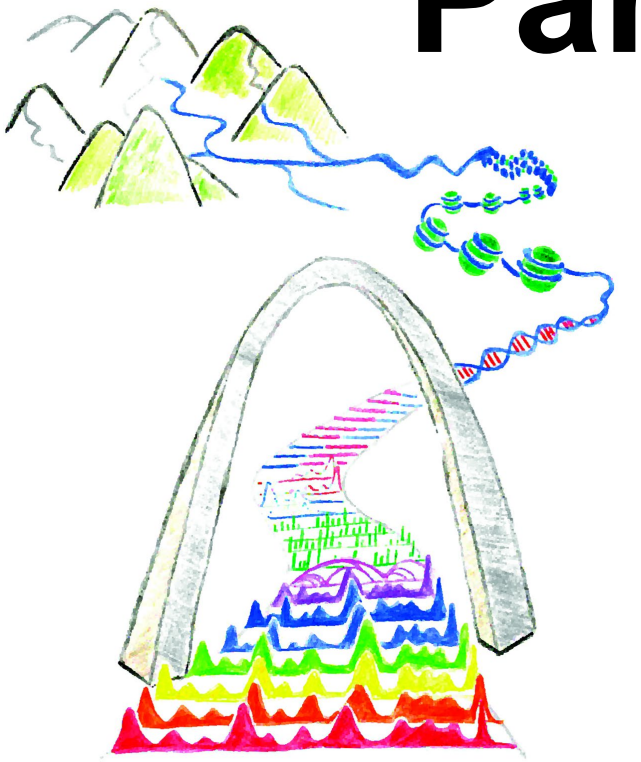
Genome and Pangenome Assembly



Juan F. Macias-Velasco, PhD

Wang lab

Department of Genetics

Washington University in St. Louis

Chad Tomlinson

Tina Lindsay

Outline

- Linear genome assembly
  - Definition of long read sequencing and its applications.
  - Explanation of genome assembly and how long reads improve assemblies.
  - Examples of commonly used long read genome assembly algorithms.
  - Overview of the pb-assembly (Falcon-Unzip) assembly algorithm.

- Pangenome assembly
  - Justification
  - Definition of genome graphs
  - Different approaches to construction
  - Utility of genome graphs

Evolution of DNA Sequencing

First Generation

Sanger Sequencing
Maxam and Gilbert
Sanger Chain-termination

- Infer nucleotide identity using dNTPs then visualize with electrophoresis
- 500-1000 bp fragments

Second Generation
(Next Generation Sequencing)

454, Solexa,
Ion Torrent
Illumina

- High throughput from the parallelization of sequencing reactions
- ~50-500 bp fragments

Third Generation

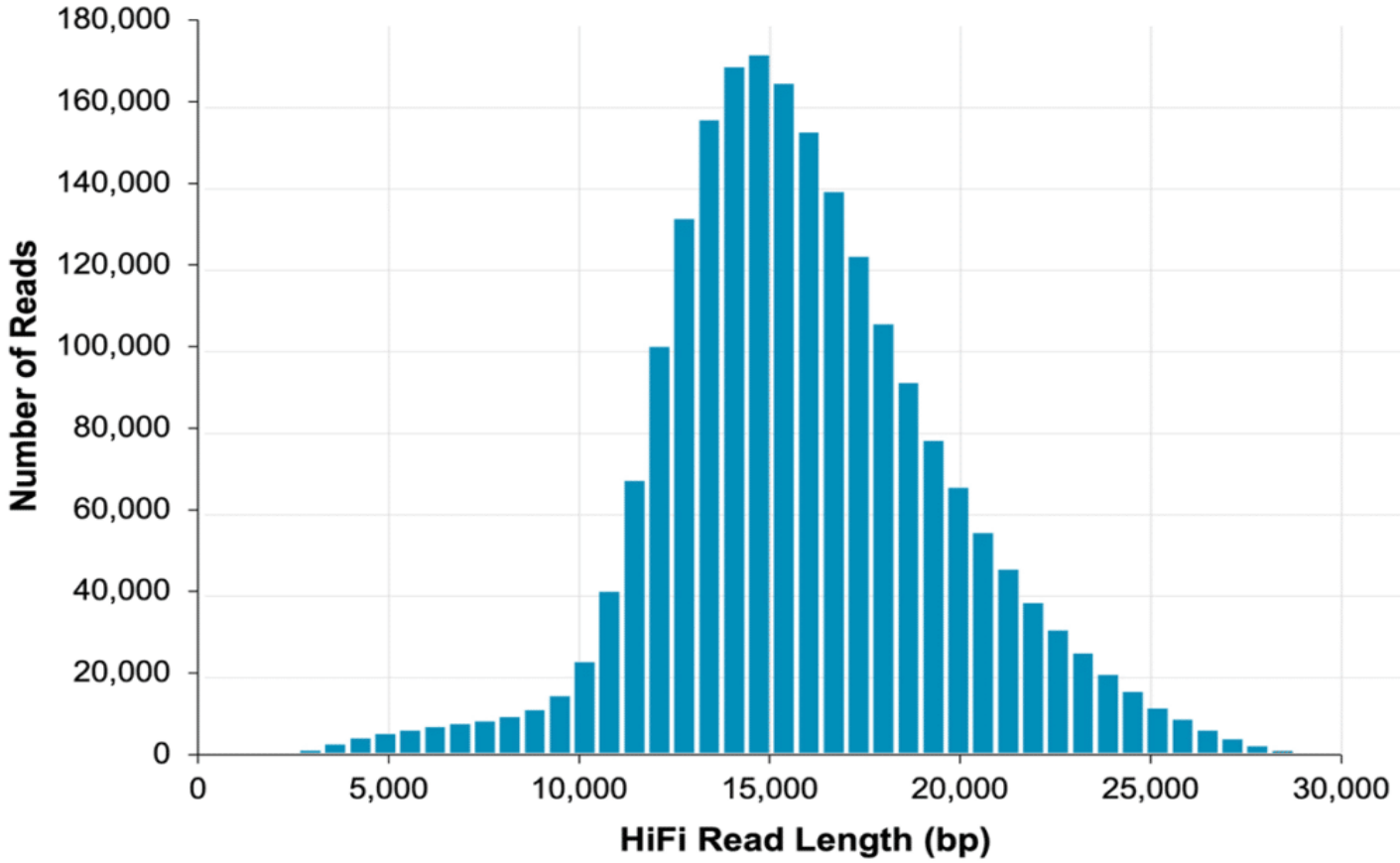
PacBio
Oxford Nanopore

- Sequence native DNA in real time with single-molecule resolution
- Tens of kb fragments, on average

Short-read sequencing

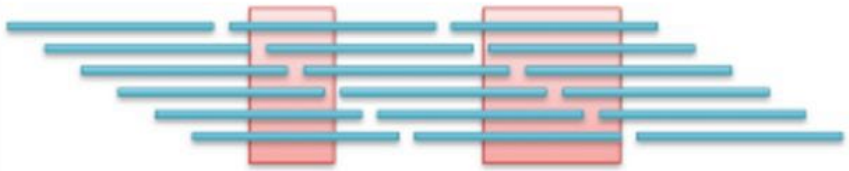
Long-read sequencing

PacBio HiFi Read Length Distribution



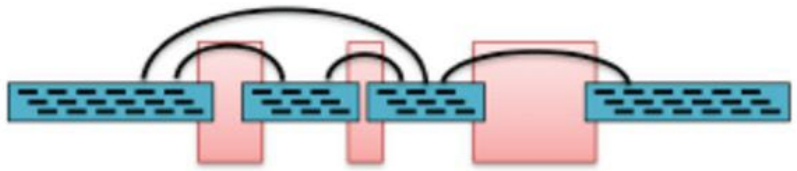
Long Read Analysis Applications

a) De novo Assembly



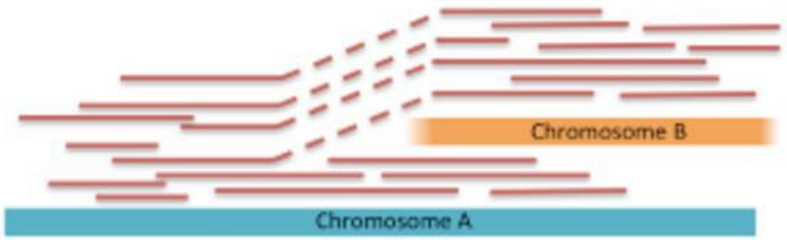
Reconstruct the genome sequence directly from the sequenced reads (blue). Longer reads will span more repetitive elements (red), and produce longer contigs.

b) Chromosome Scaffolding



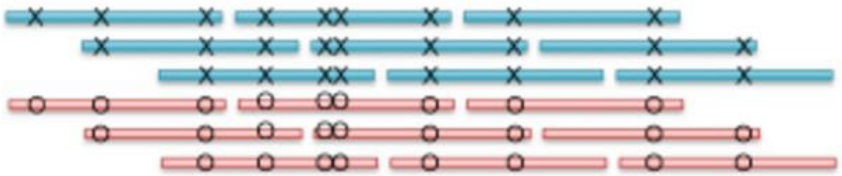
Order and orient contigs (blue) assembled from overlapping reads (black) into longer pseudo-molecules. Longer spans are more likely to connect distantly spaced contigs, especially those separated by long repeats (red).

c) Structural Variation Analysis



Identify reads/spans (red) that map to different chromosomes or discordantly within one. The longer the read/span, the more likely to capture the SV, and will have improved mappability to resolve SVs in repetitive element.

d) Haplotype Phasing



Link heterozygous variants (X/O) into phased sequences representing the original maternal (red) and paternal (blue) chromosomes. Longer reads and longer spans will be able to connect more distantly spaced variants.

