PRINCESS: Framework for Comprehensive Detection and Phasing of SNP and SVs

Fritz Sedlazeck

May, 9, 2019

Baylor College of Medicine

HGSC
HUMAN GENOME SEQUENCING CENTER

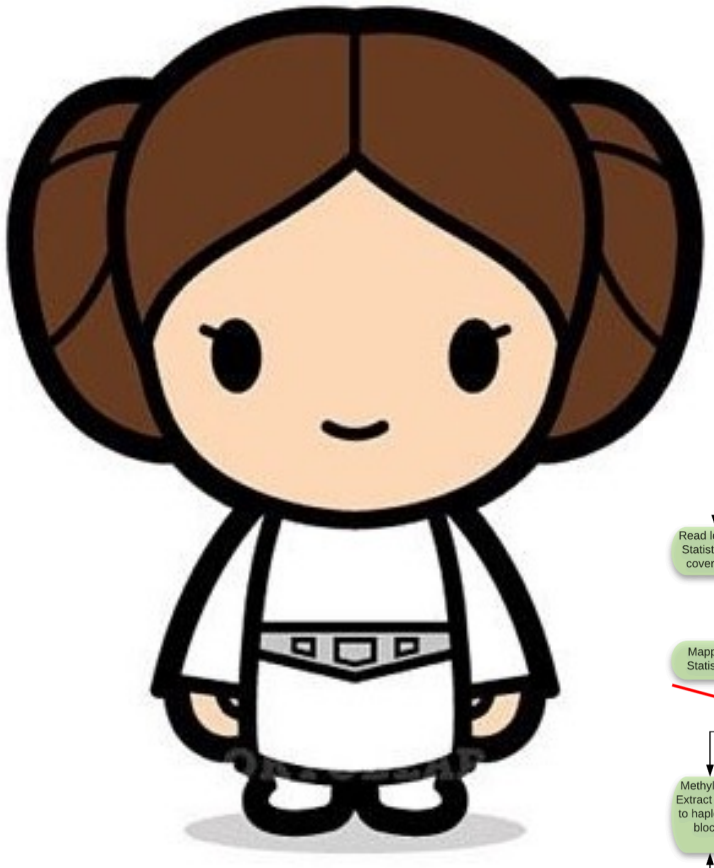
SMRT Leiden 2019

- Billy: HiFi SNP calling
- Jana: Phasing using long re
- Evan: SVs are important ar missed.
- ?? : Methylation using Pac

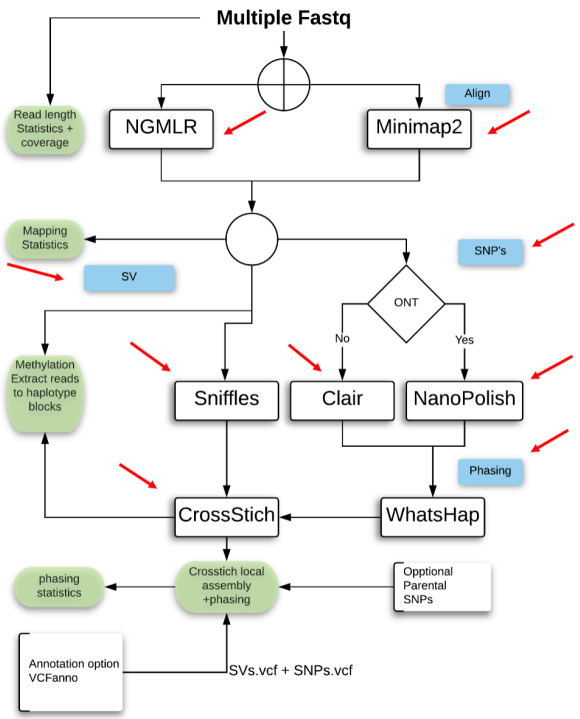
- Liz: Iso Seq

Upcoming: Princess

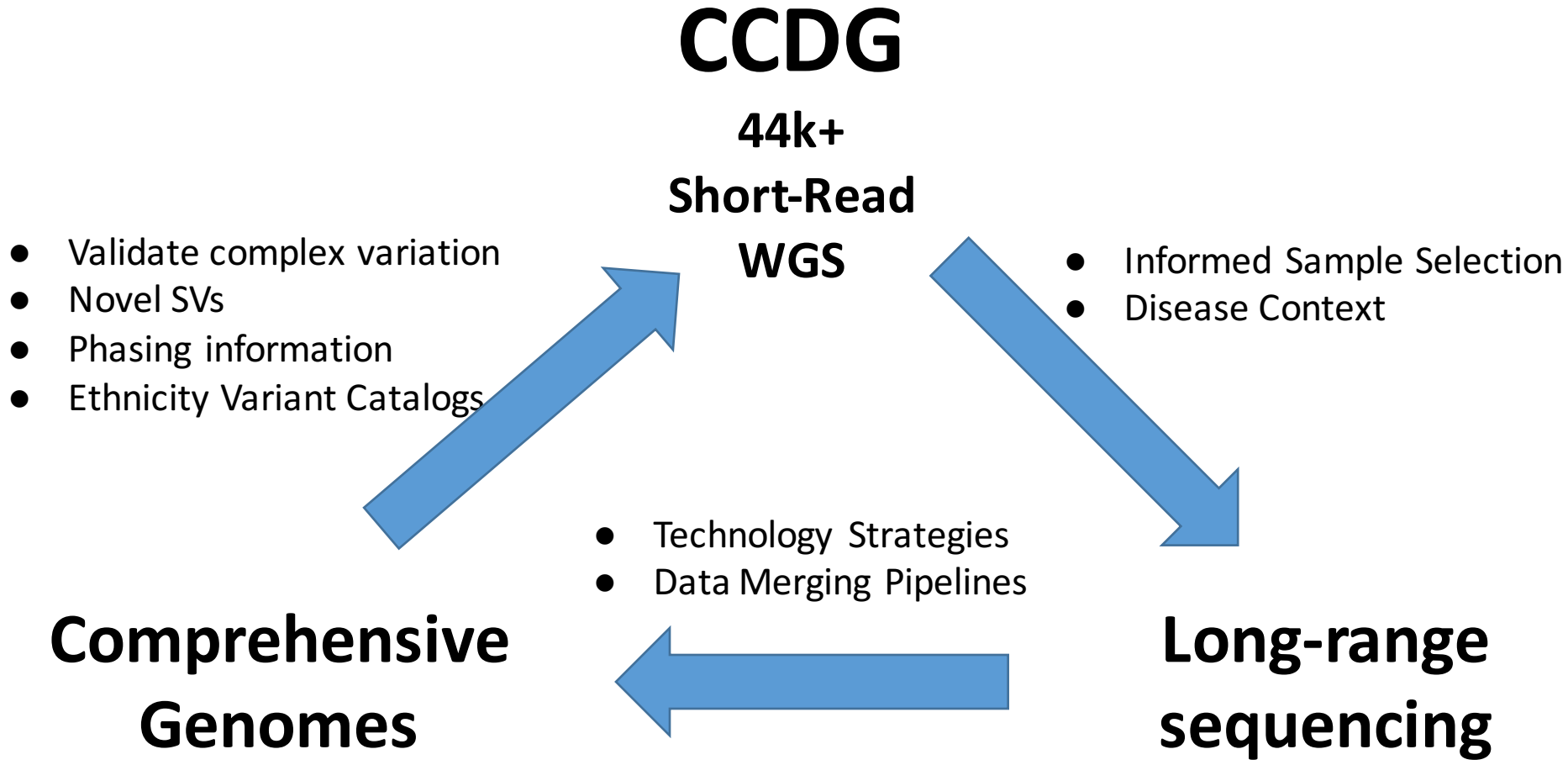
- Snakemake pipeline to:
  - Call SNPs
  - Call SVs
  - Phase SNPs + SVs + Methylation(?)
- Includes helpful options
  - Parameter estimation
  - Parental phasing
  - Summary statistics and plots
    - Read length
    - Mapping statistics
    - Variation statistics
    - Phasing length statistics



