ginkgo

SINGLE-CELL COPY-NUMBER ANALYSIS

VIZBI 2015 WORKSHOP

Robert Aboukhalil
raboukha@cshl.edu

What is Ginkgo?

Single-cell CNV analysis

Phylogenetic analysis

Data visualization tools

Share with collaborators
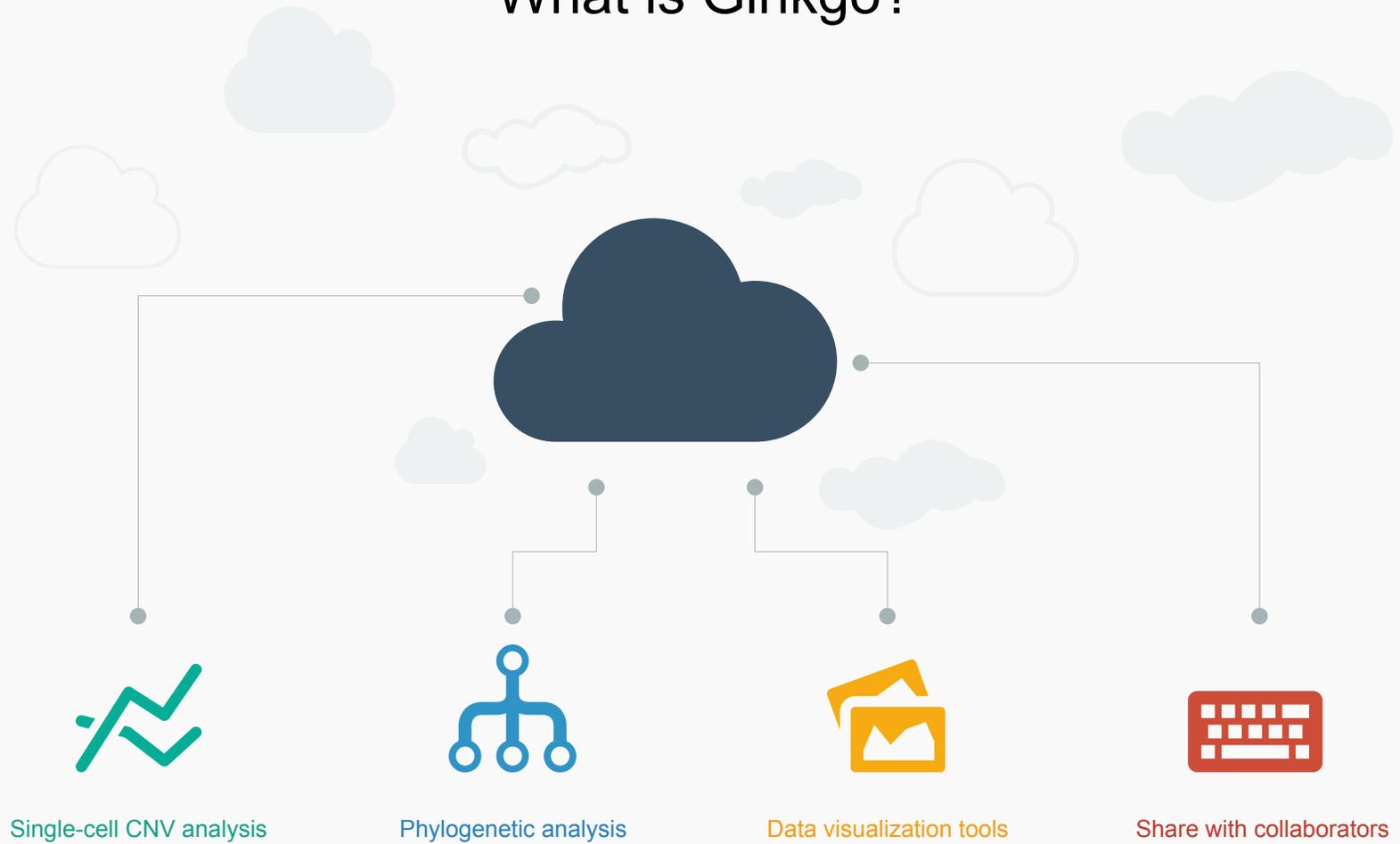
Outline

Experimental approaches

Computational approaches for CNV analysis

Demo

Hands-on

Single-cell sequencing
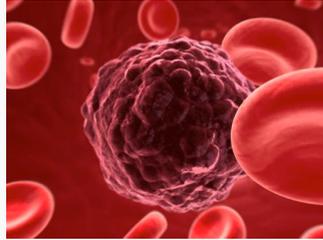
Circulating tumor cells

Neuronal mosaicism
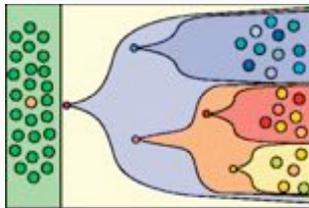
Clonal evolution in tumors
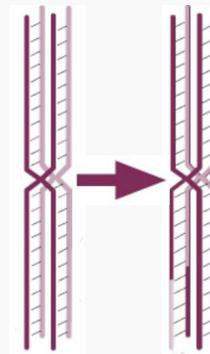
Recombination/ crossover in germ cells

Single-cell vs. bulk sequencing

ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

▶ Standard genome sequencing

Loads of DNA

A sample containing thousands to millions of cells is isolated.

