Detecting Copy Number Variations from Next-Generation Sequencing Data via a Bayesian Procedure

Yu-Chung Wei and Guan-Hua Huang

國立交通大學
National Chiao Tung University

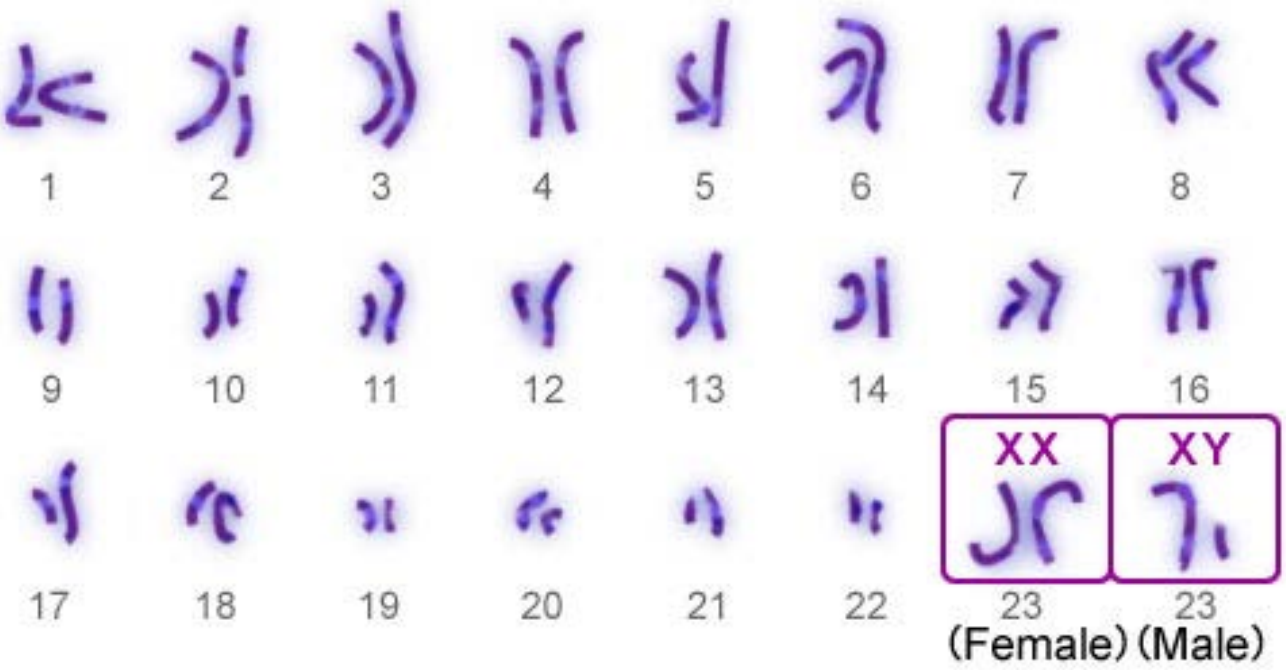
Detecting Copy Number Variations from Next-Generation Sequencing Data via a Bayesian Procedure

Yu-Chung Wei and Guan-Hua Huang

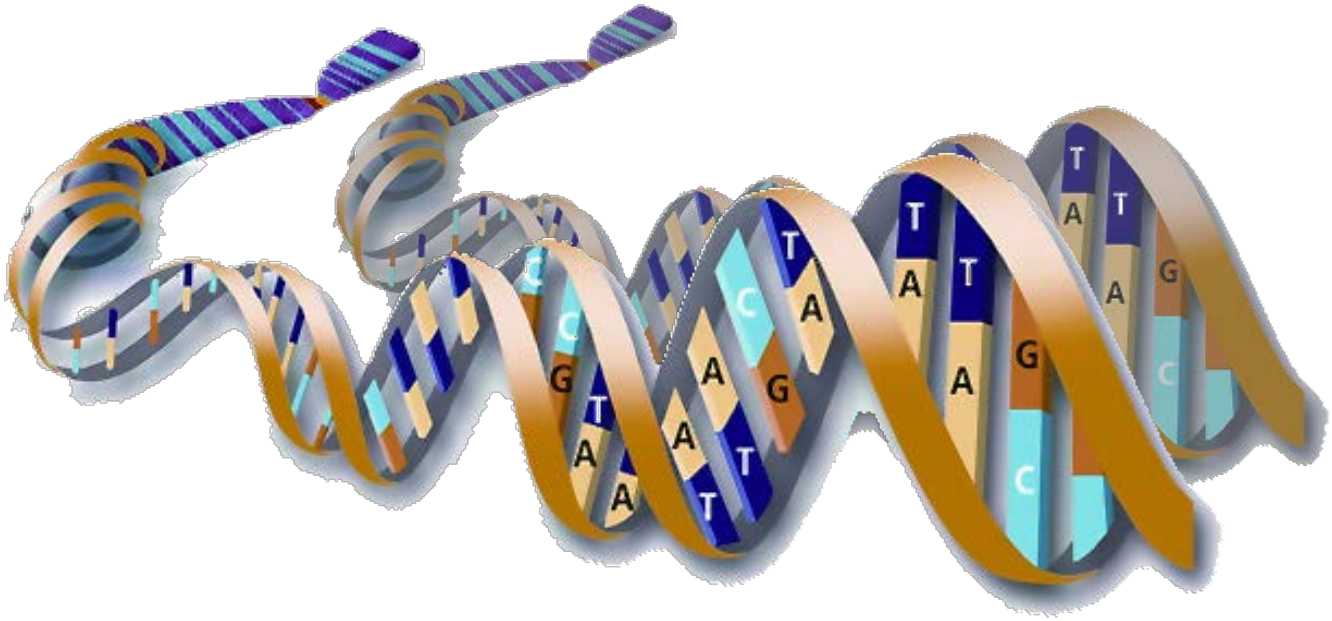
國立交通大學
National Chiao Tung University

- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

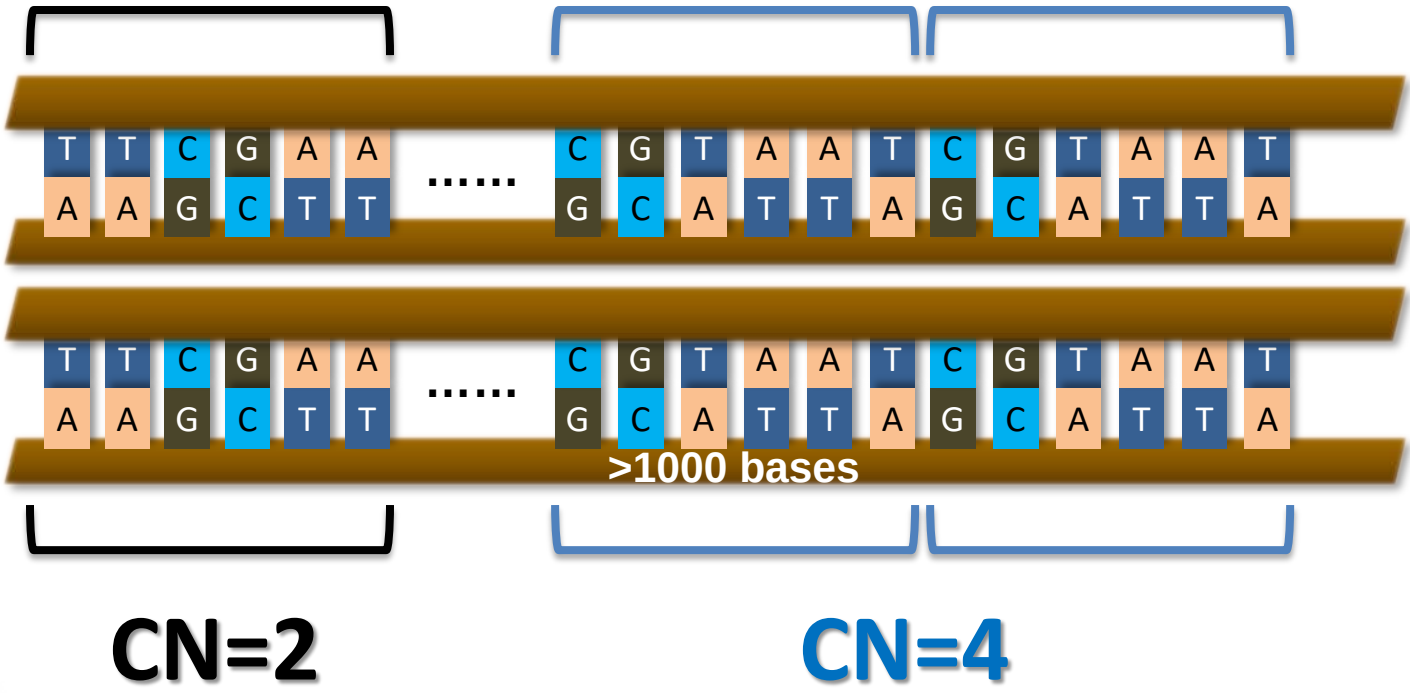
- **Copy Number Variations**
- **Sequencing Read Depths**
- **Bayesian Procedure**
  - •Model
  - •RJMCMC
- **NTUH data**



- **Copy Number Variations**
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data



- **Copy Number Variations**
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data



- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

Copy number variations (CNVs) and human genetic diseases

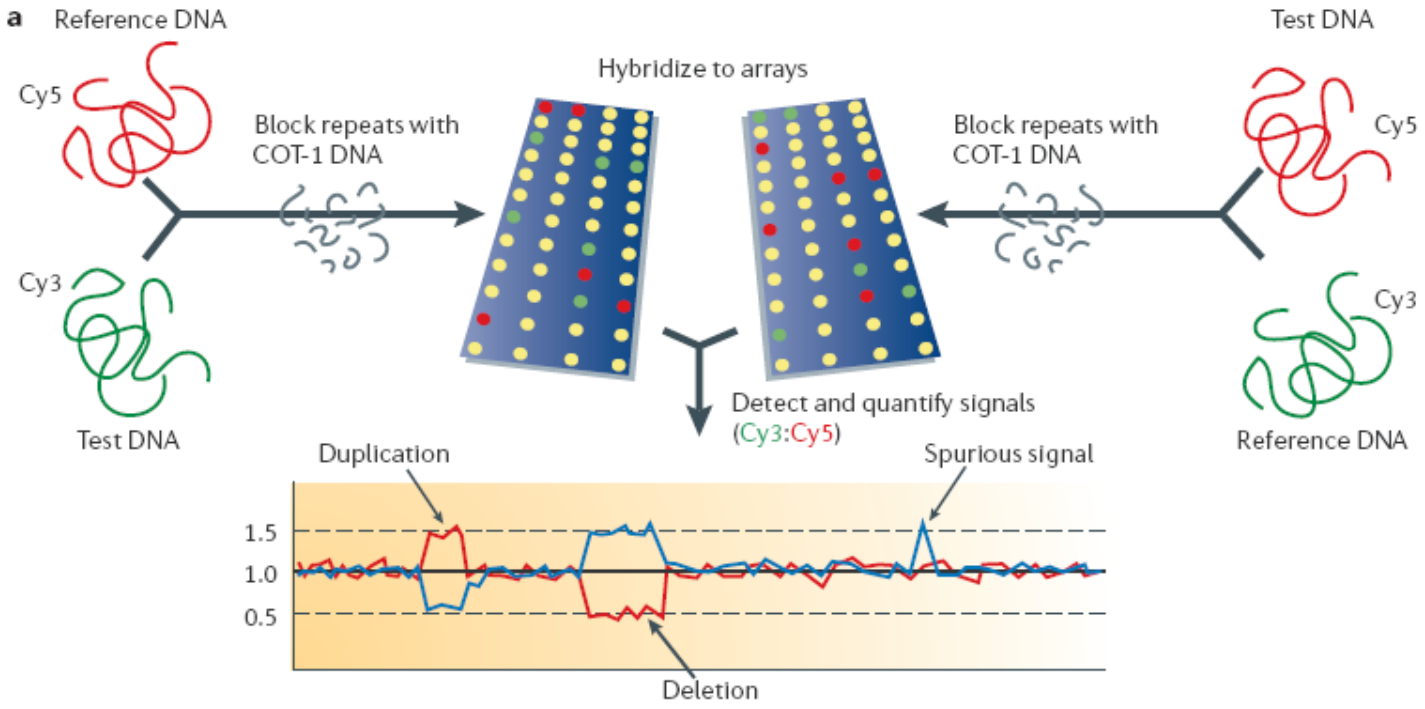
- CNVs account for roughly 12% of the human genome
- **Down syndrome**: a genetic disorder caused by the presence of a third copy of chromosome 21
- **Mental disorders**, including autism, schizophrenia: about 1% with rare DNA deletions in chromosome 15q13.3, 16p11.2, or 1q21.1

Breast cancer: 20-30% with HER-2 gene amplification and over-expression

- **Copy Number Variations**
- **Sequencing Read Depths**
- **Bayesian Procedure**
  - **Model**
  - **RJMCMC**
- **NTUH data**

Detecting CNVs

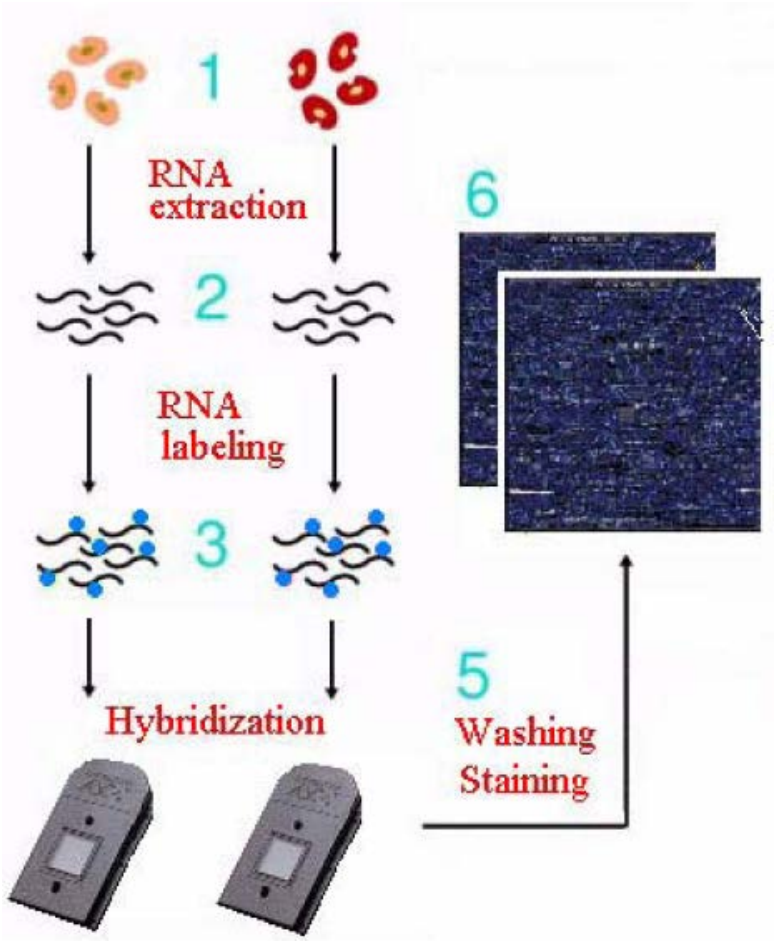
Array-based comparative genome hybridization (array-CGH)



- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

Detecting CNVs

Spotted oligonucleotides on Affymetrix SNP arrays



- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

Limitations of hybridization-based microarray approaches

- Hybridization-based microarray approaches: array-CGH and SNP arrays
- Microarrays are limited to
  - detect copy-number differences of sequences present in the reference assembly used to design the probes,
  - provide no information on the location of duplicated copies,
  - be generally unable to resolve breakpoints at the single-base-pair level

- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

Sequencing-based computational approaches

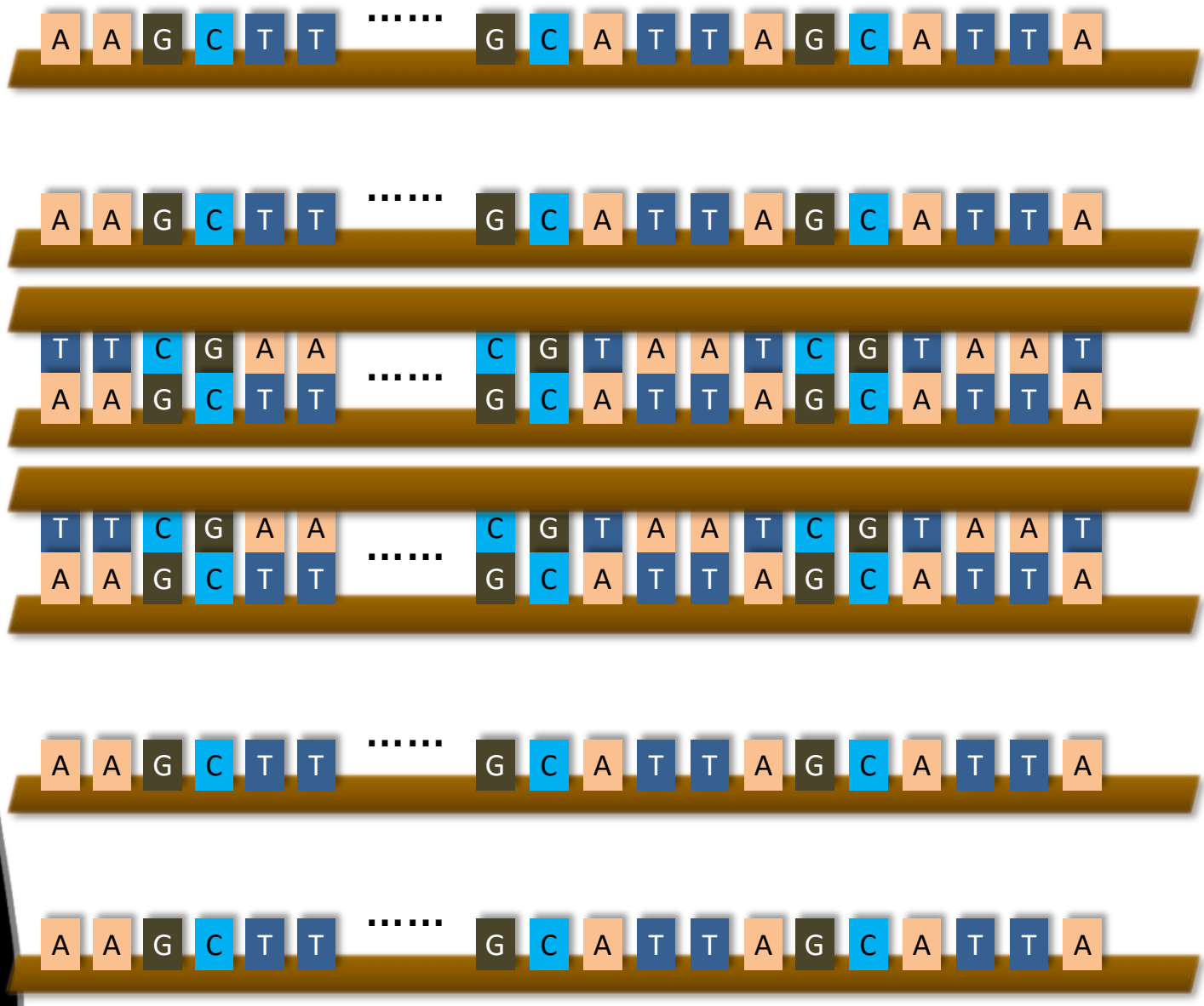
- The advent of next-generation sequencing (NGS) technologies promises to revolutionize copy number variation (CNV).
- NGS approaches can map CNVs with much greater accuracy than hybridization-based microarray approaches.
- However, NGS approaches present substantial computational and bioinformatics challenges.

- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

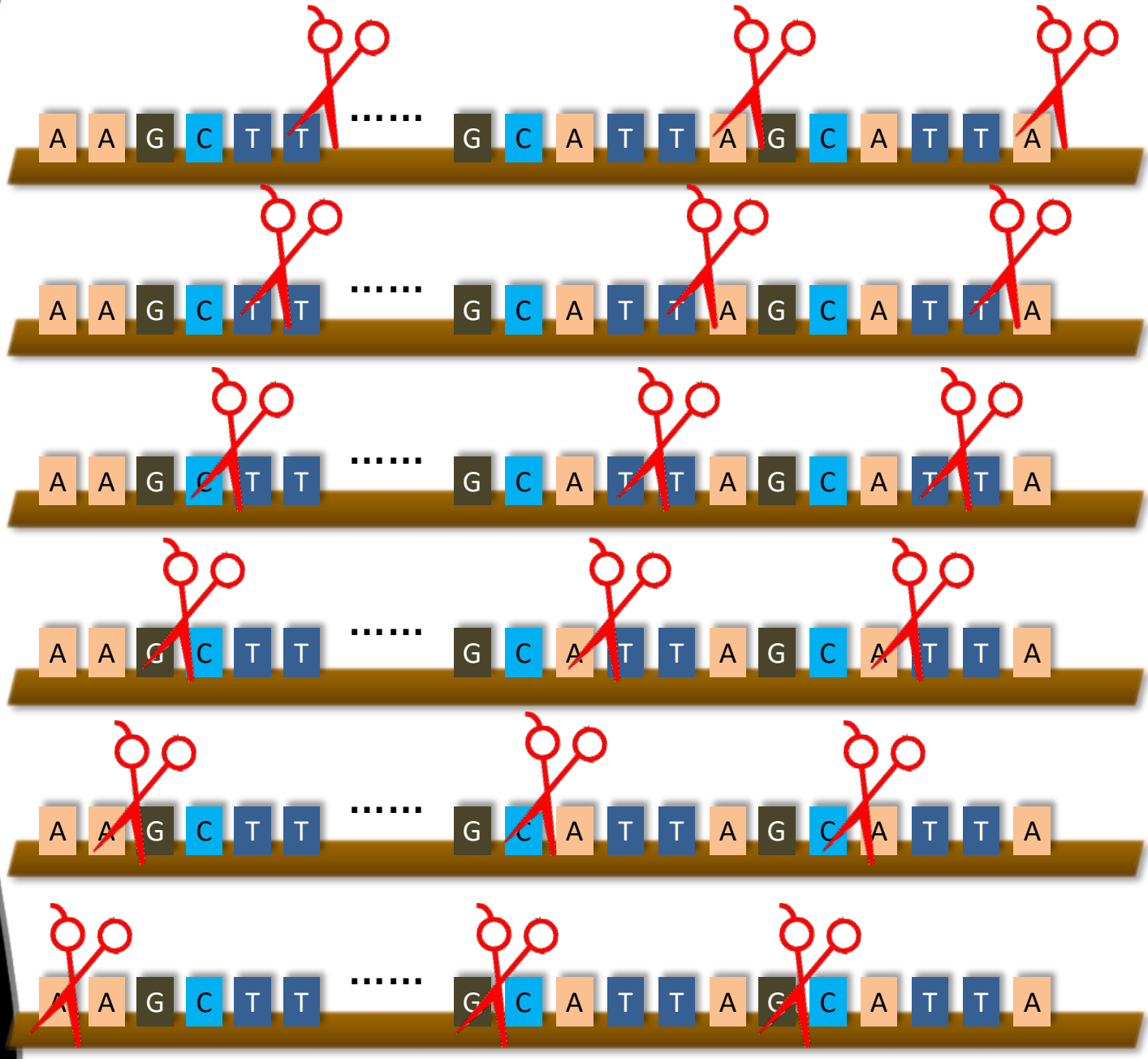
Sequencing-based computational approaches

- There are four general types of NGS strategy, all of which focus on mapping sequence reads to the reference genome and subsequently identifying CNVs:
  - read-pair (paired-end reads),
  - read-depth,
  - split-read,
  - sequence assembly.

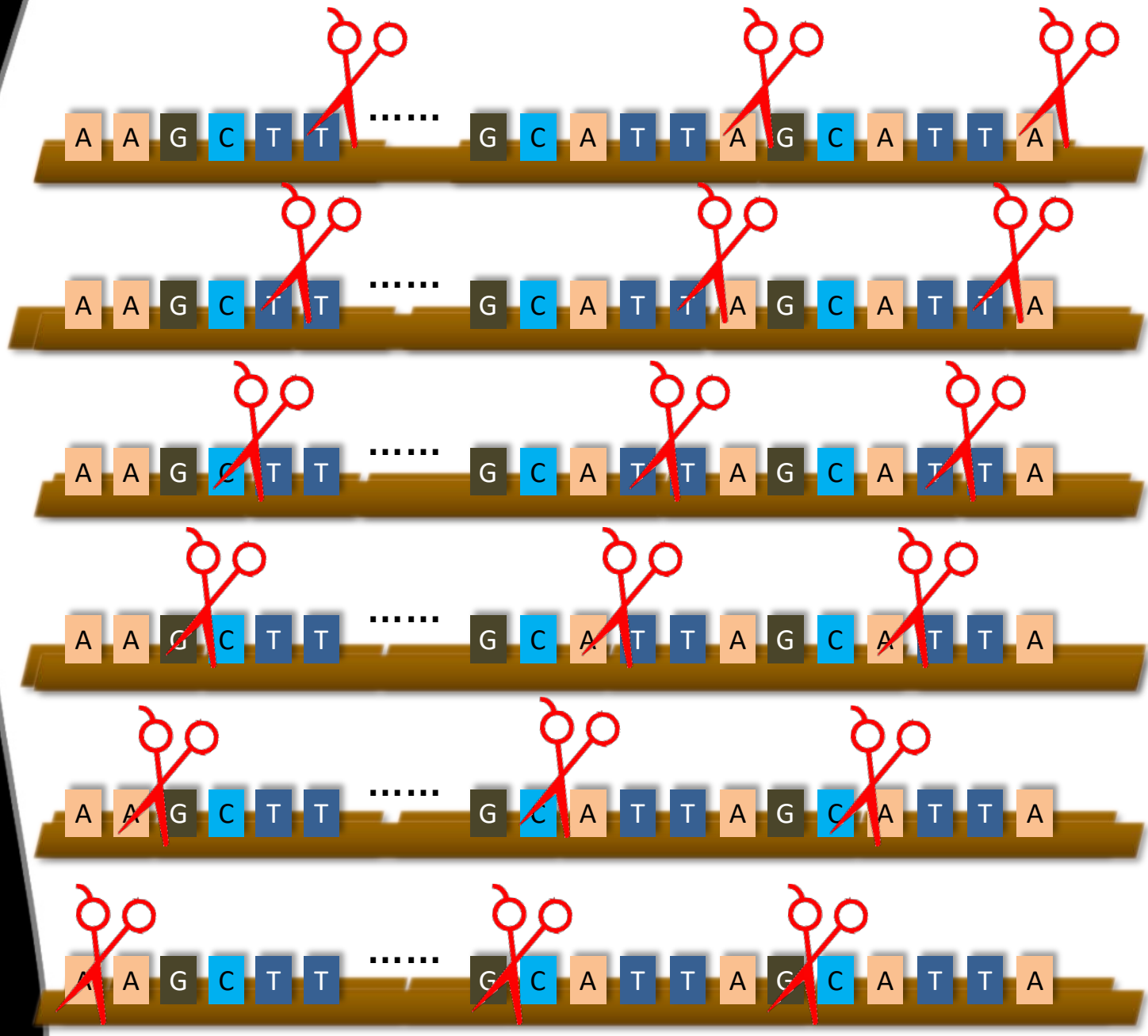
- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data