Medhat Mahmoud

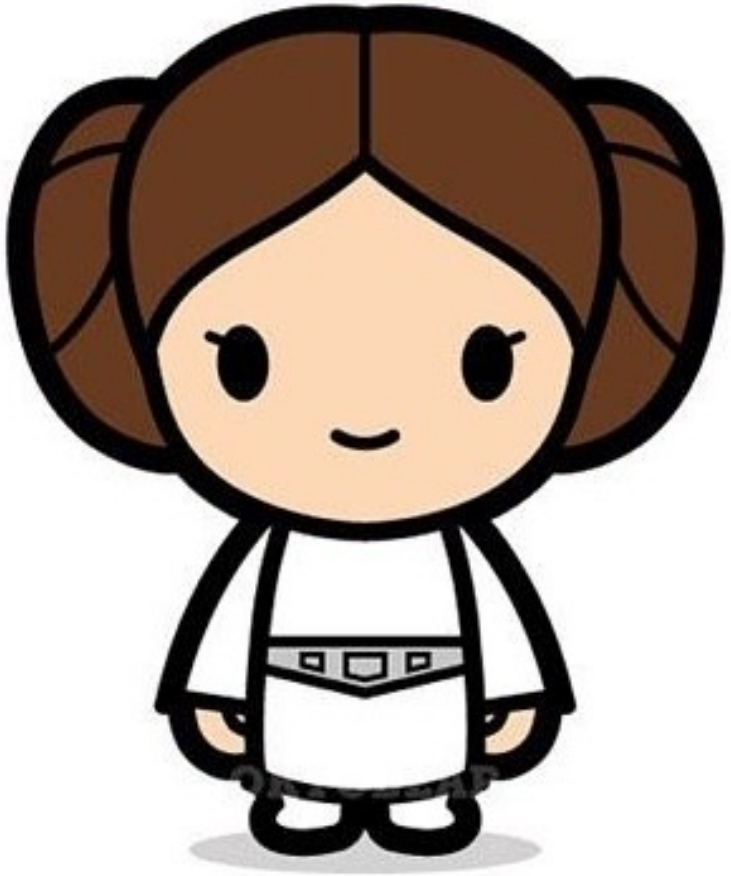


Upcoming application in large cohorts

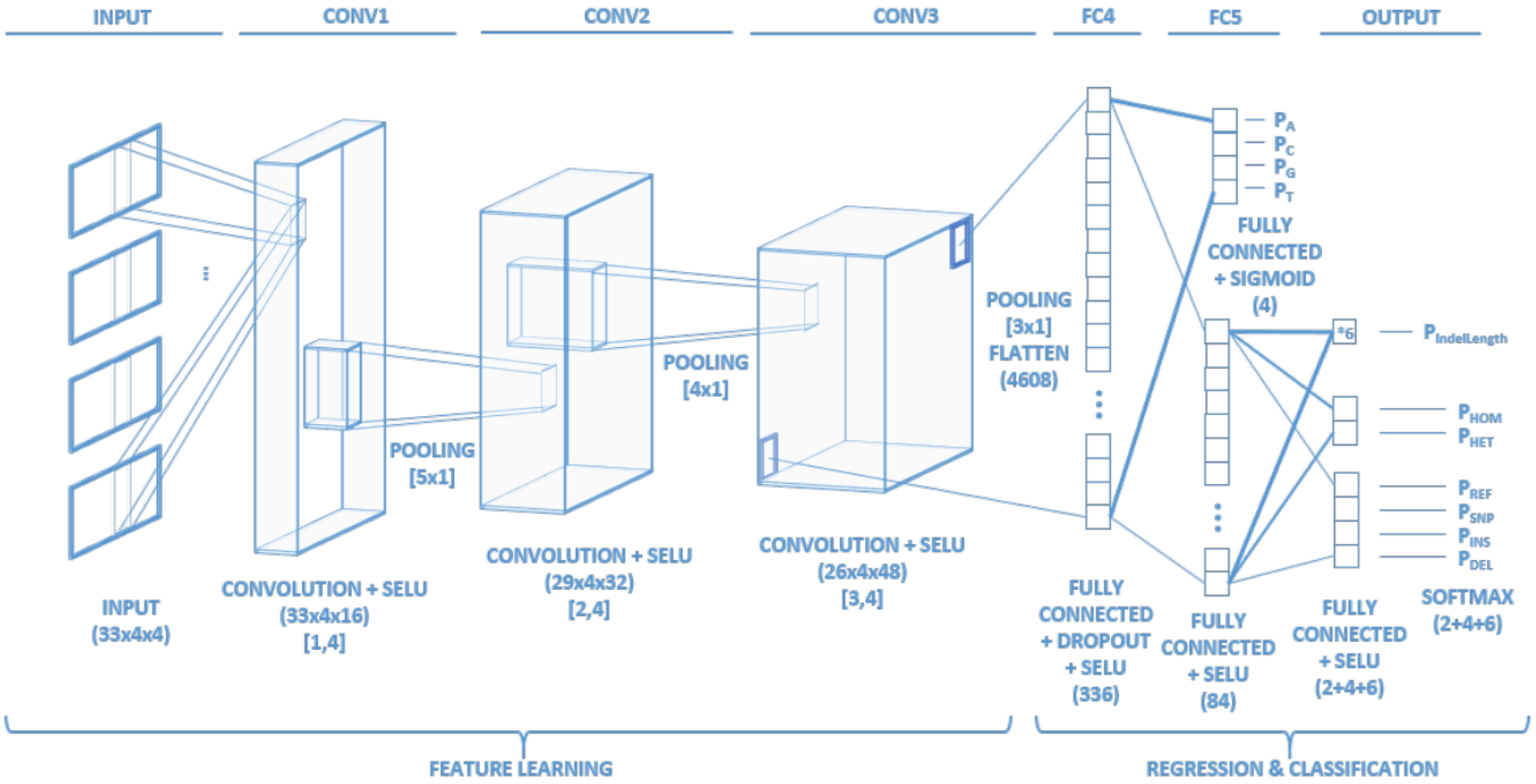