Long Read Sequencing is getting cheaper

Sequencing technology	Platform	Data type	Read length (kb)		Read accuracy (%)	Throughput per flow cell (Gb)		Estimated cost per Gb (US$)	Maximum throughput per year (Gb)[a]
			N50	Maximum		Mean	Maximum		
Pacific Biosciences (PacBio)	RS II[b]	CLR	5–15	>60	87–92	0.75–1.5	2	333–933[c]	4,380
	Sequel	CLR	25–50	>100		5–10	20	98–195[d]	17,520
	Sequel II	CLR	30–60	>200		50–100	160	13–26[e]	93,440
		HiFi	10–20	>20	>99	15–30	35	43–86[e]	10,220
Oxford Nanopore Technologies (ONT)	MinION/GridION	Long	10–60	>1,000	87–98	2-20	30	50–500[f]	21,900 (MinION) 109,500 (GridION)
		Ultra-long	100–200	>1,500		0.5–2	2.5	500–2,000[f]	913 (MinION) 4,563 (GridION)
	PromethION	Long	10–60	>1,000		50–100	180	21–42[f]	3,153,600
Illumina	NextSeq 550	Single-end	0.075–0.15	0.15	>99.9	16–30	>30	50–63[g]	>47,782
		Paired-end	0.075–0.15 (×2)	0.15 (×2)		32–120	>120	40–60[g]	>70,080
	NovaSeq 6000	Single-end	0.05–0.25	0.25		65–3,000	>3,000	10–35[h]	>1,194,545
		Paired-end	0.05–0.25 (×2)	0.25 (×2)					

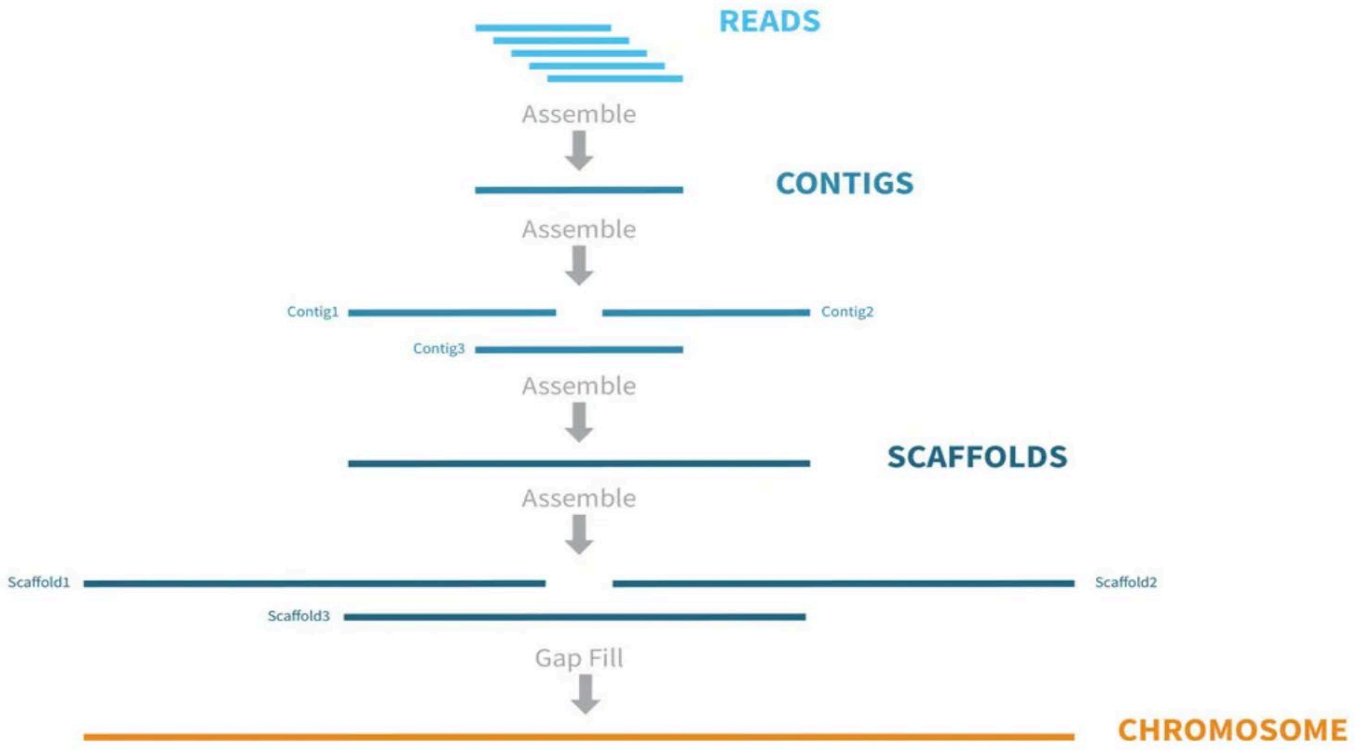
Goals of Genome Assembly

- **Maximize and evaluate completeness of genome sequence for an organism.**
  - As few gaps in the sequence as possible
  - As accurate of a sequence as possible

- **Assemble and organize the sequence.**
  - Into contigs
  - Into scaffolds
  - Into chromosomes

De Novo Genome Assembly Process and Outputs

- Algorithms are used to find accurate overlaps between reads. The consensus of a set of overlapping read sequences is called a contig.
- Contigs are segments of the genome that have gaps in between. In most instances, we do not know how the contigs are to be ordered and oriented in respect to one another.
- We can use additional methods to order and orient contigs, fill in gap regions, and chain contigs together into larger sequences called scaffolds.
- Scaffolding can be an additional step performed by the assembly algorithm, but it usually involves a separate process using BioNano, Hi-C, Oxford Nanopore data. or other data types.
- Additional methods can be used to organize scaffolds into more complete chromosomal level assemblies.
- Contigs and scaffolds are typically available in FASTA and/or FASTQ formats.



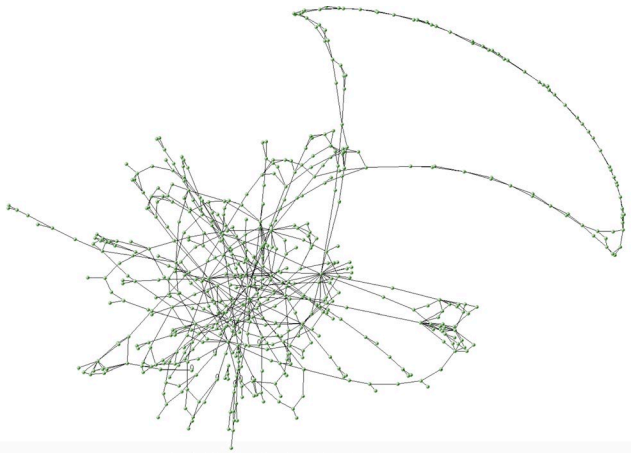
Two main assembly methods

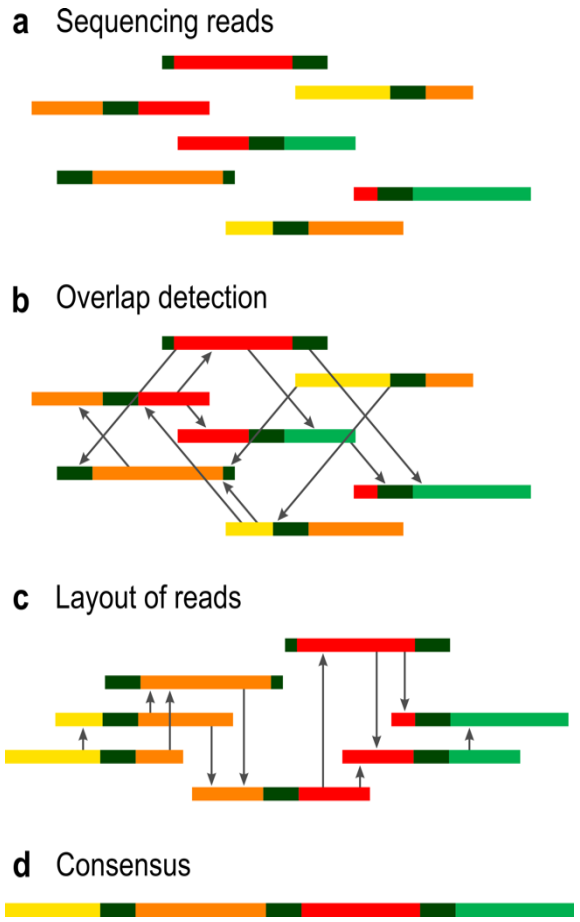
- **Overlap Layout Consensus Assemblers** – Mainly used for long reads
  - Construct overlap graph directly from reads, eliminating redundant reads; trace path for assembly.
  - Examples: pb-assembly (Falcon), Canu, Hifiasm

- **de Bruijn graph-based Assemblers** – Mainly used for short reads
  - Construct k-mer graph from the reads; original reads are discarded.
  - Trace a path through the graph to arrive at the assembly.
  - Examples: MEGAHIT, ABySS

Overlap Layout Consensus Graph

- All read vs. all read alignment and identify all possible overlaps between reads.
- The overlap relationship between reads is captured in a large assembly graph.
- The graph is refined to correct errors and simplify.
- Find the best path through the graph and traverse each node in the graph once.
- Output the consensus of the path as the assembly.





Challenges in Genome Assembly

- **Repeats**
  - Repetitive regions can make it more difficult to determine read placement leading to misassemblies, which are errors in the construction of the assembly.
  - Long reads can help this by traversing the longest repeats in the genome.

- **Sequencing Errors or Assembly Errors**
  - Base errors in sequencing reads or consensus errors introduced by assemblers can confound the assembly process.

- **Heterozygosity**: Presence of different alleles at the same loci in homologous chromosomes
  - Alleles from the same locus are more likely to be mistaken as sequences from different loci.
  - The assembler may incorporate two different contigs that actually represent the same regions of the genome.
  - This might be desirable if you are seeking haplotype separated diploid or polyploid assemblies.

- **Contamination**
  - Contamination in the sequencing data can lead to contigs of contamination in your final assembly.
  - Adapter contamination of the read data can impact assembly results.