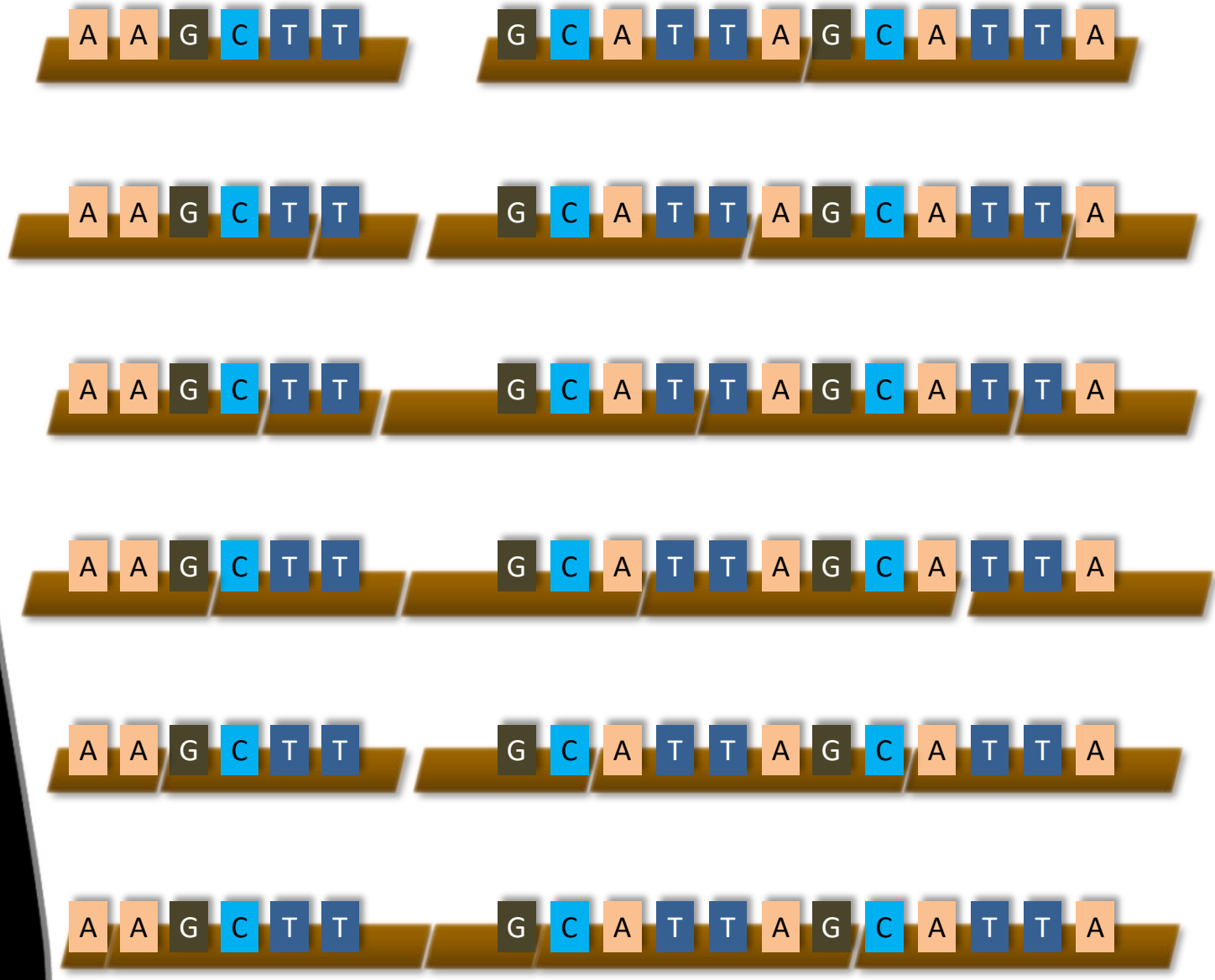
- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data



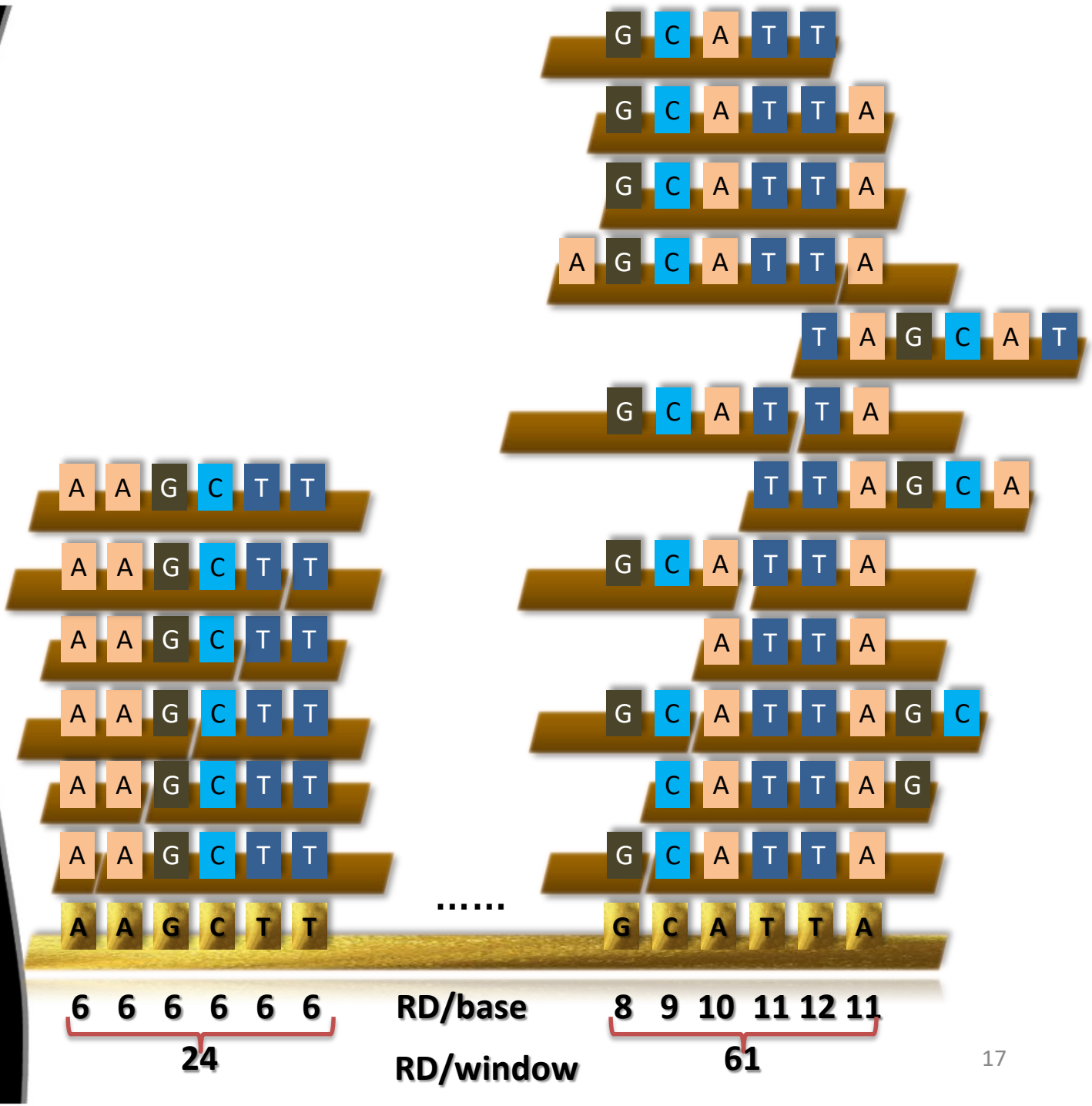
- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data



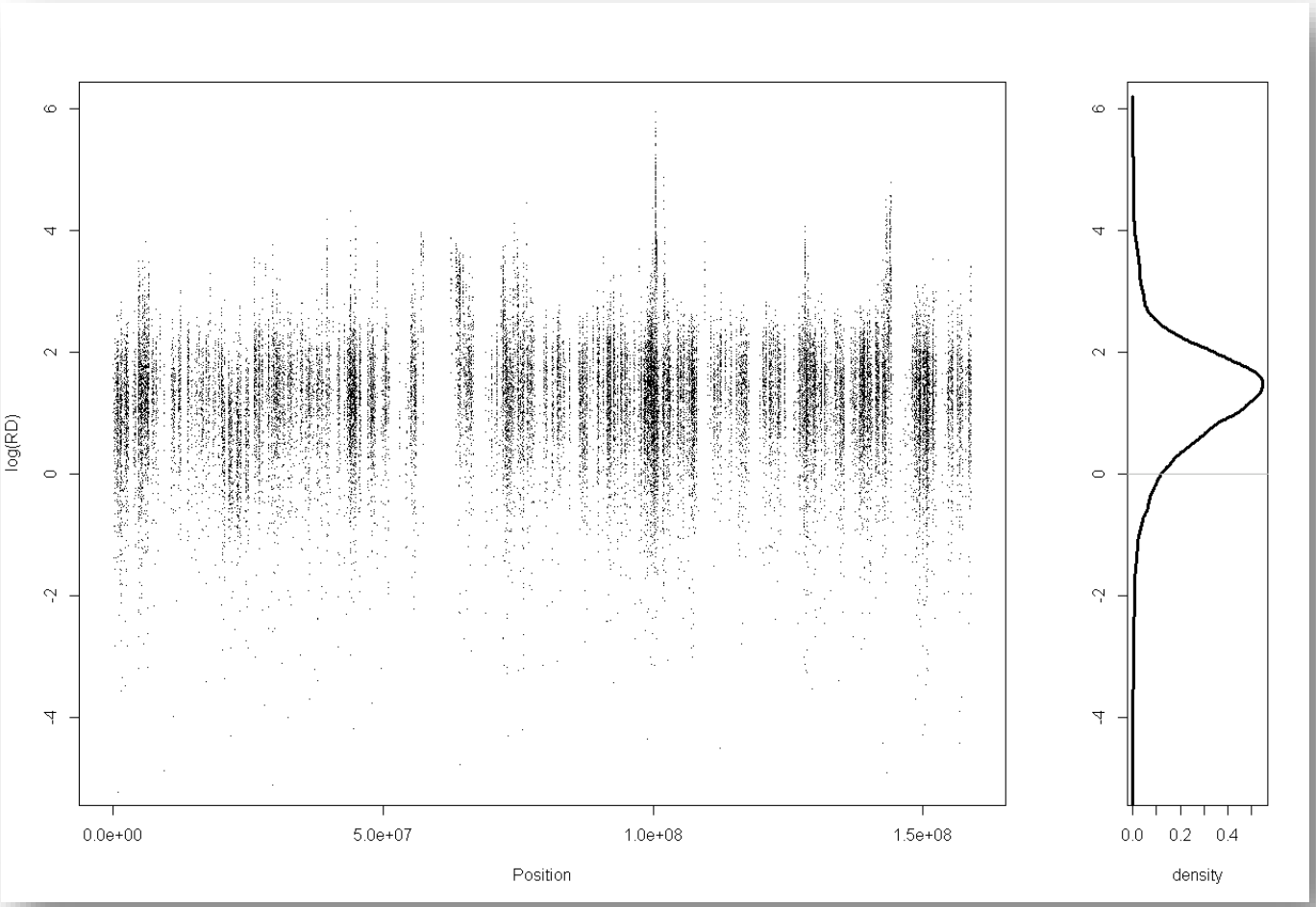
- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data



