Estimating Rearrangement Evolution in Cancer with New Sequencing Technologies

Chris Greenman

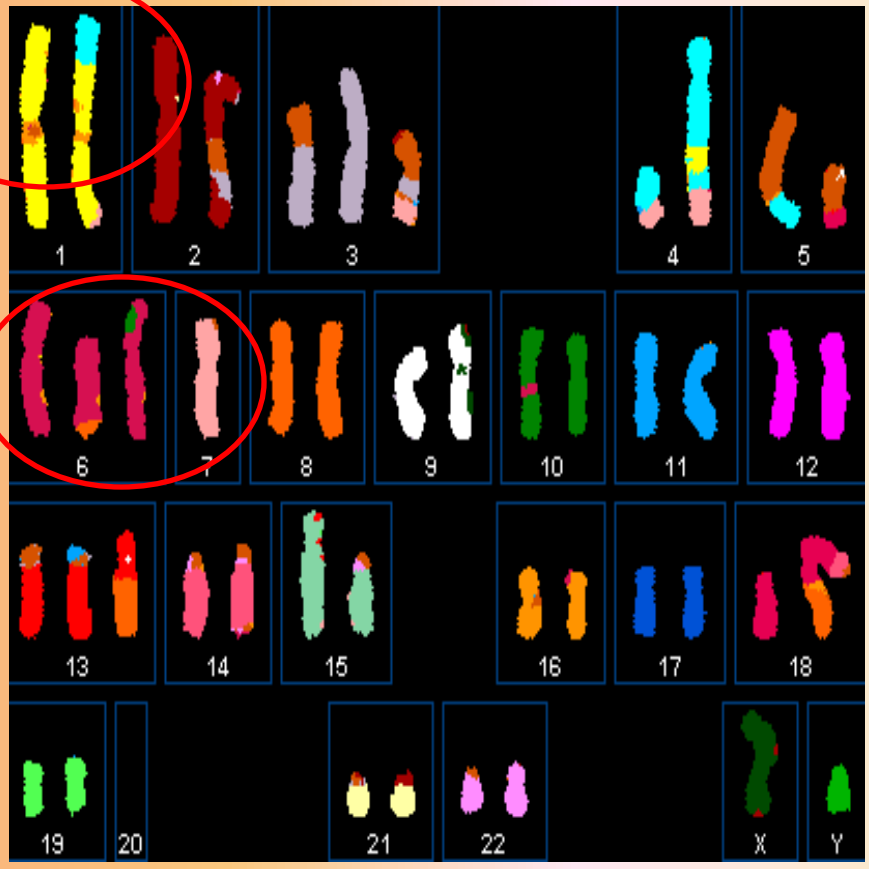
Cancer Genome Project

wellcome trust sanger institute

Cancer Genomes

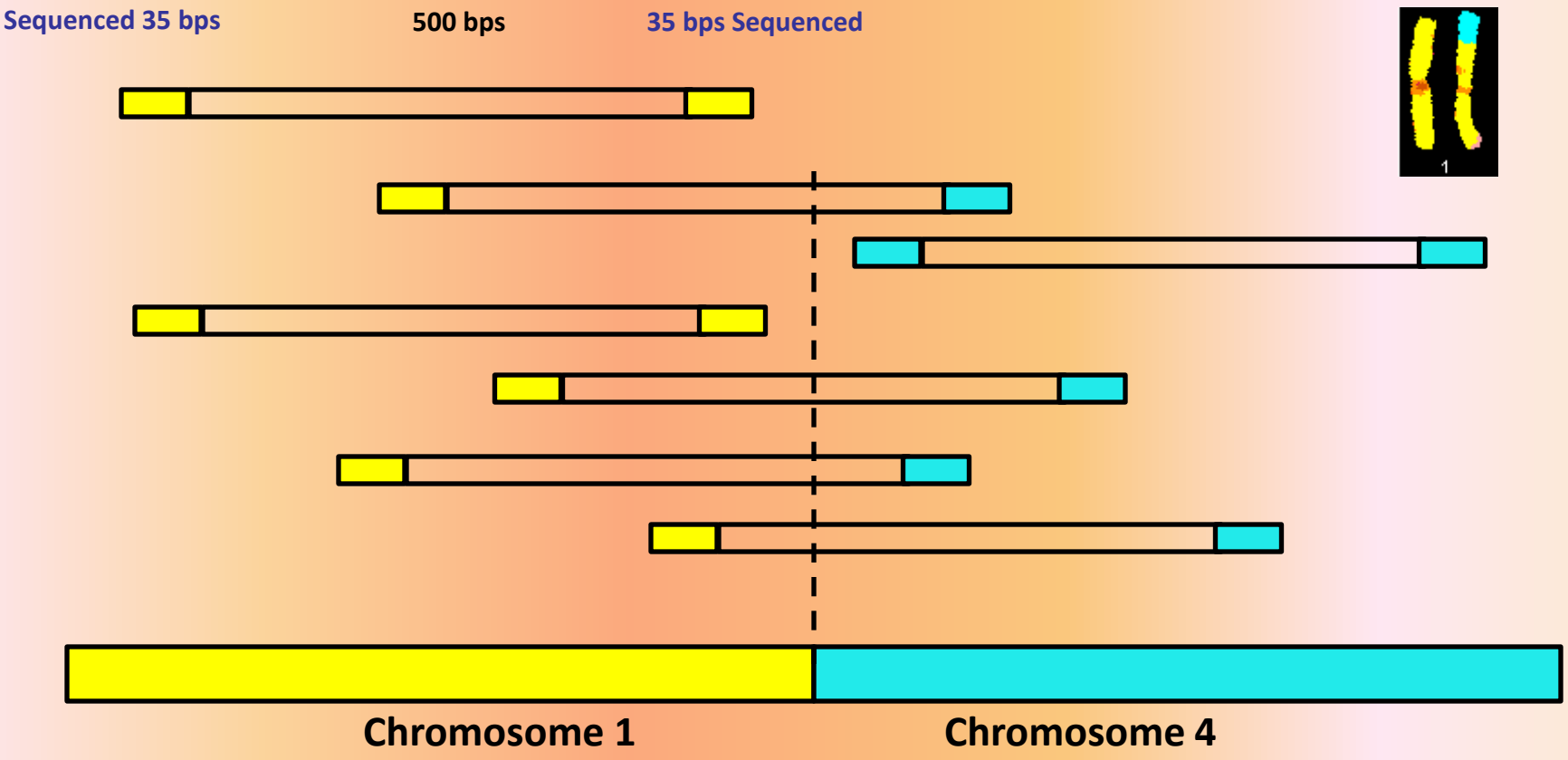
- Cancer – mutations give growth advantage
- Somatic Cancer Mutations
  - Rearrangements
  - Single Nucleotide Substitutions
  - Copy Number Variants
- What do cancer chromosomes look like?