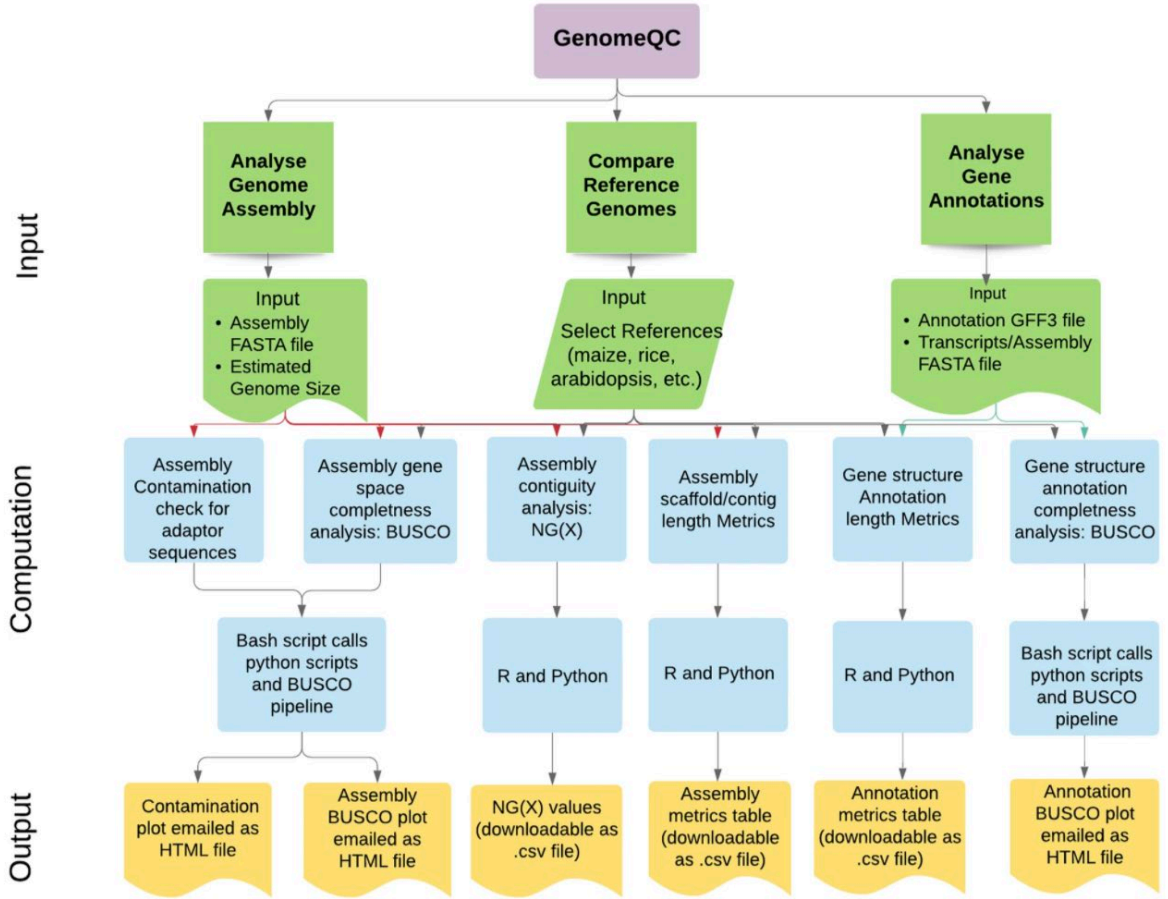
Methods to QC Genome Assemblies

- **Assembly scaffold/contig length metrics**
  -N50 length, avg. length, # contigs/scaffolds

- **Compare assembly to Reference Genome**
  -Identify misassemblies
  -Identify real SNPs and SVs (indels, translocations, duplications)

- **Evaluate against an optical map**: Ordered, genome wide high resolution restriction map
  -Identify misassemblies
  -Scaffold contig assemblies

- **Busco Analysis**: Evaluation of the assembly against a set of single-copy orthologs present in 90% of species of a particular group.
  -What percentage of core genes does your assembly contain?
  -Measure of assembly completeness.



N50 Contig/Scaffold Length

N50 size

Def: 50% of the genome is in contigs larger than N50

Example: 1 Mbp genome

N50 size = 30 kbp
(300k+100k+45k+45k+30k = 520k >= 500kbp)

Note:
N50 values are only meaningful to compare when base genome size is the same in all cases

Assemblers for PacBio Data

- **De novo Assemblers**
  - Pb-assembly (Falcon/Falcon-Unzip)
  - Canu/HiCanu

- **Trio Based Assemblers:** Use short reads from parents to partition child reads by maternal/paternal haplotypes prior to assembly.
  - TrioCanu
  - Hifiasm

- **Reference Assisted Assemblers:** Using the reference to assist in building the reads into contigs/scaffolds. This can bias assembly results towards the reference used.
  - RefKA

- **Hybrid Assemblers :** Use short and long reads. More effective for assembly of complex genomes (e.g. Plants)
  - Masurca
  - PBcR

PB Assembly (Falcon/Falcon-Unzip) of HiFi Data

- **Falcon Assembly:**
  -Input is PacBio HiFi data in FASTA format.

  -Output is haplotype-fused assembly in FASTA format and set of associated contigs (SVs from primary contig assembly).

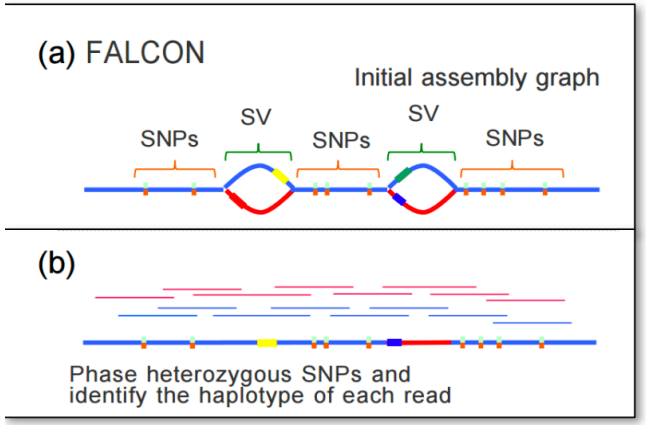
- **Falcon-Unzip:**
  -Align reads to contigs and phases the reads using heterozygous SNPs.

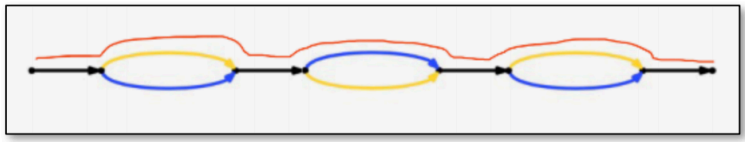
  -SNPs are used to separate the haplotypes into partially phased primary contigs and fully phased haplotigs.

  -There are switch errors in the output between maternal and paternal haplotypes.

  -Output is a FASTA file of primary contigs and a FASTA file of haplotigs.

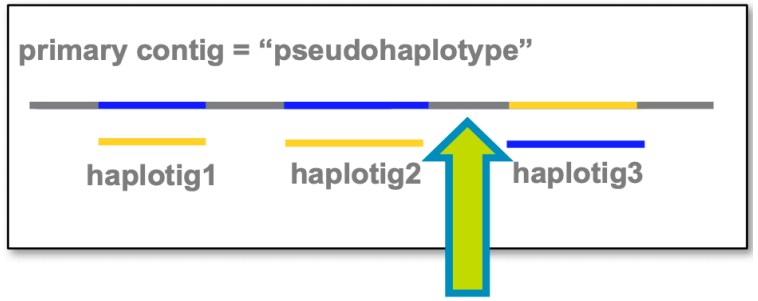
- PacBio data for diploid individual (no trio)
- Phase PacBio reads using SNPs identified in initial assembly graph
- Output phased and collapsed regions in high contiguity contigs





Weisenfeld et al. 2017

PSEUDOHAPLOTYPE AND HAPLOTIGS

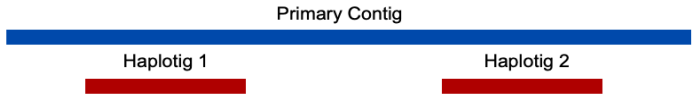


"Phase/Haplotype Switch"

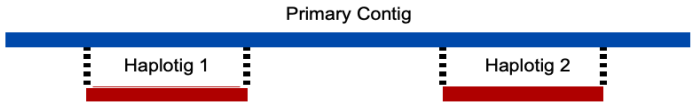
Falcon-Phase

FALCON-PHASE WORKFLOW

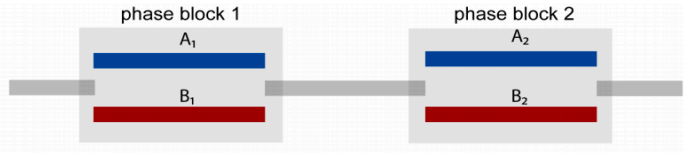
Input: FALCON-Unzip assembly & HiC data



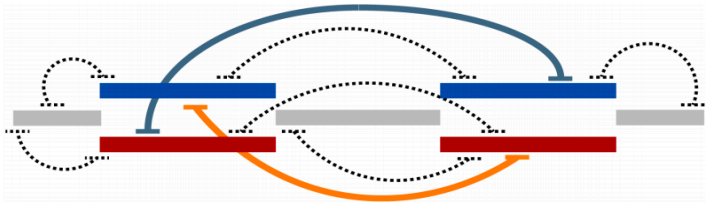
1. Identify **haplotig placement** on primary contigs



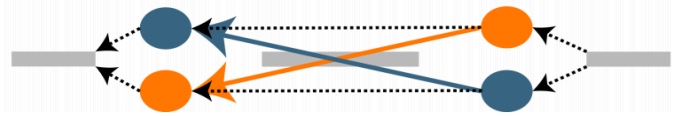
2. **Mince** primary contigs: separate **haplotig pairs** and **collapsed haplotypes**



3. **Map** paired HiC reads to minced contigs and generate normalized **contact matrix**