Princess

1. SNP calling
2. SVs calling
3. Phasing of SNP + SVs
4. Methylation (planed)



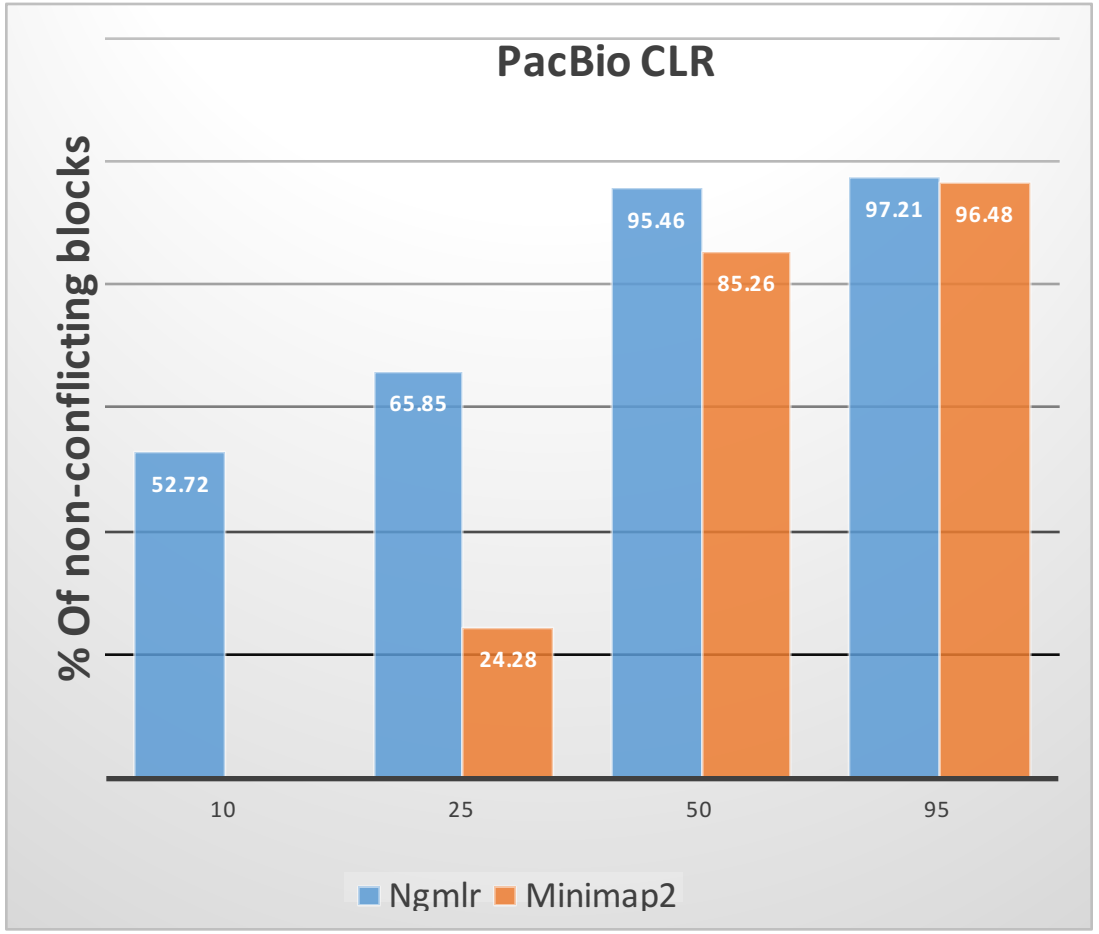
Clair (upcoming)