- What rearrangements took place?
- What order did they occur?
- ‘When’ did they occur?

Lung Cancer NCI-H209 SKY



~15,000 somatic mutations

New Sequencing Data

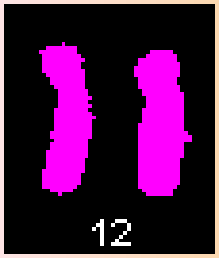


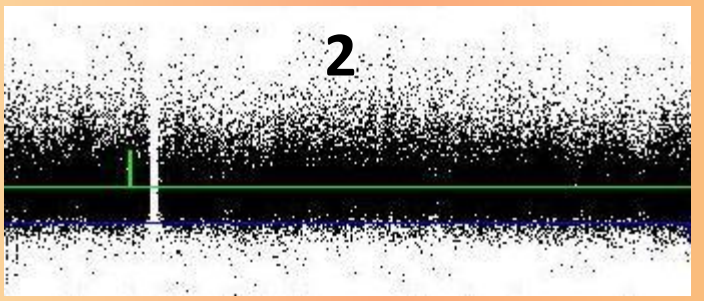
- Aberrant reads link rearranged segments
- Describe breakpoints to base pair level

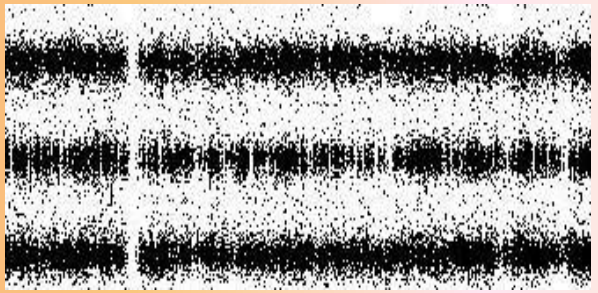
Integral Allelic Copy Number (PICNIC)

