Visualize segmental duplication regions with Paraviewer

Abstract # 4109W

Jonathan R. Belyeu, Xiao Chen, Tom Mokveld, Guillerme de Sena Brandine, Zev Kronenberg, Christopher T. Saunders, Michael A. Eberle

PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Paraviewer enhances interpretation of Paraphase results

Paraphase

- Many medically relevant genes fall in regions where high sequence homology (paralogs/pseudogenes) limits reference-based variant calling.
- Paraphase[1] phases haplotypes for genes in the same family, determines copy numbers, and reports phased variant calls.
- Paraphase reports variant calls in 160 segmental duplications, including 11 medically relevant regions.

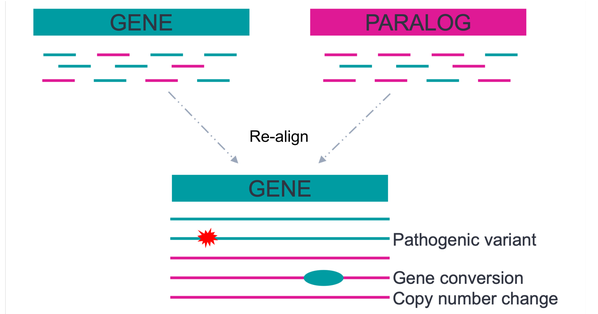

Figure 1: Paraphase enables variant calling in repetitive gene regions. Reads are re-aligned from multiple copies of highly identical genes, variants are called in individual haplotypes, and copy number changes are identified.

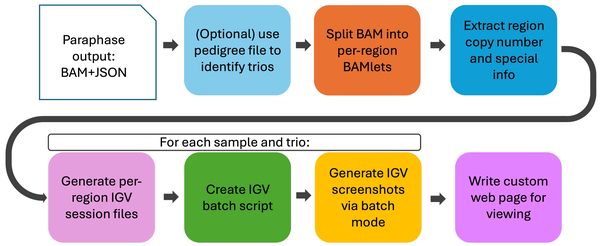

Figure 2: Paraviewer helps view and prioritize Paraphase results.

- Runs on outputs from Paraphase.
- Identifies copy number and putative pathogenic variants.
- Generates IGV screenshots (for a quick look at alignments).
- Creates simple web-based table for user review.
  - Each region has clickable interactive IGV.js session for a deeper dive.
  - Includes numerous options for filtering and viewing.
  - If pedigree file included, allows viewing by trio or sample.

Paraviewer with WGS data

Platinum Pedigree

- We ran Paraphase on a set of 22 WGS HiFi samples from the 4-generation Platinum Pedigree, a public multi-technology sequencing data cohort without known genetic disease.[2]
- Many samples in the pedigree harbor copy number alterations, including in medically relevant genes (see **Figure 3A-B**).
- Paraviewer allows review of these regions to identify unaltered gene copies (see **Figure 3C**). The current full set of 160 Paraphase regions are included.

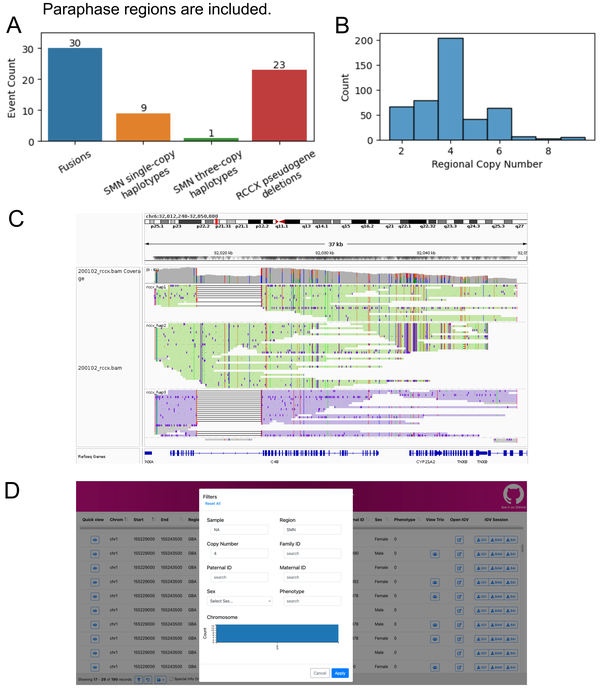

Figure 3: Paraviewer simplifies visual inspection of Paraphase results.
A. Copy number of Platinum Pedigree samples in Paraphase regions.
B. Counts of region-region fusions and copy number changes in *SMN*+*RCCX* illustrate types of common variability.
C. Sample 200102: 1-copy deletion in *RCCX*. Paraviewer shows two gene copies: hap1 and hap3 (spanning *TNXB*). One pseudogene copy (hap2, truncated at *TNXB*); other pseudogene copy is lost.
D. Viewer filtering tool, with filters for Sample, Region, Copy Number, etc.

