value
  - Proof of sustainability in clinical practice
  - Benchmarking and accreditation
  - Establishing „wet lab“ SOPs and bioinformatic pipeline for streamlined enrollment in other university medical centers

*lonGER Study is supported by ONT

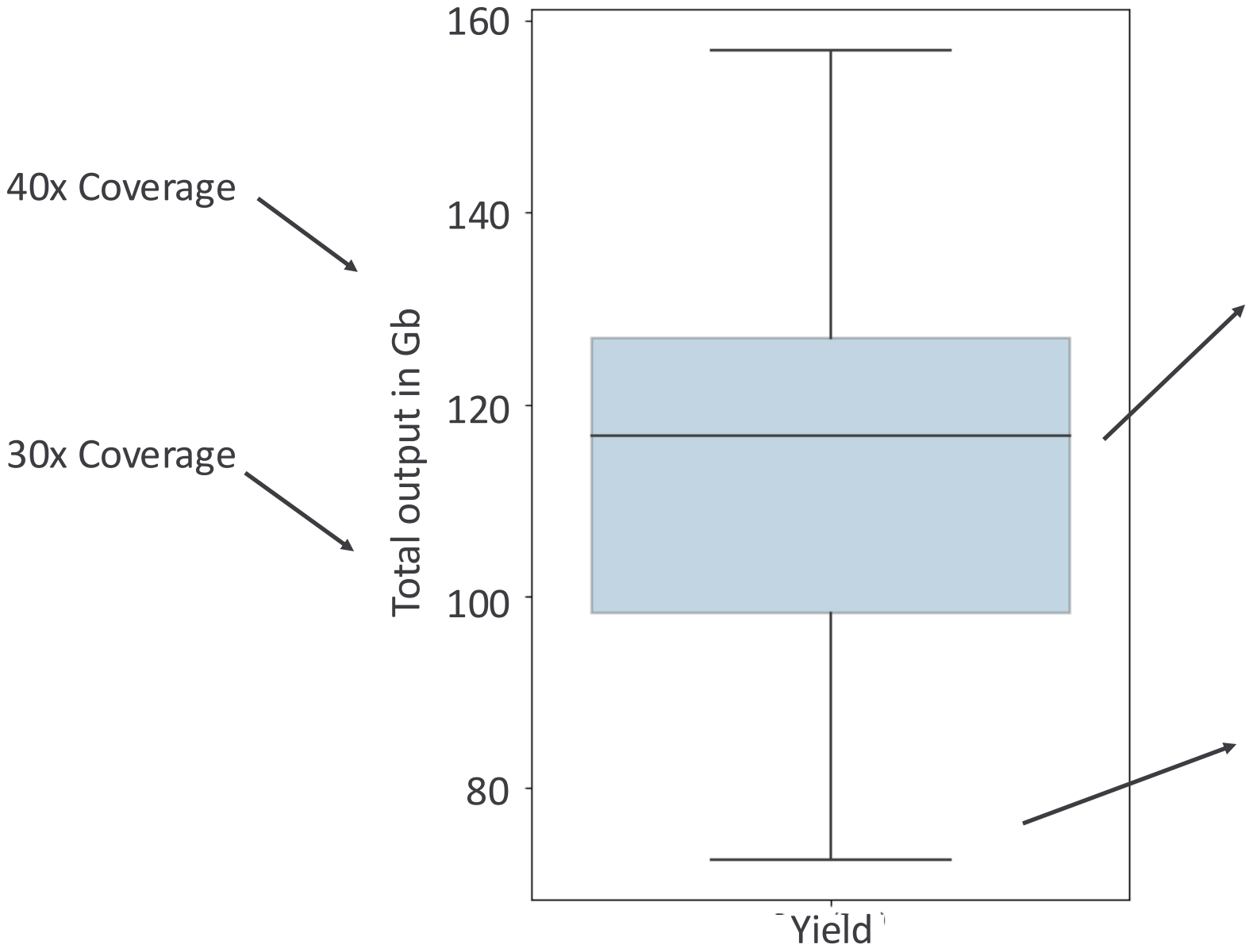
lonGER Pilot Study: Overview

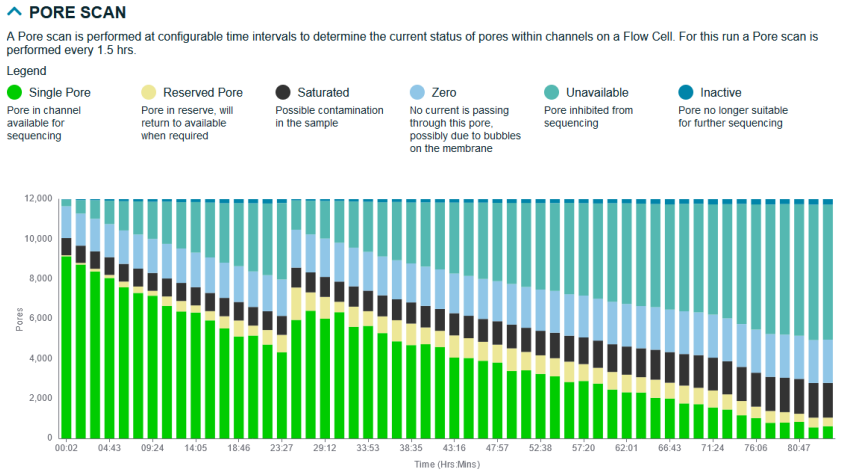
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

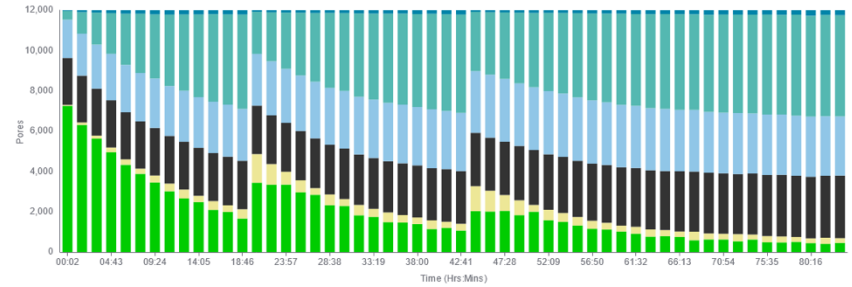
- 2 Genome in a Bottle (**GiaB**) cell lines sequenced at each site (total of 12 FC)
- 68 patient samples (1 FC per genome)
- R10.4.1 with V14 chemistry

- Clinical use cases
  - Repeat Expansions (e.g. FGF14, SCA03 …)
  - Low complexity regions (e.g. RPGR, KANSL1)
  - Pseudogenes (e.g. SMN1/SMN2)
  - Pathogenic haplotypes (e.g. OPN1-Cluster)
  - Imprinting disorders (e.g. FMR1, SNRPN)

Yield per Flowcell







Quality Control for Nanopore Sequencing

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Sample	N50 – Pass (bp)	Bases – Pass (Gb)	avg. read depth	20x coverage (%)	Runtime: Analysis
TUEB_04	19603	121.5	37.24	99.80	30:57
TUEB_05	11907	111.1	33.89	98.93	15:39
TUEB_06	15948	91.4	27.91	90.81	24:50
TUEB_07	18110	113.9	34.82	99.37	31:22
TUEB_08	10611	146.7	44.95	99.99	30:35
TUEB_09	14586	98.2	30.03	96.42	21:49
TUEB_10	13570	98.2	29.87	96.11	24:09
TUEB_11	13767	98.1	29.98	95.98	17:37
TUEB_12	14302	131.0	40.24	99.92	29:57
TUEB_13	9931	126.9	38.30	99.79	25:21
TUEB_14	9943	72.6	22.27	73.41	21:07
TUEB_15	18946	157.0	48.20	99.99	32:52
TUEB_16	12949	85.9	26.23	87.03	19:56
TUEB_17	14697	121.4	37.16	99.58	29:19
TUEB_18	15428	123.7	38.00	99.60	24:50
TUEB_19	14479	150.1	45.92	99.98	31:27
TUEB_20	10428	116.7	35.59	99.46	29:48
Average	***13725.1***	***115.2***	***35.6***	***96.0***	***~26h***

Genome in a Bottle Reference Data for Germline NGS

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

NIST Search NIST Menu

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 .

Benchmark (or "High-confidence") variant calls and regions:
We developed an integration pipeline to utilize sequencing data generated by multiple technologies to generate variant calls and regions for use in benchmarking and validating variant calling pipelines. Currently, benchmark VCF and BED files for small variants are available for GRCh37 and GRCh38 under each genome at https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/
GIAB's versions of GRCh37 and GRCh38 reference fasta files, including a new GRCh38 reference in collaboration with the GRC that masks false duplications in GRCh38, are at https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/references/.

New benchmarks for difficult variants and regions:
Structural variants: Currently available for HG002 on GRCh37 and in Challenging Medically Relevant Gene

ORGANIZATIONS
Material Measurement Laboratory
Biosystems and Biomaterials Division
Biomarker and Genomic Sciences Group

NIST STAFF
Justin Zook
Nathanael David Olson
Justin Wagner
Nathan Dwarshuis
Megan Cleveland
Lindsay Harris
Sierra Miller

FORMER STAFF
Marc Salit
Jennifer McDaniel
Lesley Chapman

CONTACT
Justin Zook
justin.zook@nist.gov
(301) 975-4133

DATES
Started: August 2012

PROJECT STATUS
ONGOING

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

Benchmark (or "High-confidence") variant calls and regions:

We developed an integration pipeline to utilize sequencing data generated by multiple technologies to generate variant calls and regions for use in benchmarking and validating variant calling pipelines. Currently, benchmark VCF and BED files for small variants are available for GRCh37 and GRCh38 under each genome at https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/

New benchmarks for difficult variants and regions:

Structural variants: Currently available for HG002 on GRCh37 and in Challenging Medically Relevant Gene benchmark below

Small variants in more difficult regions: v4.2.1 is available for all 7 GIAB samples on GRCh37 and GRCh38 (manuscript).

MHC: Included in v4.2.1 small variant benchmark for HG001-HG007 (Manuscript describing MHC benchmark)

273 Challenging Medically Relevant Genes small variant and SV benchmarks in HG002 and Preliminary benchmark for T2T-CHM13v1.0

GiaB Technology

https://www.nist.gov/programs-projects/genome-bottle

Published: 08 July 2015

Genome in a bottle—a human DNA standard

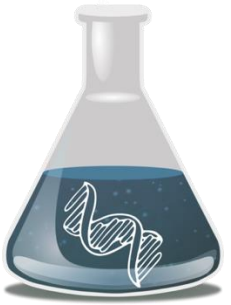
Nature Biotechnology **33**, 675 (2015) | Cite this article

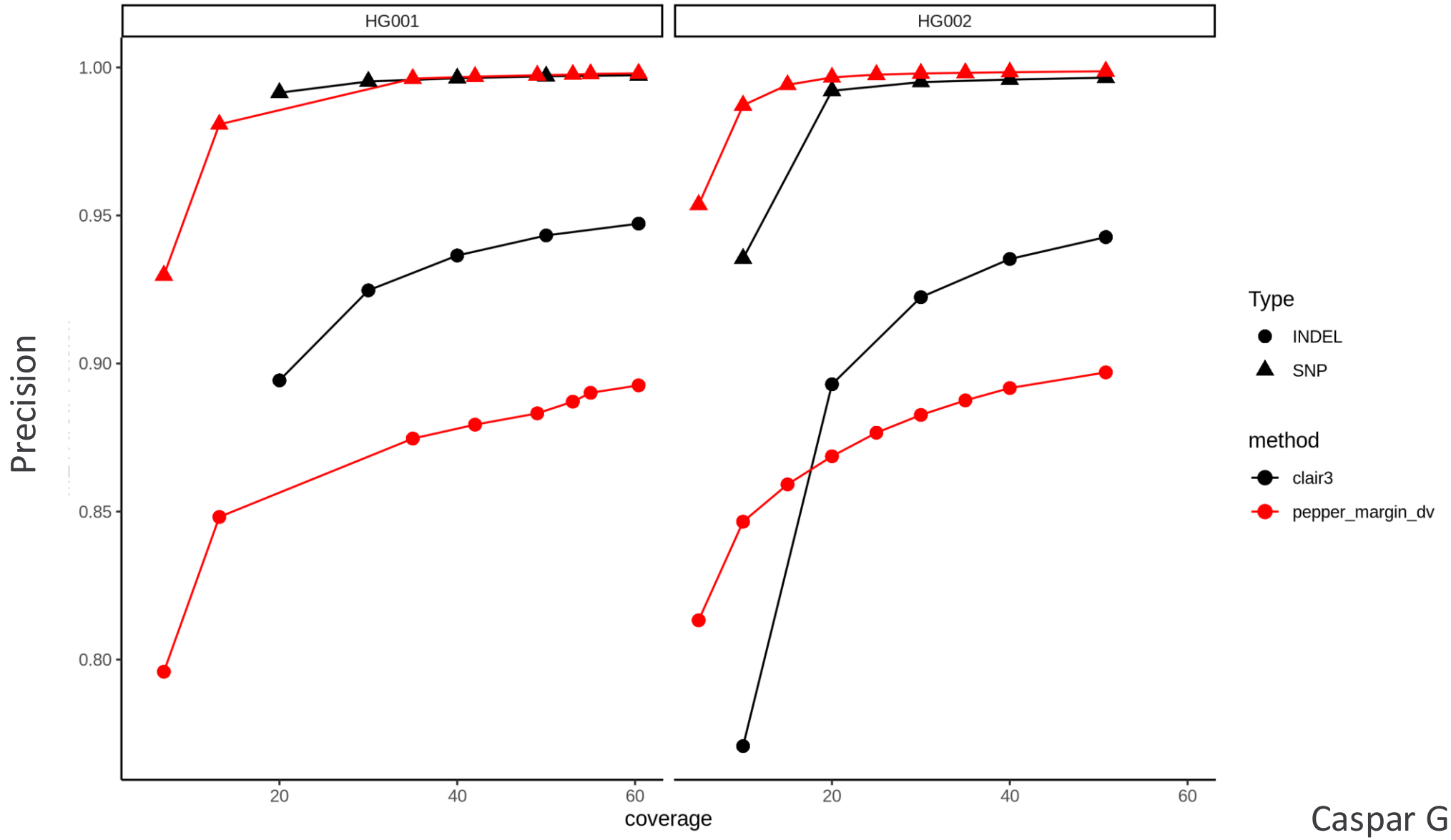
NA12878:HG001 https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/NA12878/

Sequencing Platform	Sequence	Alignment
Illumina WGS 2x150bp 300X	HG001	bwamem: HG001:hg19 (downsampled30x) novoalign: HG001
Illumina HiSeq Exome	HG001 HG001:trimmed_fastq	bwamem: HG001:hg19
Illumina TruSeq Exome		bwamem: HG001:hg19
10X Genomics		bwamem: HG001:hg19 bwamem: HG001:hg19 (size_selected)
10X Genomics ChromiumGenome		LongRanger2.0: HG001:hg19-hg38 LongRanger2.1: HG001:hg19-hg38
CompleteGenomics		CGAtools: HG001:hg19
Ion Proton 1000x Exome		TMAP: HG001:hg19
NA12878 SOLiD5500W		LifeScope: HG001:hg19
BGI BGISEQ500	sequence.index	alignment.index
BGI MGISEQ	sequence.index	alignment.index
BGI stLFR	sequence.index	alignment.index
PacBio 40x	HG001:hdf5	
PacBio SequelII CCS 11kb	sequence.index	alignment.index
Ultralong_OxfordNanopore	-	minimap2: HG001