Copy Number

Allelic Ratio



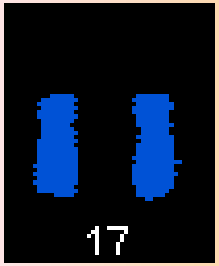


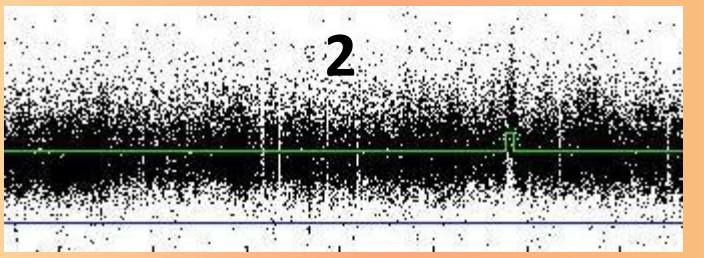


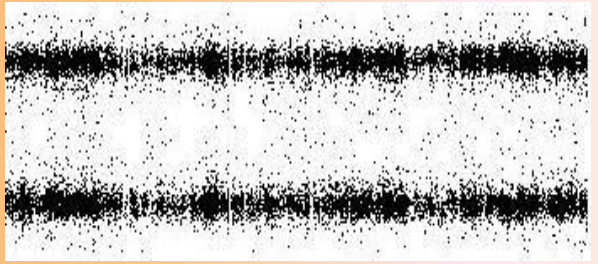
AA

AB

BB

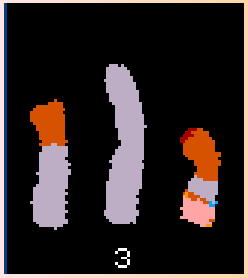


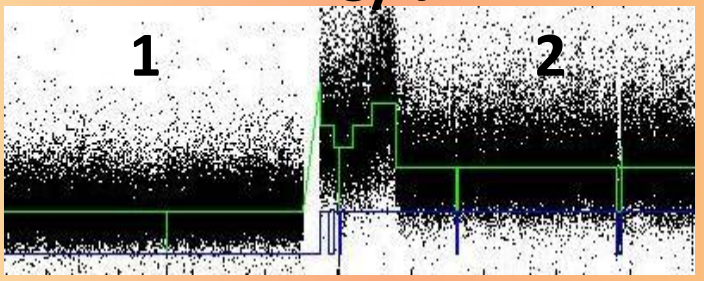