4. **Phase** haplotigs along each primary contig

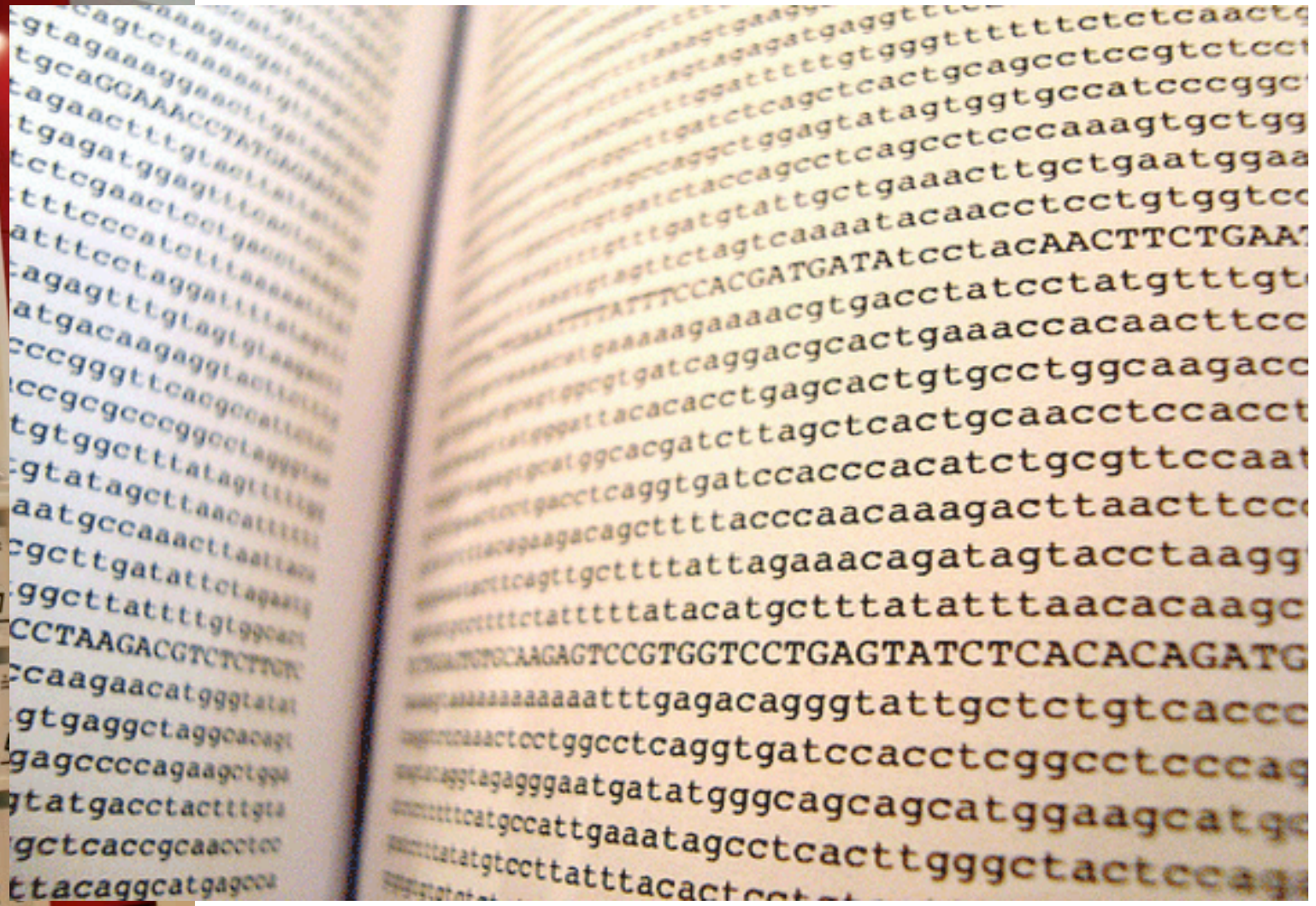


Output: phased full length haplotigs

Pangenomes

… and why you should consider using one

Human genome, the book of our lives




nature
the human genome


nature
THE HUMAN GENOME AT TEN
Growing pains of the genomics age
THE FAINT YOUNG SUN
A climate paradox revisited
QUANTUM MECHANICS
Controlling objects you can see
SLEEPING SICKNESS
Drug target for a neglected disease


HUMAN
PANGENOME

The three issues of the human reference genome (hg38)
Whose genome is it?

- It is a mosaic genome.
- It is not complete.
- Not a single cell on this planet has this genome.

What is the reference human genome?

Lander: So the genes from which most of the work was done come from Buffalo, New York.

Krulwich: From Buffalo, New York?

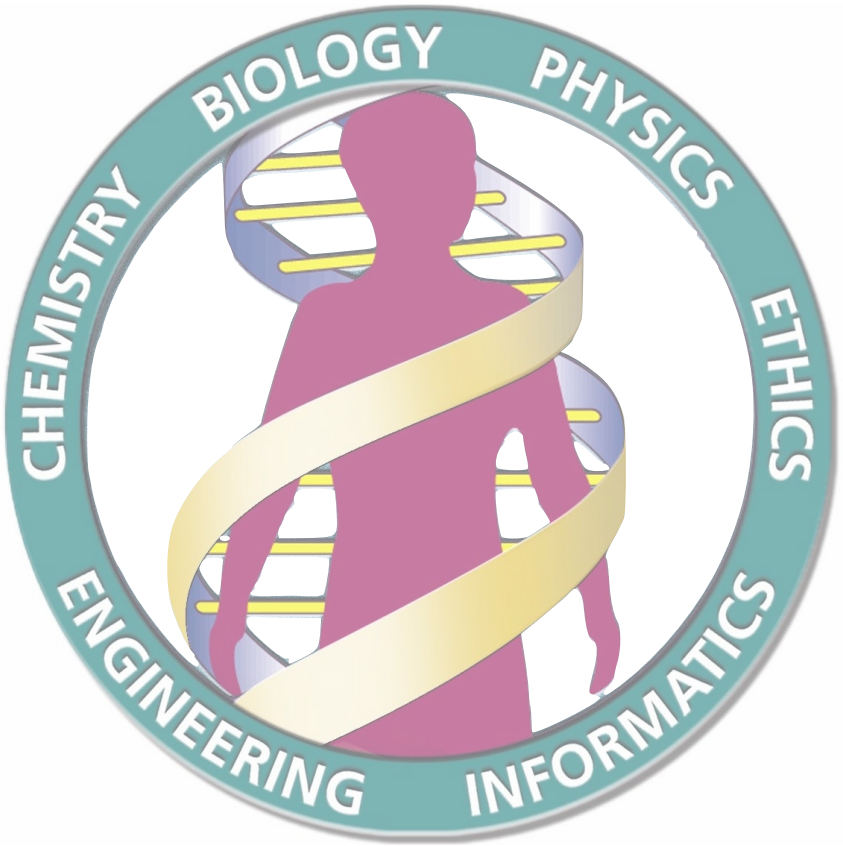
Lander: Yes. **It's mostly a guy from Buffalo and a woman from Buffalo.** But that's because the laboratory that was making--

...

Eric Lander
NOVA interview, 2001

Lander: The laboratory that prepared the large DNA libraries that were used was a laboratory in Buffalo. And so they put an ad in the Buffalo newspapers, and they got random volunteers from Buffalo, and they got about 20 of them. They then erased all the labels and chose at random this sample and that sample and that sample. So nobody knows who they are. We don't have any links back to who they are, and that's deliberate.

A Need to Modernize the Human Reference Genome



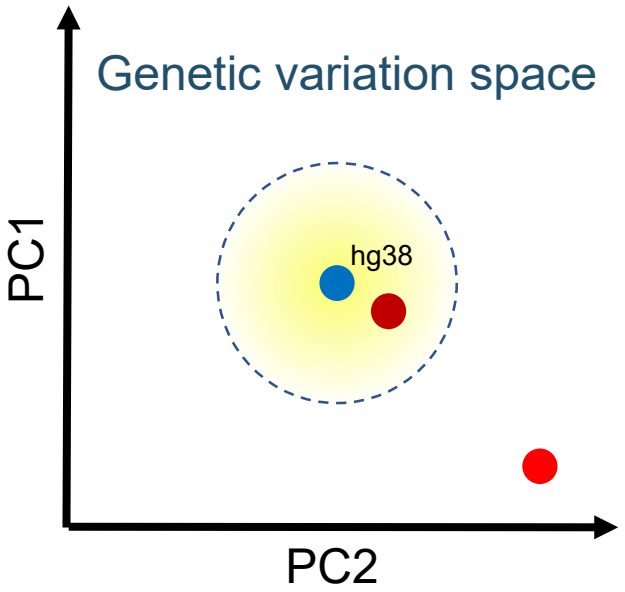
- **The human reference genome** is a foundational resource in human genetics and like most technology-driven resources, is overdue for an upgrade.