SNV and InDel Precision

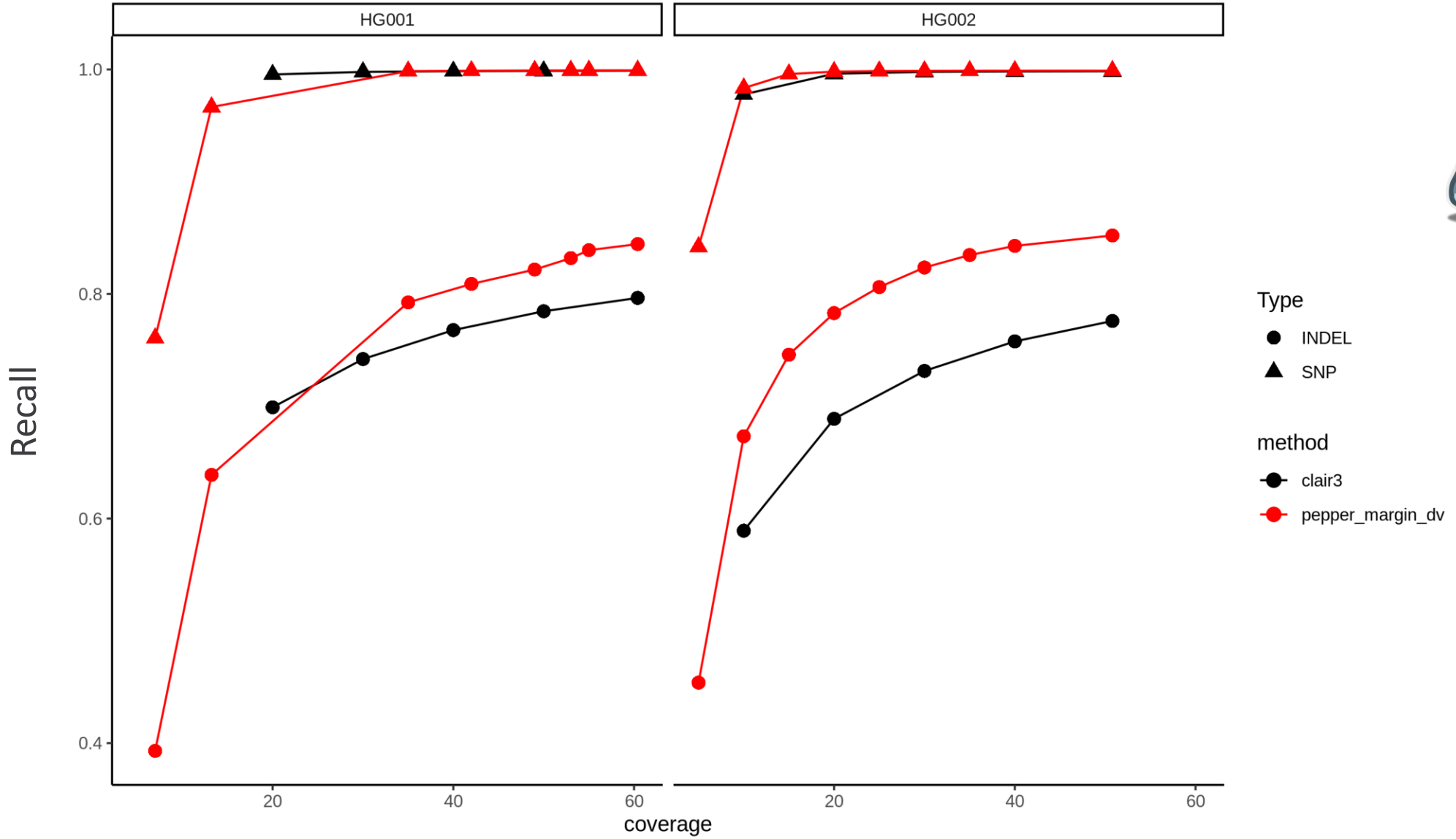
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN





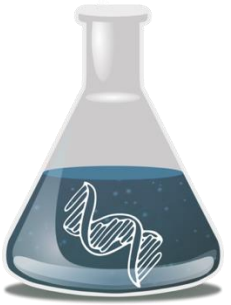
Caspar Gross

SNV and InDel Recall



Higher InDel Quality in Coding Regions

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Genome-in-a-Bottle Reference NA12878 (HG001), 40x coverage

SNV/INDEL validation (whole genome):

	recall/sensitivity	precision	genotyping accuracy
SNVs	0.999	0.999	0.999
INDELs	**0.794**	0.966	0.998

SNV/INDEL validation (only protein-coding exons):

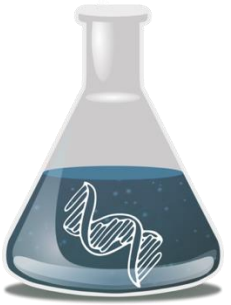
	recall/sensitivity	precision	genotyping accuracy
SNVs	1.000	0.999	1.000
INDELs	**0.936**	0.978	0.998

Results with Dorado 5-base model, including cytosine methylation calls

Leon Schütz, Marc Sturm

How much Nanopore read coverage is needed?

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Genome-in-a-Bottle Reference NA12878 (HG001)

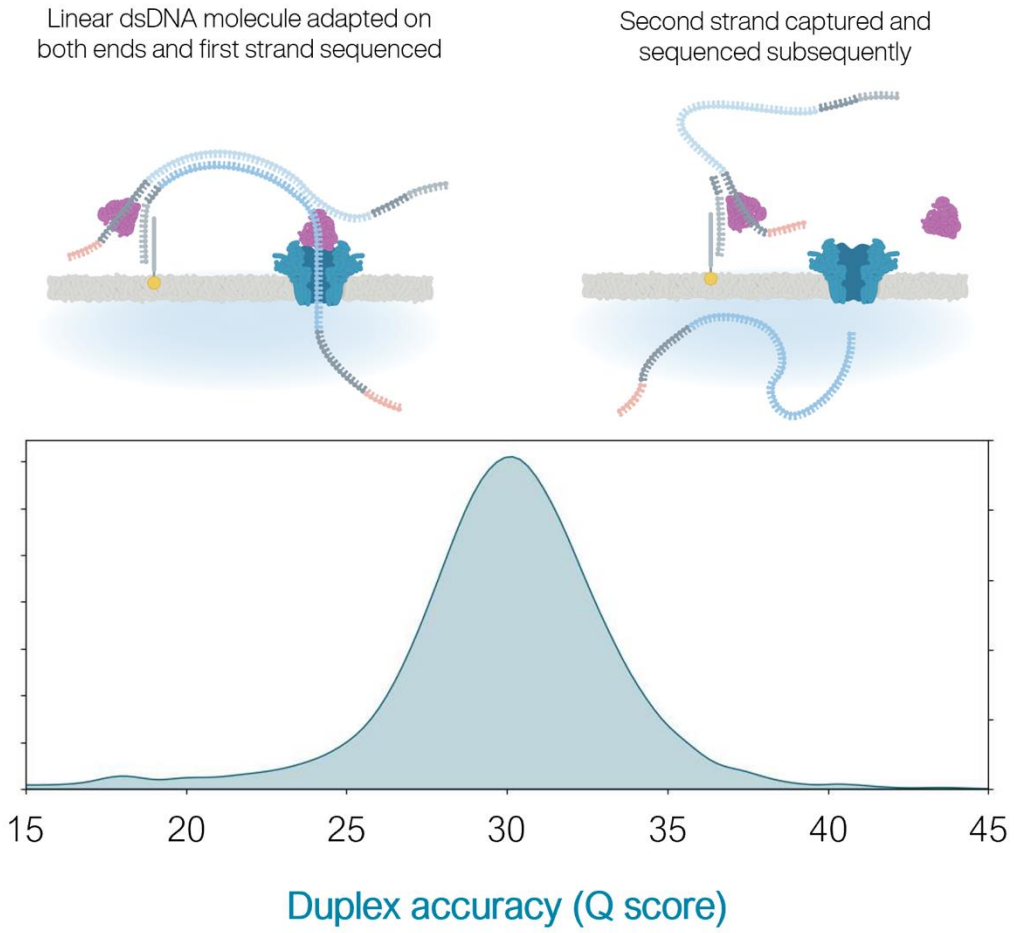
SNV/INDEL validation (only protein-coding exons):

Coverage	SNV			INDEL		
	sensitivity	Precision	genotyping	sensitivity	Precision	genotyping
50x	0.999	0.996	1.000	0.940	0.967	0.998
40x	0.999	0.994	1.000	0.932	0.944	0.996
30x	0.999	0.990	0.999	0.921	0.935	0.998
20x	0.998	0.989	0.999	0.900	0.848	0.996
10x	0.986	0.888	0.994	0.812	0.494	0.998

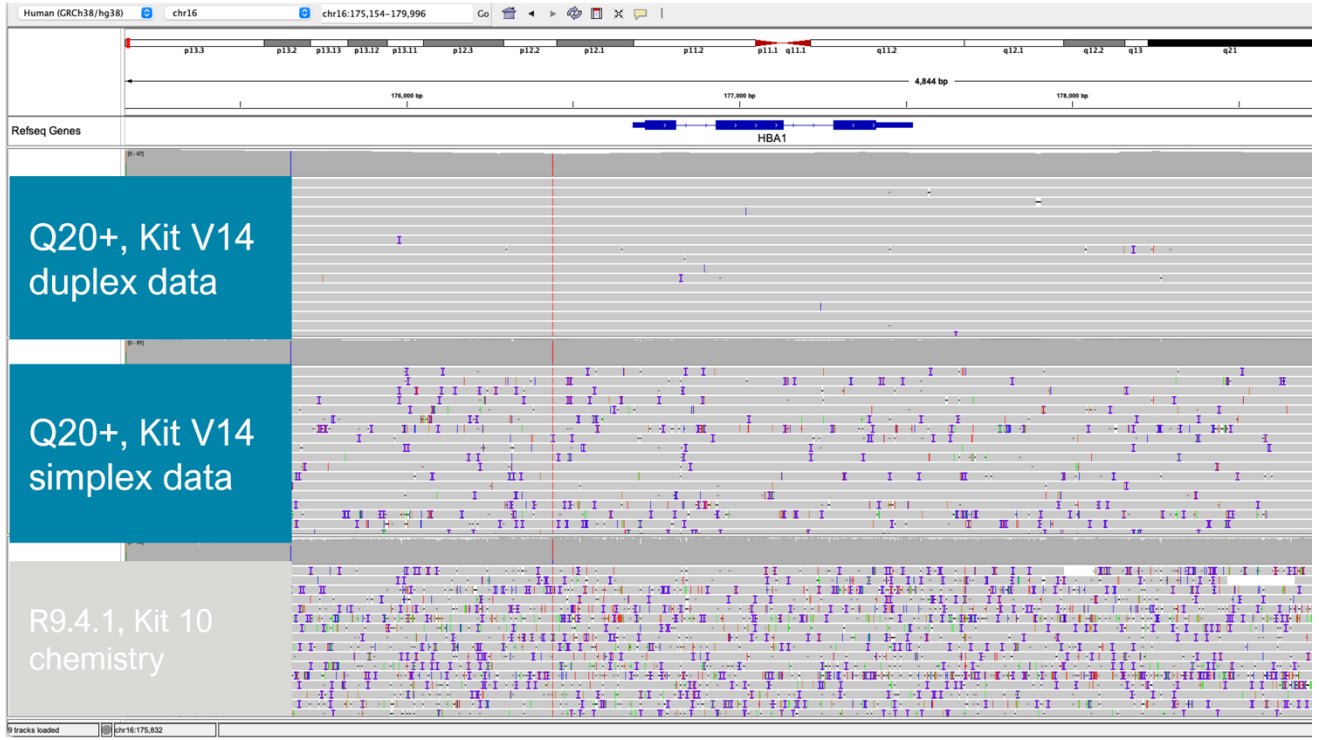
Can we further improve InDel calling: Duplex Reads!

Leon Schütz, Marc Sturm

Duplex Reads



Alignments in IGV

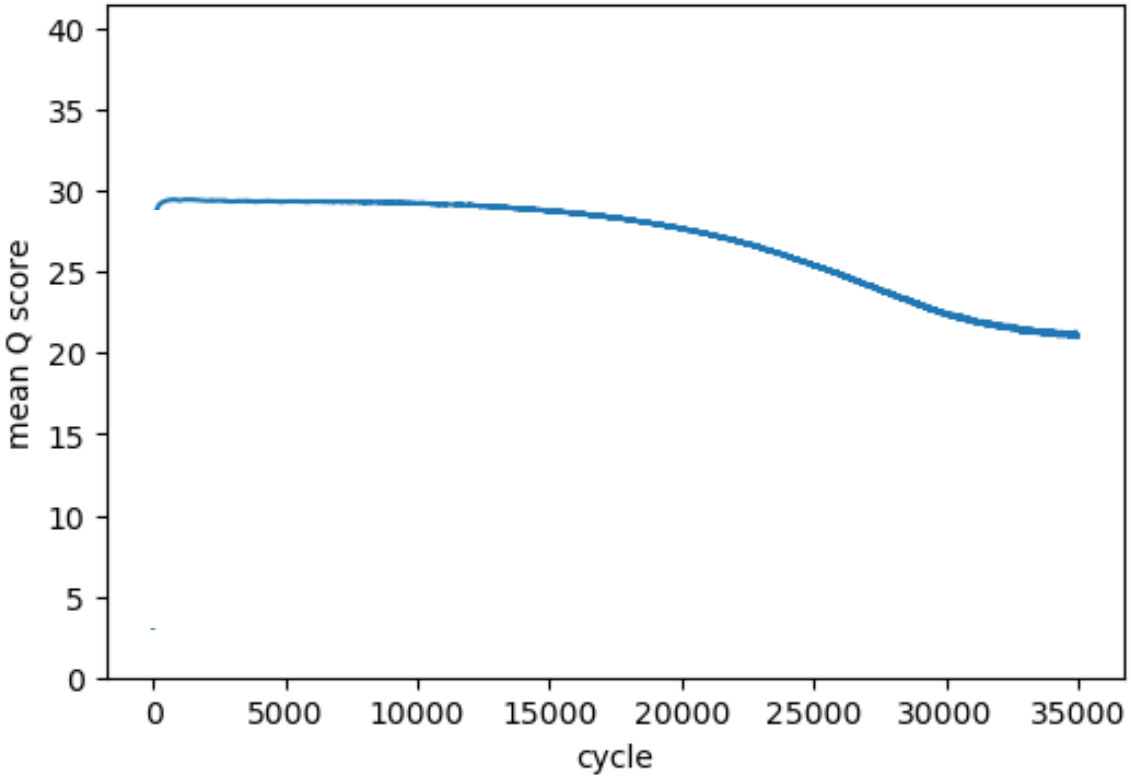


https://nanoporetech.com/q20plus-chemistry

Duplex Read Quality of Sample HG001 (NA12878)

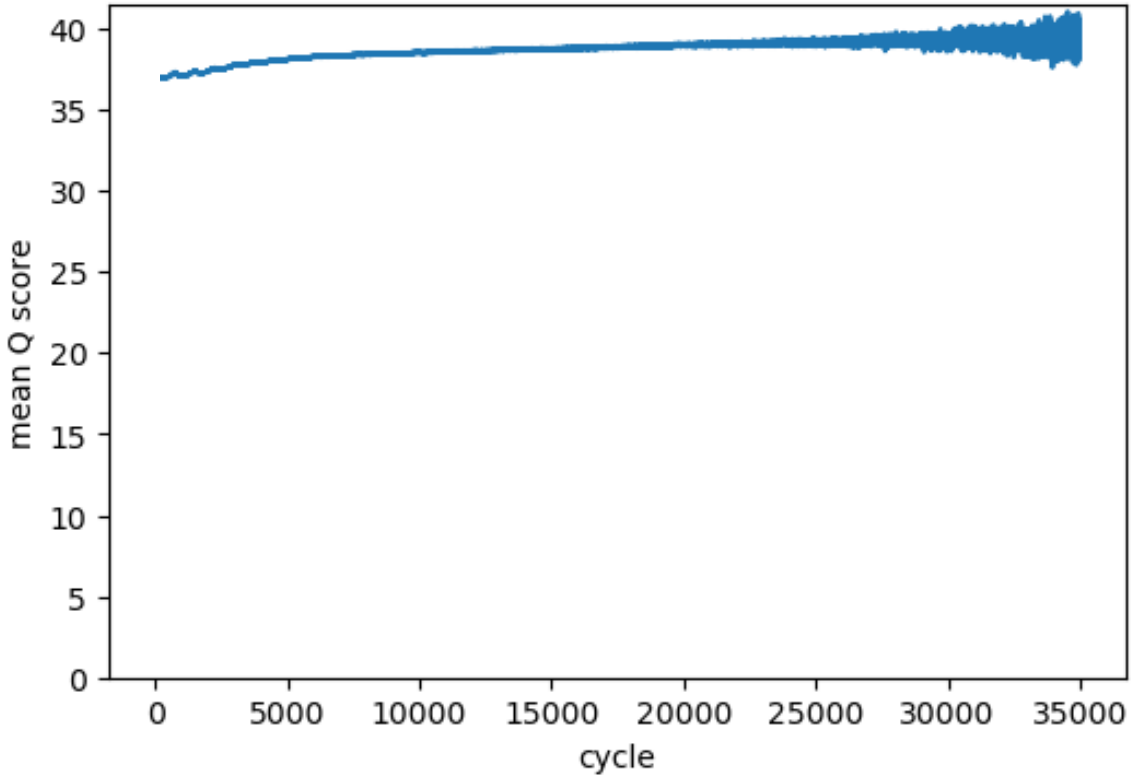
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Simplex Reads