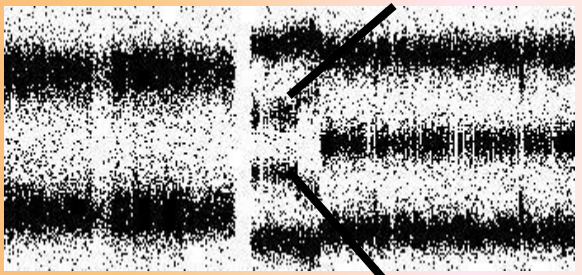


AA

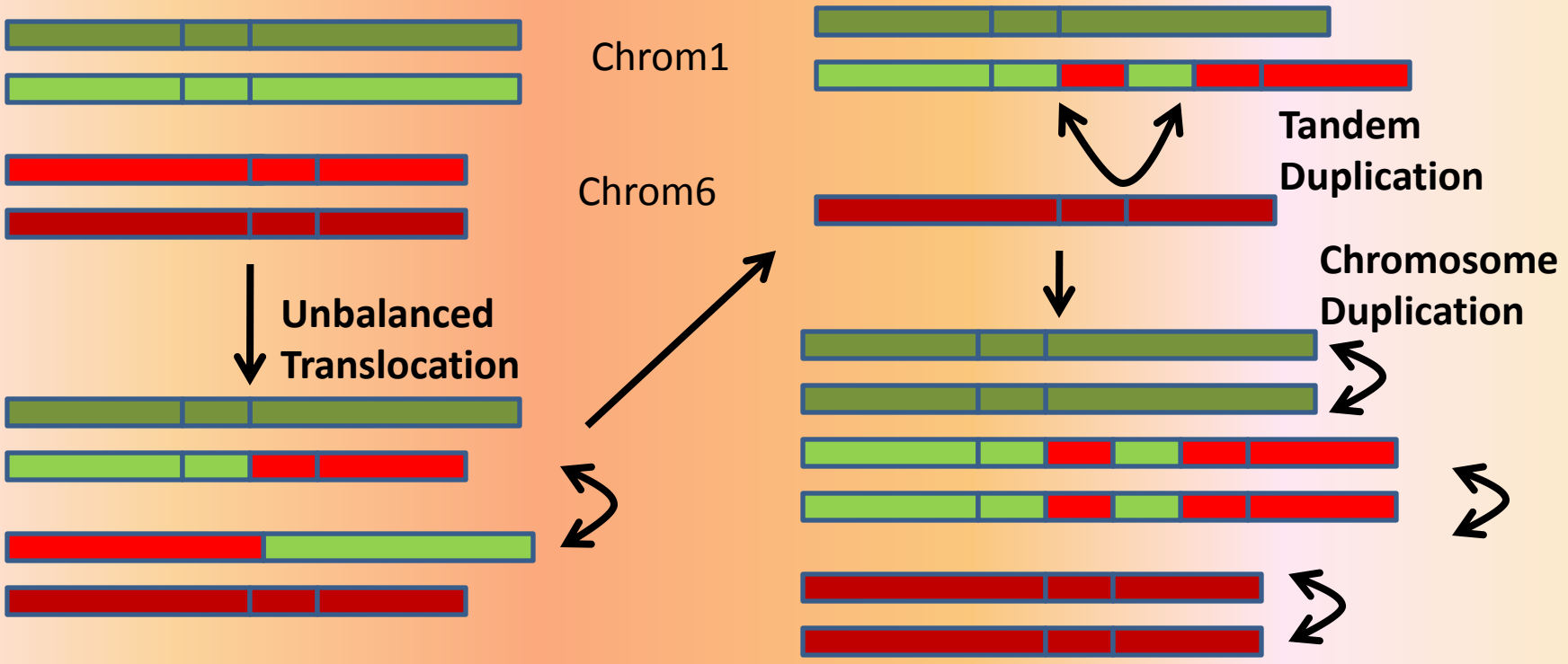
BB







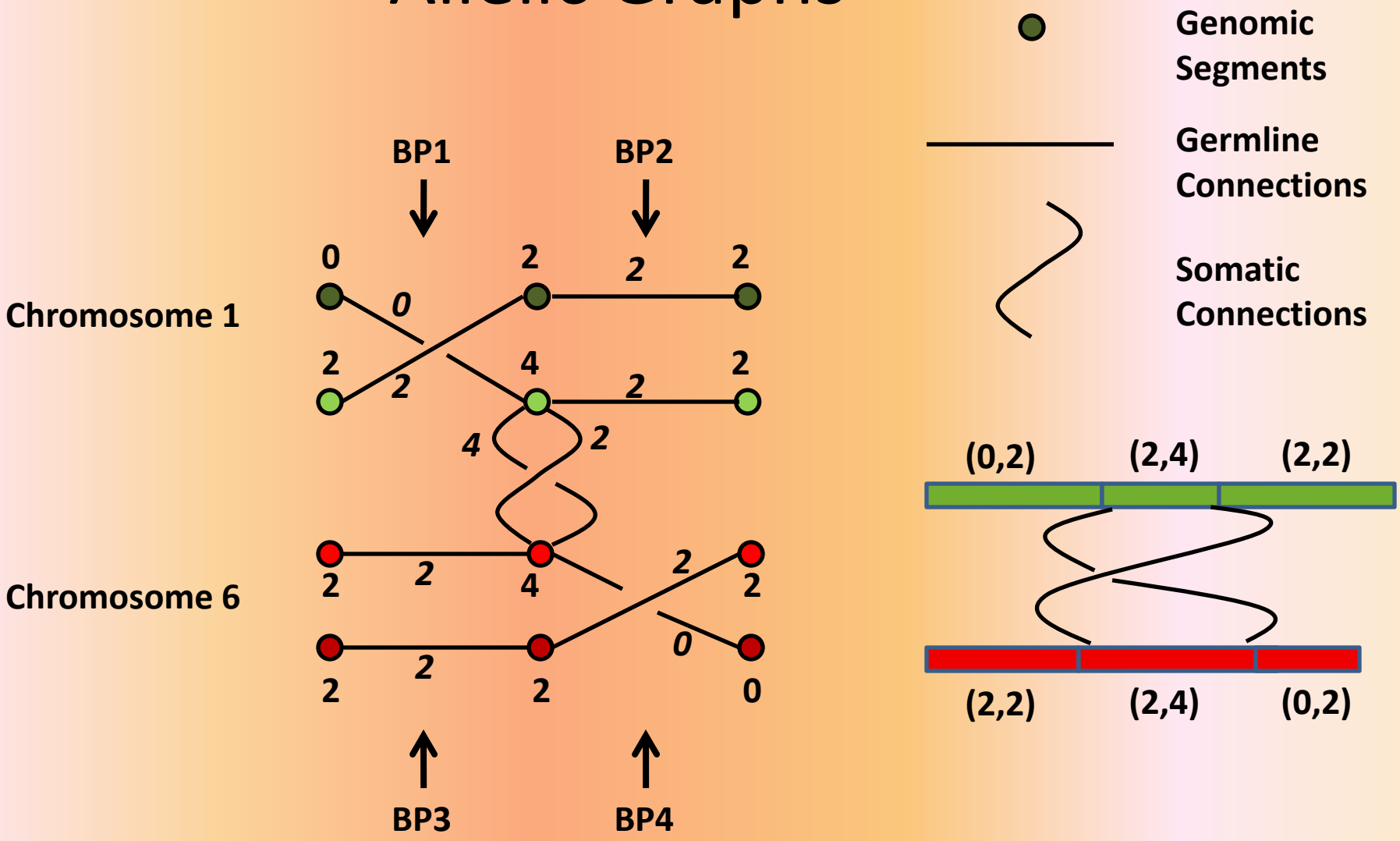
A Real Example



- They occur uniquely in one parental chromosome
- Copy number conserved in other parent