DNA is extracted from all the nuclei.

DNA is broken into fragments and then sequenced.

The sequences are assembled to give a common, 'consensus' sequence.

GENOME ✓ COMPLETE

▶ Single-cell sequencing

Hardly any DNA

DNA amplification

A single cell is difficult to isolate, but it can be done mechanically or with an automated cell sorter.

The DNA is extracted and amplified, during which errors can creep in.

Amplified DNA is sequenced.

Errors introduced in earlier steps make sequence assembly difficult; the final sequence can have gaps.

GENOME ○ ERROR

Single-cell vs. bulk sequencing

ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

▶ **Standard genome sequencing**

Loads of DNA

A sample containing thousands to millions of cells is isolated.

DNA is extracted from all the nuclei.

DNA is broken into fragments and then sequenced.

The sequences are assembled to give a common, 'consensus' sequence.

GENOME ✓ COMPLETE

▶ **Single-cell sequencing**

Hardly any DNA

DNA amplification

A single cell is difficult to isolate, but it can be done mechanically or with an automated cell sorter.

The DNA is extracted and amplified, during which errors can creep in.

Amplified DNA is sequenced.

Errors introduced in earlier steps make sequence assembly difficult; the final sequence can have gaps.

GENOME ◯ ERROR

Whole Genome Amplification (WGA) methods

DOP-PCR (Degenerate Oligonucleotide Primed PCR)
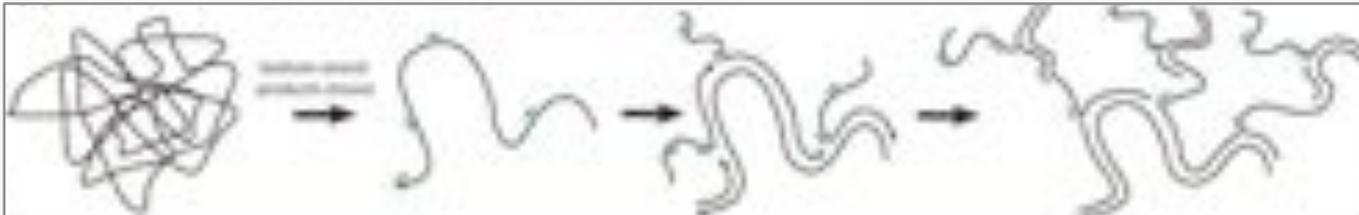
MDA (Multiple Displacement Amplification)

MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

Whole Genome Amplification (WGA) methods

DOP-PCR (Deg

MDA (Multiple I

MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

bit.ly/1EkCCOO

Cold Spring Harbor Laboratory

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

New Results

Interactive analysis and quality assessment of single-cell copy-number variations

Tyler Garvin , Robert Aboukhalil , Jude Kendall , Timour Baslan , Gurinder S Atwal , James Hicks , Michael Wigler , Michael Schatz

doi: http://dx.doi.org/10.1101/011346

Paul Blainey, FEMS Microbiol Rev. 2013

Copy-number variant analysis

Tyler Garvin
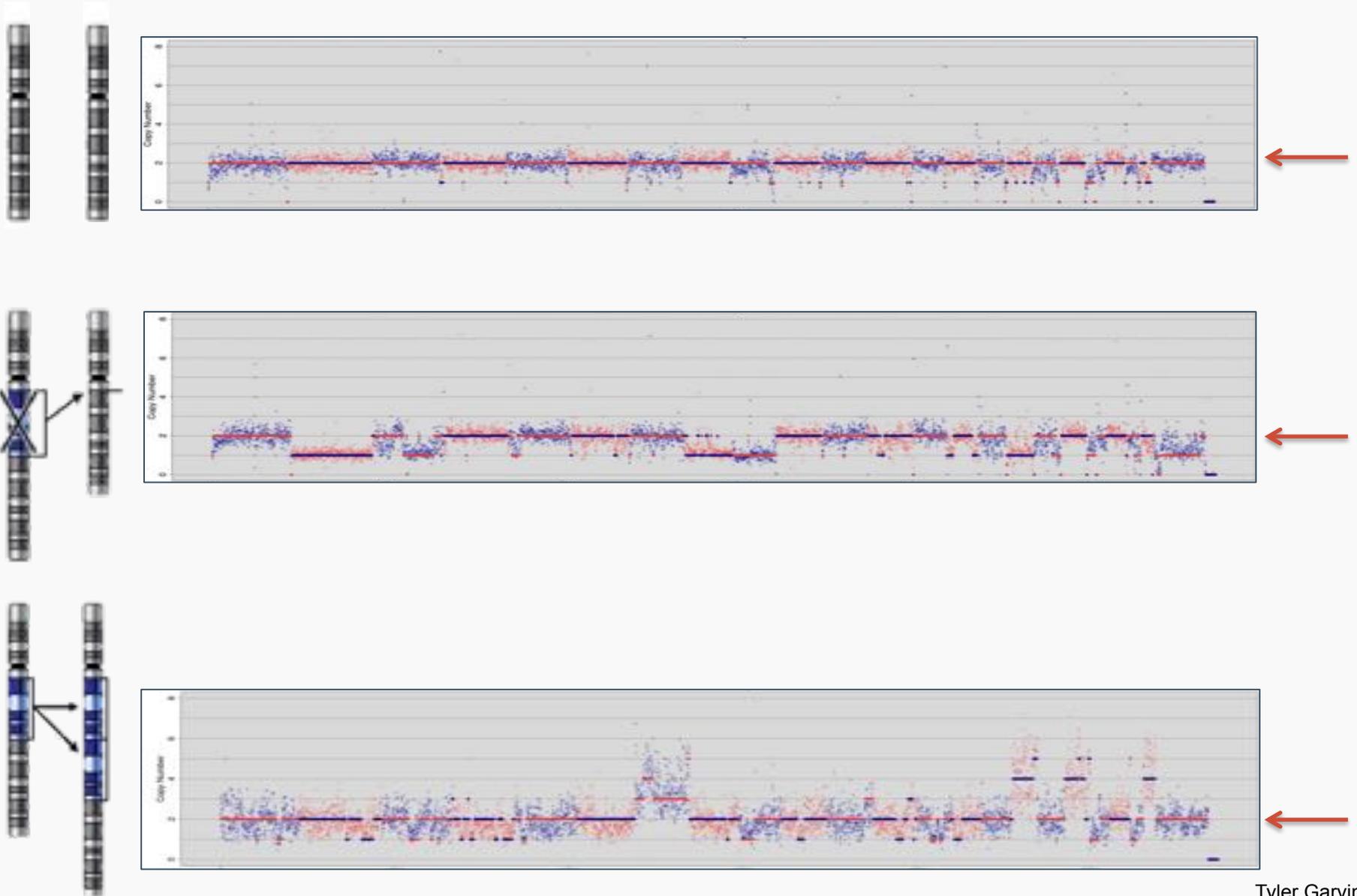
Copy-number variant analysis