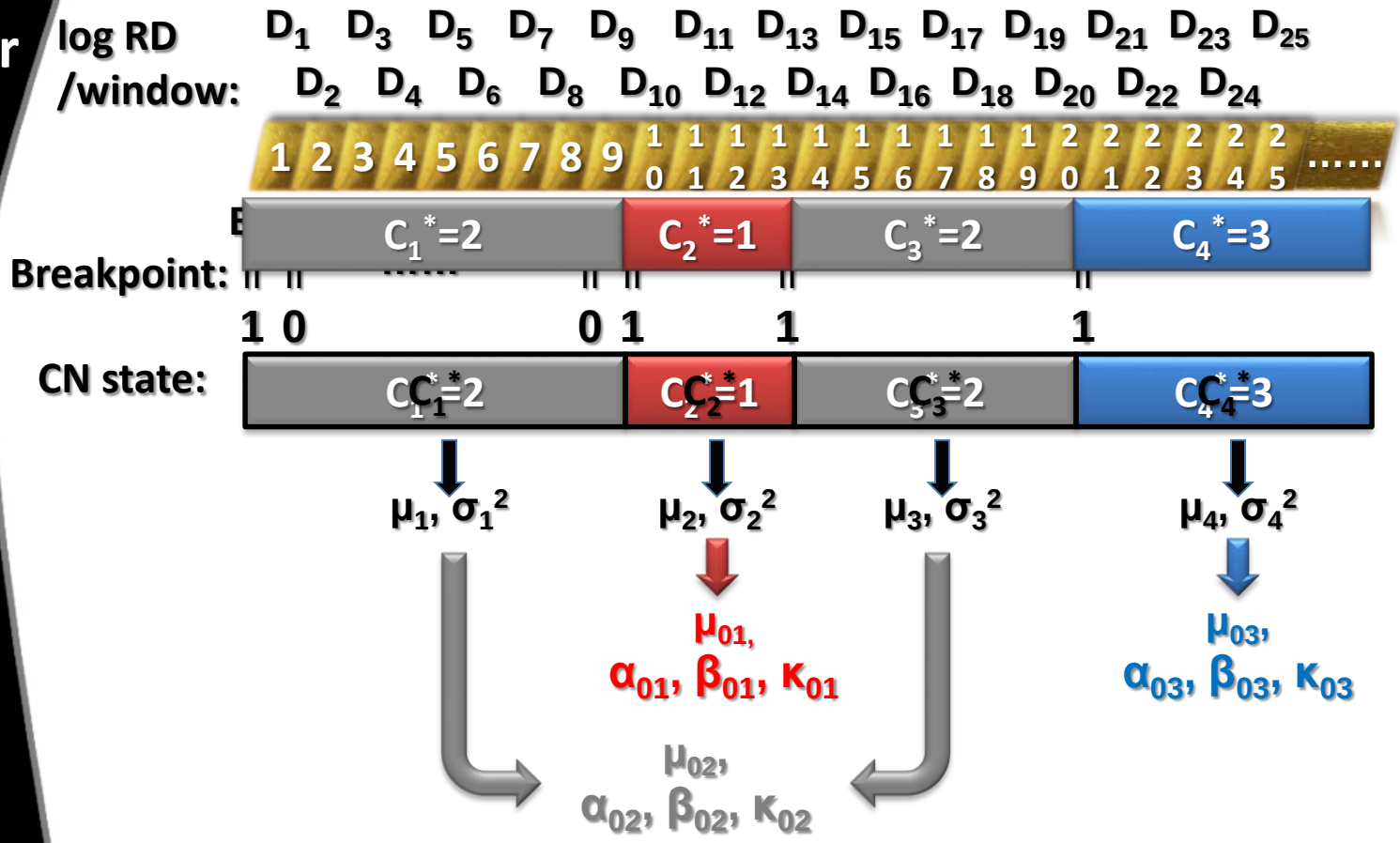
- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data



- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data



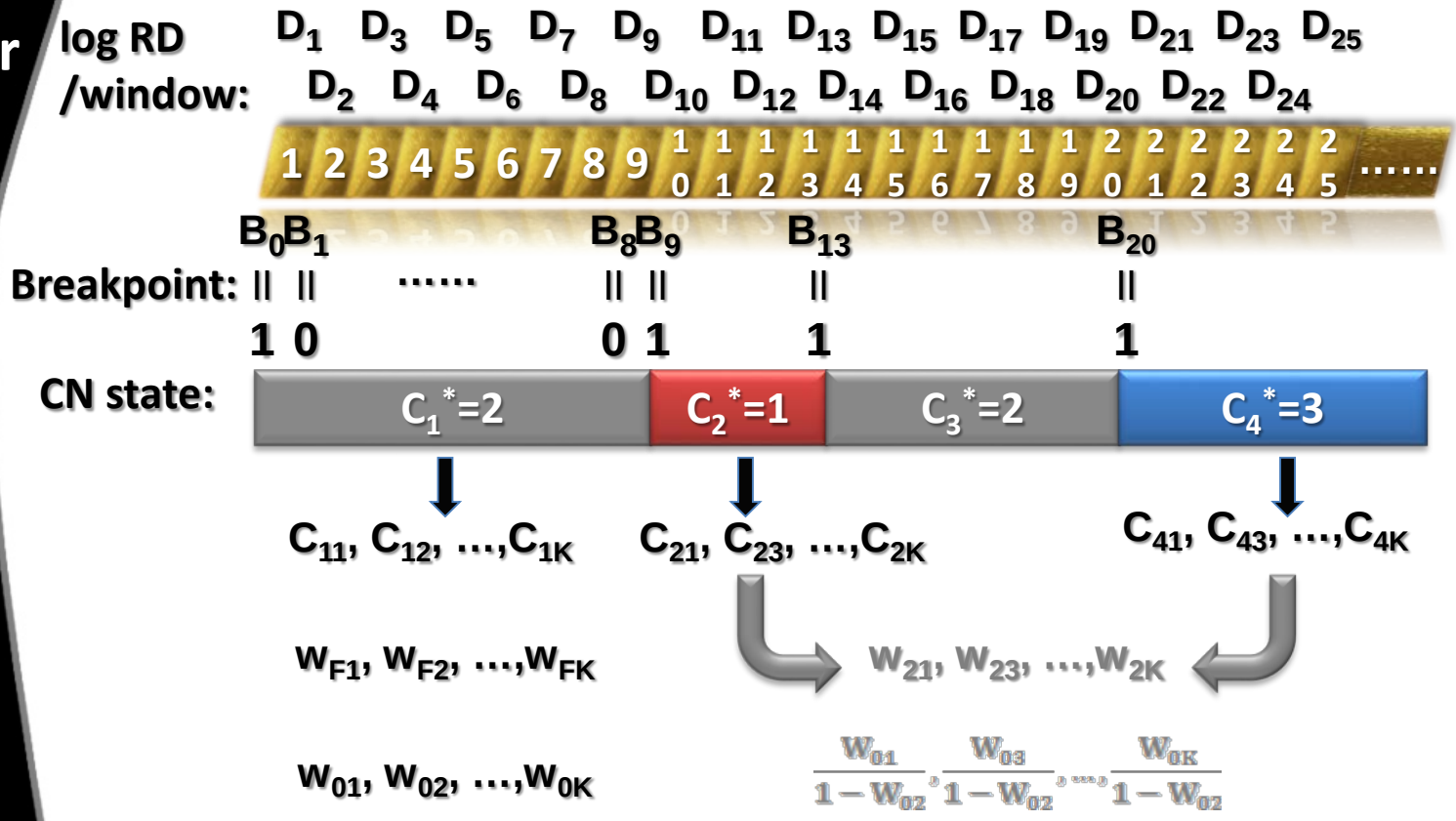
- Copy Number Variations
- Sequencing Read Depths
- **Bayesian Procedure**
  - **•Model**
  - •RJMCMC
- NTUH data



$$P(\underset{\sim}{C},\underset{\sim}{B}|\underset{\sim}{D})$$

$$\propto P(\underset{\sim}{D}|\underset{\sim}{C},\underset{\sim}{B})\times P(\underset{\sim}{C},|\underset{\sim}{B})\times P(\underset{\sim}{B})$$