Paraviewer PureTarget Carrier Pipeline (PTCP) mode

- PacBio's new PureTarget Carrier Panel includes medically relevant gene targets[3].
- Seven of these gene targets (*CYP21A2*, *F8*, *GBA*, *HBA*, *HBB*, *RPGR*, and *SMN1*) are supported by Paraviewer for enhanced visualization.
- We created a Paraviewer review site for 19 public samples with known carrier status for variants in the *SMN1* and *CYP21A2* regions.

Region	Variant	Samples
SMN1	2 copy loss (affected)	NA09677, NA10684, NA00232, NA03813
SMN1	1 copy loss (carrier)	HG00281, HG00324, HG01686, HG01773, HG01892, HG00346, HG01612, HG03583, NA20787
CYP21A2	1 copy gene→pseudogene conversion	HG00544
CYP21A2	V282L point mutation	HG01887, NA19030
CYP21A2	Q319X point mutation	HG04115, HG01071, HG02717

Table 1: PureTarget Carrier Pipeline samples with *CYP21A2* or *SMN1* variants. 19 total samples were included in this Paraviewer analysis, including 4 with complete loss of *SMN1*, 9 with single copy loss of *SMN1*, and 6 with *CYP21A2* carrier status from gene conversions or point mutations.

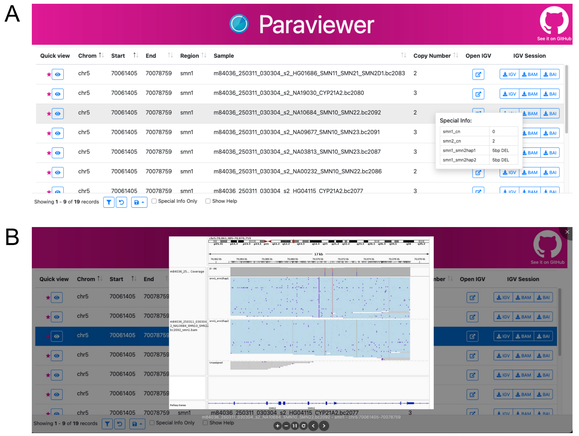

Figure 4: 2-copy deletion of *SMN1*.
A. Sample NA10684 has two copies of *SMN2* and total loss of *SMN1*, shown in hover-text by the Paraviewer region table.
B. In-browser IGV shows two *SMN2* haplotypes, none for *SMN1*. A small number of unphased reads are also shown.

Try it out!

Platinum Pedigree subset demo site:
https://pacificbiosciences.github.io/ParaviewerWGSDemo

- 22 WGS HiFi family samples - with pedigree.
- 5 regions:
  - *GBA* (Gaucher disease, expected CN=4)
  - *NEB* (6, nemaline myopathy, expected CN=6)
  - *PMS2* (2, Lynch Syndrome, expected CN=2)
  - *RCCX* (CAH/congenital adrenal hyperplasia., expected CN=4)
  - *SMN1* (SMA/spinal muscular atrophy, expected CN=4)
- **Site highlights:**
  - WGS Paraviewer results
  - Non-disease/control samples
  - Trio and single-sample review

Public PureTarget Carrier Pipeline demo site:
https://pacificbiosciences.github.io/ParaviewerPTCPDemo

- 19 PureTarget HiFi samples.
- 2 regions:
  - *CYP21A2* (CAH/congenital adrenal hyperplasia., expected CN=4)
  - *SMN1* (SMA/spinal muscular atrophy, expected CN=4)
- **Site highlights:**
  - PureTarget Paraviewer results
  - Genomic disease cases and carrier genomes
  - Single-sample review

Learn more about PureTarget
CoLab Theater 1: Wed 3:15-3:45pm

Paraviewer is now available for early users at
https://github.com/PacificBiosciences/paraviewer

Check out these other posters: 2051F (Platinum Pedigree, Zev Kronenberg), 4050F (Mitorsaw, Matt Holt), 4104F (D4Z4 (Xiao Chen).

References:

1. Chen, X., *et al.* Genome-wide profiling of highly similar paralogous genes using HiFi sequencing. *Nat Commun* (2025).
2. Kronenberg, Z., *et al.* The Platinum Pedigree: a long-read benchmark for genetic variants. *Nat Methods* (2025).
3. https://www.pacb.com/technology/puretarget/