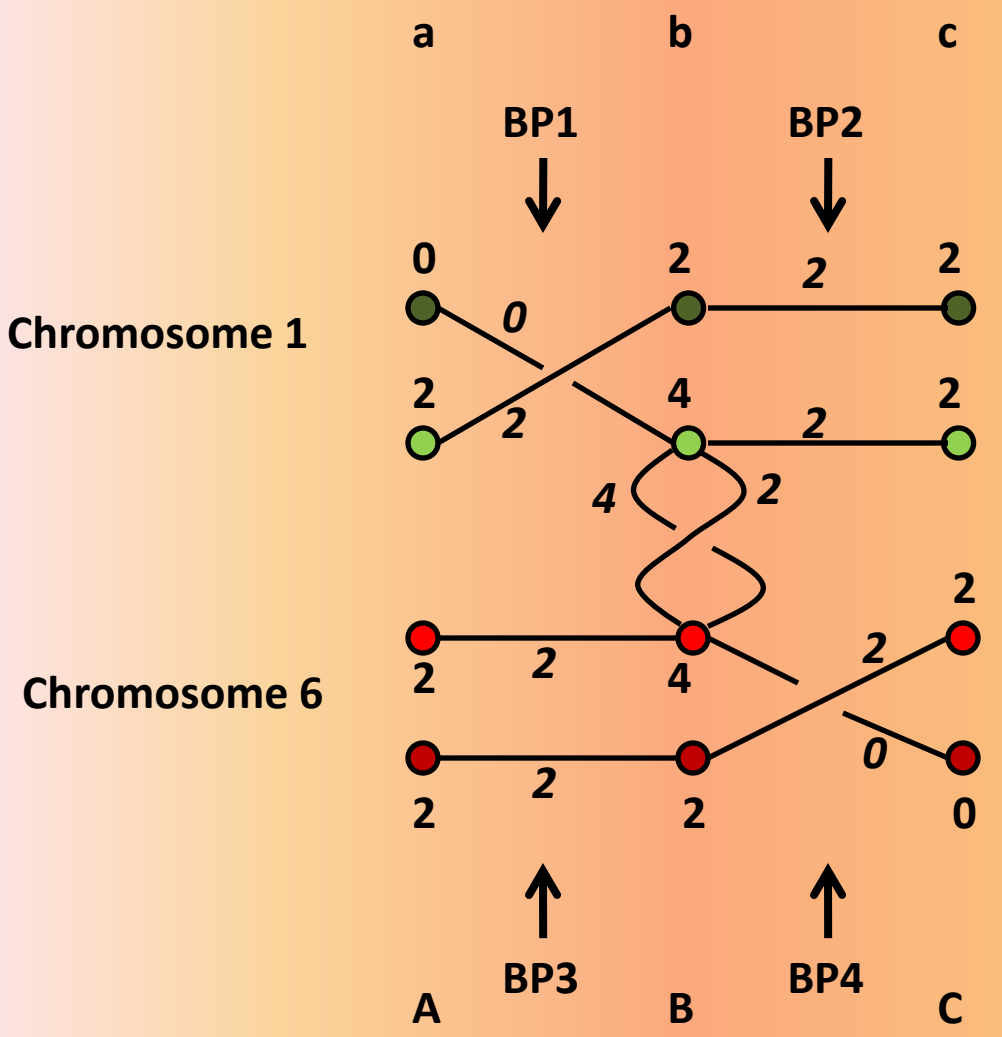
Segment	1	2	3
	2	2	2
	2	4	0
	0	4	2
	2	2	2