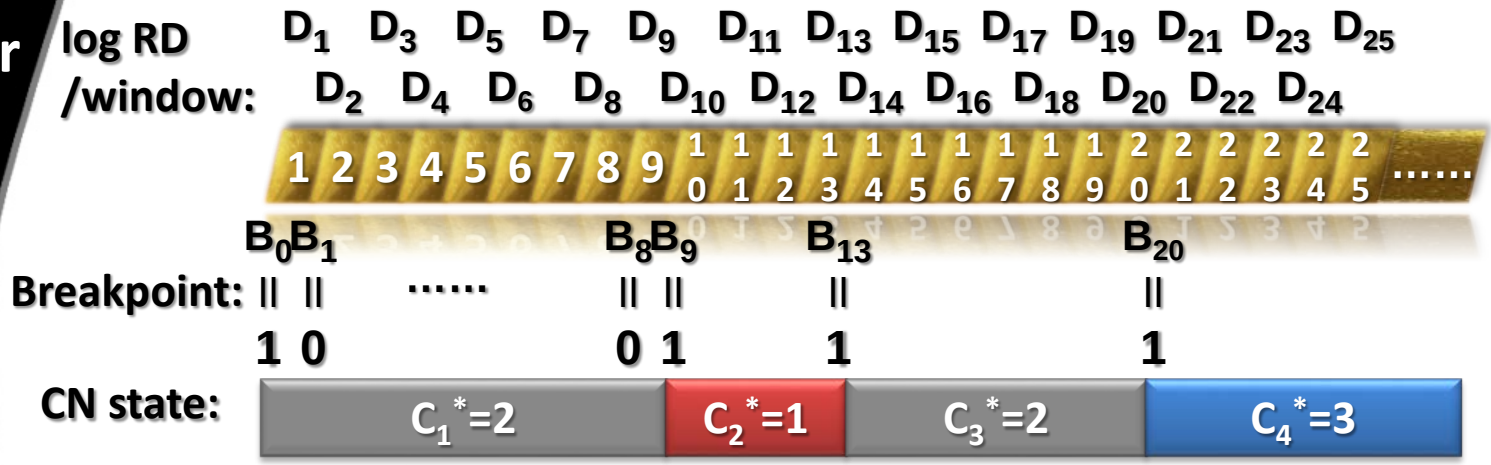
- Copy Number Variations
- Sequencing Read Depths
- **Bayesian Procedure**
  - •Model
  - •RJMCMC
- NTUH data



$$P\left(\underset{\sim}{C},\underset{\sim}{B}\middle|\underset{\sim}{D}\right)$$

$$\propto P\left(\underset{\sim}{D}\middle|\underset{\sim}{C},\underset{\sim}{B}\right)\times P\left(\underset{\sim}{C}\middle|\underset{\sim}{B}\right)\times P\left(\underset{\sim}{B}\right)$$

- Copy Number Variations
- Sequencing Read Depths
- **Bayesian Procedure**
  - **Model**
  - RJMCMC
- NTUH data

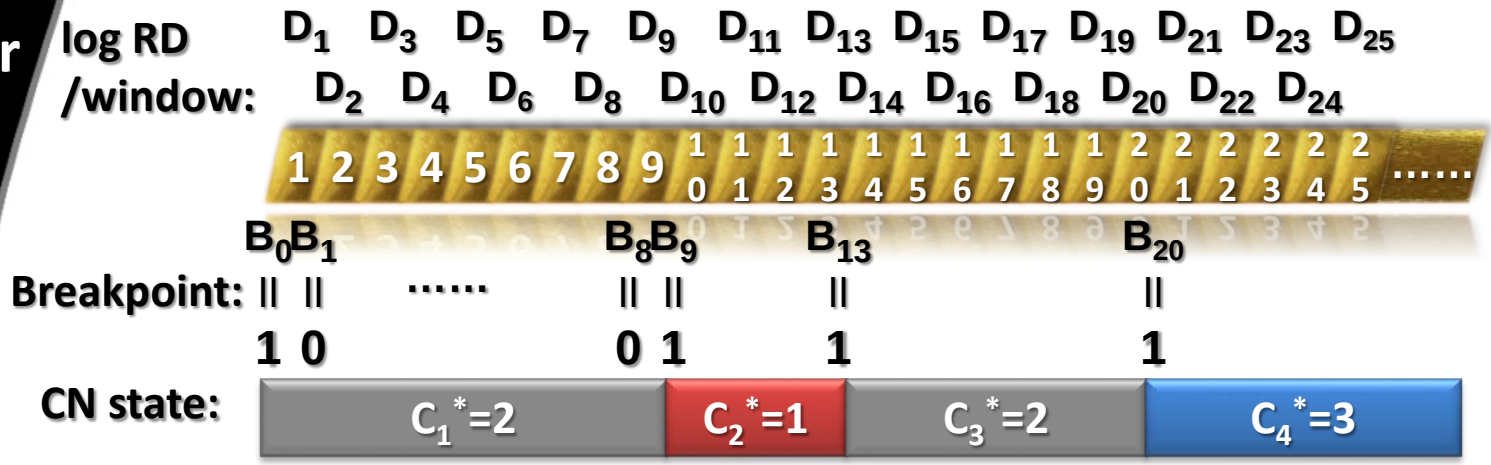


$$B_i \sim \text{Binomial}(\lambda)$$

$$P(\underset{\sim}{C}, \underset{\sim}{B} \mid \underset{\sim}{D})$$

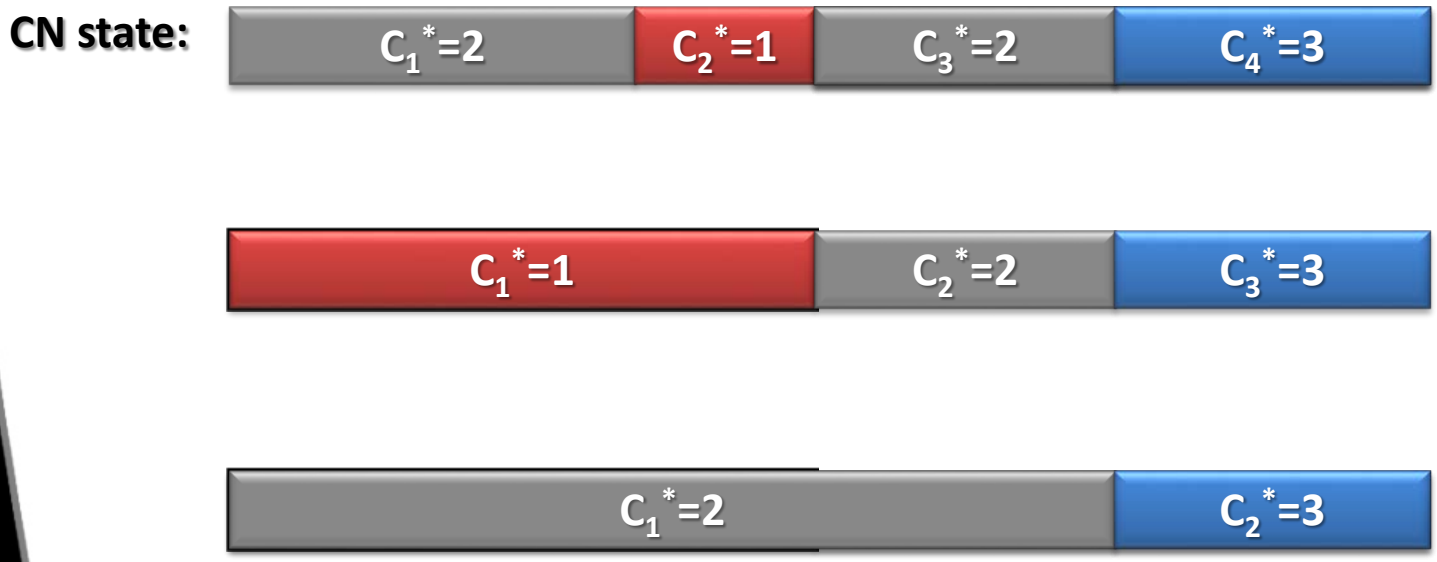
$$\propto P(\underset{\sim}{D} \mid \underset{\sim}{C}, \underset{\sim}{B}) \times P(\underset{\sim}{C} \mid \underset{\sim}{B}) \times P(\underset{\sim}{B})$$

- Copy Number Variations
- Sequencing Read Depths
- **Bayesian Procedure**
  - **•Model**
  - •RJMCMC
- NTUH data