Low coverage allows us to study copy-number variants

Copy-number variant analysis

NNNNNN

Divide genome into "bins" with ~50 – 100 reads / bin

Copy-number variant analysis
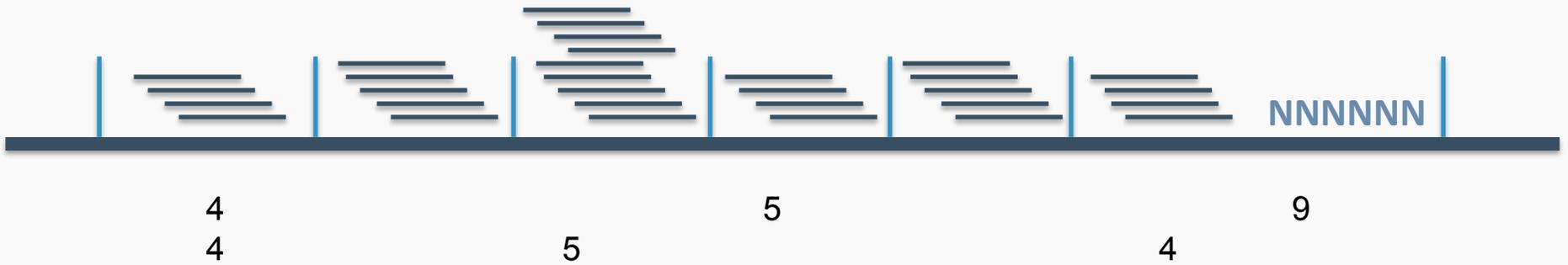
NNNNNN

4   5

4   5   4   9

Copy-number variant analysis

4
4

5

5

9

4

NNNNNN

Copy-number variant analysis

NNNNNN

4
4

5

9

5

4
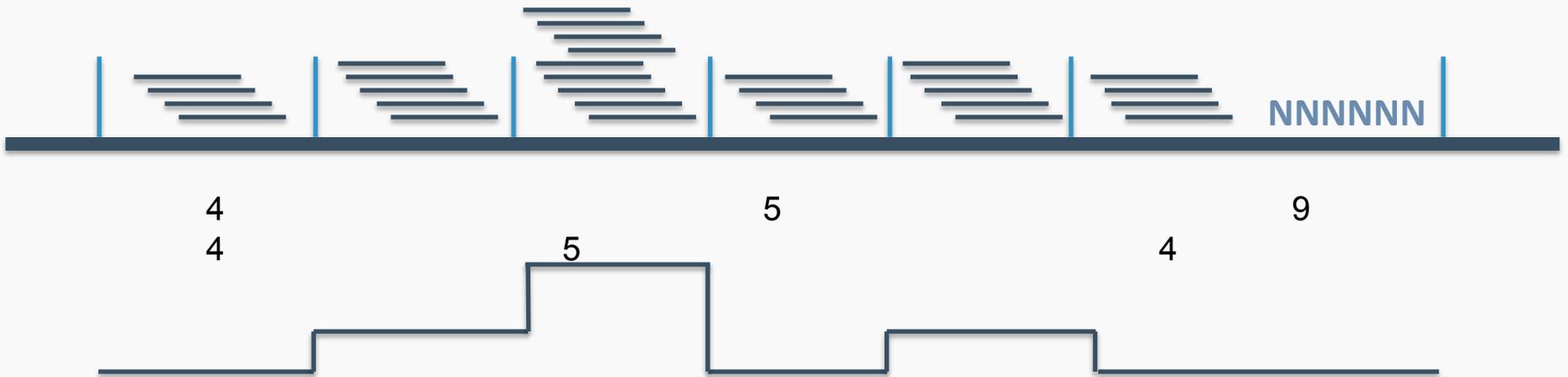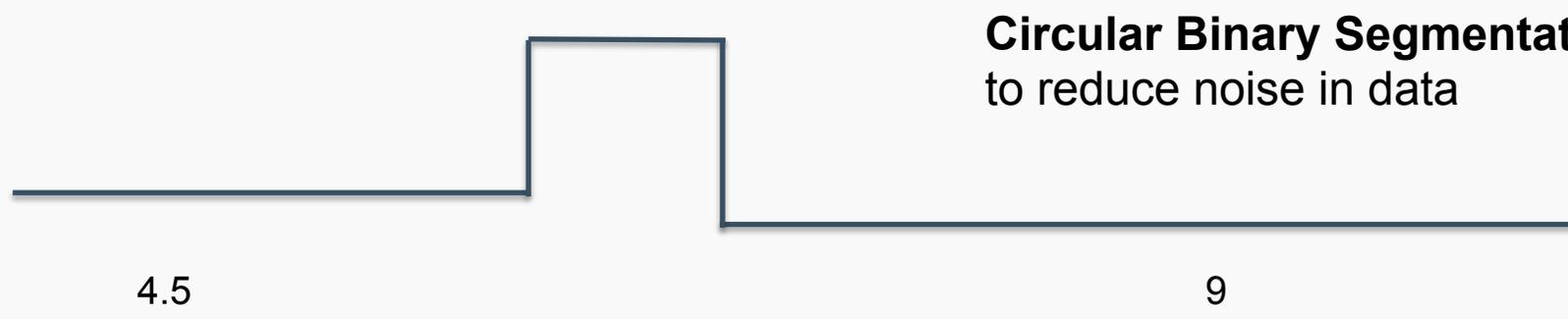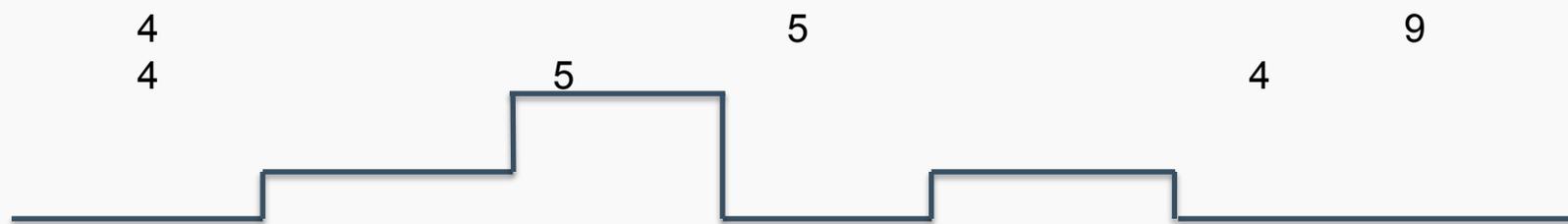