- Q20 read percentage: 97.09
- Q30 read percentage: 58.73

Duplex Reads



- Q20 read percentage: 99.96
- Q30 read percentage: **89.68**

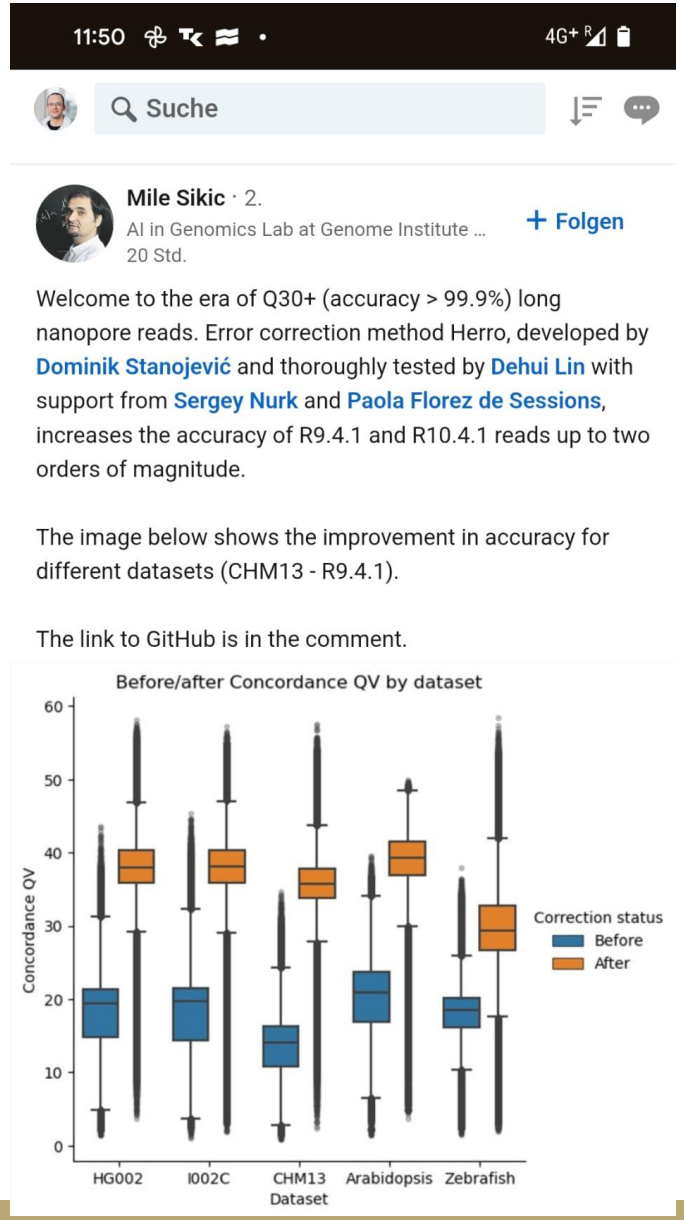
Duplex Reads Increase Quality

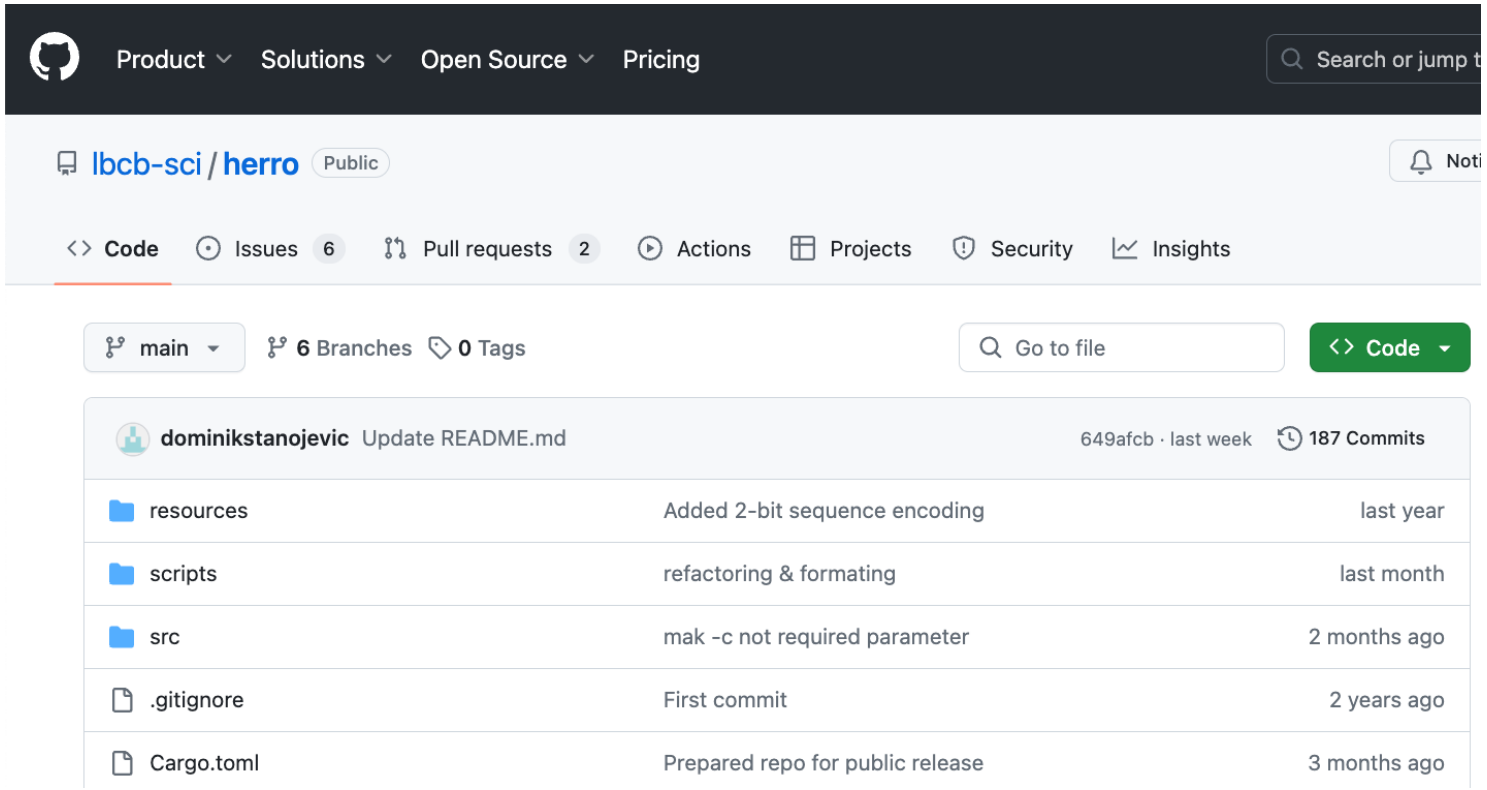
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

	Simplex	Duplex
Error Rate	2.03%	0.54%
Mismatch Rate	1.06%	0.32%
Insertion Rate	0.45%	0.12%
Deletion Rate	0.52%	0.10%
Soft-clip Rate	99.85%	12.20%

TODO: test improvement for indel calling by pooling duplex reads from multiple flowcells (HG001, HG002)

Haplotype Aware Error Correction?



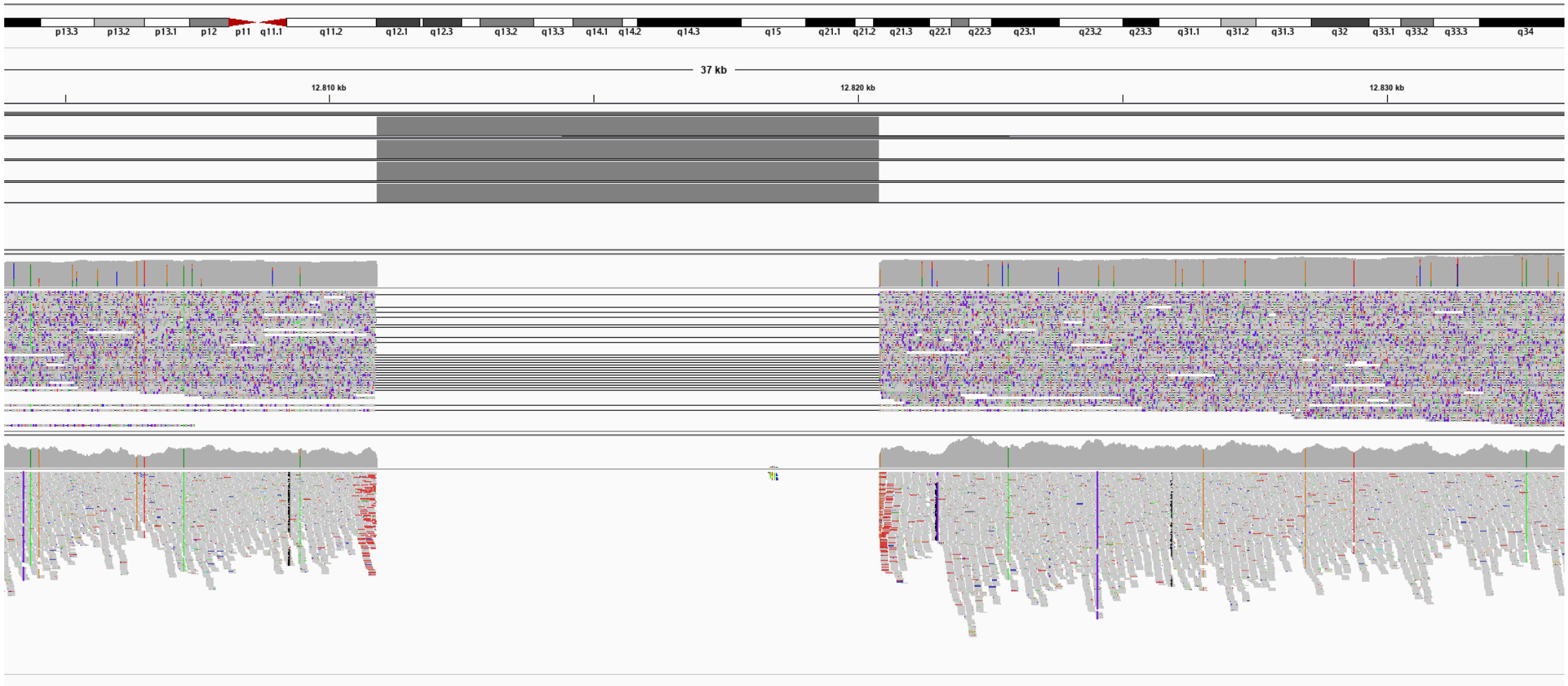


HERRO

HERRO (Haplotype-aware ERRor cOrrection) is a highly accurate, haplotype-aware, deep-learning tool for error correction of Nanopore R10.4.1, Kit 14 reads (length of ≥ 10000bp is recommended). An experimental model for R9.4.1 data is also provided for download.

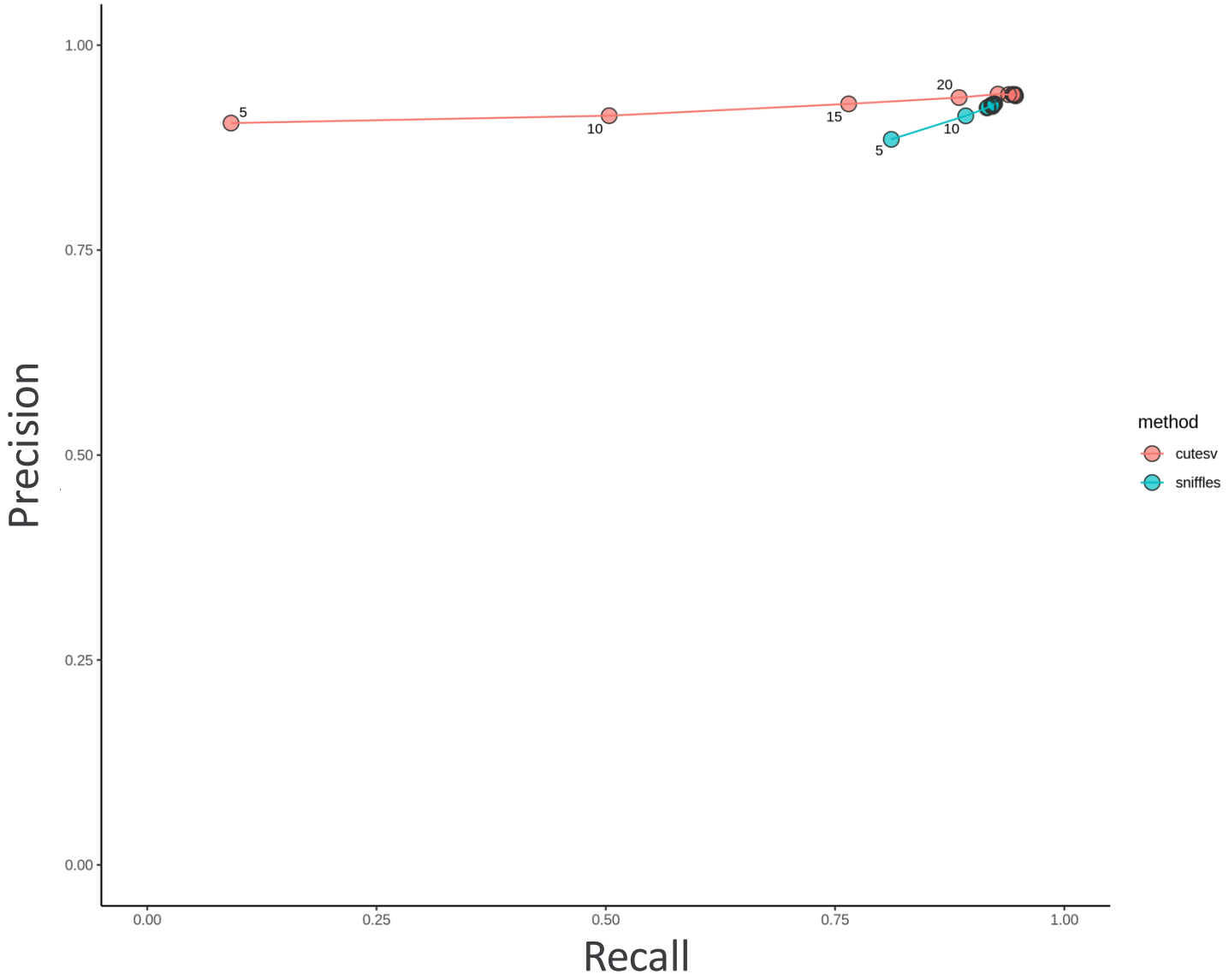
Structural Variant Detection

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

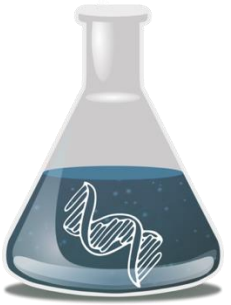


*Bionano: optical genome mapping

Structural Variants: High Recall and Precision



Sniffles with 50x:
- Precision: 92.2
- Recall: 91.8



CuteSV with 50x:
- Precision: **93.8**
- Recall: **94.7**

Sniffles with 5x:
- Precision: 81.1
- Recall: 84.7

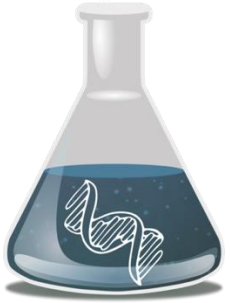
Caspar Gross

Structural Variants with Sniffles

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

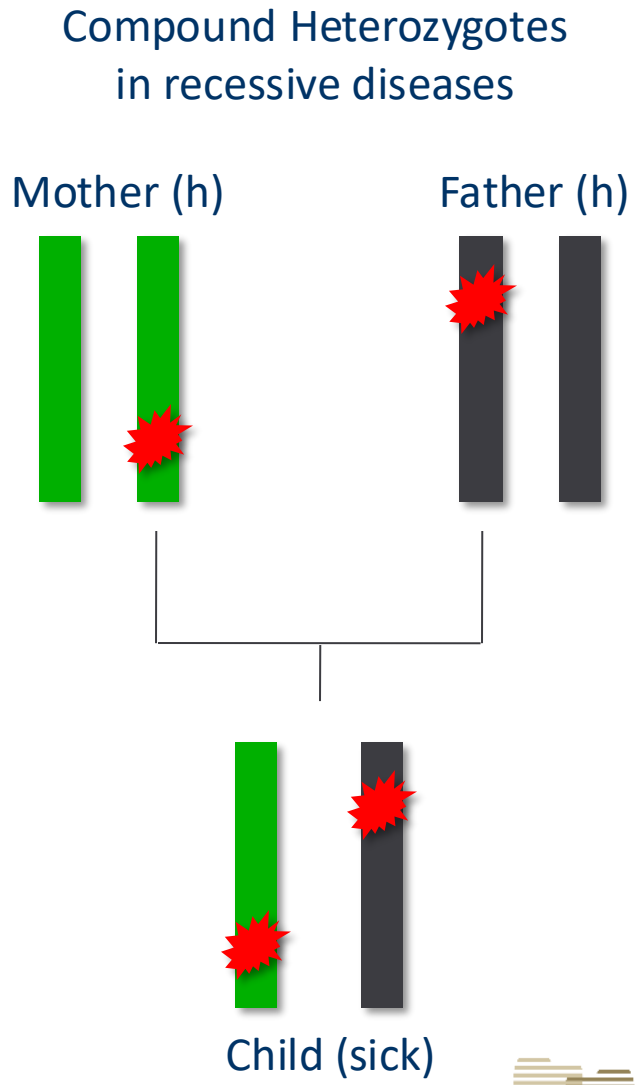
Coverage	True Pos	False Neg	False Pos	Precision	Recall
50x	17286	757	670	0.9219	0.9176
40x	17317	735	661	**0.9223**	**0.9187**
30x	17296	735	662	**0.9204**	**0.9169**
20x	17122	846	656	0.9162	0.9070
10x	16149	1474	676	0.8935	0.8557