Allelic Graphs



- Find Edge counts = integer programming problem

Chromosomes are Graph Walks



Case I

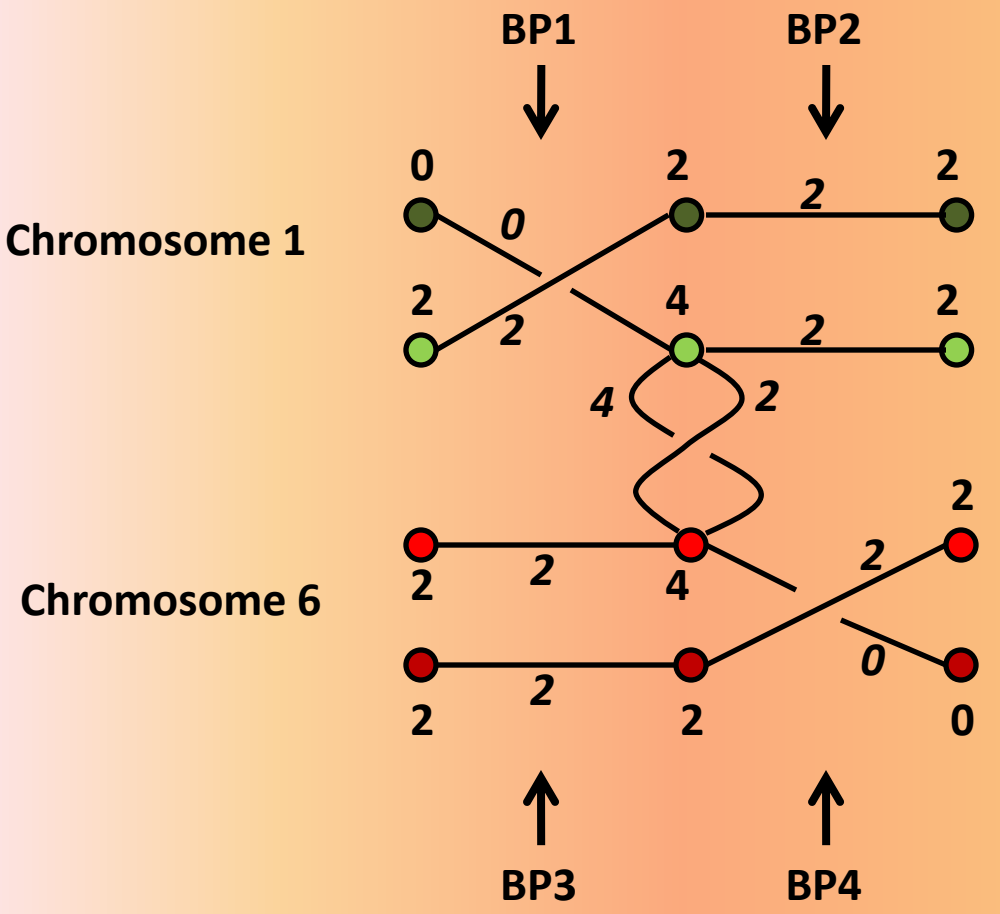
- 2 x abc
- ABbc, ABbBbBbc
- 2 x ABC

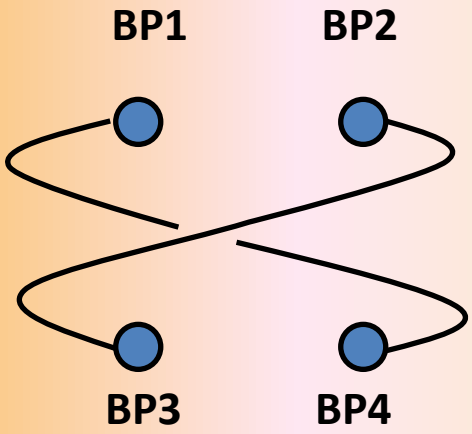
Case II

- 2 x abc
- 2 x ABbBbc
- 2 x ABC

- Paths are chromosomes
- Which rearrangements?
- Which order?

Somatic Graphs





$$G_{14}^{-+} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = G_{41}^{+-trans}$$

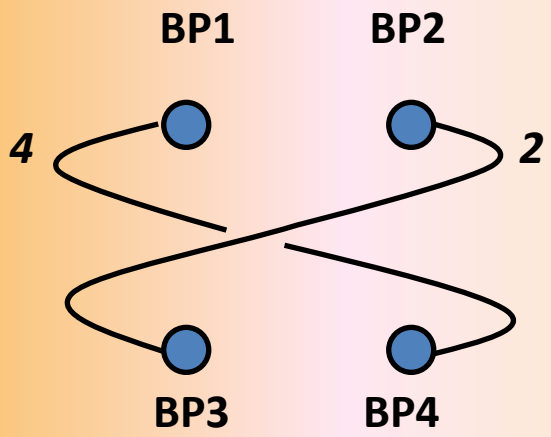
$$G_{14}^{++} = G_{41}^{++trans} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$G_{14}^{--} = G_{41}^{--trans} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Rearrangement Dictionary

Rearrangement Classifications

- Somatic graph component = Rearrangement
- Rearrangement Types
  - Inversion (INV)
  - Tandem Duplication (TD)
  - Inverted Duplication (ID)
  - Breakage Fusion Bridge (BFB)
  - Chromosomal Arm Loss (CAL)
  - Translocation (T)
  - Unbalanced Translocation (UT)
  - Insertion (INS)
  - Deletion (DEL)
- G matrices are rearrangement 'signatures'
- Both events are TD, UT or DELs
- Edge counts =>
  - BP1,4 1st
  - BP2,3 2nd,
  - chromosomal duplication 3rd