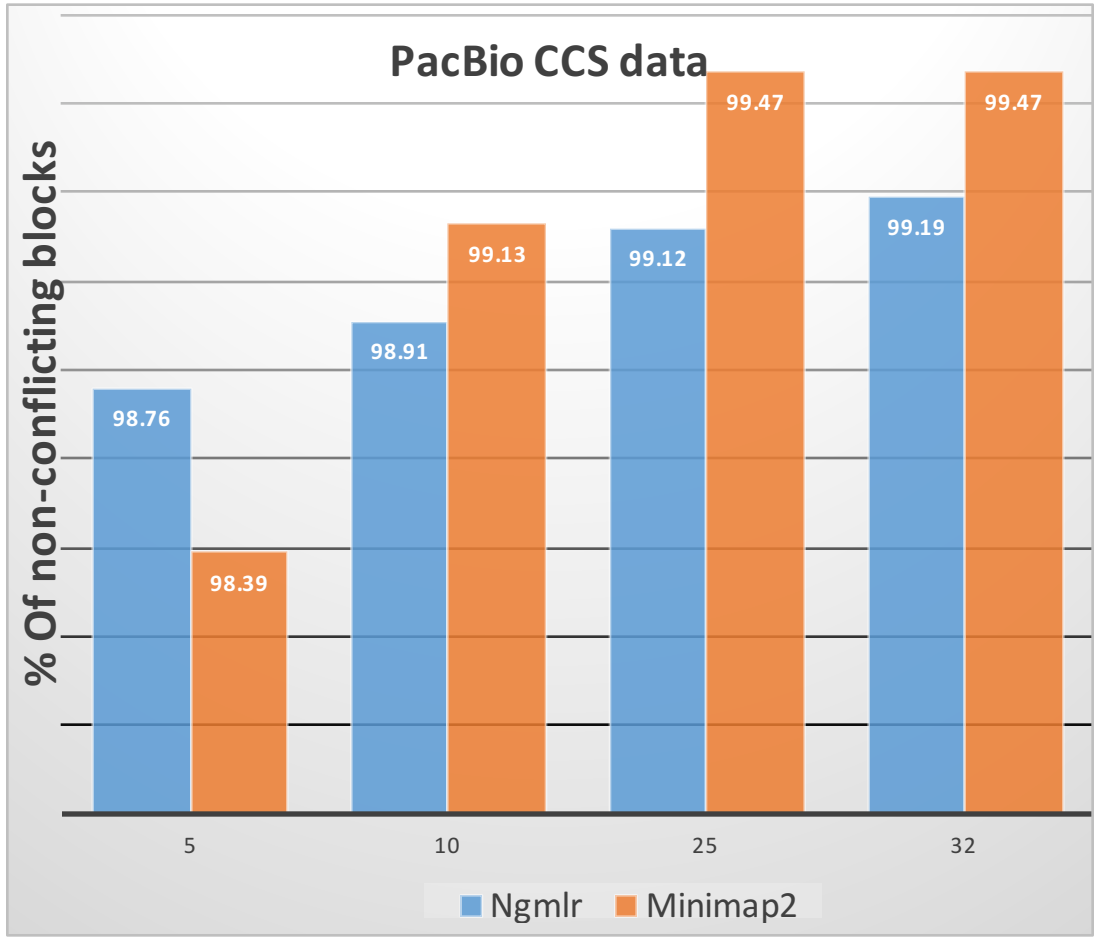


Ruibang Luo

Clairovante: Ruibang et. al (2019)