- 30-40x coverage is sufficient for high quality SV calling with sniffles
- At > 40x coverage cuteSV outperforms sniffles
- With short reads, recall is less than 50%
- Genome-wide: ˜10.000 SV with short and 23.000 SV with long reads

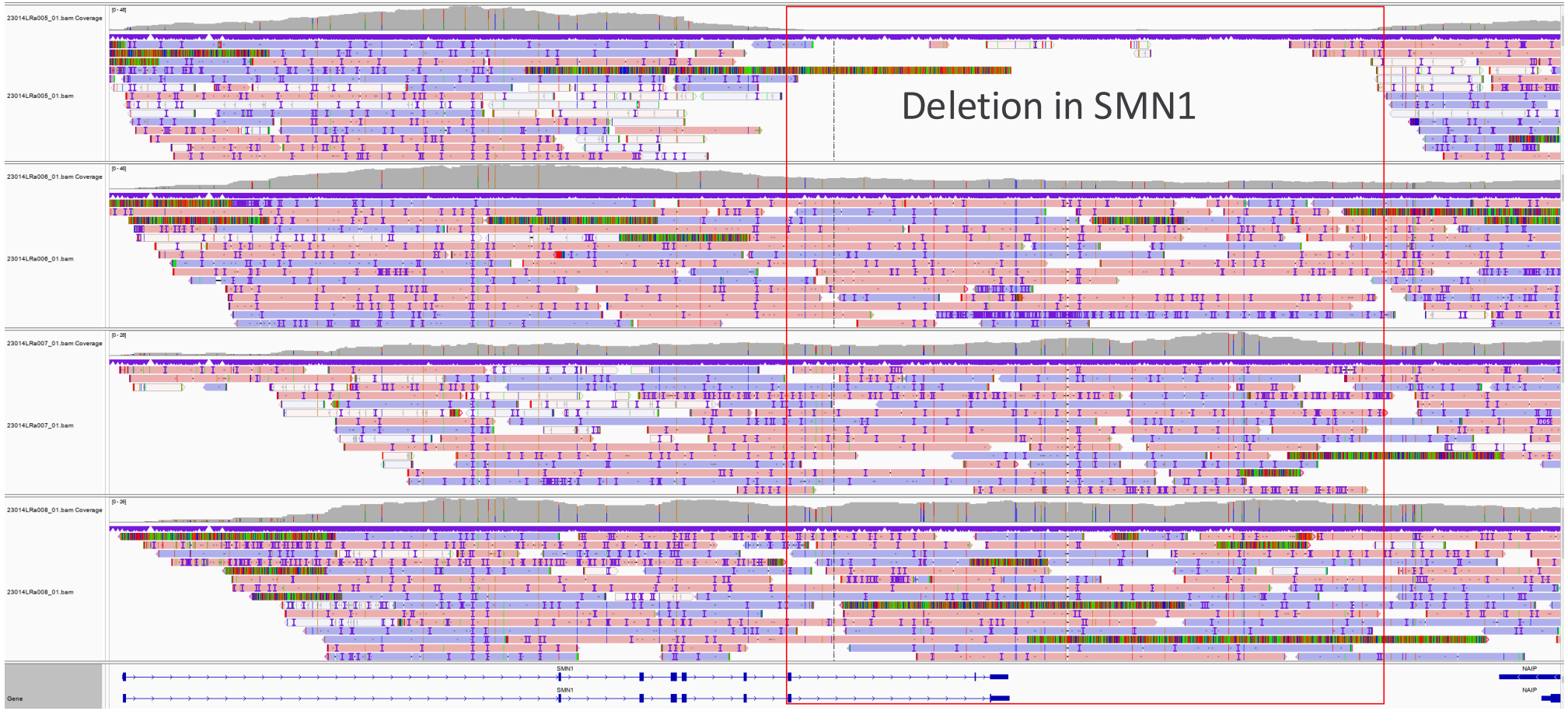


Examples for Improved Clinical Diagnostics

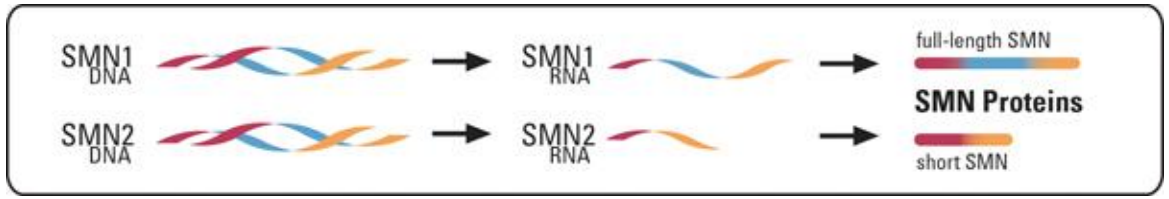
1. **Causal Compound Heterozygotes** when sequencing only the index case (thanks to Haplotype Phasing)
2. **Structural variants and Mobile elements**: doubling the recall for SVs
3. **Duplicate gene** analysis (or other ‘dark regions of the genome’)
4. **Repeat Expansions**
5. **DNA Methylation, Imprinting**



Duplicate Genes: SMN1 and SMN2 in Spinal Muscular Atrophy



Tü-Case-8



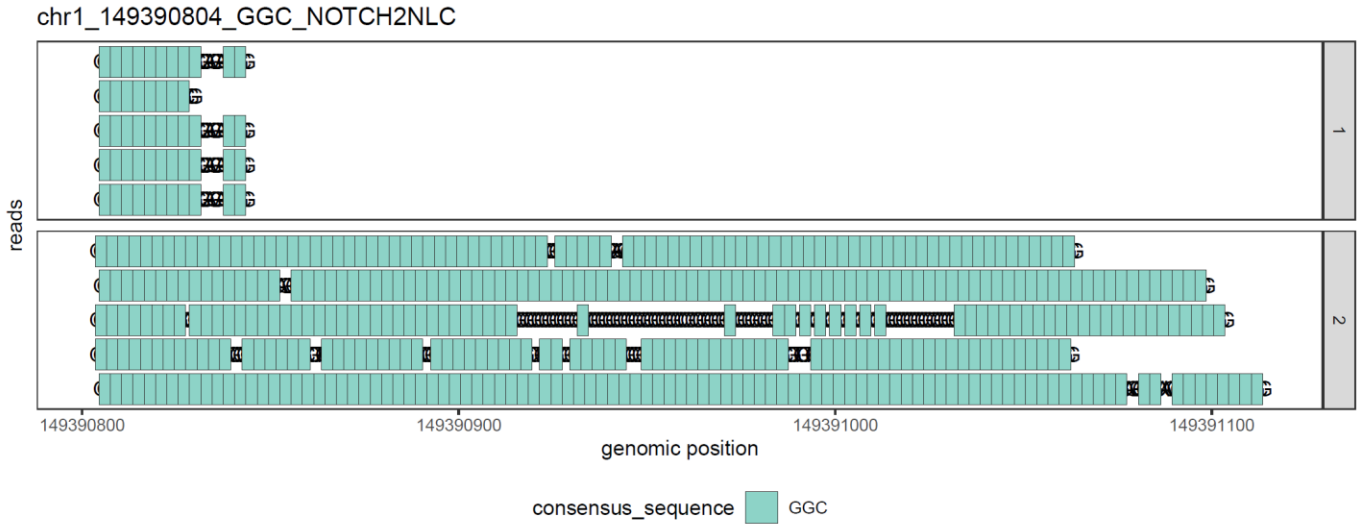
https://www.mda.org/disease/spinal-muscular-atrophy/causes-inheritance

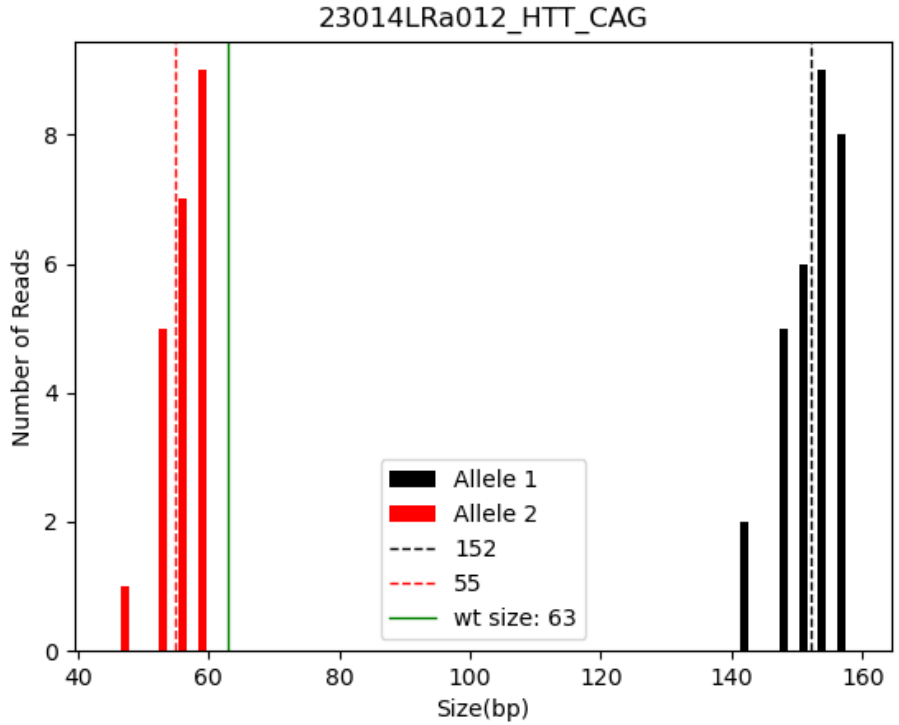
Repeat Expansion Detection

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Algorithm:

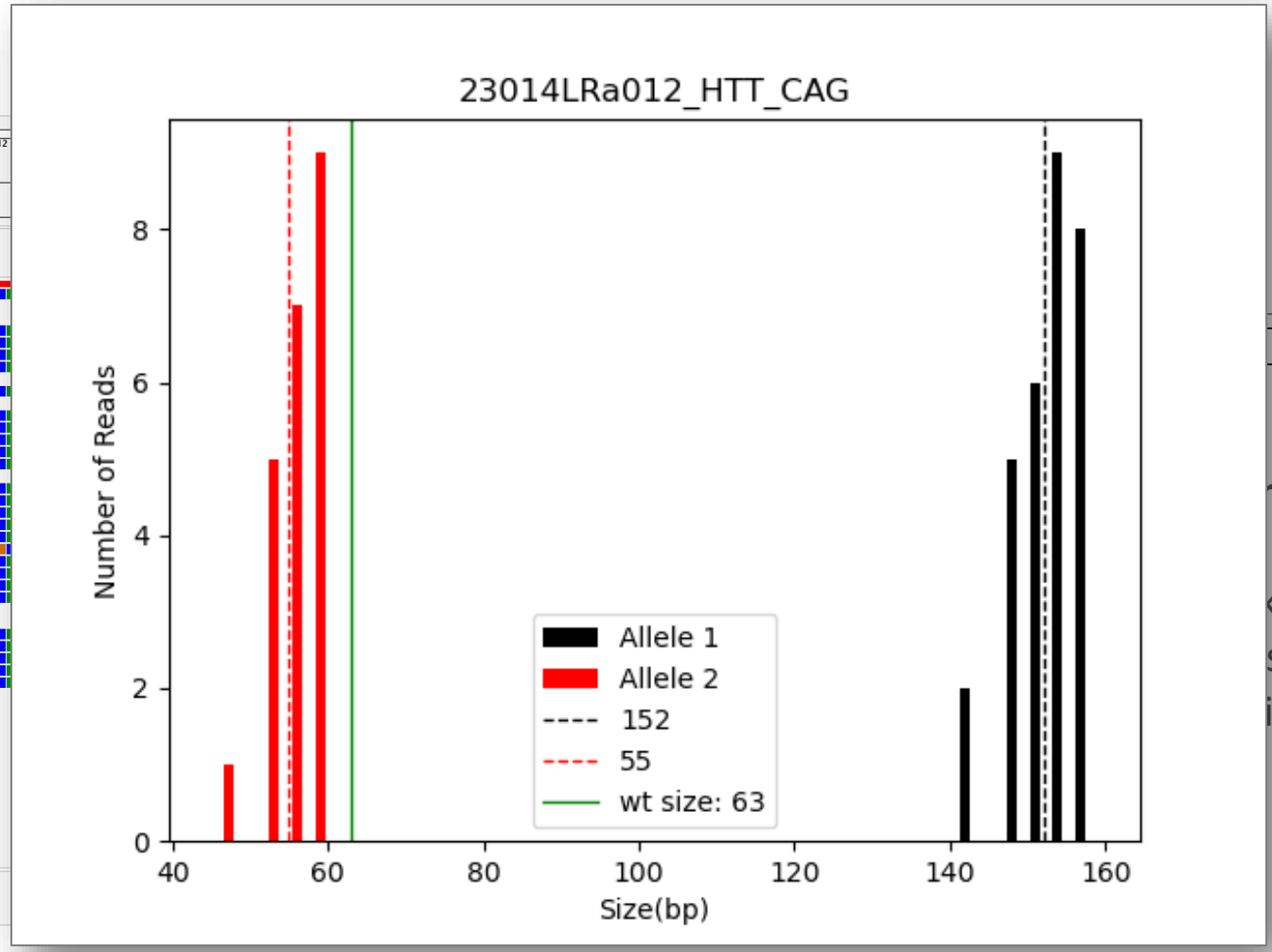
- Phase repeat haplotypes of cases and controls
- Generate consensus repeat-length for each haplotype
- Measure minimum length causing disease
- We currently use the Straglr tool with in-house visualization

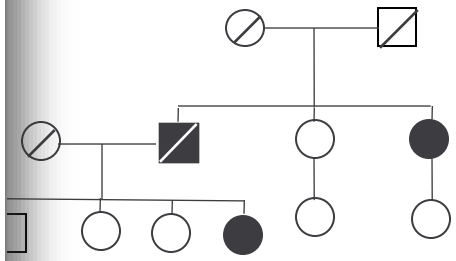




Caspar Gross, Thomas Braun,
Chia Ying Ko, Leon Schütz

Example: Huntington Disease (Tübingen Case 15)





n: Huntington disease

kinetic movement disorder
sturbance
ine motor coordination

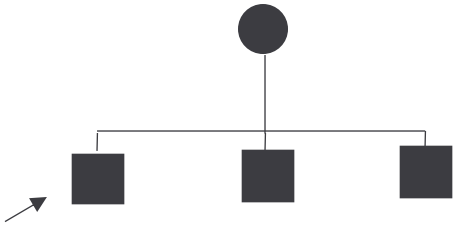
Extended allele: 53 CAG Repeats in HTT

Example: Ataxia (SCA3, Tübingen Case 16)