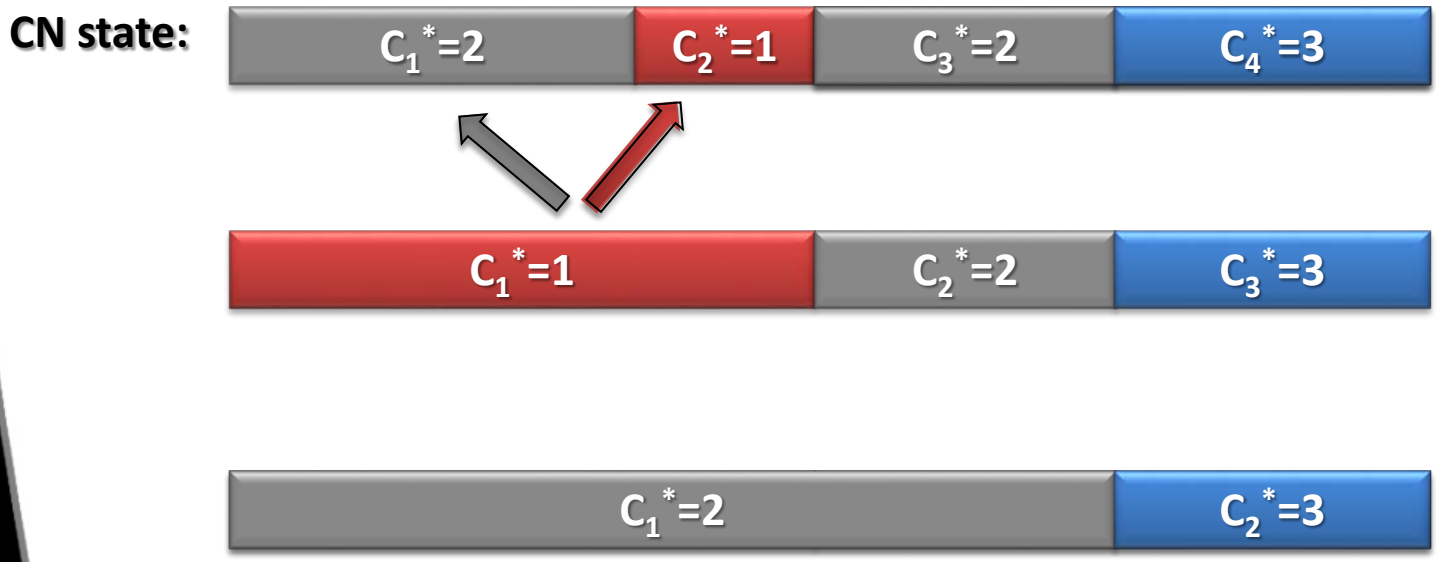
- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

1. Merge



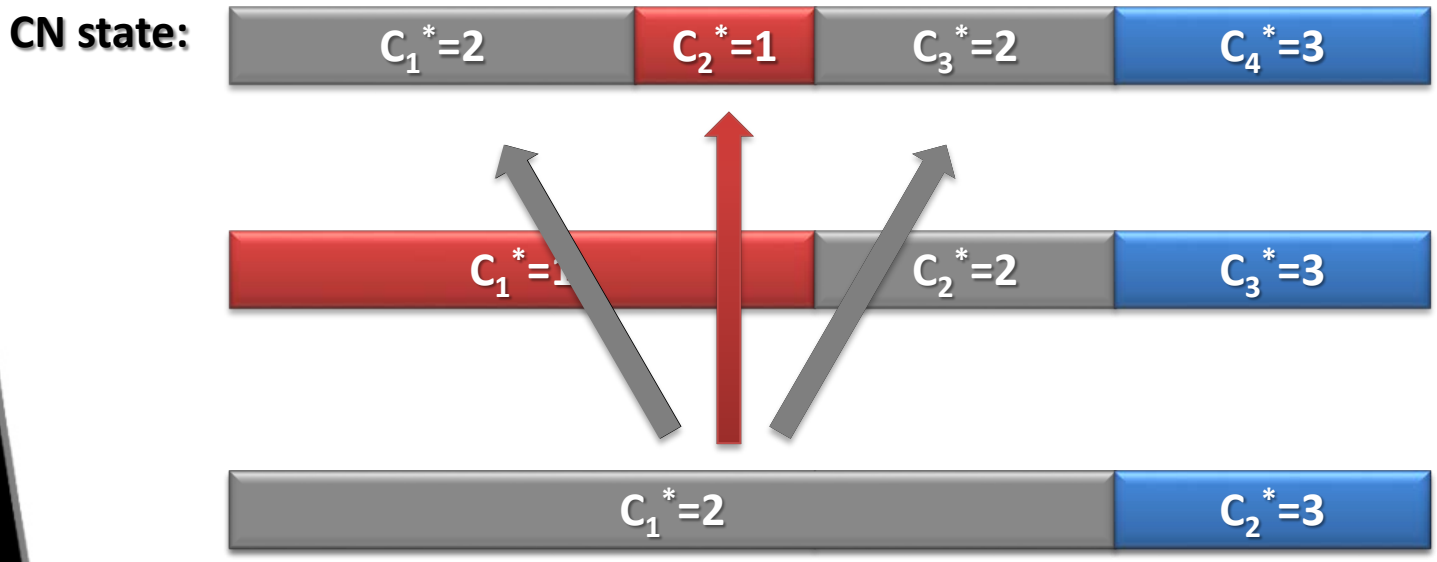
- Copy Number Variations
- Sequencing Read Depths
- **Bayesian Procedure**
  - •Model
  - •RJMCMC
- NTUH data

2. Split



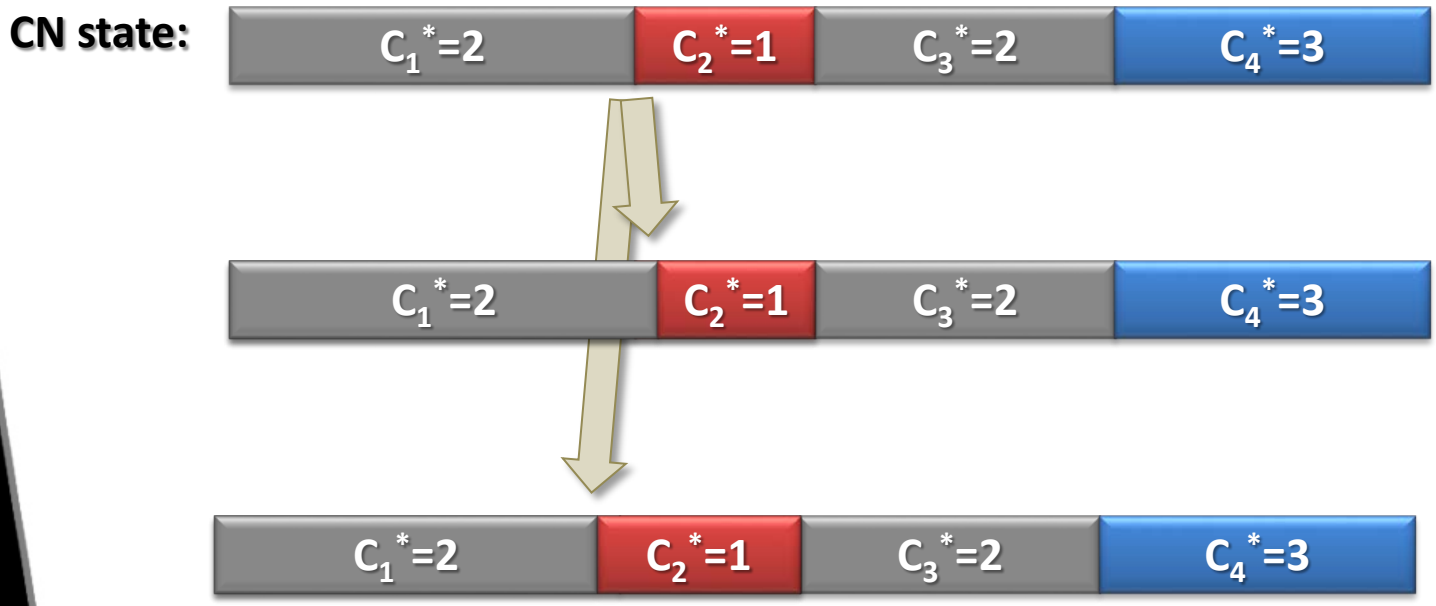
- Copy Number Variations
- Sequencing Read Depths
- **Bayesian Procedure**
  - •Model
  - •RJMCMC
- NTUH data