- The current structure is a **linear haplotype, largely representing a single individual.**

- **Mapping limitations of short reads and inherent reference biases** means we have missed more than 70% of structural variants in traditional whole-genome sequencing studies

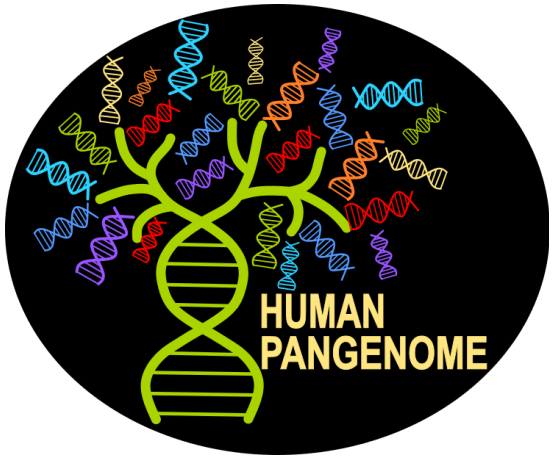
The streetlamp effect

The streetlamp effect



Goals

- Improve upon the limitations of reference genomes



Perspective | Published: 20 April 2022

The Human Pangenome Project: a global resource to map genomic diversity

Ting Wang, Lucinda Antonacci-Fulton, Kerstin Howe, Heather A. Lawson, Julian K. Lucas, Adam M. Phillippy, Alice B. Popejoy, Mobin Asri, Caryn Carson, Mark J. P. Chaisson, Xian Chang, Robert Cook-Deegan, Adam L. Felsenfeld, Robert S. Fulton, Erik P. Garrison, Nanibaa' A. Garrison, Tina A. Graves-Lindsay, Hanlee Ji, Eimear E. Kenny, Barbara A. Koenig, Daofeng Li, Tobias Marschall, Joshua F. McMichael, Adam M. Novak, the Human Pangenome Reference Consortium + Show authors

Nature **604**, 437–446 (2022) | Cite this article

47k Accesses | **35** Citations | **356** Altmetric | Metrics

New Results

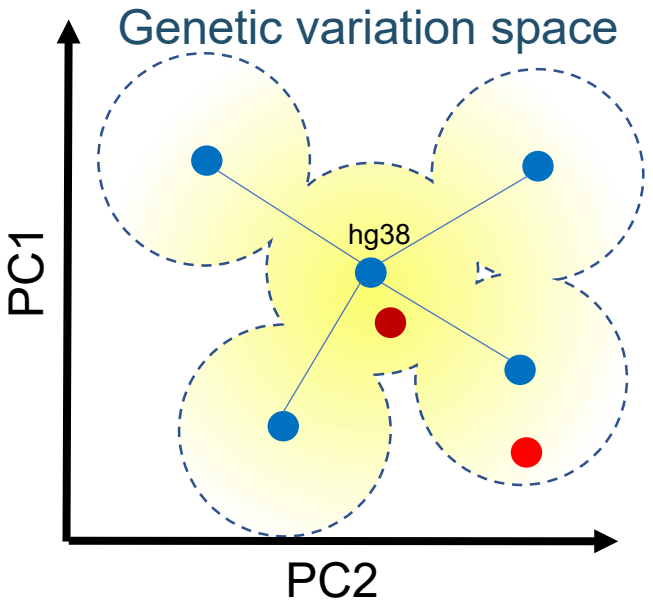
Follow this preprint

A Draft Human Pangenome Reference

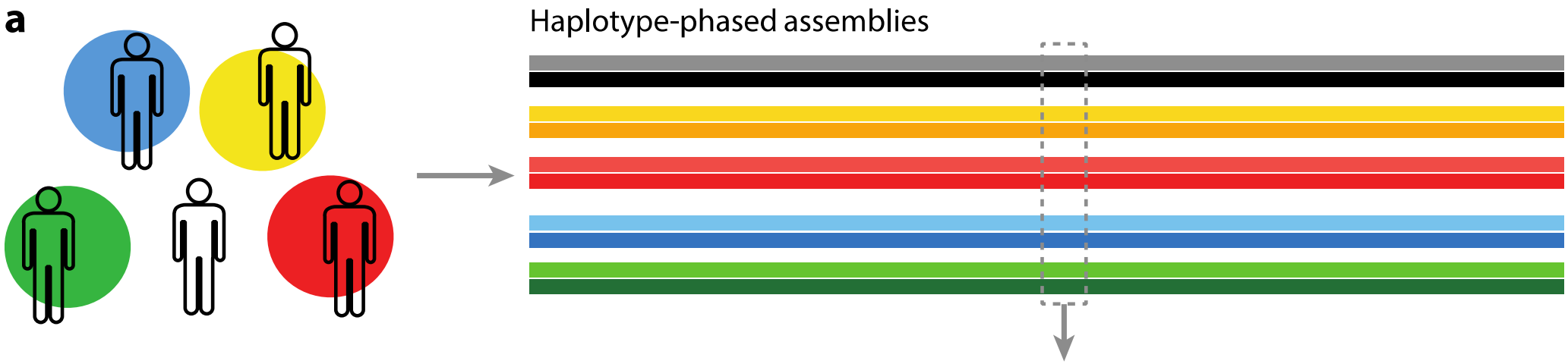
Wen-Wei Liao, Mobin Asri, Jana Ebler, Daniel Doerr, Marina Haukness, Glenn Hickey, Shuangjia Lu, Julian K. Lucas, Jean Monlong, Haley J. Abel, Silvia Buonaiuto, Xian H. Chang, Haoyu Cheng, Justin Chu, Vincenza Colonna, Jordan M. Eizenga, Xiaowen Feng, Christian Fischer, Robert S. Fulton, Shilpa Garg, Cristian Groza, Andrea Guarracino, William T Harvey, Simon Heumos, Kerstin Howe, Miten Jain, Tsung-Yu Lu, Charles Markello, Fergal J. Martin, Matthew W. Mitchell, Katherine M. Munson, Moses Njagi Mwaniki, Adam M. Novak, Hugh E. Olsen, Trevor Pesout, David Porubsky, Pjotr Prins, Jonas A. Sibbesen, Chad Tomlinson, Flavia Villani, Mitchell R. Vollger, Human Pangenome Reference Consortium, Guillaume Bourque, Mark JP Chaisson, Paul Flicek, Adam M. Phillippy, Justin M. Zook, Evan E. Eichler, David Haussler, Erich D. Jarvis, Karen H. Miga, Ting Wang, Erik Garrison, Tobias Marschall, Ira Hall, Heng Li, Benedict Paten

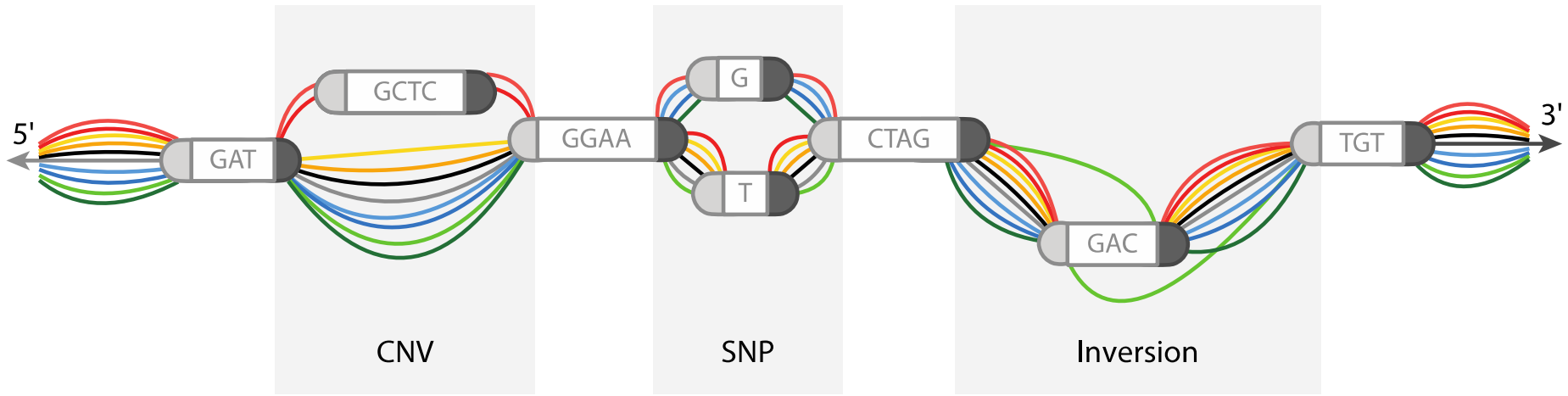
doi: https://doi.org/10.1101/2022.07.09.499321

The streetlamp effect



The Pangenome Graph





The Pangenome Graph

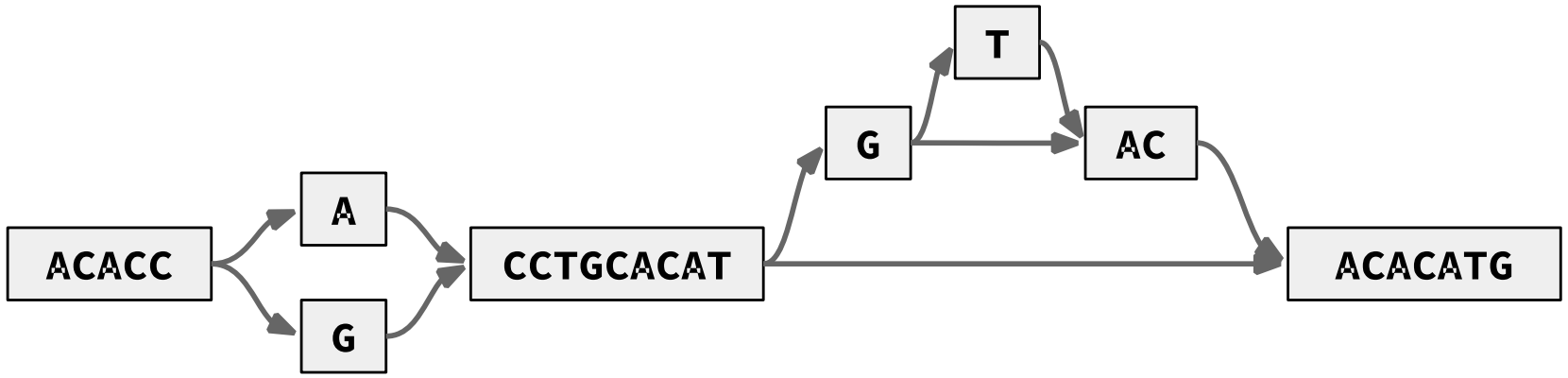
High-quality assemblies

```
ACACCACCTGCACATGACACACATG
ACACCACCTGCACATACACATG
ACACCACCTGCACATGACACACATG
ACACCACCTGCACATACACATG
ACACCACCTGCACATGACACACATG
ACACCGCCTGCACATGACACACATG
ACACCGCCTGCACATGTACACACATG
ACACCGCCTGCACATGACACACATG
```