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Indication: SCA 3

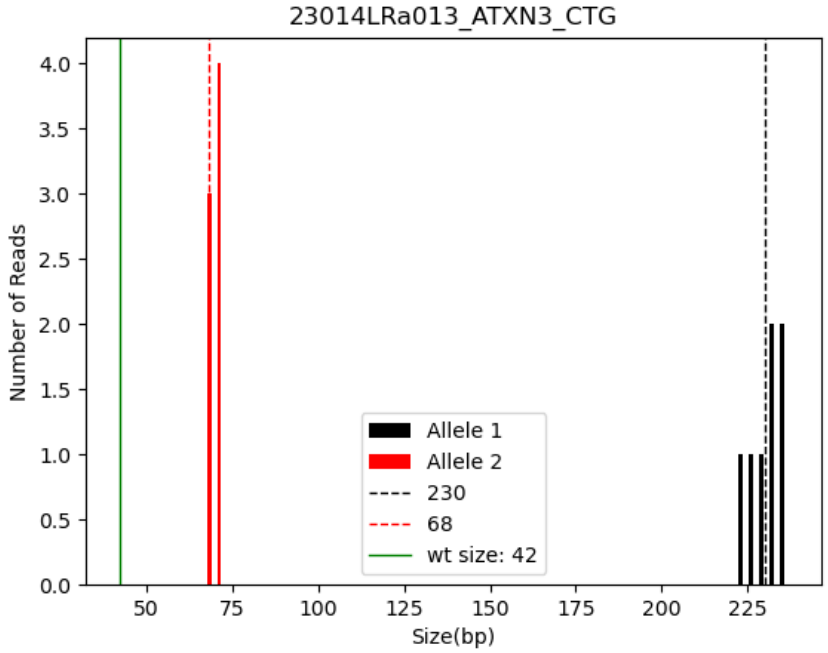
- ataxia
- dysarthria
- dysphagia
- cognitive impairment
- ophthalmoparesis
- parkinsonism



Extended allele: 77 CAG Repeats in ATXN3

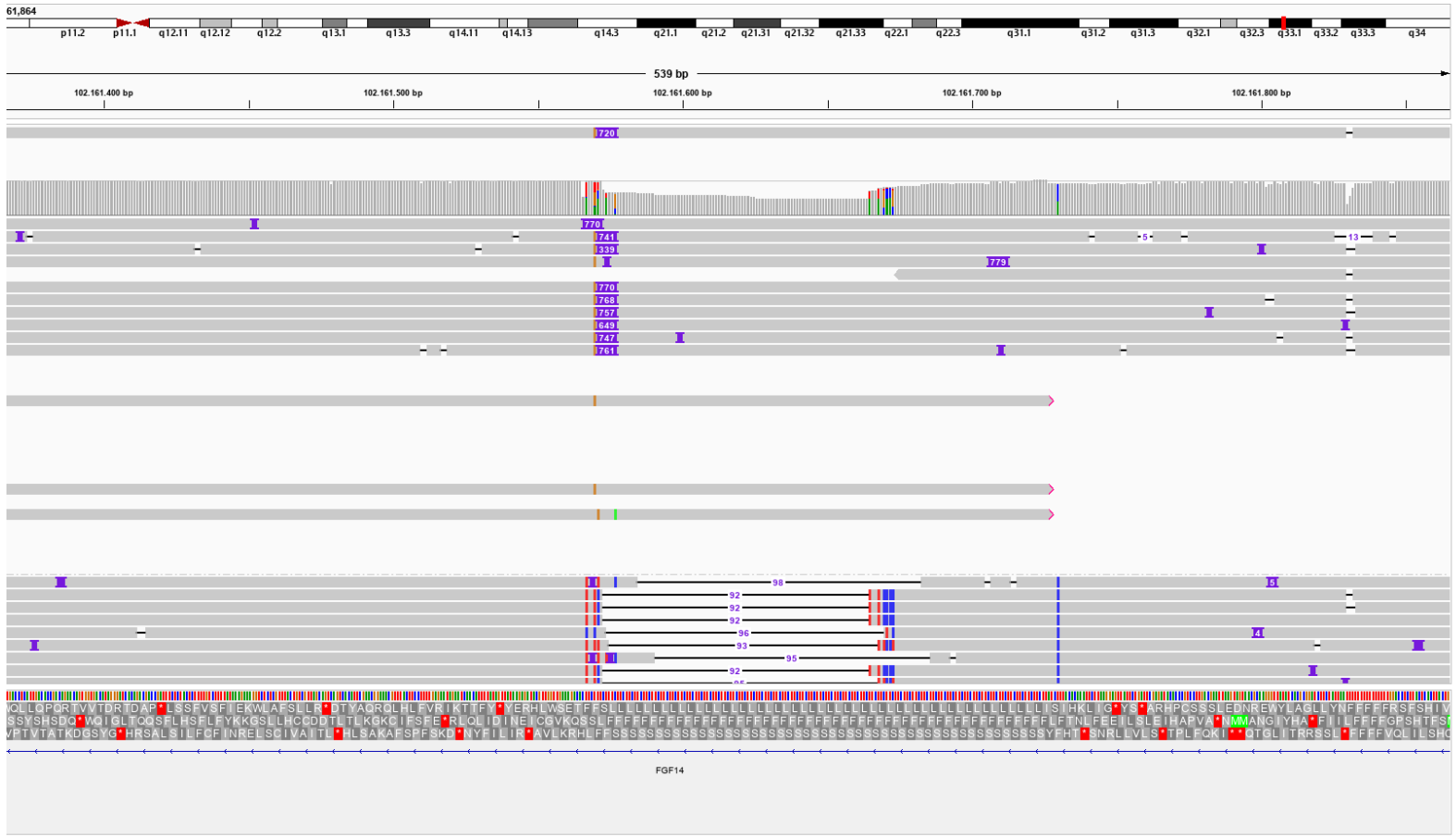
ONT Long Read Analysis

- ATNX3: $(CAG)_{23}$ / $(CAG)_{77}$



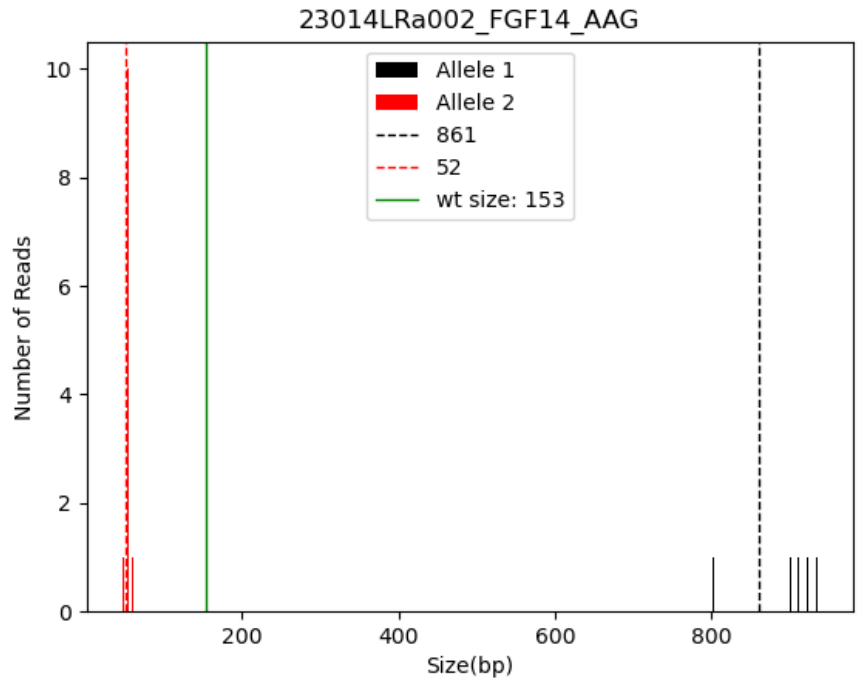
Ataxia with FGF14 Repeat Expansion (Tü Case 5)

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



ONT Long Read Analysis

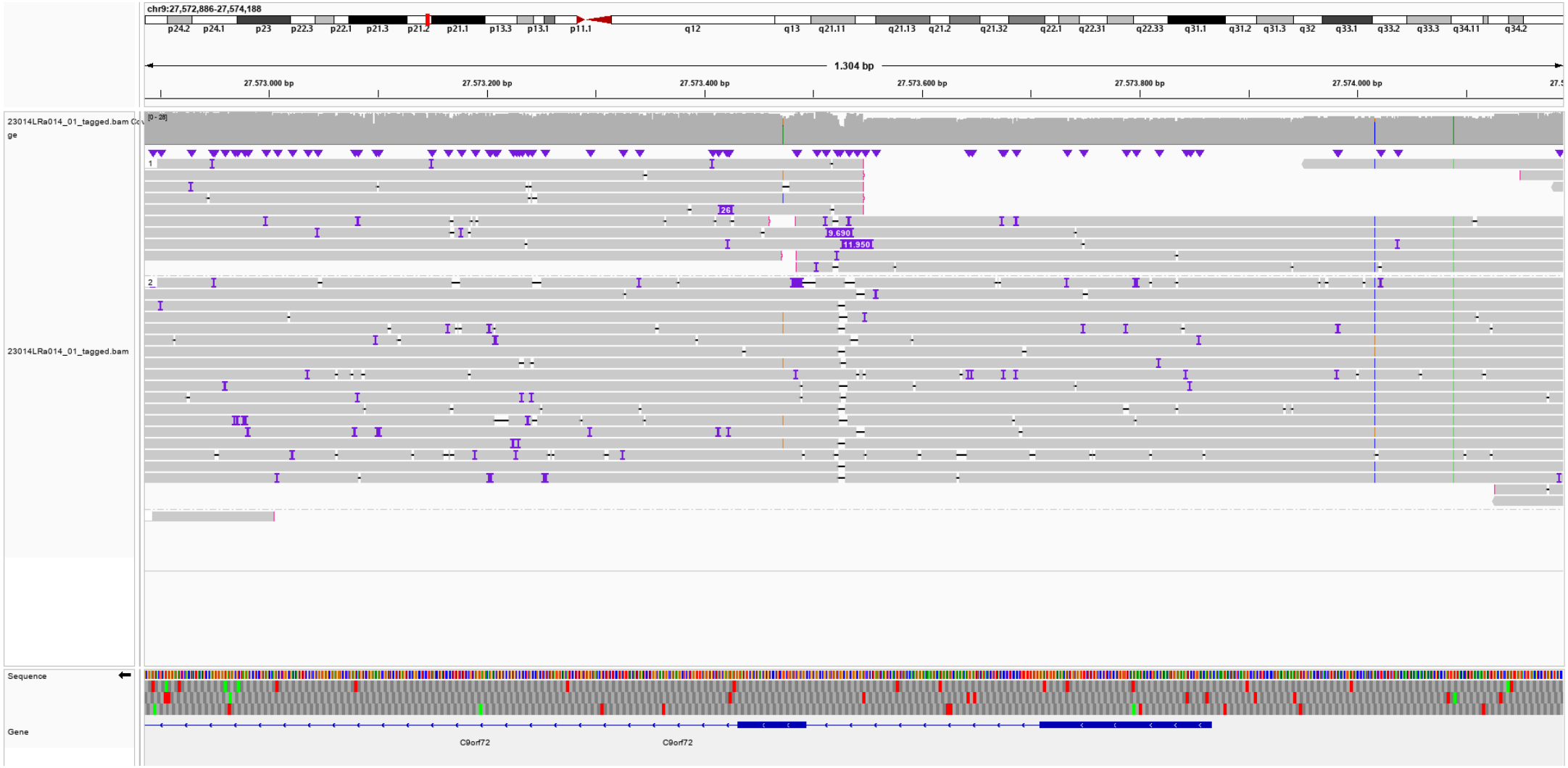
- FGF14: $(GAA)_{309}$



GAA repeat in FGF14:

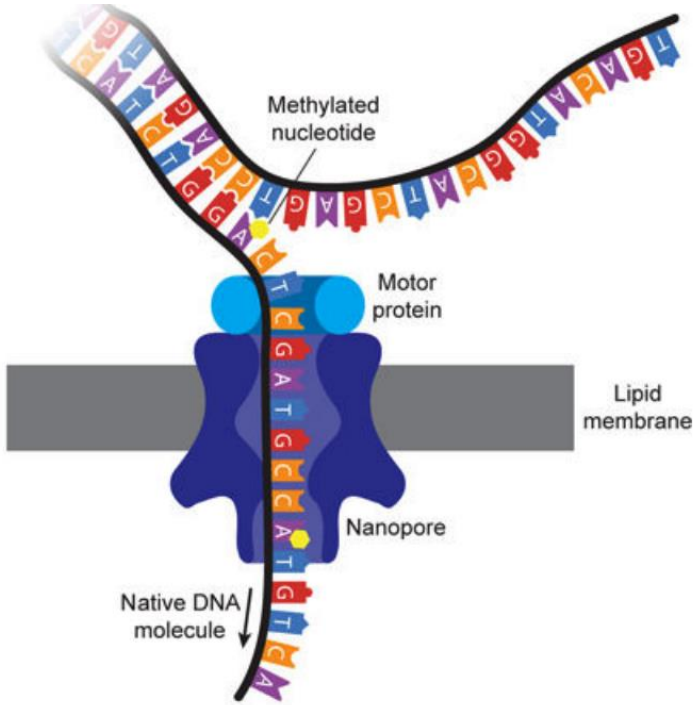
- **allele 1: ~ 30 repeats,**
- **allele 2: >300 repeats**

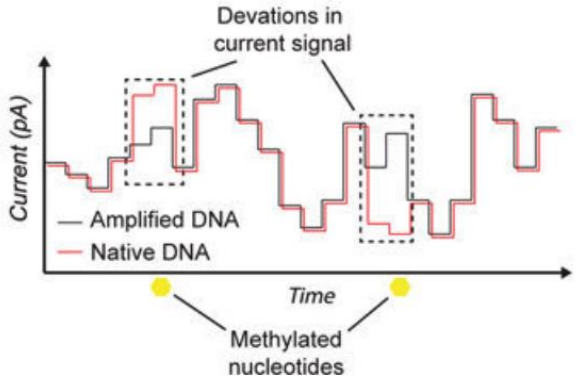
Amyotrophic lateral sclerosis: *c9orf72 Repeat Expansion*



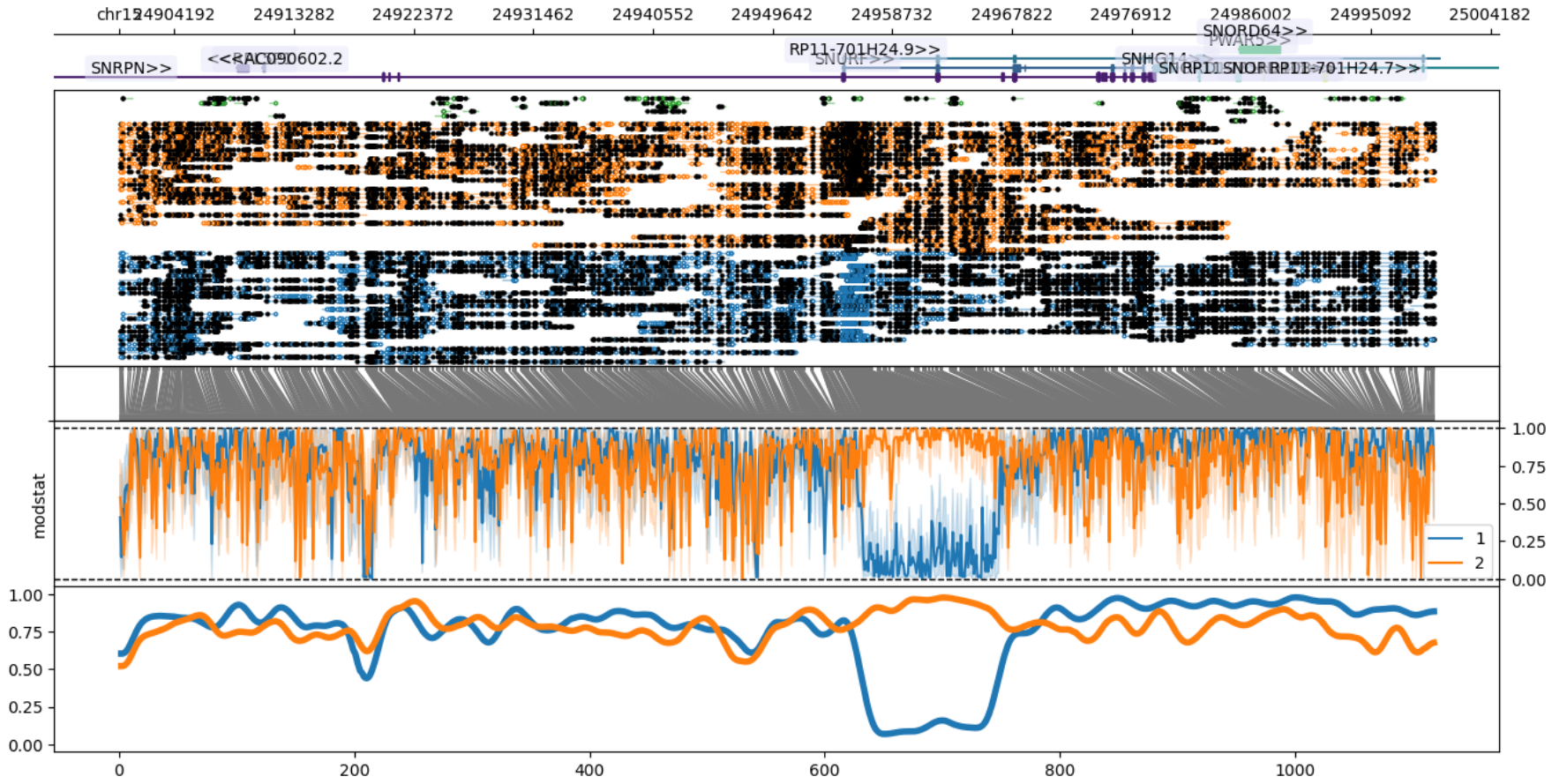
Very long G4C2 hexanucleotide repeat expansion (around 10kb)