Circular Binary Segmentation (CBS
to reduce noise in data

4.5

9

Circular Binary Segmentation

- Recursively divide up the genome until identify segments that have probe distribution different than neighbors

BioDiscovery
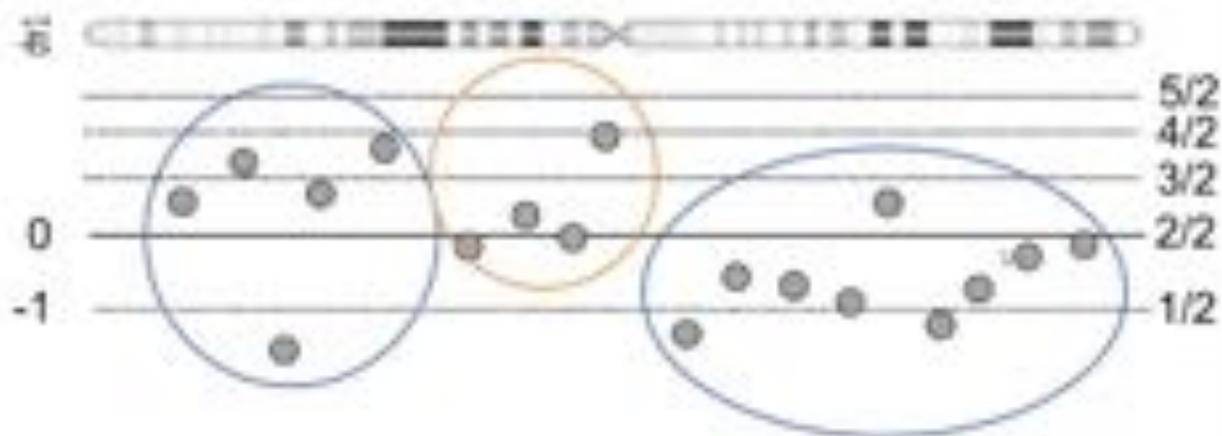
Circular Binary Segmentation

- Recursively divide up the genome until identify segments that have probe distribution different than neighbors