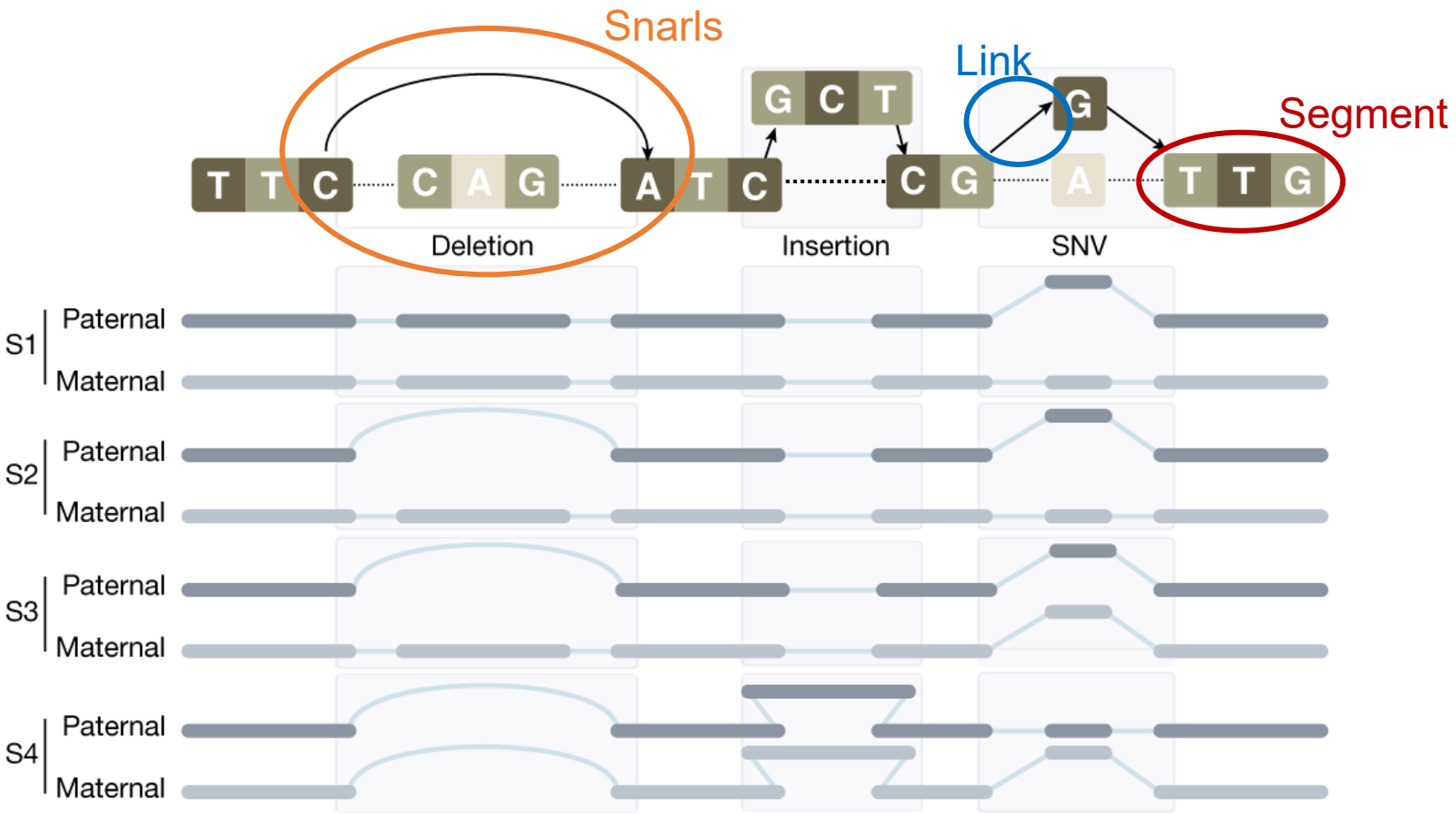
Alignment

```
ACACCACCTGCACATGACACACATG
ACACCACCTGCACAT---ACACATG
ACACCACCTGCACATGACACACATG
ACACCACCTGCACAT---ACACATG
ACACCACCTGCACATGACACACATG
ACACCGCCTGCACATGACACACATG
ACACCGCCTGCACATGTACACACATG
ACACCGCCTGCACATGACACACATG
```

Pangenome Graph



Genome graphs

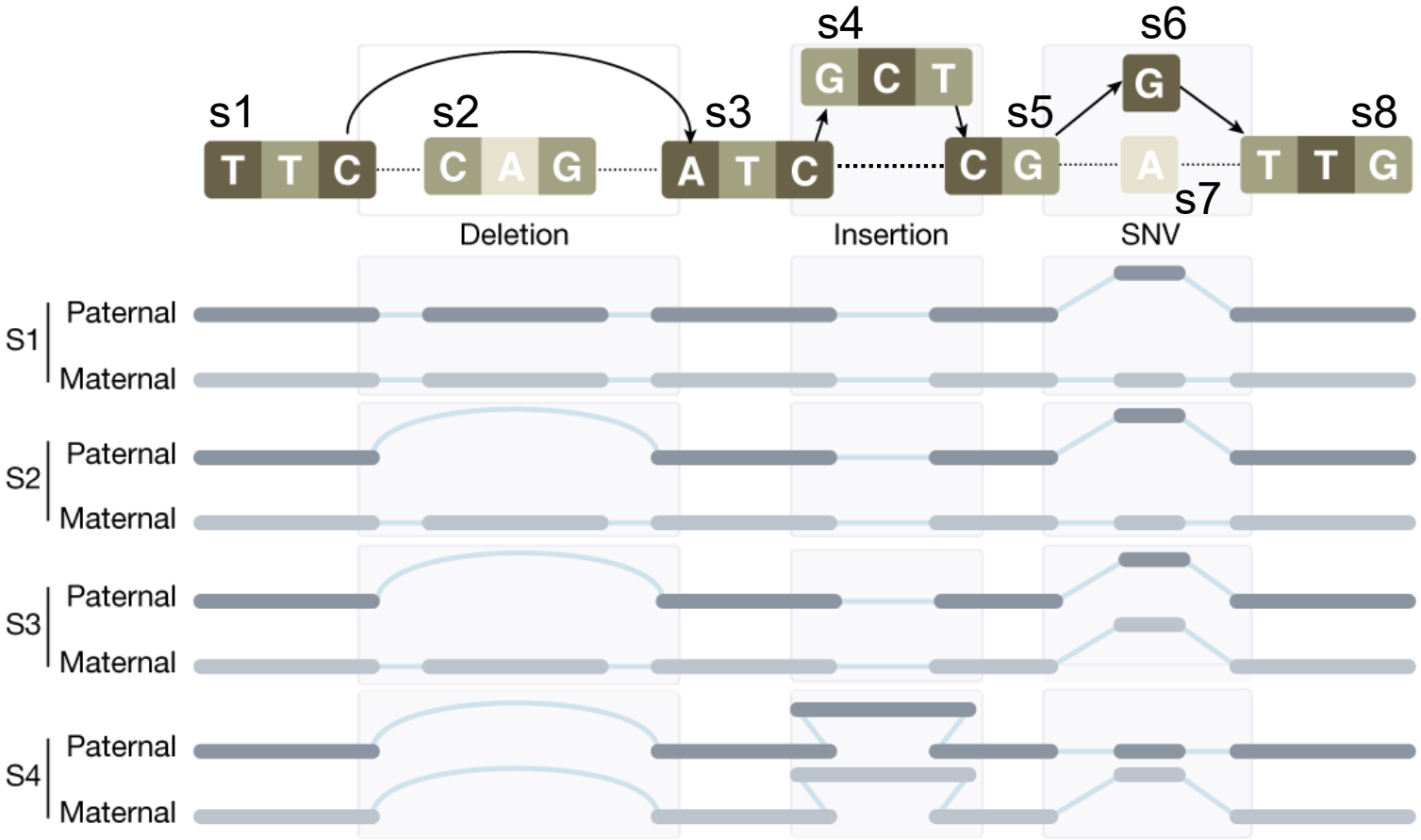


Redundant vocabulary:

Node = Vertex = Segment

Link = Edge

Genome graphs



s1 > s2 > s3 > s5 > s6 > s8

s1 > s2 > s3 > s5 > s7 > s8

s1 > s3 > s5 > s6 > s8

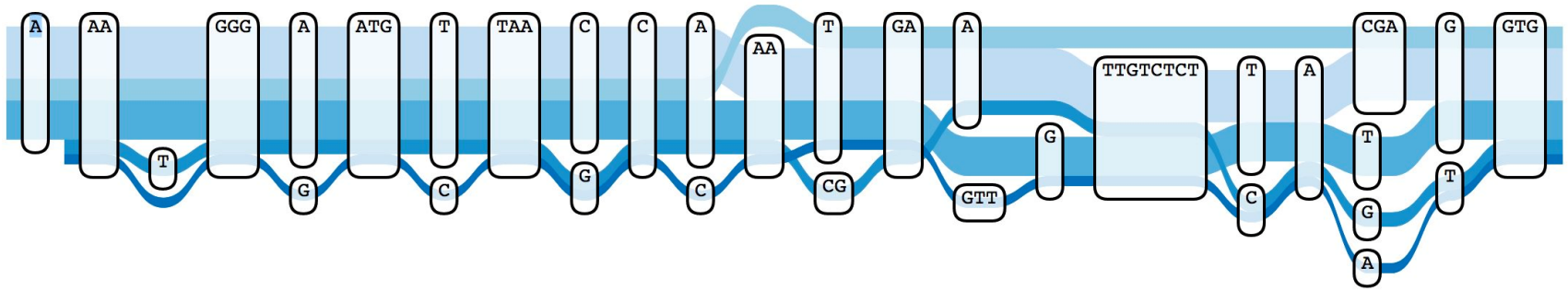
s1 > s2 > s3 > s5 > s7 > s8

Redundant vocabulary:

Node = Vertex = Segment

Link = Edge

The Pangenome Graph



Alignments of high-quality assemblies were performed using three different methods, pioneered by our team:

Minigraph (Li et al., 2020),
Minigraph-Cactus (MC)
PanGenome Graph Builder (PGGB).