Haplotype-Phased DNA Methylation Calling





Plot: MethylArtist

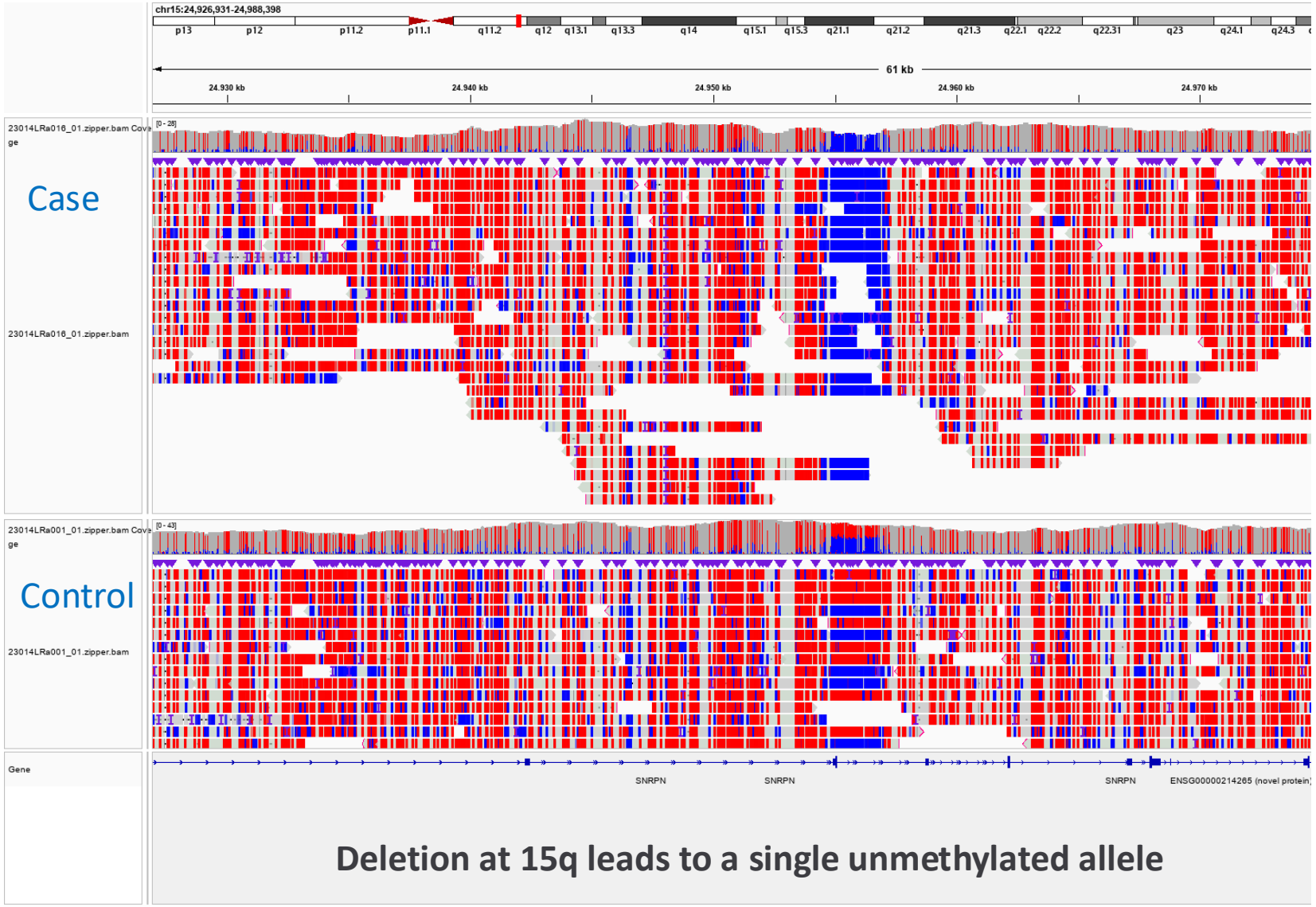


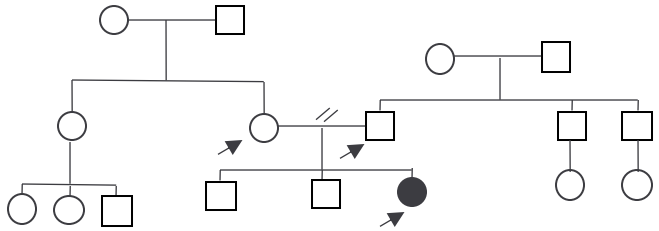
Imprinting Locus

Caspar Gross

Autism Spectrum Disorder: Hypomethylation of SNRPN

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN





Phenotypes:

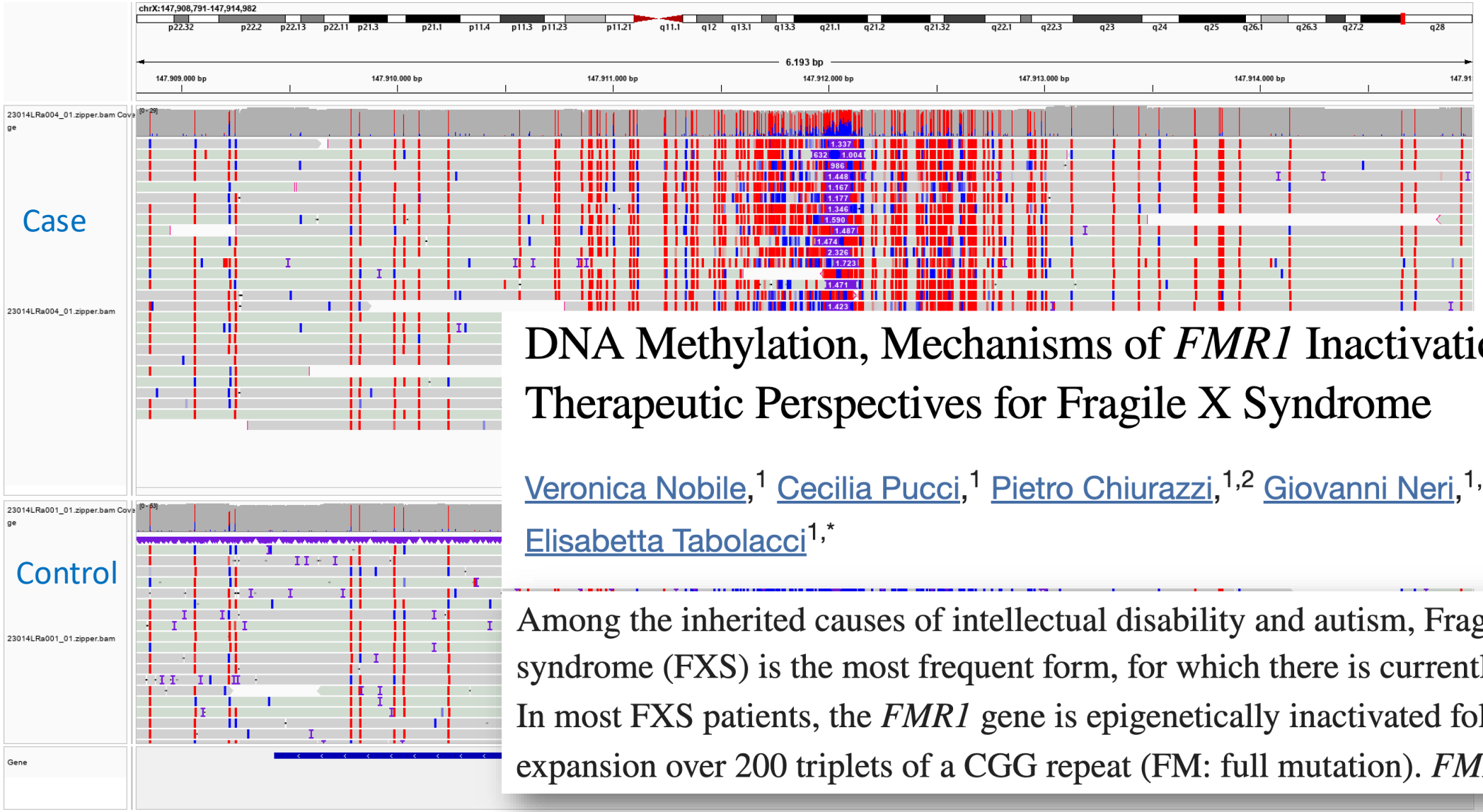
- developmental delay
- muscular hypotonia
- microcephaly
- behavioral abnormalities

Deletion at 15q leads to a single unmethylated allele

Fragile X Syndrome: FMR1 Inactivation by Methylation

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

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

Discovery of a Novel Repeat Expansion Genes

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

nature genetics

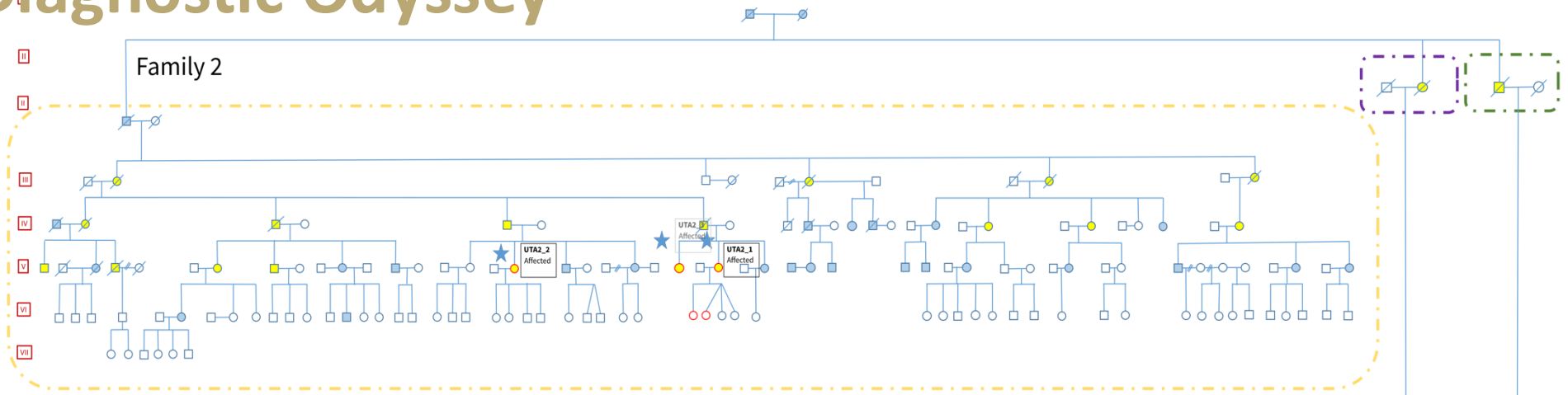
Letter | Published: 29 April 2024

A GGC-repeat expansion in *ZFHX3* encoding polyglycine causes spinocerebellar ataxia type 4 and impairs autophagy

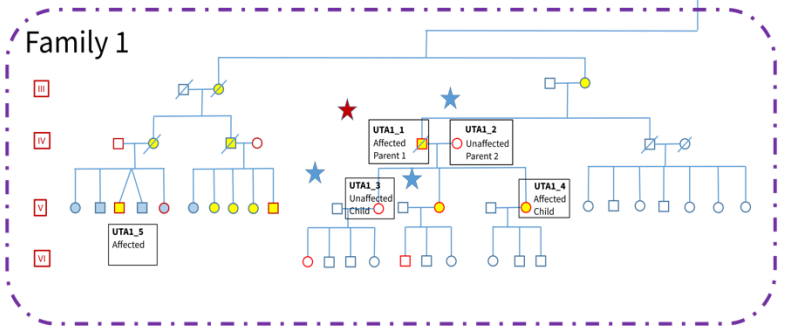
Karla P. Figueroa, Caspar Gross, Elena Buena-Atienza, Sharan Paul, Mandi Gandelman, Naseebullah Kakar, Marc Sturm, Nicolas Casadei, Jakob Admard, Joohyun Park, Christine Zühlke, Yorck Hellenbroich, Jelena Pozojevic, Saranya Balachandran, Kristian Händler, Simone Zittel, Dagmar Timmann, Friedrich Erdlenbruch, Laura Herrmann, Thomas Feindt, Martin Zenker, Thomas Klopstock, Claudia Dufke, Daniel R. Scoles, Arnulf Koeppen, Malte Spielmann, Olaf Riess ✉, Stephan Ossowski, Tobias B. Haack & Stefan M. Pulst ✉

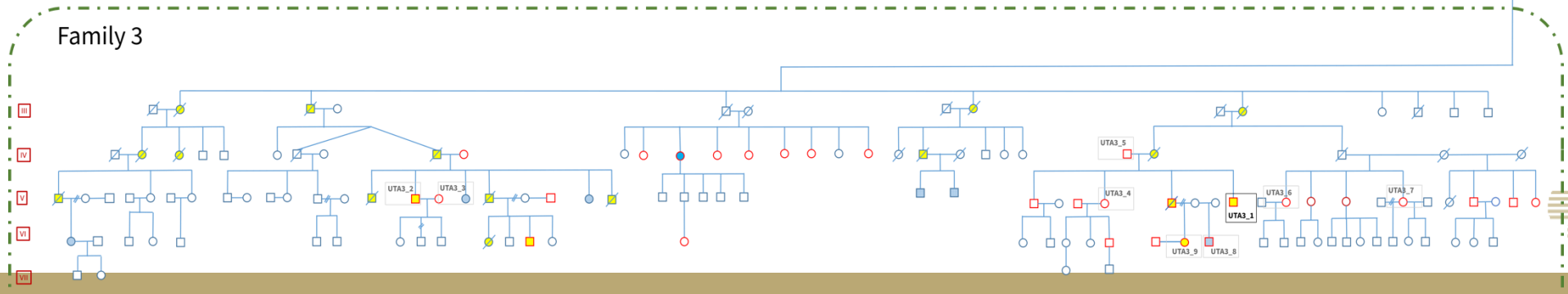
Spinocerebelar Ataxia Type 4: Ending a 25-Year Diagnostic Odyssey

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



- Multi-Generation family from Utah (Swedish ancestry) with dominant SCA4 (many cases)
- Linkage on 16q, but causal gene unknown
- 6 cases + 2 Controls selected for long-read sequencing (PacBio-HiFi and ONT-Nanopore)

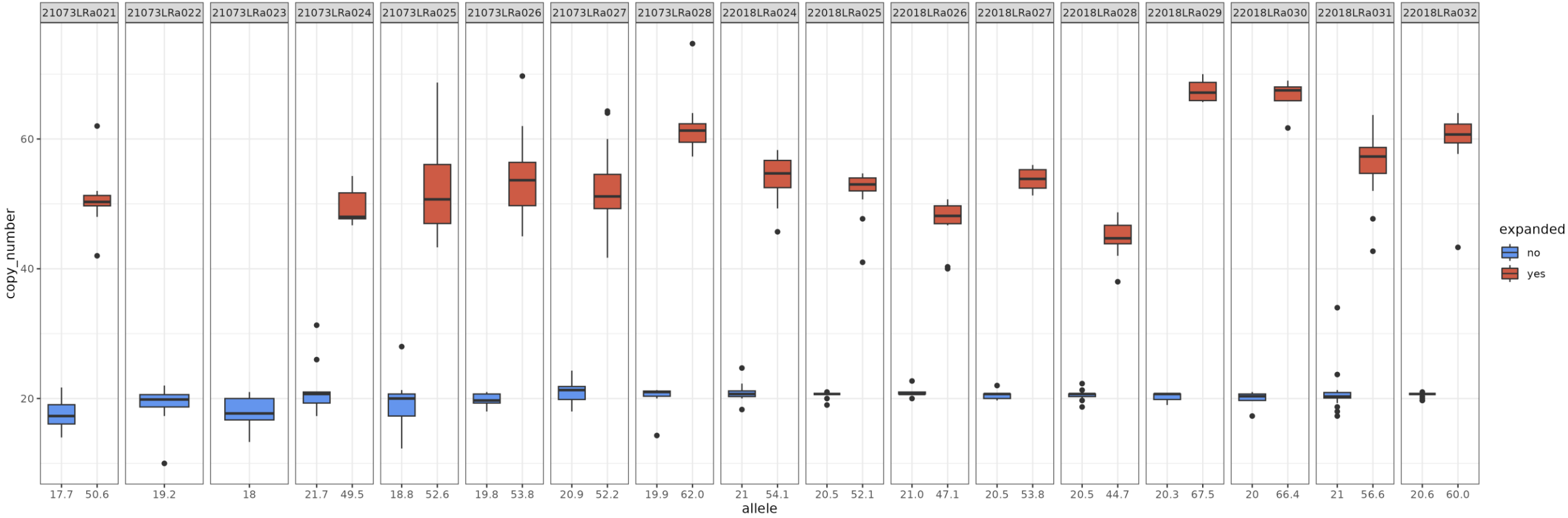




Repeat Length Analysis using Nanopore Reads with Straglr

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Minimap2 + Straglr + in-house visualization