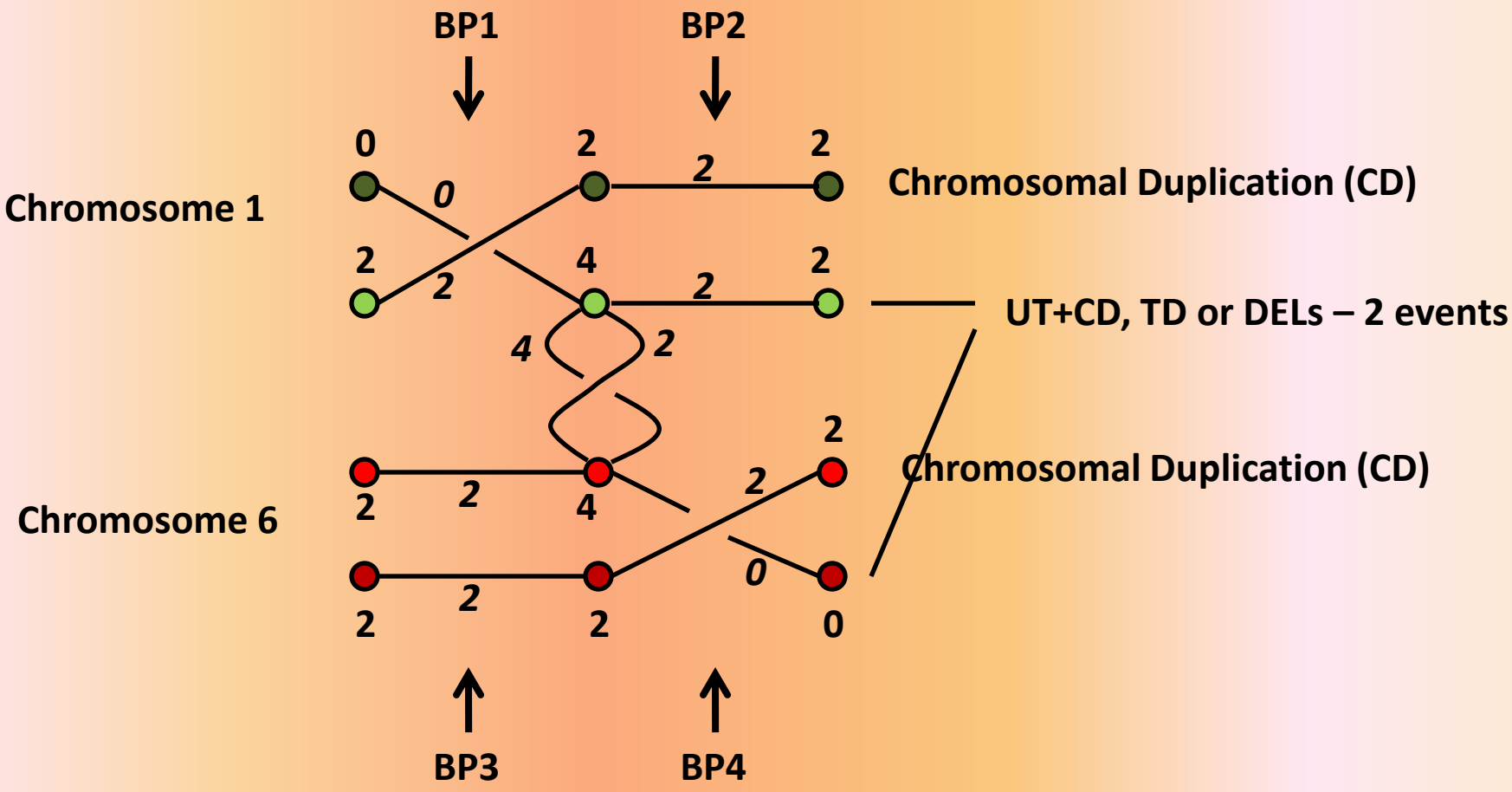


$$G_{14}^{-+} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = G_{41}^{+-trans}$$

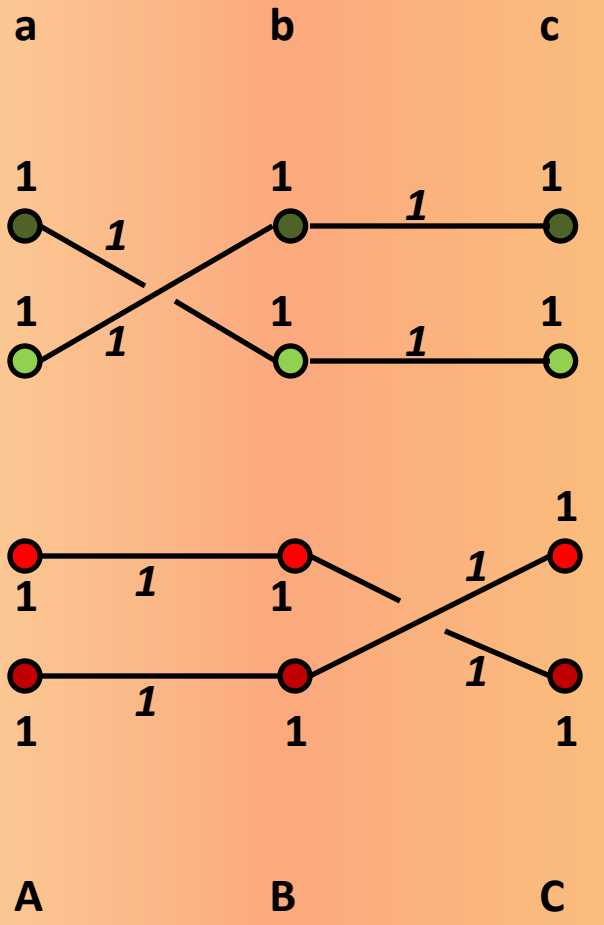
$$G_{14}^{++} = G_{41}^{++trans} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$G_{14}^{--} = G_{41}^{--trans} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Simulations