3. Trifid



- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

4. Boundary Change



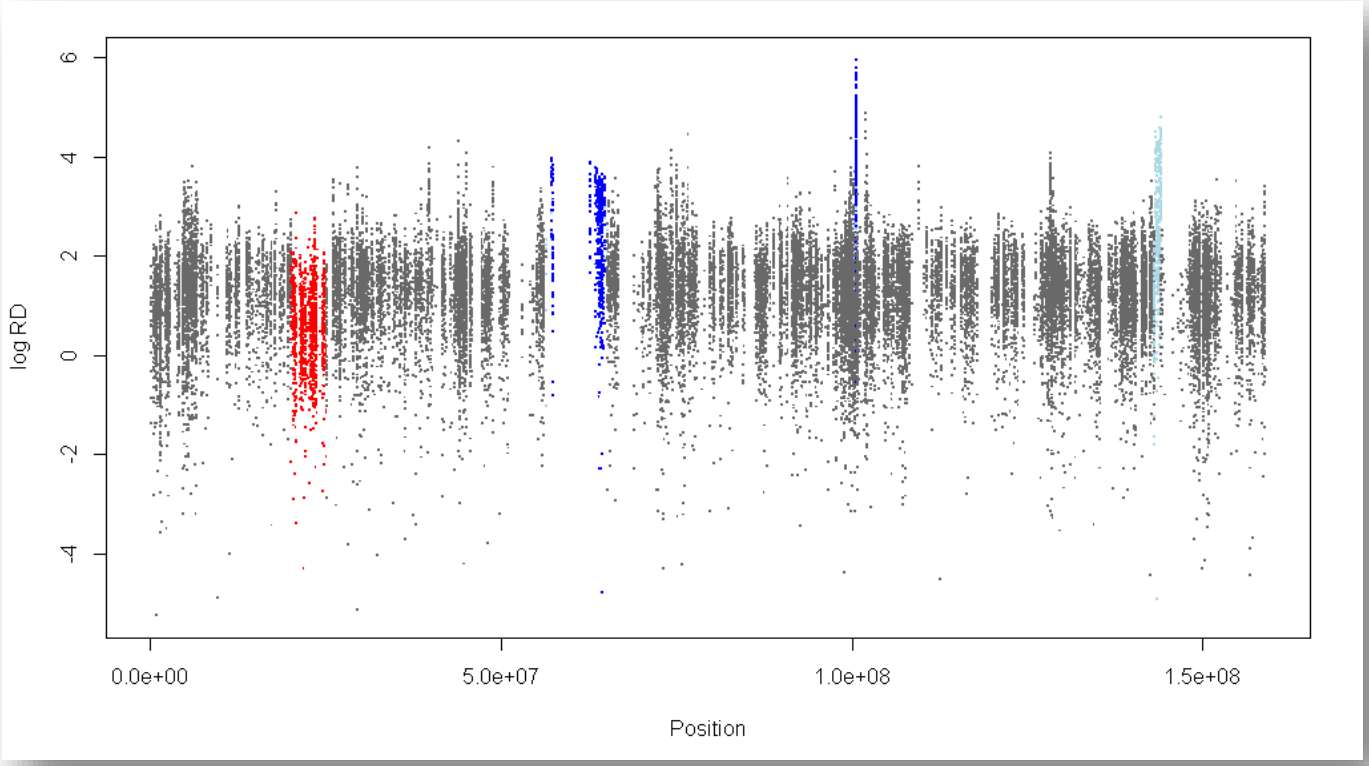
- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

NTUH Department of OB/GYN

- **Silver-Russell syndrome**: a growth disorder, have a small, triangular face with distinctive facial features
- Illumina/Solexa (NGS technology)
- Targeted exon region (protein coding regions)
- Chromosome 7
- 32387 windows

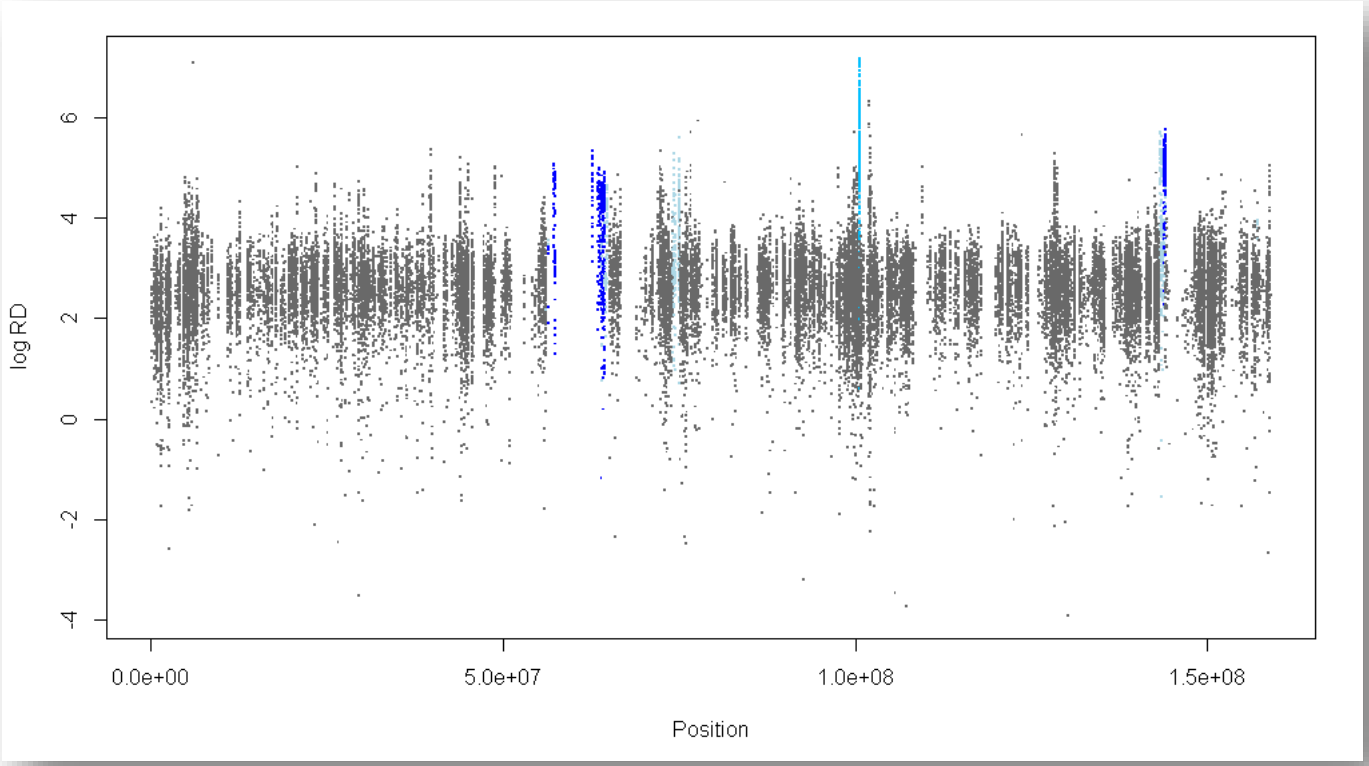
- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

Sample 1:



- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

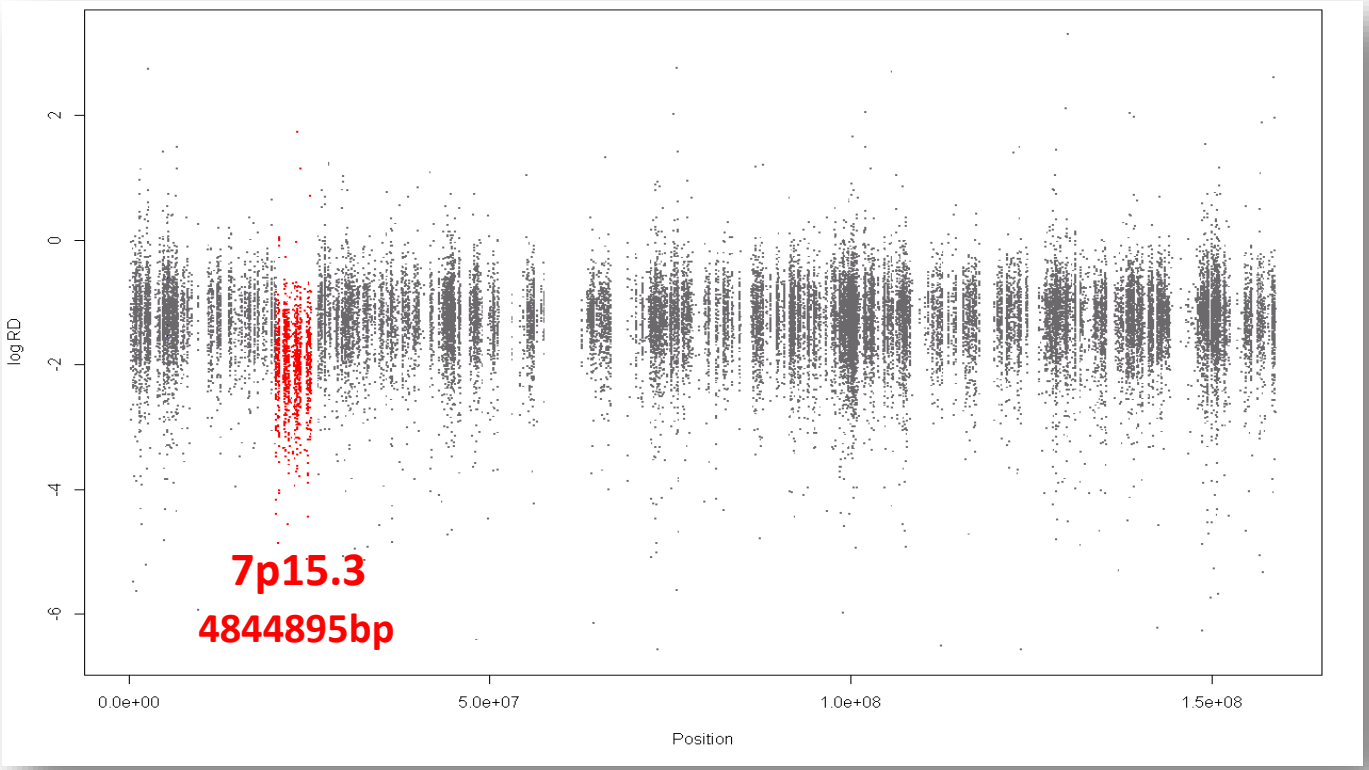
Sample 2:



● :CN1 ● :CN2 ● :CN3 ● :CN4 ● :CN5

- Copy Number Variations
- Sequencing Read Depths
- Bayesian Procedure
  - •Model
  - •RJMCMC
- NTUH data

Sample 1 vs. 2:



● :Deletion ● :Normal ● :Duplication