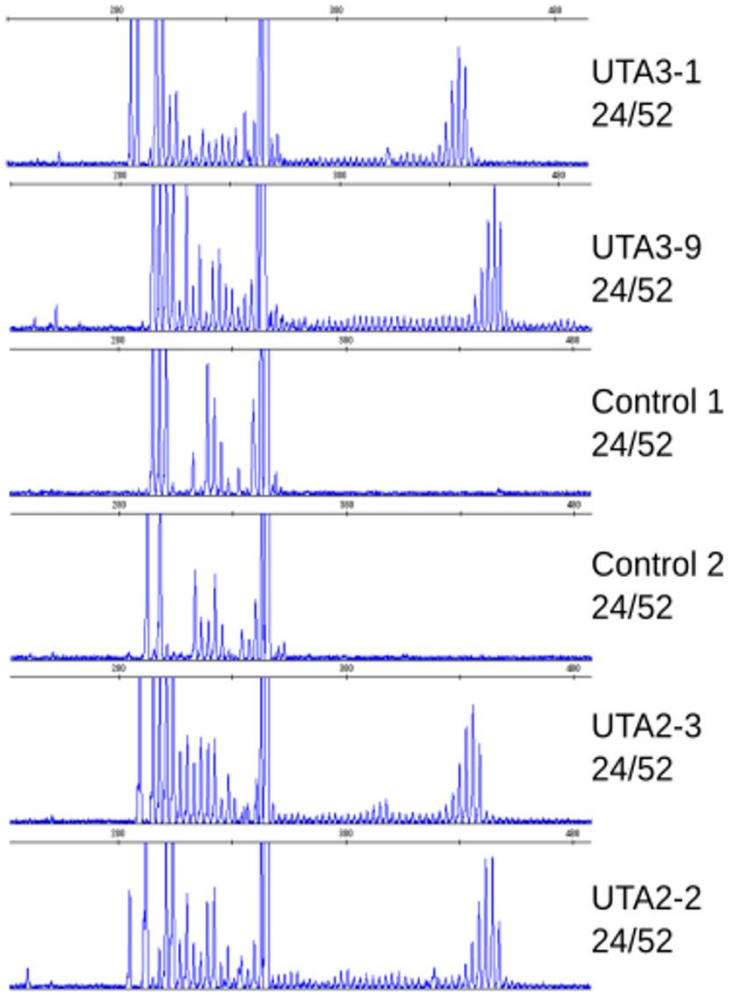
- We identified a heterozygous GGC repeat expansion in gene ZFHX3 in all cases
- Pathogenic repeat length is >42 repeats

SCA4: Pathogenic GGC Repeat Expansion in ZFHX3

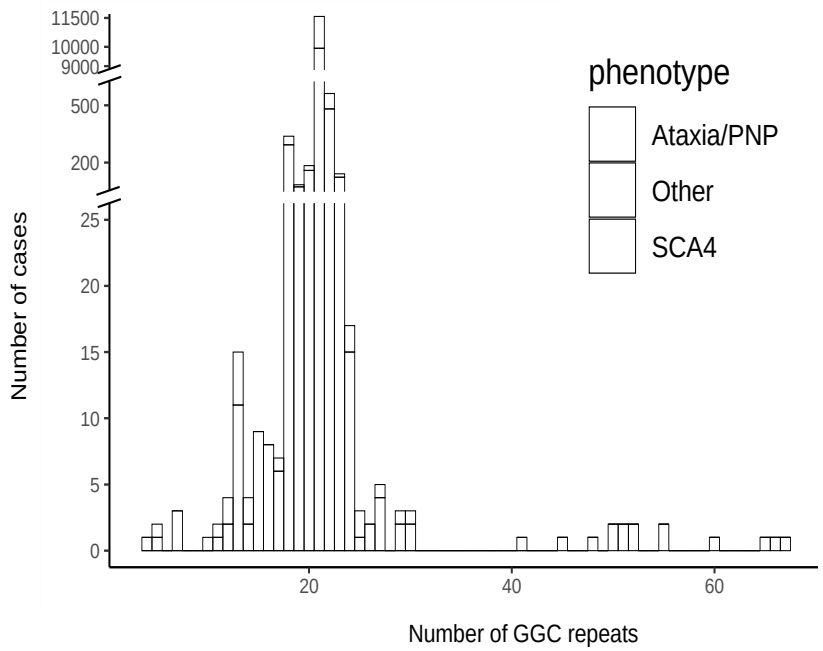
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Discovery: HiFi + TRGT

Validation (PCR amplification)



Pathogenic repeat length



News & Views | Published: 29 May 2024

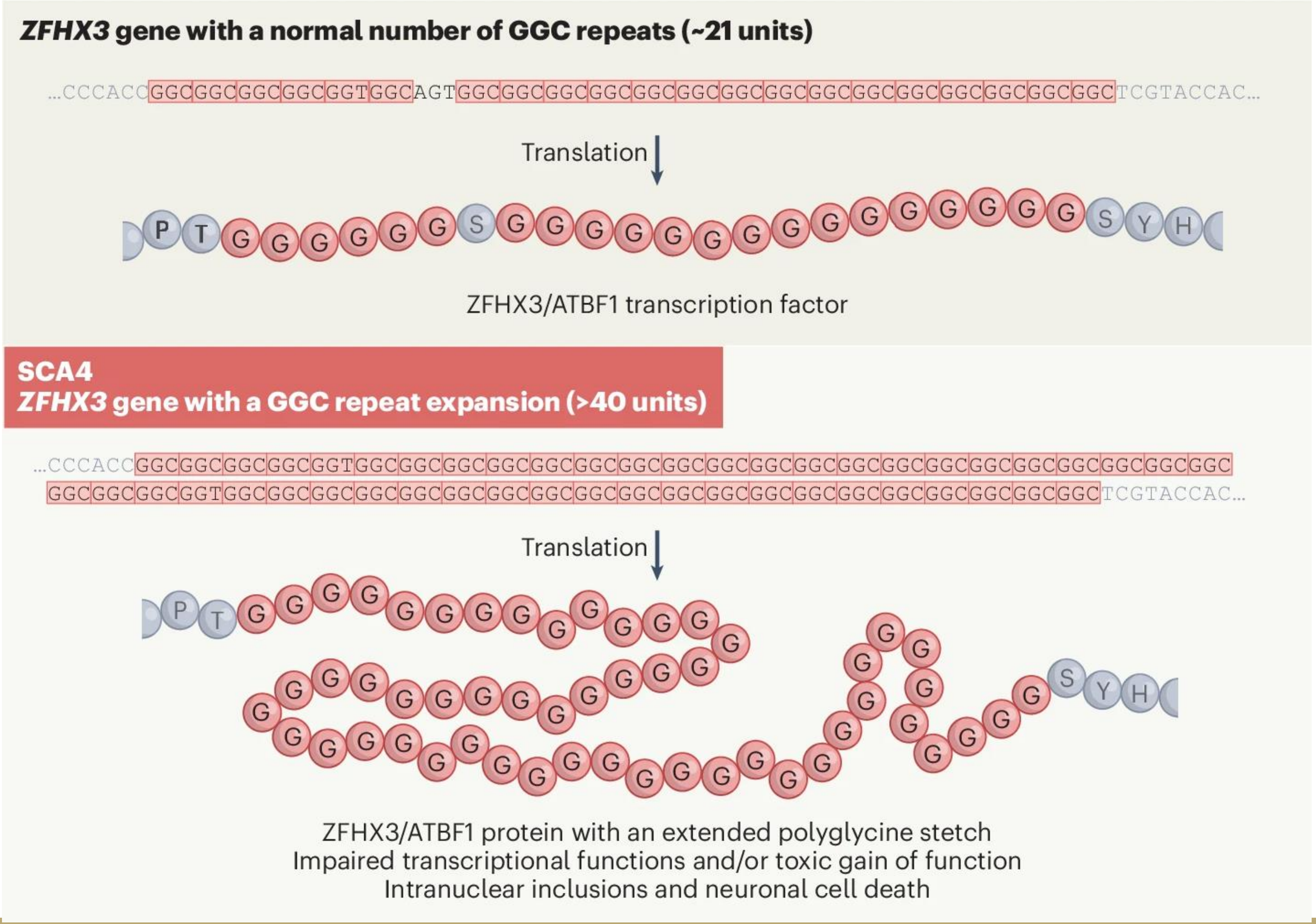
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Neurodegeneration

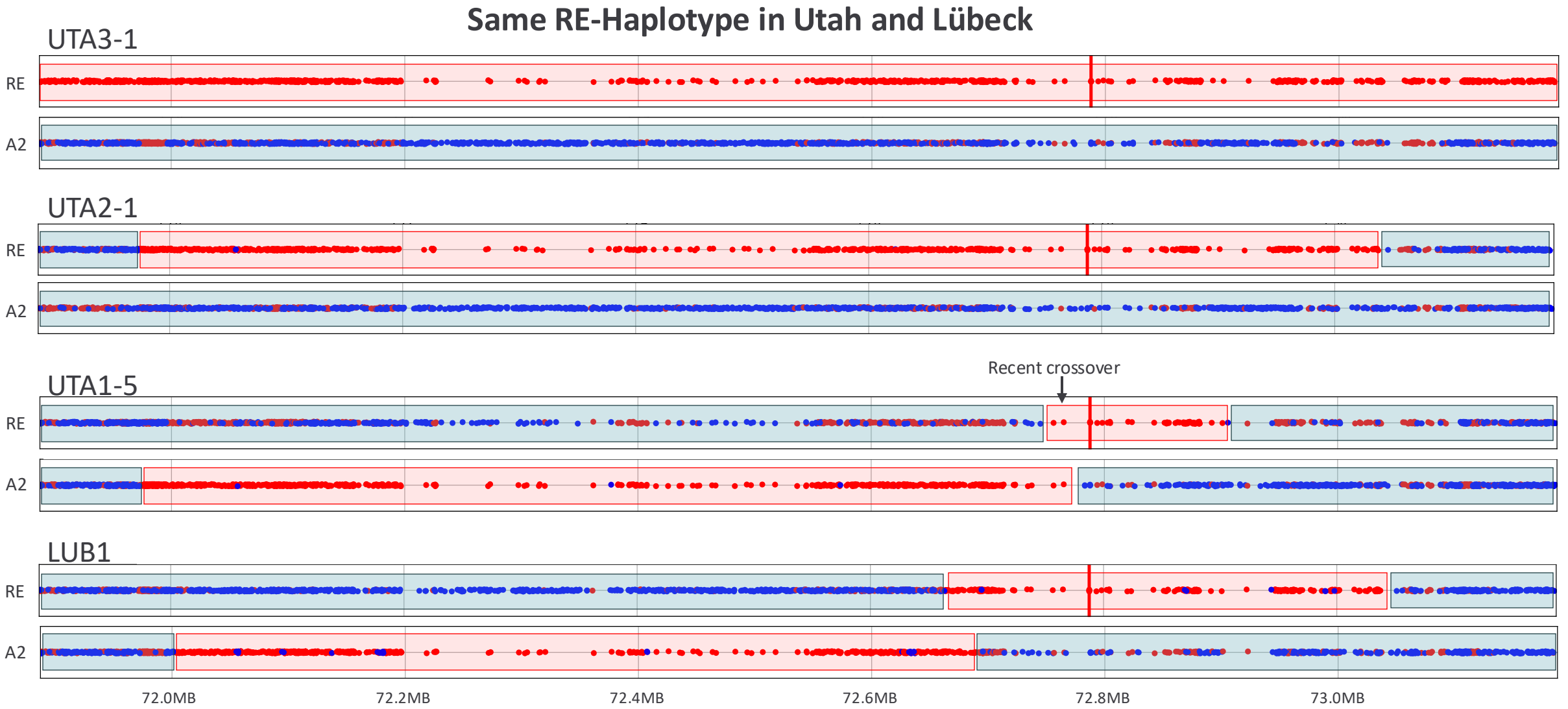
An unexpected polyglycine route to spinocerebellar ataxia

Nicolas Charlet-Berguerand

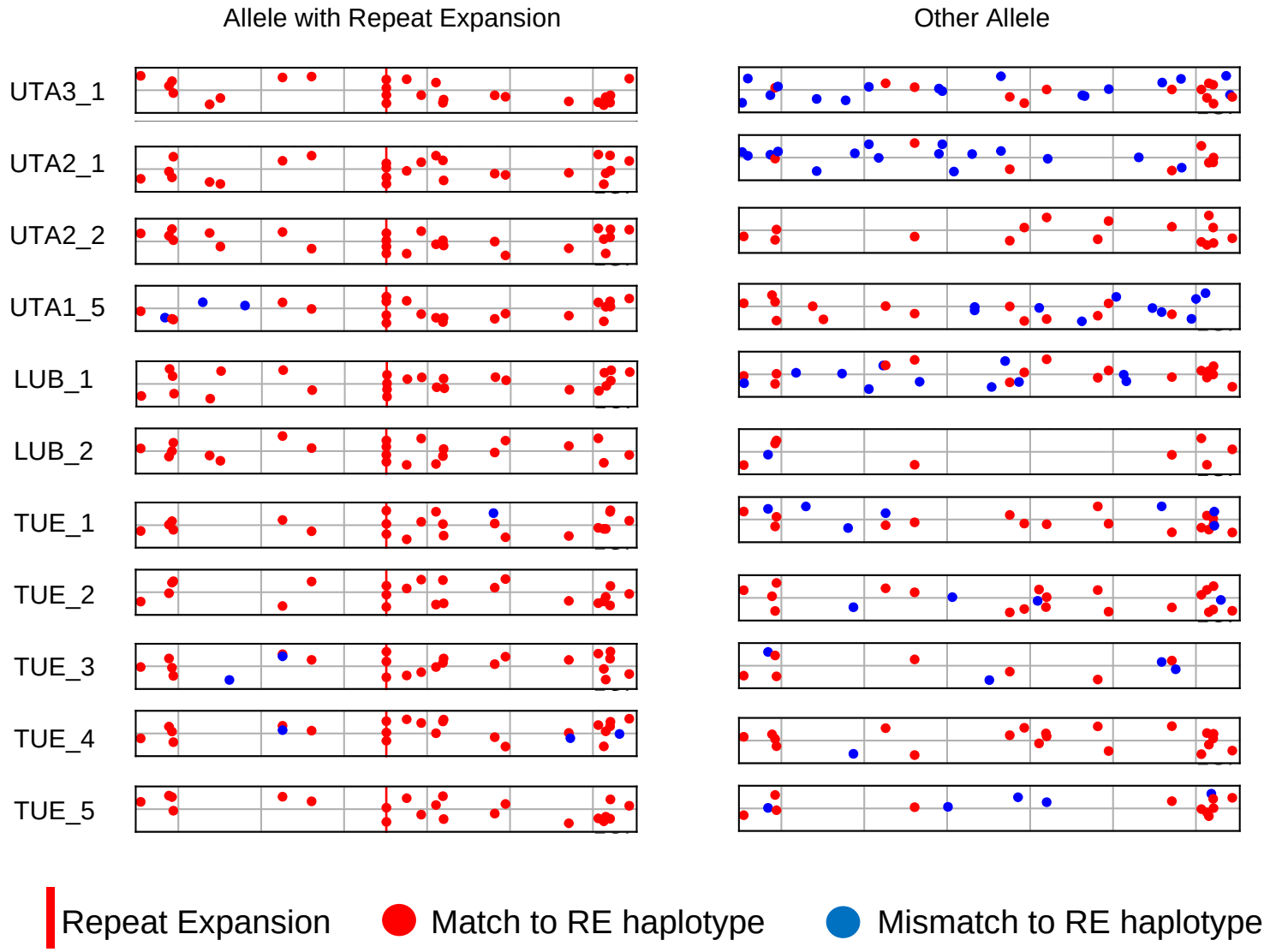
Nature Genetics (2024) | Cite this article