Building pangenome – approaches

- Iterative building – *affected by order of genome addition*
  - Minigraph
  - Minigraph-Cactus
- Simultaneous building
  - Pangenome Graph Builder

Building pangenome – iterative

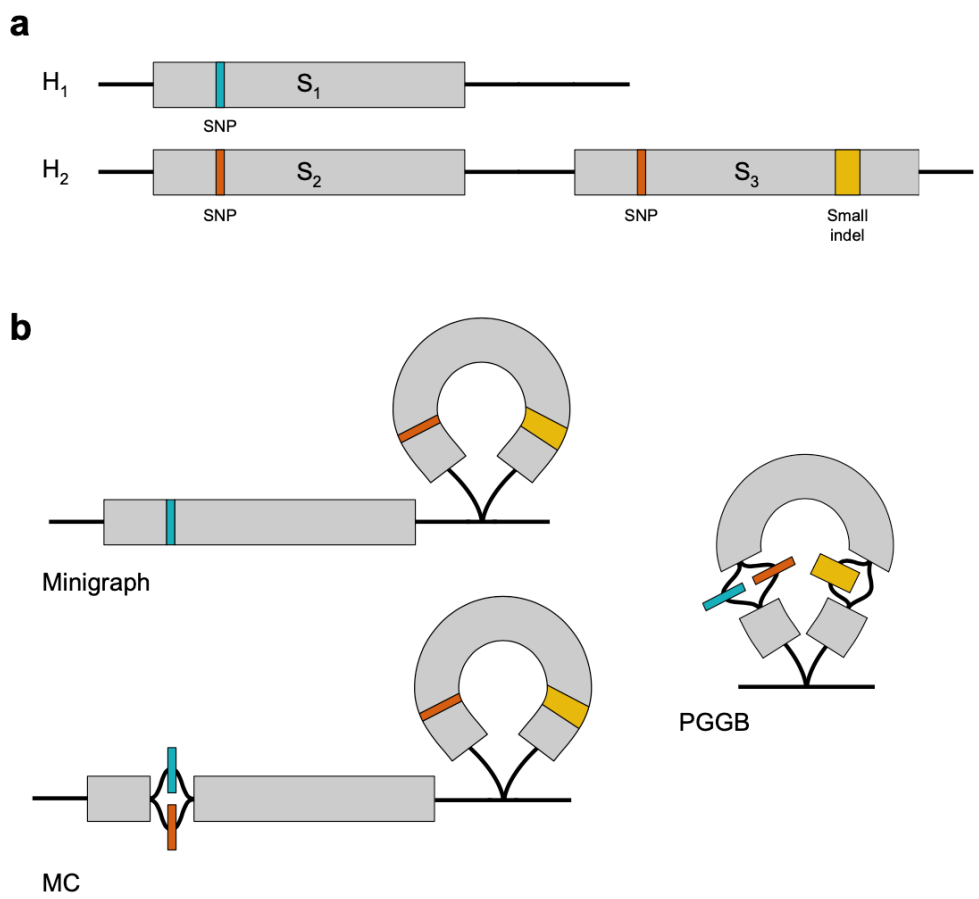
- Minigraph
  - Only considers structural variants >= 50bp
  - Does not consider sequence variants

- Minigraph-Cactus (MC)
  - 1. builds Minigraph pangenome
  - 2. performs base level sequence alignment of syntenic regions with Cactus.
  - 3. incorporates sequence variants into graph

Building pangenome – simultaneous

- PGGB
  - First Performs everything to everything alignment
  - Compresses into a graph by partial order alignment

Building pangenome – approaches



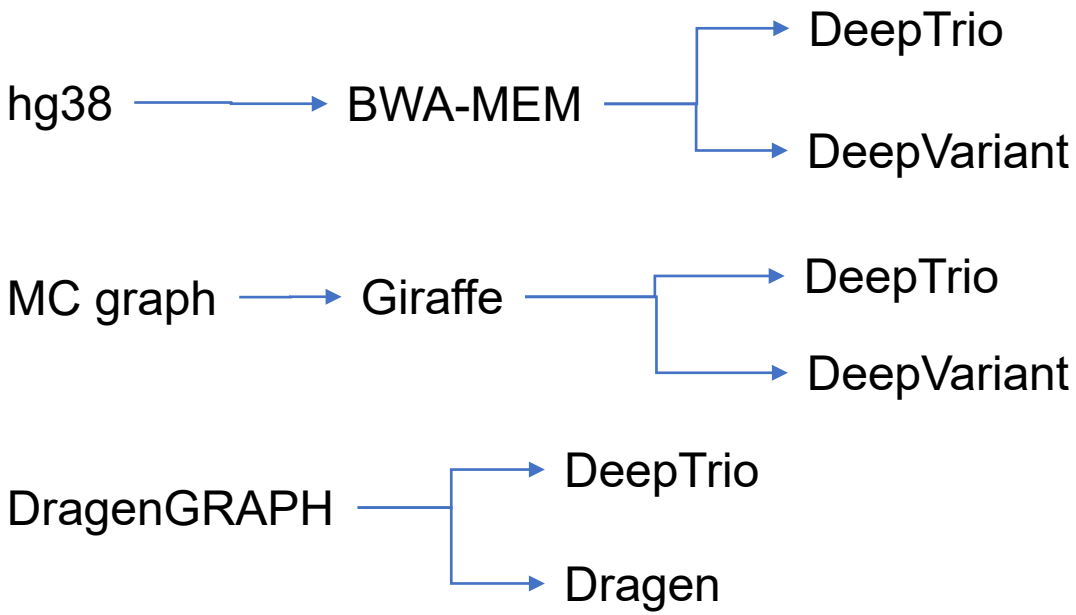
- Iterative building
  - Minigraph
    - Pros: Usable, Simple
    - Cons: Only captures structural variants, biased by order
  - Minigraph-Cactus
    - Pros: Usable, Captures all types of variation
    - Cons: Intentionally reduced complexity, Not a lossless graph, biased by order
- Simultaneous building
  - Pangenome Graph Builder
    - Pros: Lossless graph, unbiased by order, lends itself best to universal coordinate system
    - Cons: Much larger in size to the point of rendering it unusable for many applications
      - *"the inclusion of heterochromatic sequences in the PGGB graph made read mapping impractically slow"*

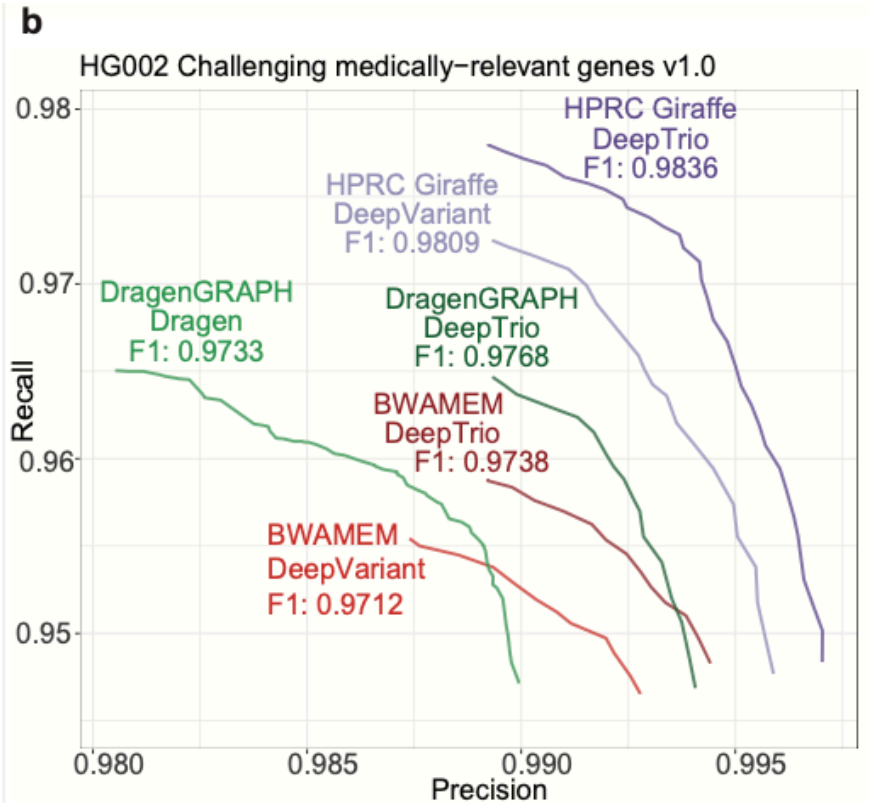
What can I use a genome graph to do?