Circular Binary Segmentation
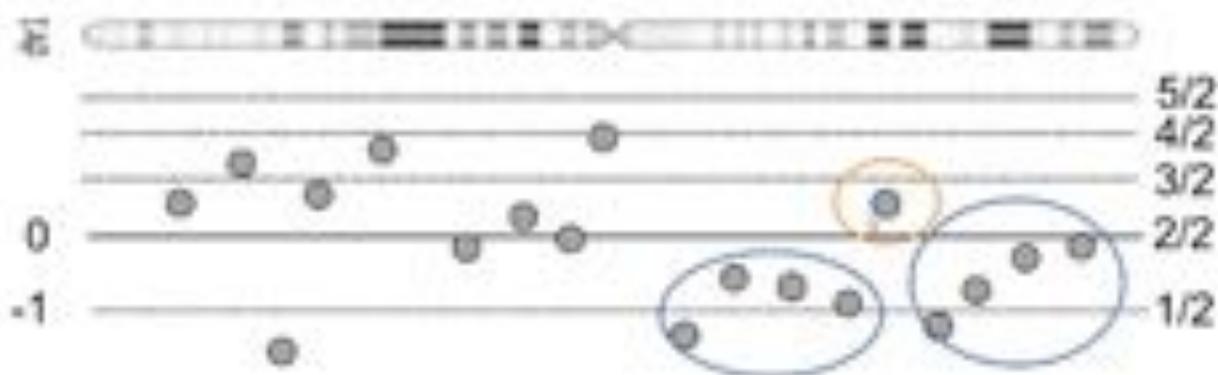
- Recursively divide up the genome until identify segments that have probe distribution different than neighbors