Independent Discovery in 2 Cases of Lübeck (Spielman Lab)



Characteristics SNVs Only Found in RE-Linked Haplotype

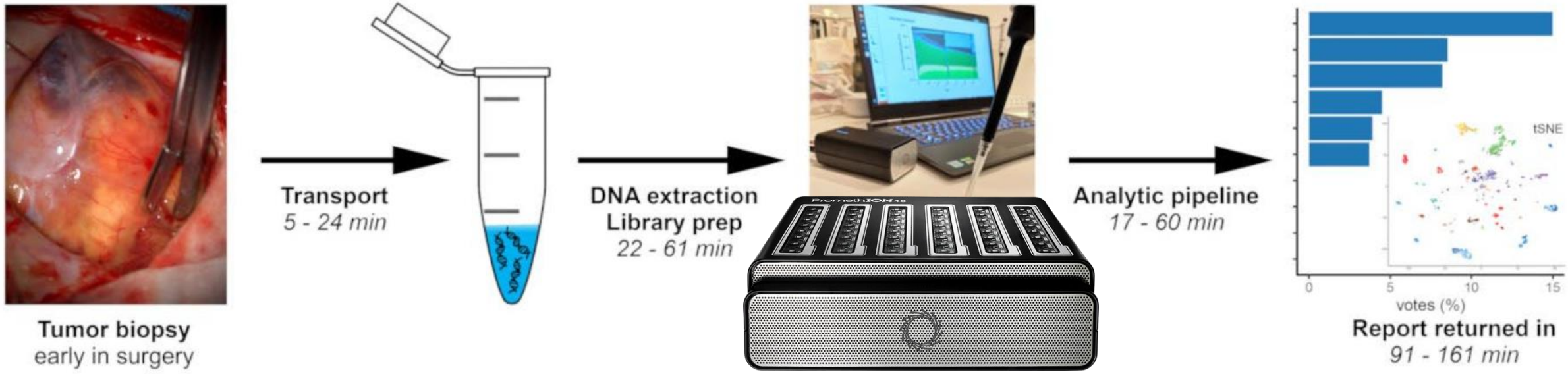


6 Ultra-rare SNVs are specific to the Repeat Expansion Allele

Screening in 6,495 srGS Datasets from U. Tübingen using Unique SNVs & ExpansionHunter: 7 more cases identified

				Characteristc variants flanking repeat expansion					
Sample	Disease status	Repeat length	Seq-Tech	chr16:72734457 A>G	chr16:72737703 T>C	chr16:72787719 A>G	chr16:72787737 A>G	chr16:72787739 T>C	chr16:72787743 A>G
UTA2	Unaffected	normal	lr	no	no	no	no	no	no
UTA3	Unaffected	normal	lr	no	no	no	no	no	no
UTA1	Affected	expanded	lr	yes	yes	yes	yes	yes	yes
UTA4	Affected	expanded	lr	yes	yes	yes	yes	yes	yes
UTA5	Affected	expanded	lr	yes	yes	yes	yes	yes	yes
UTA6	Affected	expanded	lr	yes	yes	yes	yes	yes	yes
UTA7	Affected	expanded	lr	yes	yes	yes	yes	yes	yes
UTA8	Affected	expanded	lr	yes	yes	yes	yes	yes	yes
LUB1	Affected	expanded	lr	yes	yes	yes	yes	yes	yes
LUB2	Affected	expanded	lr	yes	yes	yes	yes	yes	yes
TUB1	Affected	expanded	sr	yes	no	yes	yes	yes	yes
TUB2	Affected	expanded	sr	yes	no	yes	yes	yes	yes
TUB3	Affected	expanded	sr	yes	no	yes	yes	yes	yes
TUB4	Affected	expanded	sr	yes	no	yes	yes	yes	yes
TUB5	Affected	expanded	sr	yes	no	yes	low quality	low quality	yes
TUB6	Affected	expanded	sr	yes	no	yes	yes	yes	yes
TUB7	Affected	expanded	sr	yes	no	yes	yes	yes	yes

LeOPARD: Intraoperative classification of tumors

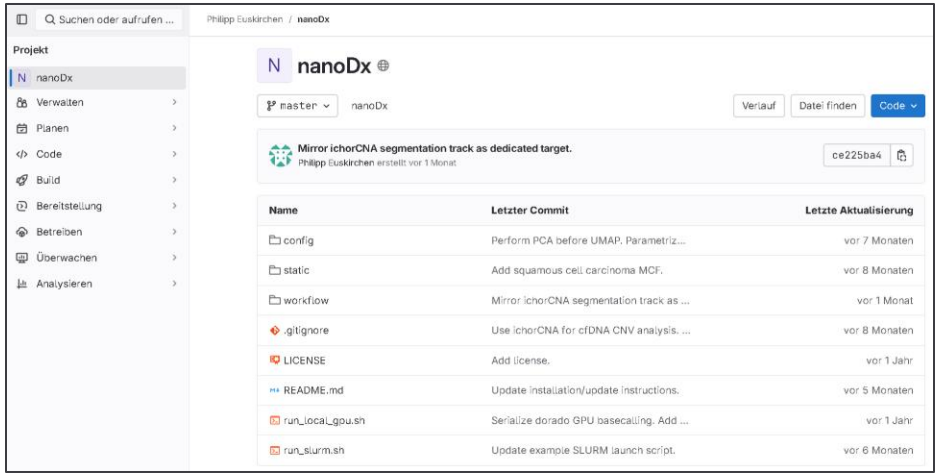


Goal:

- Classification of tumor subtypes within 2 hours
- Adjustment of surgery strategy during surgery based on results

Coordination: Philipp Euskirchen, Charité Berlin

nanoDX pipeline: methylation-based classification of tumors



https://gitlab.com/pesk/nanoDx

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Follow this preprint

crossNN: an explainable framework for cross-platform DNA methylation-based classification of cancer

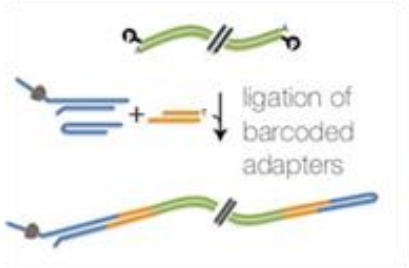
Dongsheng Yuan, Robin Jugas, Petra Pokorna, Jaroslav Sterba, Ondrej Slaby, Simone Schmid, Christin Siewert, Brendan Osberg, David Capper, Pia Zeiner, Katharina Weber, Patrick Harter, Nabil Jabareen, Sebastian Mackowiak, Naveed Ishaque, Roland Eils, Sören Lukassen, Philipp Euskirchen

doi: https://doi.org/10.1101/2024.01.22.24301523

nanoDX:

- Uses low-pass Nanopore whole genome sequencing data
- Classifies tumor types based on methylation profile
- Two neural networks available: Brain-tumors (Capper et al.) and Pan-Cancer

nanoDX pipeline: round robin test



DNA QC

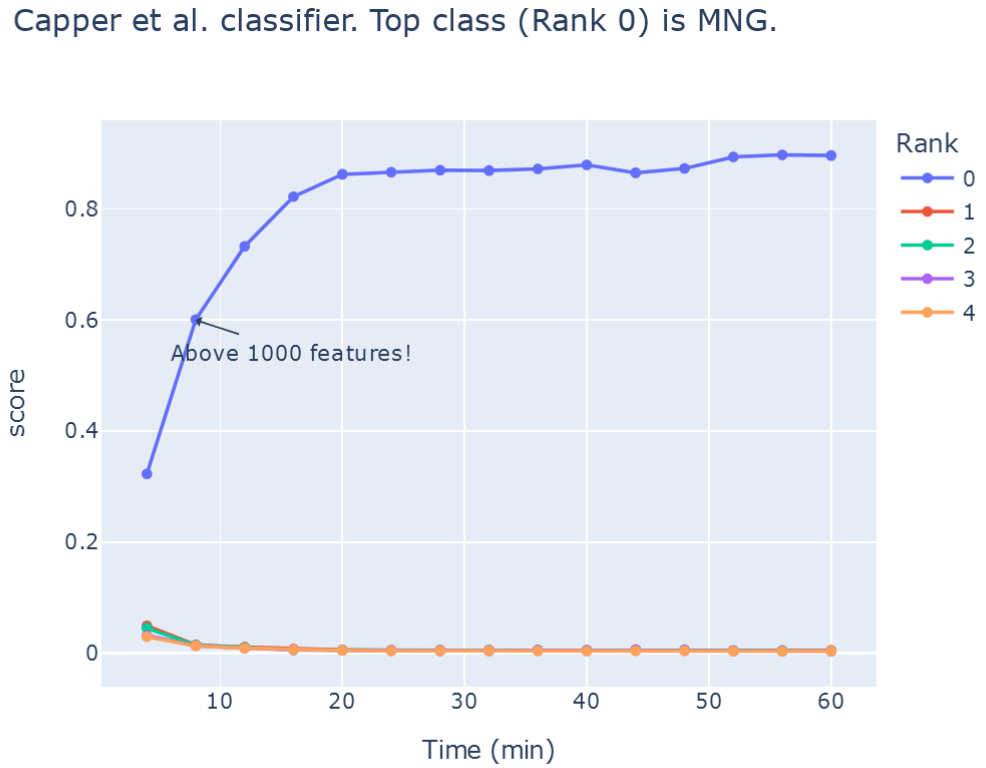
(5 min)

library construction

(15 min)

sequencing and analysis

(6 h)



- Samples: DNA derived from brain tumor tissue
- Realtime analysis using nanoDx (with neural network for brain tumors)
- crossCNN updates results every minute

Tübingen: Julian Broche, Miguel Camarena Sainz

nanoDX pipeline: Clinical Report

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

nanopore low-pass whole genome sequencing report

Sample ID: bams_ALL

Quality control metrics

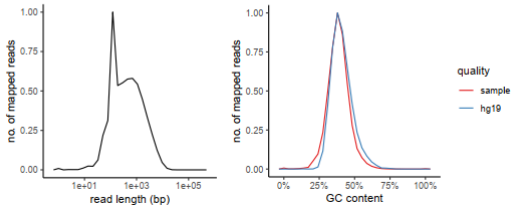
Barcode statistics Demultiplexing statistics are shown for quality control only using a maximum of randomly subsampled 100,000 reads. All reads are considered for downstream analysis.

```
Adapters detected in 811 of 100000 reads
  NBD104/NBD114    811: |                    |   0.81 %
           none  86275: |    ################ |  86.28 %
Barcodes detected in 811 of 100000 adapters
      barcode01    810: |                    |   0.81 %
      barcode16      1: |                    |   0.00 %
           none  86275: |    ################ |  86.28 %
12914 reads were skipped due to the min. length filter.
Demultiplexing finished in 149.50s
```

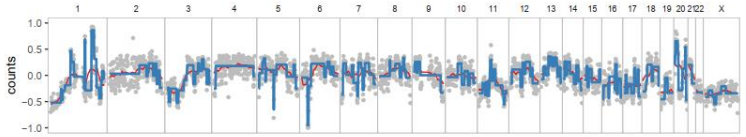
Read statistics

```
General summary:
Average percent identity:            96.9
Fraction of bases aligned:            1.0
Mean read length:                 1,191.0
Mean read quality:                   14.9
Median percent identity:             97.9
Median read length:                 573.0
Median read quality:                 16.6
Number of reads:              1,225,509.0
Read length N50:                  2,543.0
STDEV read length:                1,737.2
Total bases:              1,459,604,278.0
Total bases aligned:      1,435,112,548.0
Number, percentage and megabases of reads above quality cutoffs
>Q10:   1208064 (98.6%) 1443.3Mb
```

Mean genome coverage is 0.46X.



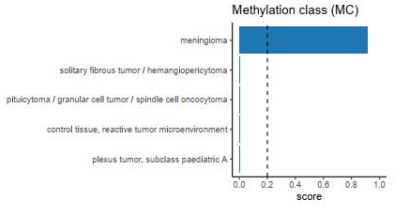
Copy number profile

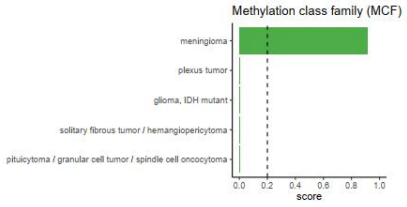


Methylation-based classification

Methylation-based classification is based on **83061** CpG sites (overlap of sites covered in this sample and the model). At the methylation class (MC) level, the sample has been classified as **meningioma**. This prediction has a confidence score of **0.911**. At the methylation class **family** (MCF) level, the sample has been classified as **meningioma**. The MCF prediction has a confidence score of **0.911**.

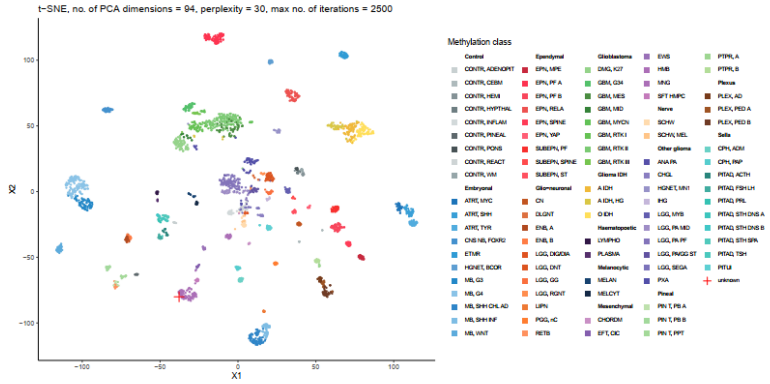
Scores for the Top 5 entities on MC and MCF level are given below. Vertical dashed lines indicate the recommended >0.2 cut-off for classification.





Dimensionality reduction plot

Dimensionality reduction plots are only intended for visual quality control, not classification, and must be interpreted with caution.



Disclaimer Methylation-based classification using nanopore whole genome sequening is a research tool currently under development. It has not been clinically validated in sufficiently large cohorts. Interpretation and implementation of the results in a clinical setting is in the sole responsibility of the treating physician.

(Report generated on 2024-10-02 09:22:41.916161. Pipeline version: v1.0rc3-9-gce225ba, ce225ba)

Thanks!

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

University Hospital Aachen
Ingo Kurth
Florian Kraft
Sebastian Gießelmann

Charité – Universitätsmedizin Berlin
Nadja Ehmke

Berlin Institute of Health
Janine Altmüller
Manuel Holtgrewe
Claudia Quedenau

Medical School Hannover
Bernd Auber
Gunnar Schmidt

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

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)

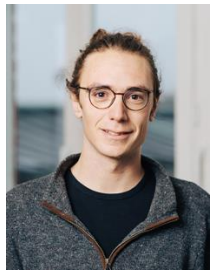
Oxford Nanopore Technologies
Manuela Saathoff
Tonya McSherry
Gerald Goh
Alexander Vogel
Anthony Doran
Alexander Rotmann

NVIDIA
Uwe Samer
Harry Clifford
Lotfi Slim

lonGER Kickoff-Meeting Frankfurt, April 2023

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Elena Buena-Atienza



Caspar Gross

Jakob Admard

Posters:

P15.088.A

P21.060.C

Talk: Tobias Haack (tomorrow)

Nanopore Experts in Tübingen

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

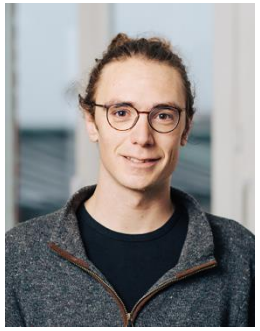
Elena Buena-Atienza

Nanopore Sequencing,

wet-lab and analysis

Marc Sturm

Lead Diagnostic Bioinformatics



Caspar Gross

Long-read bioinformatics

method development and analysis

Leon Schütz

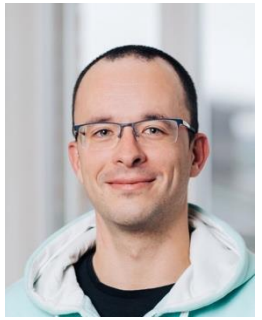
Diagnostic Nanopore Sequencing

megSAP / Gsvar Software

Jakob Admard

Long-read bioinformatics

method development and analysis



German Demidov

ERDERA Rare-Disease Consortium

Nanopore Workpackage