- Calling variants – Giraffe aligner + DeepVariant
- Genotyping structural variants – Giraffe aligner + Pangenie
- Analyze ATACseq – Giraffe + GraphPeakCaller
- Analyze RNAseq – mpmap + rpvg

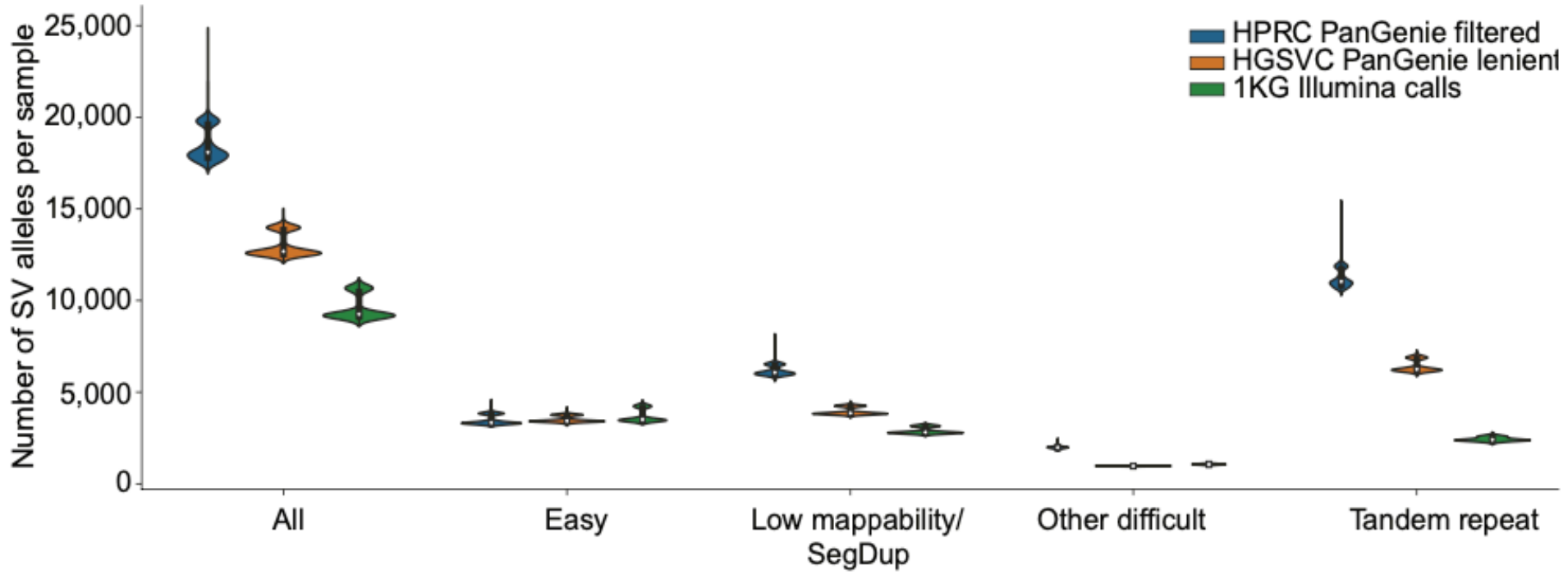
- Analyze WGBS – No tool exists
- Analyze HiC – No tool exists

Why use a genome graph? – remove some reference biases – Improved variant calling





Why use a genome graph? – remove some reference biases – Improved structural variant detection from short reads



Linear genome literature

Long-read human genome sequencing and its applications

Logsdon, G.A., Vollger, M.R. & Eichler, E.E. Long-read human genome sequencing and its applications. *Nat Rev Genet* **21,** 597–614 (2020). https://doi.org/10.1038/s41576-020-0236-x

Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome

Wenger, A.M., Peluso, P., Rowell, W.J. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* **37,** 1155–1162 (2019). https://doi.org/10.1038/s41587-019-0217-9

Long-read sequencing in deciphering human genetics to a greater depth

Midha, M.K., Wu, M. & Chiu, KP. Long-read sequencing in deciphering human genetics to a greater depth. *Hum Genet* **138,** 1201–1215 (2019). https://doi.org/10.1007/s00439-019-02064-y

Introduction to Genome Assembly

Bioinformatics Workbook (Online)

Andrew Severin - Author

https://bioinformaticsworkbook.org/dataAnalysis/GenomeAssembly/Intro_GenomeAssembly.html

Phased Diploid Genome Assembly with Single Molecule Real-Time Sequencing

Chin CS, Peluso P, Sedlazeck FJ, et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*. 2016;13(12):1050-1054. doi:10.1038/nmeth.4035

Extended haplotype phasing of *de novo* genome assemblies with FALCON-Phase

Zev N. Kronenberg, Arang Rhie, Sergey Koren, Gregory T. Concepcion, Paul Peluso, Katherine M. Munson, Stefan Hiendleder, Olivier Fedrigo, Erich D. Jarvis, Adam M. Phillippy, Evan E. Eichler, John L. Williams, Tim P.L. Smith, Richard J. Hall, Shawn T. Sullivan, Sarah B. Kingan

bioRxiv 327064; doi: https://doi.org/10.1101/327064

GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations

Manchanda, N., Portwood, J.L., Woodhouse, M.R. *et al.* GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations. *BMC Genomics* **21,** 193 (2020). https://doi.org/10.1186/s12864-020-6568-2

Genome graph literature

The Human Pangenome Project: a global resource to map genomic diversity

https://www.nature.com/articles/s41586-022-04601-8

A Draft Human Pangenome Reference

https://www.biorxiv.org/content/10.1101/2022.07.09.499321v1

Genome graphs detect human polymorphisms in active epigenomic state during influenza infection

https://www.biorxiv.org/content/10.1101/2021.09.29.462206v2.full

The design and construction of reference pangenome graphs with minigraph

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02168-z

Pangenome Graph Construction from Genome Alignment with Minigraph-Cactus

https://www.biorxiv.org/content/10.1101/2022.10.06.511217v1

Building pangenome graphs

https://www.biorxiv.org/content/10.1101/2023.04.05.535718v1

Acknowledgements

Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

Angela Page
Peter Goodhand

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

David Haussler
Benedict Paten
Karen Miga
Ed Green
Mark Akeson
Miten Jain
Hugh Olsen
Erik Garrison
Marina Haukness
Mark Akeson
Jean Monlong
Glenn Hickey
Adam Novack

Genome Sciences UNIVERSITY OF WASHINGTON

Adirna Fuller
Tony Tsung Yu Lu
Xian Chang
Trevor Pesout
Ryan Lorig-Roach
Charles Markello
Jonas Sibbesen
Melissa Meredith
Kishwar Shafin
Jouni Siren
Beth Sheets
Jordan Eizenga
Julian Lucas
Brian Hannafious

CBG Max Planck Institute of Molecular Cell Biology and Genetics

THE GENOME INSTITUTE at Washington University

Ting Wang
Lucinda Fulton
Sarah Cody
Robert Fulton
Heather Lawson
Wen-Wei Liao
Nathan Stitziel
Milinn Kremitizki
Haley Abel
Eddie Belter
Derek Albracht
Chad Tomlinson
Allison Regier
Tina Lindsay
Chris Markovic

Yale University

Ira Hall
Wen-Wei Liao
Shuangjia Lu

wellcome sanger institute

EMBL-EBI

Paul Flicek
Susan Fairley
Daniel Zerbino

Icahn School of Medicine at Mount Sinai

Adam Felsenfeld
Mike Smith
Carolyn Hutter
Taylorlyn Stephan
Heidi Sof
Adam Phillippy
Sergey Koren
Arang Rhie
Chirag Jain

Eimear Kenny
Vimi Desai
Kerstin Howe
Richard Durbin

UNIVERSITY OF CAMBRIDGE

NIH National Human Genome Research Institute

National Center for Biotechnology Information NCBI

Evan Eichler
Katy Munson
Pete Audano
Mitchell Vollger
David Porubksy
Arvis Sulovari
Gene Myers

HARVARD MEDICAL SCHOOL

Valerie Schneider
Terence Murphy
Paul Kitts
Chunlin Xiao
Francoise Thibaud-Nissent

AnVIL

Barbara Koenig (UCSF)
Nanibaa' Garrison (UCLA)
Charmaine D Royal (Duke)
Bob Cook-Deegan (ASU)
Alice Popejoy (UC Davis)
Bartha Knoppers (Mc Gill)

Heng Li
Shilpa Garg
Haoyu Cheng
Xiaowen Feng

NIST

Company Partnerships

aws

Justin Zook

illumina
PACBIO
HudsonAlpha INSTITUTE FOR BIOTECHNOLOGY
Google DeepVariant
ARIMA GENOMICS
Dovetail GENOMICS
Oxford NANOPORE Technologies

Coriell Institute FOR MEDICAL RESEARCH DECODING THE GENOME

Alissa Resch
Brittany Kerr
Brittney Martinez
Ellen Kellly

THE ROCKEFELLER UNIVERSITY
Science for the benefit of humanity

TOWARDS A COMPLETE REFERENCE OF HUMAN GENOME DIVERSITY

HUMAN PANGENOME

Erich Jarvis
Olivier Fedrigo
Giulio Formenti
Sadye Paez
Lauren Shalmiyev

NIH/NHGRI Funding:
U01HG010971, U41HG010972, 3U41HG010972-03S1

Acknowledgements – Lab as whole

Current lab members

Xiaoyun Xing
Noah Basri
Ju Heon Maeng
Ivy Chen
Holden Liang
Aparna Ananda
Heather Schmidt
Daofeng Li
Xiaoyu Zhuo
Kara Quid
Benpeng Miao
Jessica Harrison
Heather Lawson
Prashant Kuntala
Yiran Hou
Alan Du
Changxu Fan
Celine St. Pierre
Juan F. Macias-Velasco
Xuan Qu

Roadmap Epigenomics
ENCODE
4D Nucleome
TaRGET
Human Pangenome Project
IGVF

Joe Costello
Brad Bernstein
John Stamatoyannopoulos
Manolis Kellis
Art Petronis
Greg Schuler
Mike Snyder
Cheryl Walker
Dana Dolinoy
Shyam Biswal
Gokhan Mutlu
Cedric Feschotte
Fred Tyson
Lisa Chadwick
Kim McAllister
Ira Hall
Erich Jarvis
Bing Ren
Joe Ecker
Martin Hirst
Feng Yue
Tanya Barrett
Barak Cohen
Marisa Bartolomei
David Aylor
Sanjay Rajagopalan
Sheng Zhong
Peter Park
David Haussler
Karen Miga
Benedict Paten
Evan Eichler

TE in cancer

Hui Shen
Josh Jang
Christoph Plass
David Brocks
Christopher Oakes
Steve Baylin
Peter Jones
Kate Chiappinelli
Albert Kim
Gavin Dunn
Sid Puram

NIHR01HG007354, R01HG007175, U01CA200060, U24ES026699, U01HG009391, U41HG010972, U24HG012070
American Cancer Society RSG-14-049-01-DMC
Emerson Collective Funds
Siteman Investigator Program

Acknowledgements

Wang lab @ MGI HPRC team members

Xiaoyu Zhou
Tina Lindsay
Chad Tomlinson
Eddie Belter
Milinn Kremitzki
Derek Albrecht

Questions?