Copy-number variant analysis

Tyler Garvin

Copy-number variant analysis

Mike Schatz

Analysis Pipeline

.FASTQ

Galaxy

Map reads

Remove
duplicates

BAM to BED

.BED

Ginkgo

Bin reads

GC bias
correction

Segment bins

Call integer copy
number state

Interactive
visualization
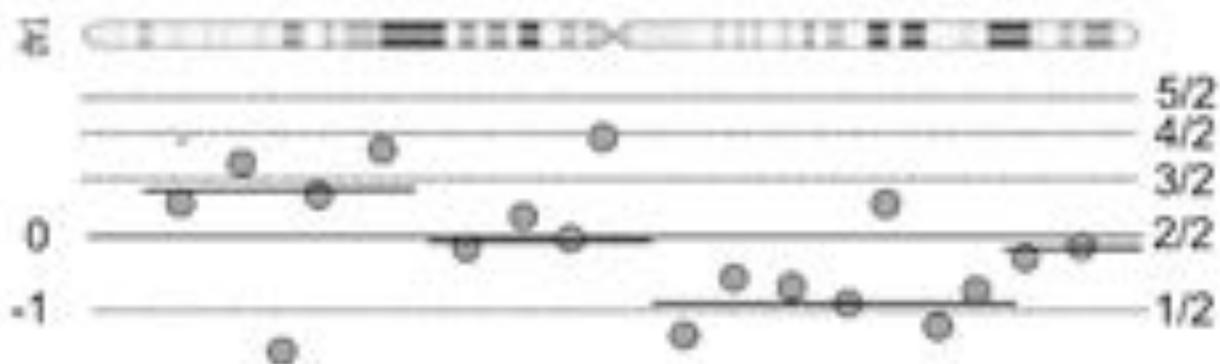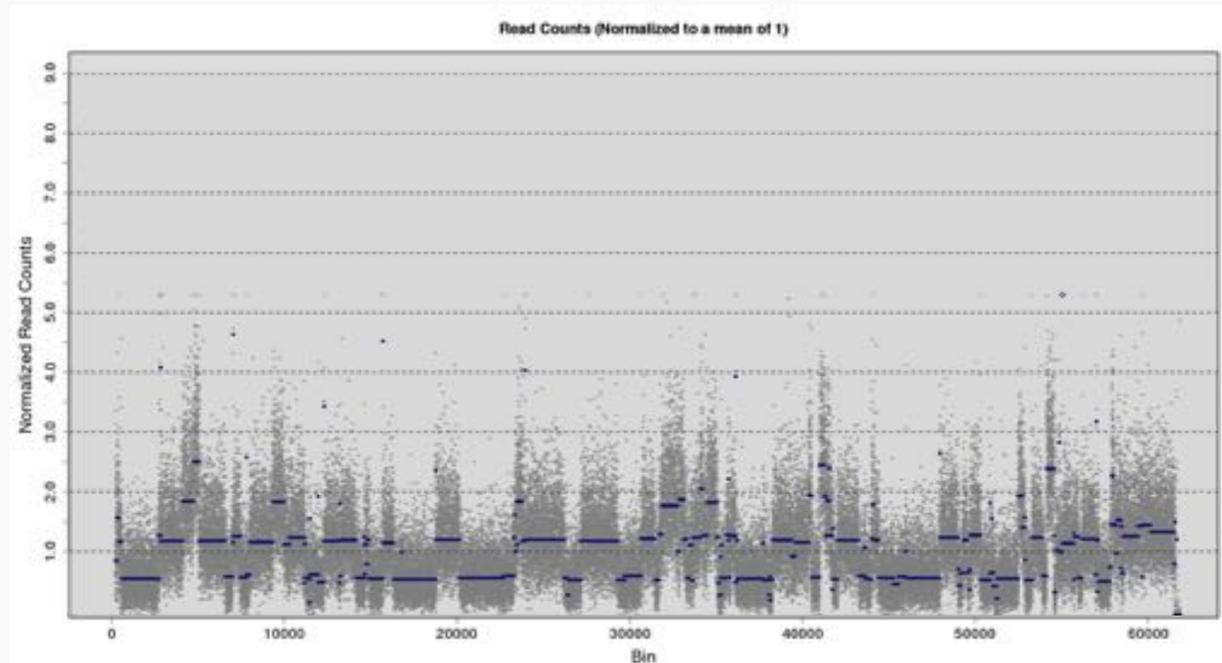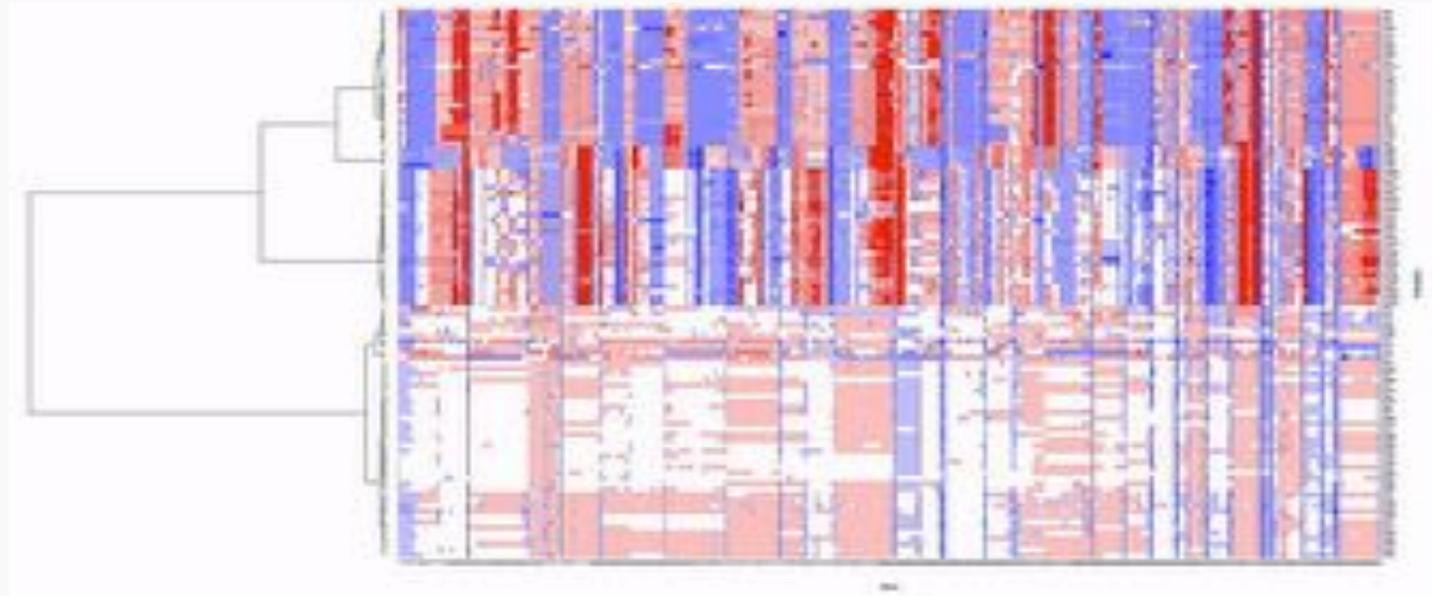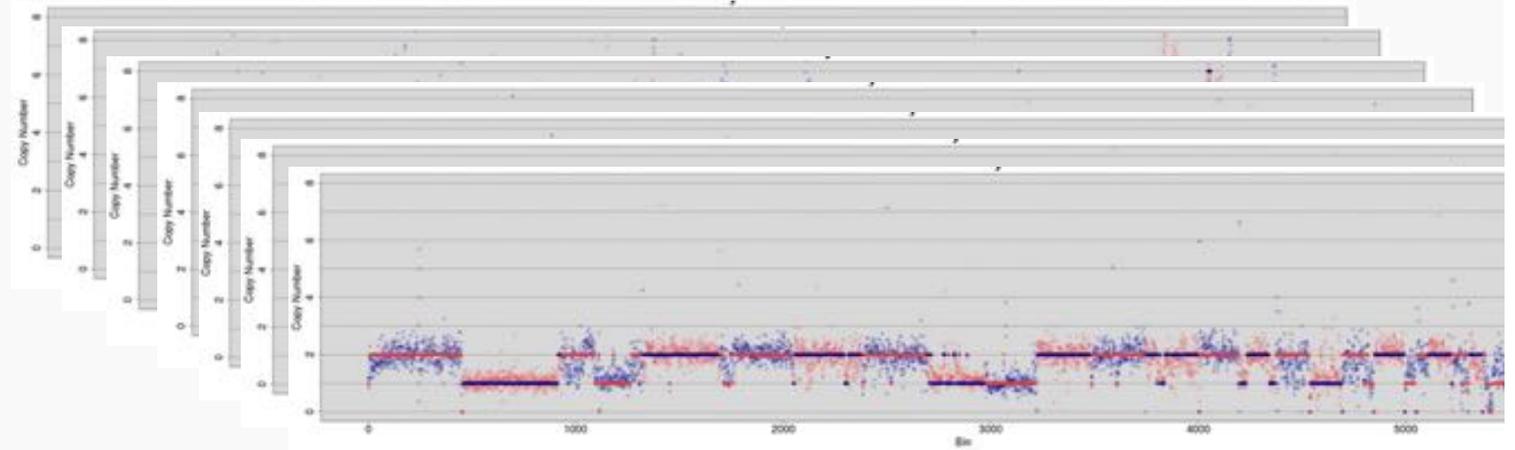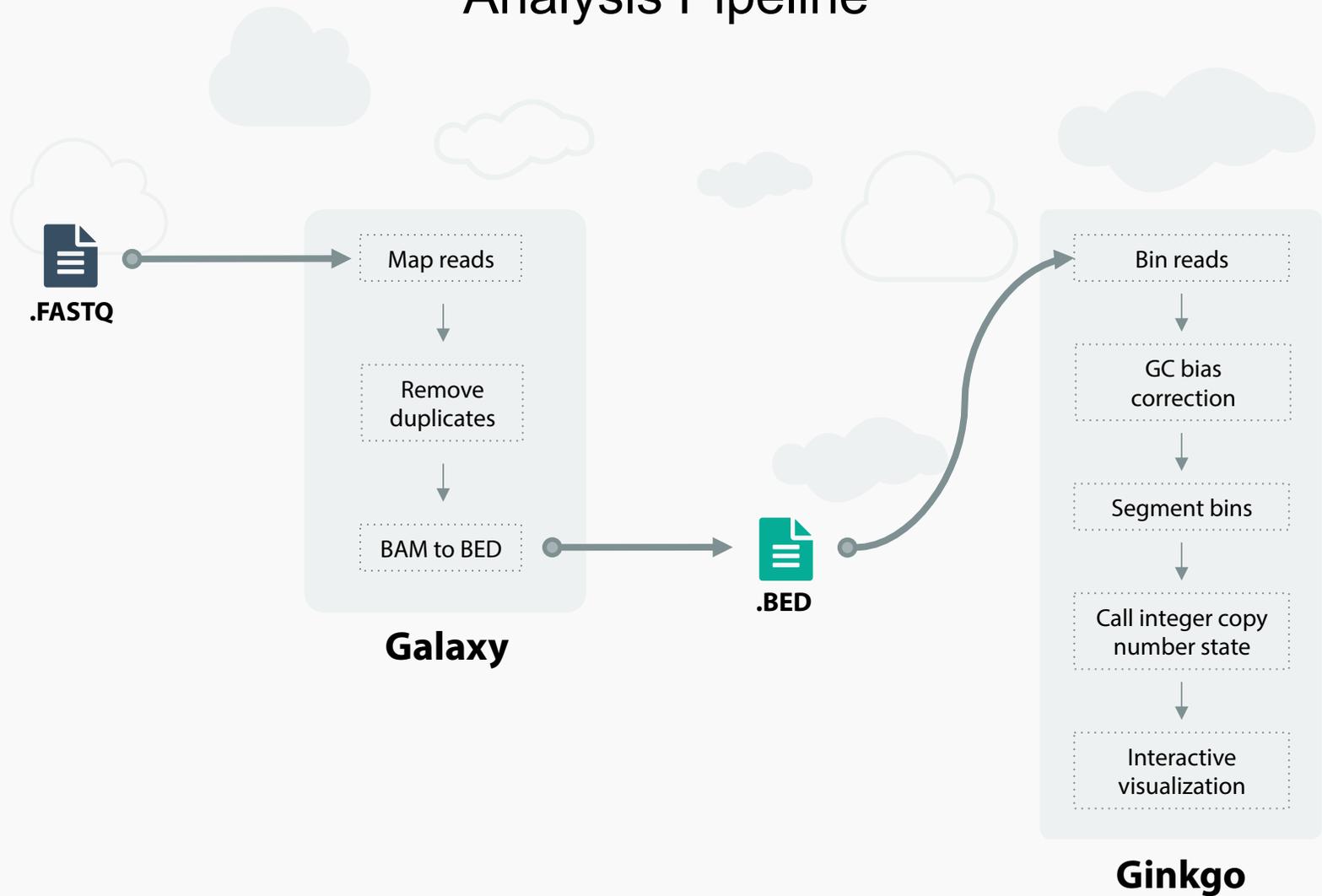
Galaxy intro