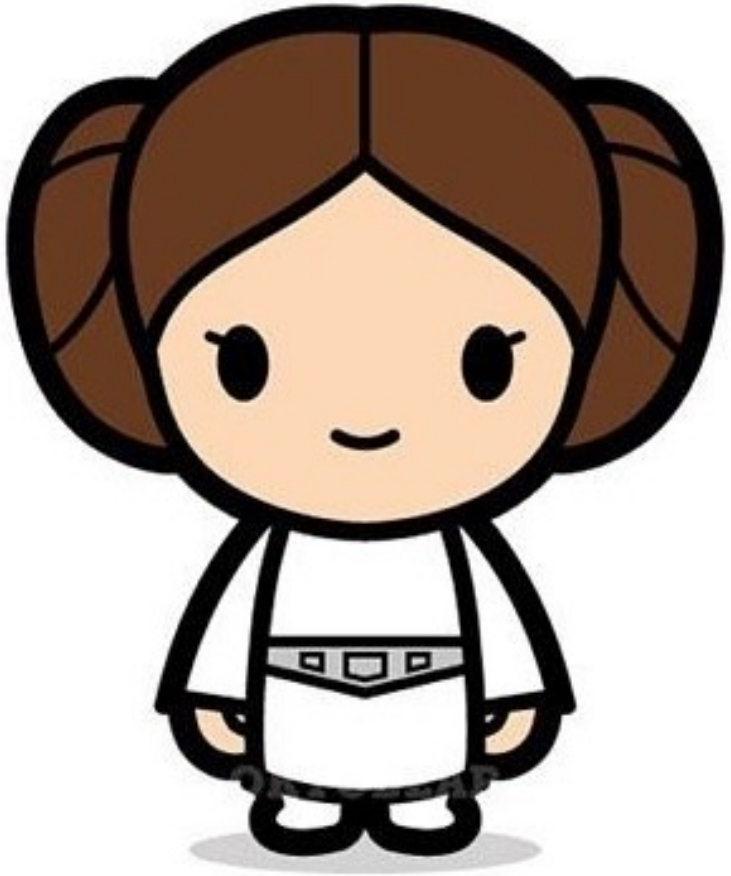
SNPs benchmarking





Princess

1. SNP calling
2. **SVs calling**
3. Phasing of SNP + SVs
4. Methylation (planed)

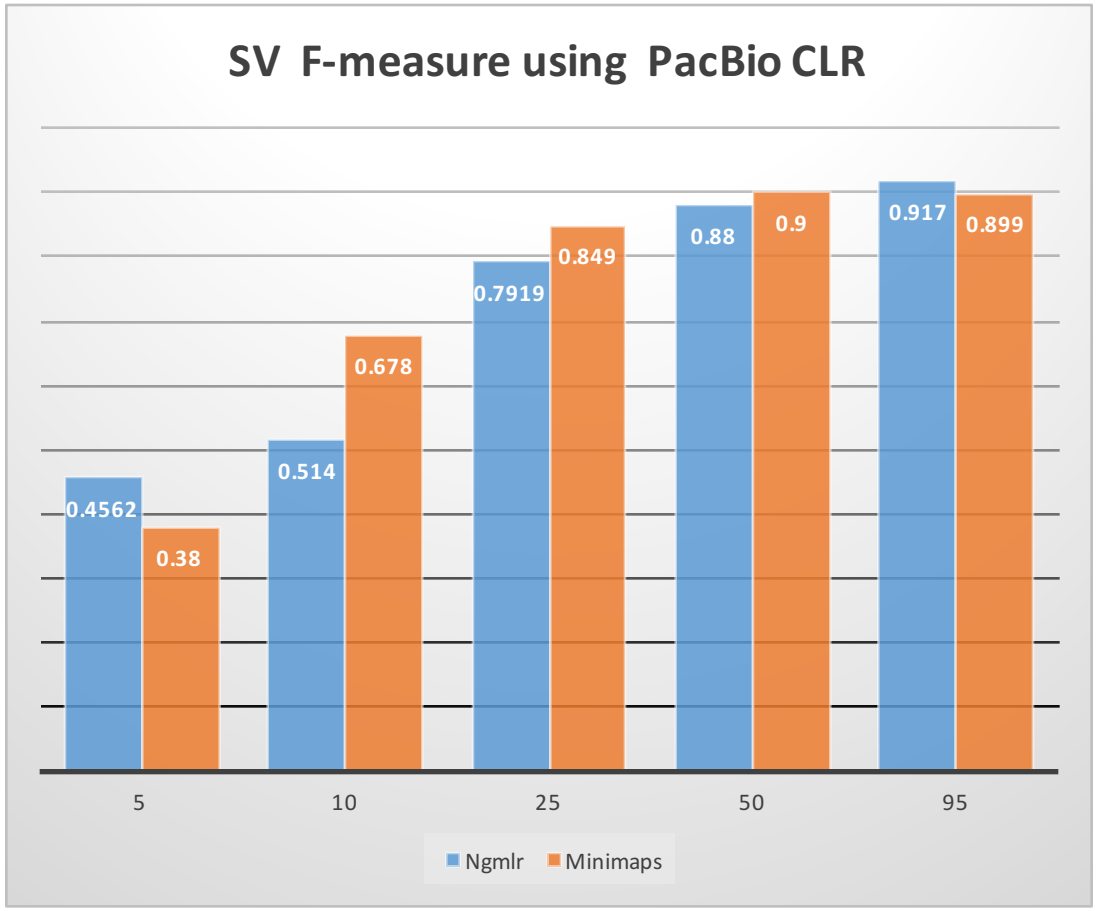


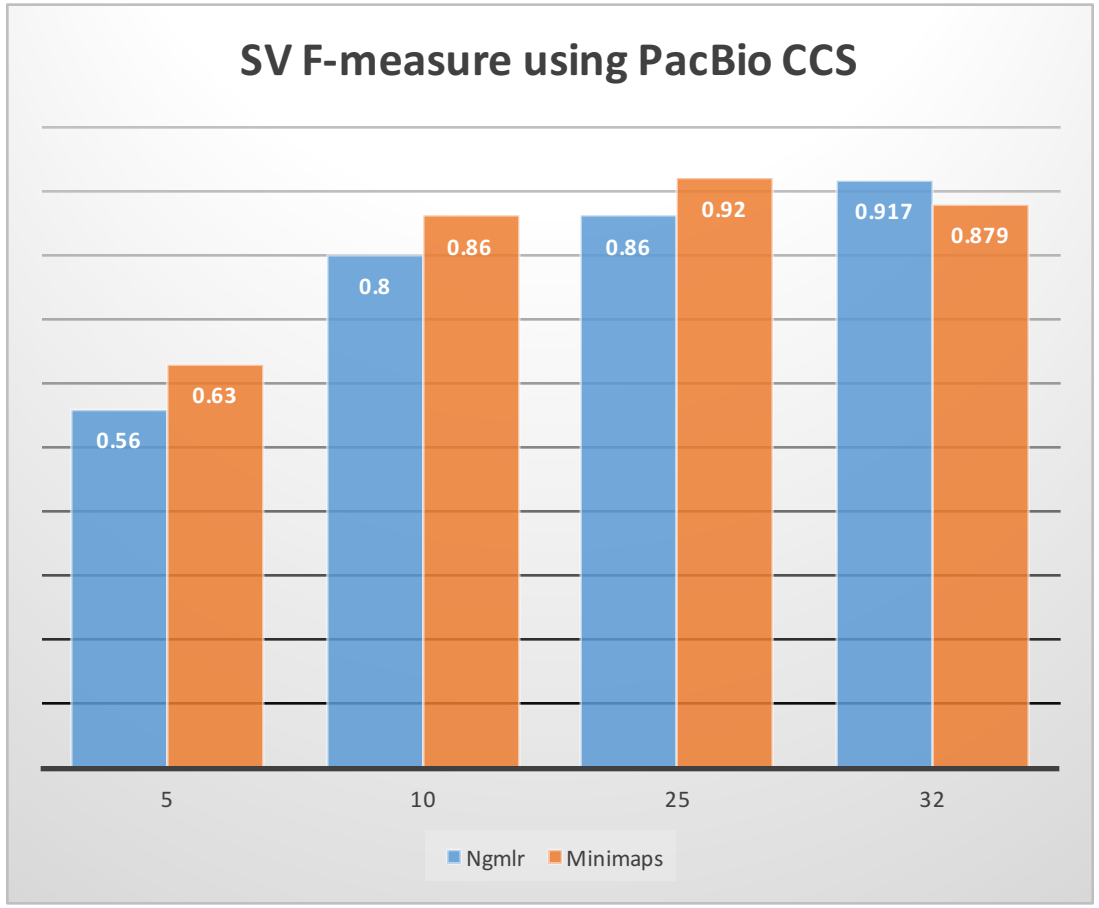
Sniffles

- Parameter estimation
- Analyzing:
  - split reads
  - alignment events
  - noisy regions
- Detect sequencing artifacts

- Optional:
  - Genotype estimation
  - Clustering/phasing of SVs
  - Sequence resolved indels