Demo – Galaxy Pipeline

Galaxy

Analyze Data Workflow Shared Data▾ Visualization▾ Cloud▾ Help▾ User▾

Tools

search tools

Get Data
Send Data
Lift-Over
Text Manipulation
Convert Formats
Filter and Sort
Join, Subtract and Group
NGS: QC and manipulation
NGS: Mapping
NGS: SAM Tools
NGS: Picard
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Graph/Display Data
Phenotype Association
snpEff
BEDTools
Genome Diversity
EMBOSS
Regional Variation
FASTA manipulation
Evolution
Multiple Alignments
Metagenomic analyses
Motif Tools
NGS TOOLBOX BETA

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

Want help?
Get answers.

Biostars
GALAXY EXPLAINED

Tweets   Follow

PyPI Recent Updates
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project [http://galaxypr... bit.ly/1O0CaYI
Retweeted by Galaxy Project
Expand

Galaxy Project @galaxyproject
#usegalaxy Histories now support advanced searching and offer multiple other improvements.
wiki.galaxyproject.org/Histories
pic.twitter.com/1JRlNYs5Q
View Photo

Yannick Wurm @ywurm...   12 Mar
Fun genome #bioinformatics #job in

Tweet to @galaxyproject

PENNSTATE   JOHNS HOPKINS   TACC   iPlant Collaborative
                        UNIVERSITY

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, and the Department of Biology and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the iPlant Collaborative at the Texas Advanced Computing Center, with support from the National Science Foundation.

History

search datasets

Unnamed history
0 bytes

Annotation
Click here to add annotation

This history is empty. You can load your own data or get data from an external source.

SRR054622

Analysis Pipeline

.FASTQ

Galaxy

Map reads

Remove duplicates

BAM to BED

.BED

Ginkgo

Bin reads

GC bias correction

Segment bins

Call integer copy number state

Interactive visualization
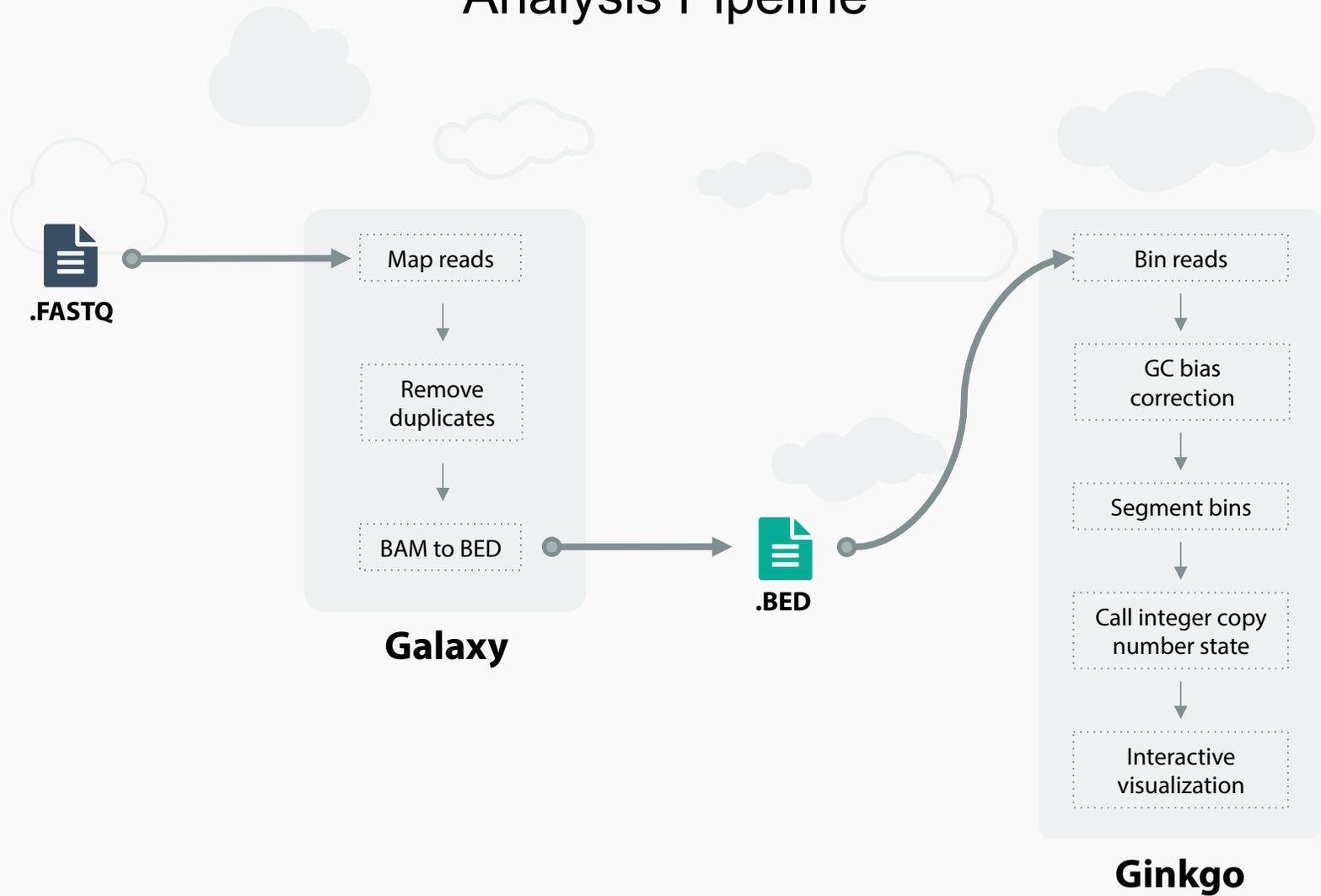
Hands-on with Ginkgo

Ginkgo: qb.cshl.edu/ginkgo

Sample Data: qb.cshl.edu/ginkgo/vizbi2015.tar