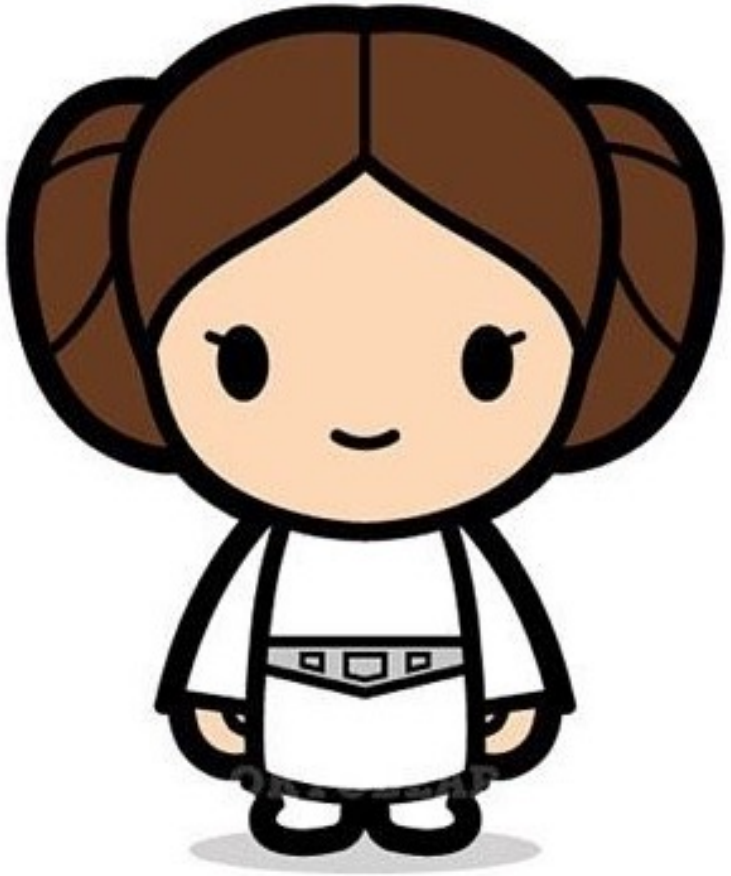
SVs benchmarking



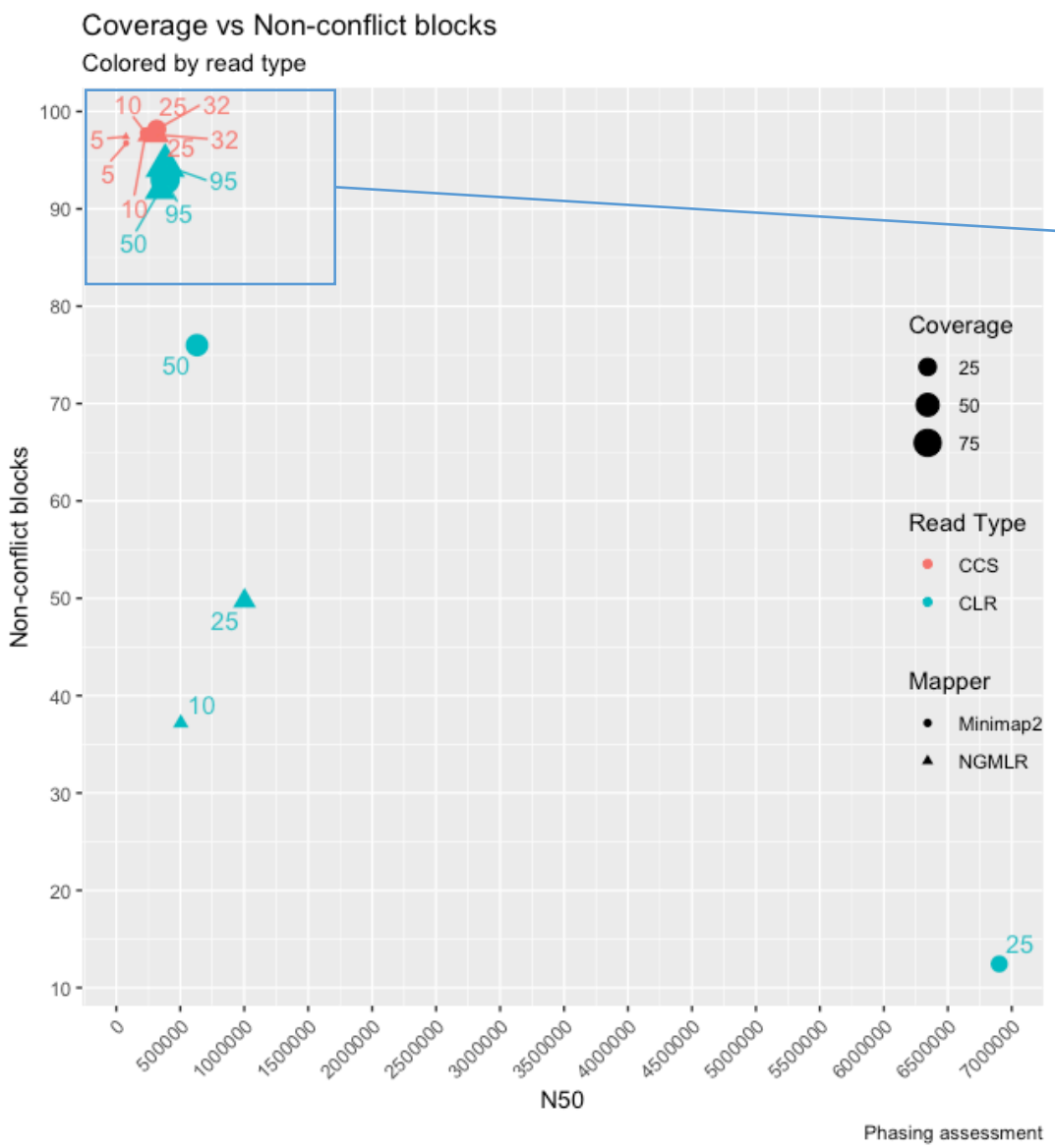


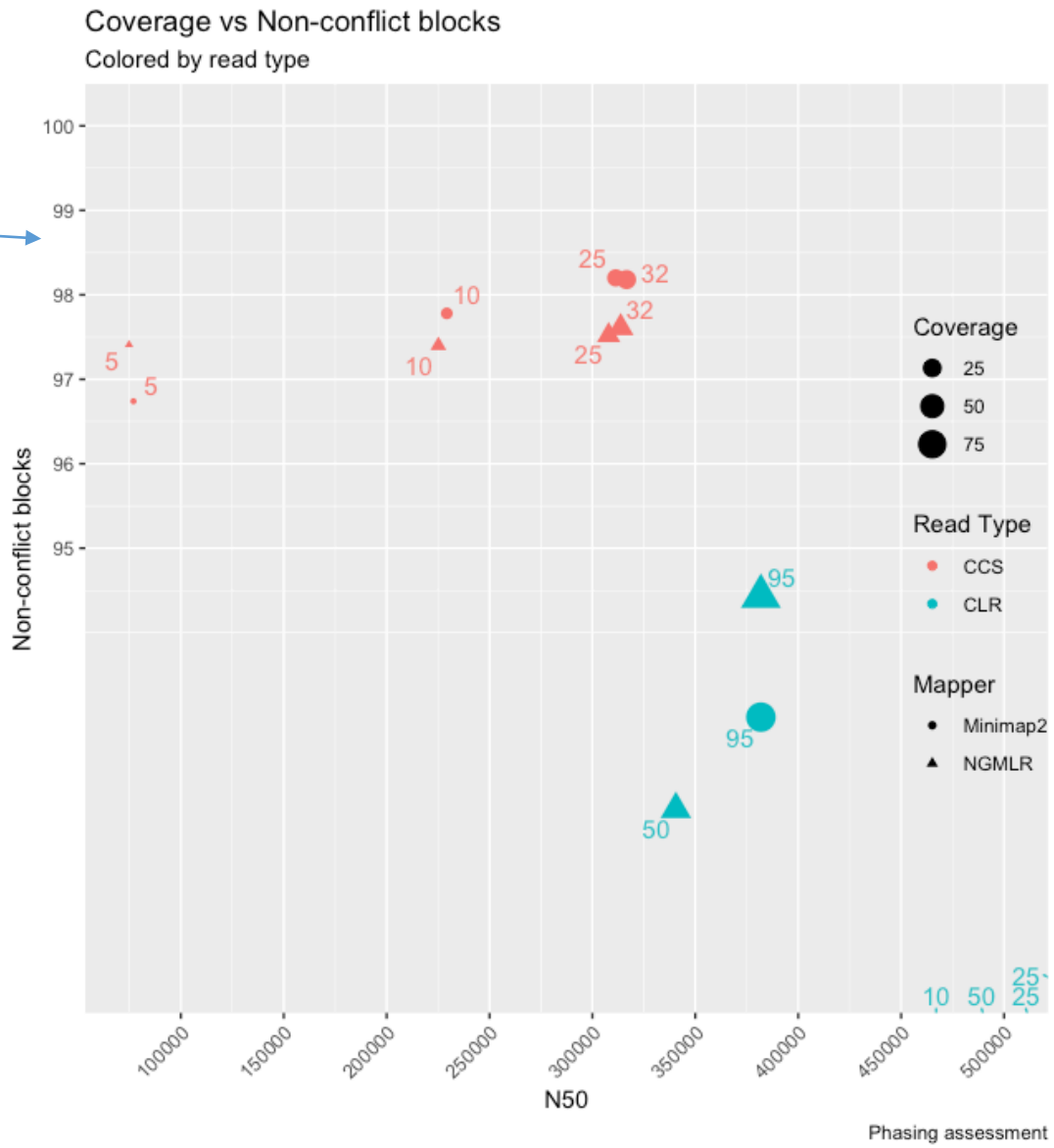
Princess

1. SNP calling
2. SVs calling
3. **Phasing of SNP + SVs**
4. Methylation (planed)



Phasing benchmarking





Sniffles: HLA-F deletion

PacBio:
HS-1011

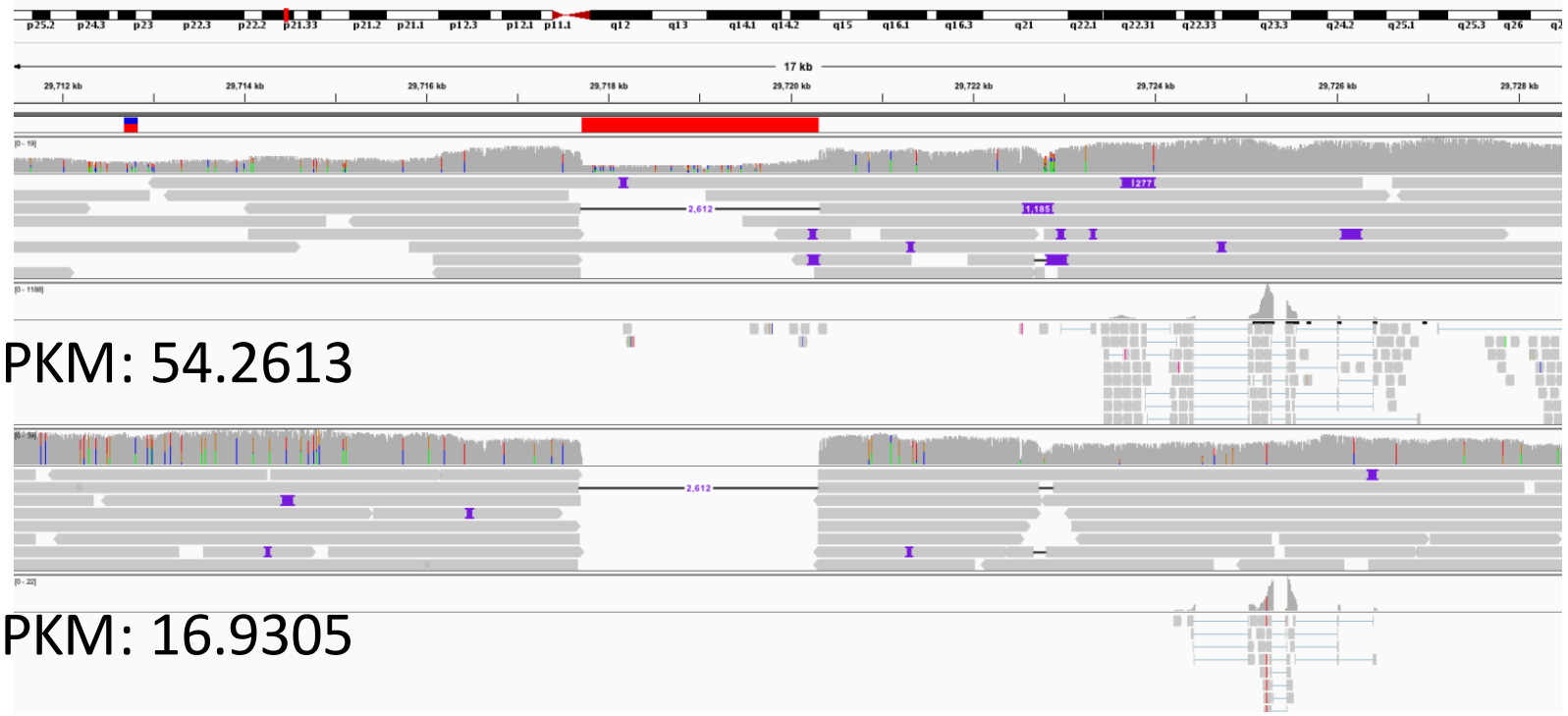
RNA-Seq
HS-1011

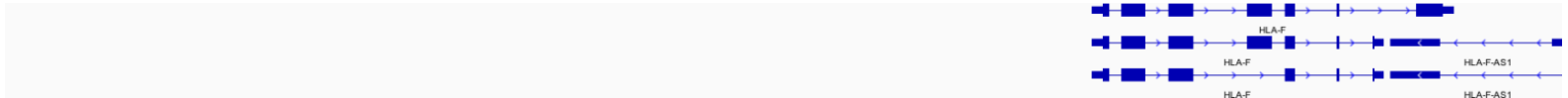
FPKM: 54.2613

PacBio:
NA12878

RNA-Seq
NA12878

FPKM: 16.9305



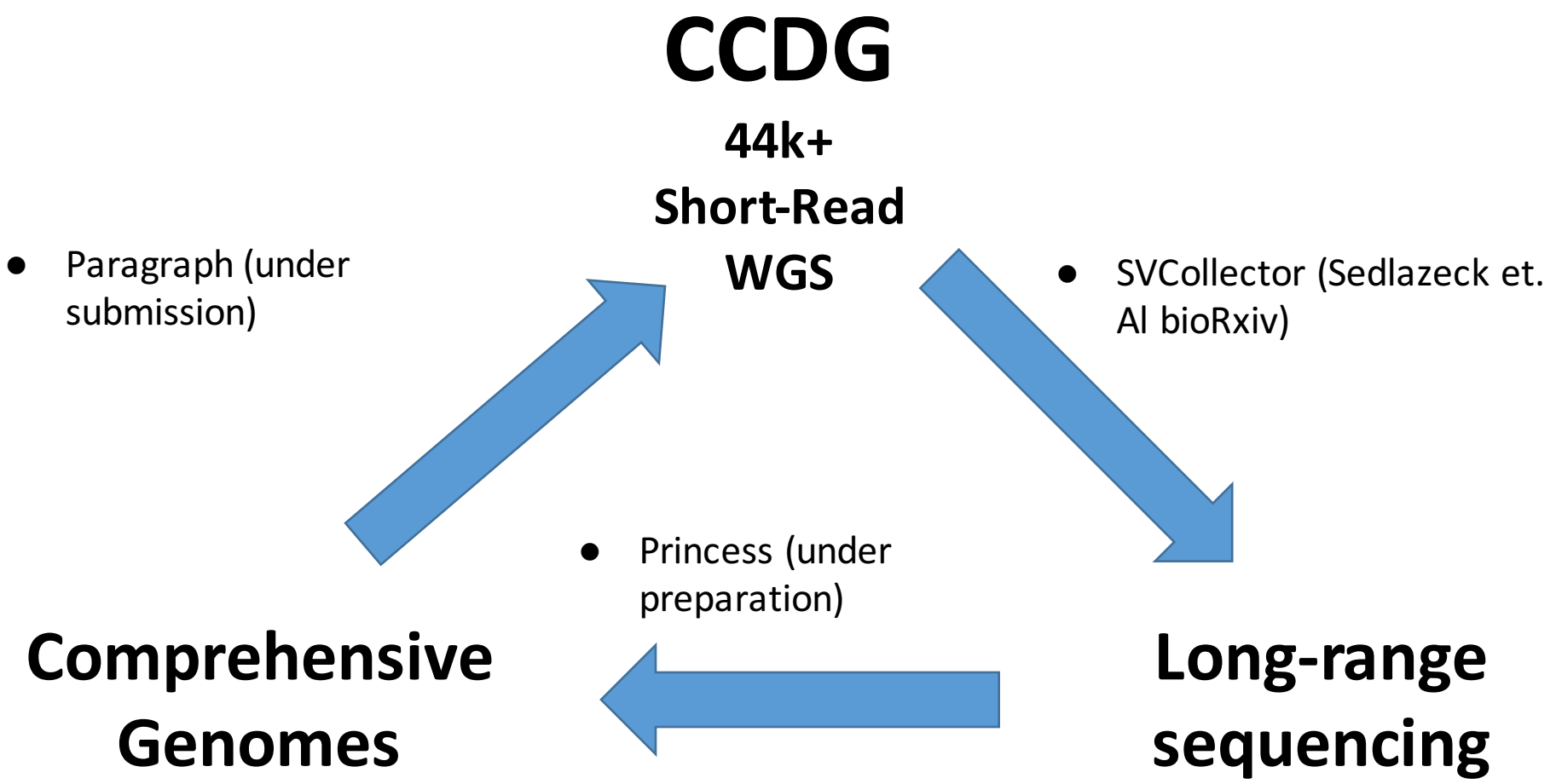


Upcoming…

- 18 matching Pacbio + 10x genomics + RNA-Seq data sets from CCDG
  - How can we apply Pacbio to population data sets?
  - How much benefit will we get?
  - Insights into phasing and SV calling onto RNA-Seq expression

- Sniffles optimization for STR regions

Upcoming application in large cohorts



Acknowledgments

HGSC
HUMAN GENOME SEQUENCING CENTER

Medhat Helmi

Richard Gibbs

JOHNS HOPKINS UNIVERSITY

Michael Schatz

Michael Kirsche

Schatz lab

香 港 大 學
THE UNIVERSITY OF HONG KONG

Ruibang Luo

NIH National Human Genome Research Institute

Princess: Fancy pipeline fully utilizing fancy data