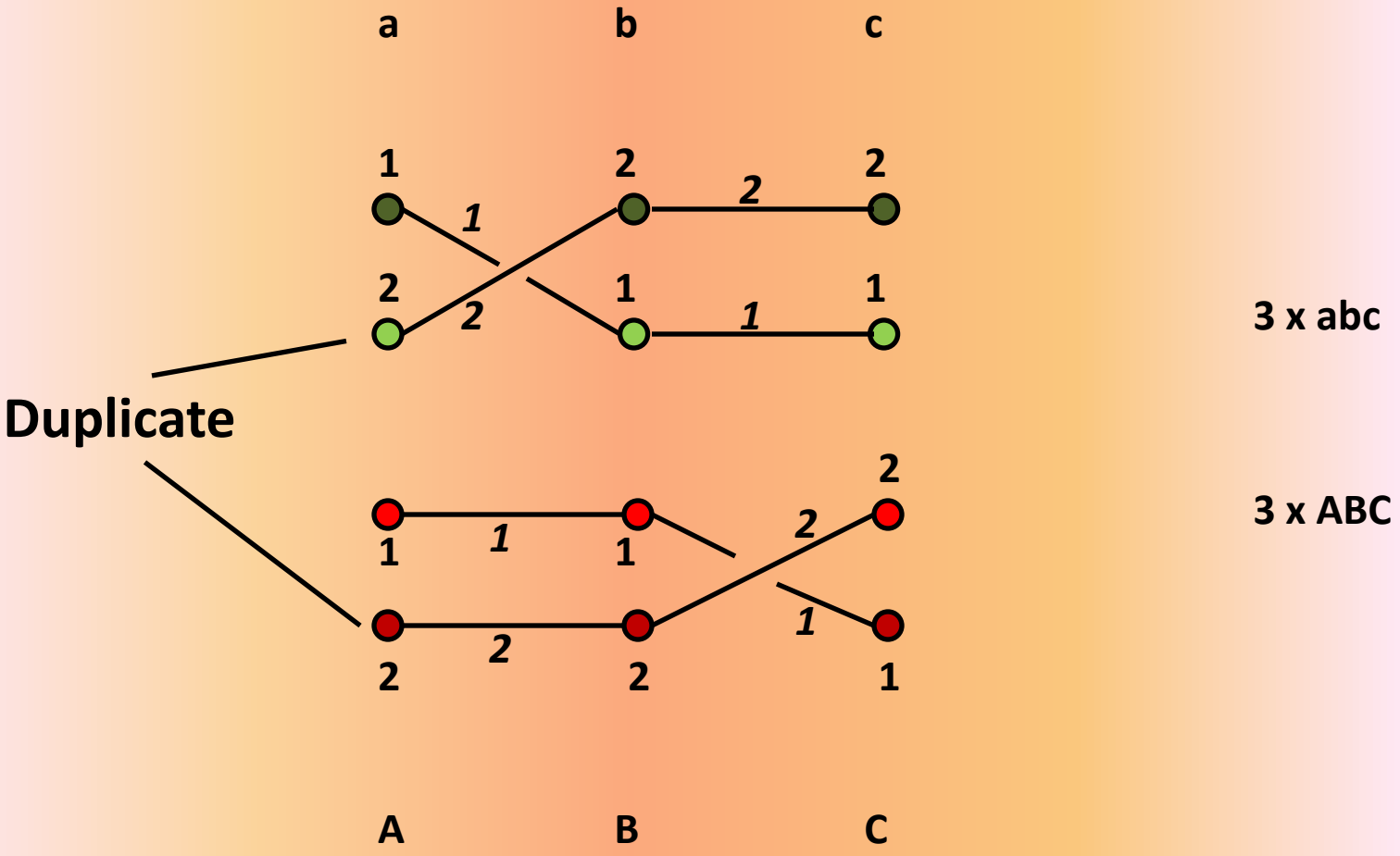
Wild Type



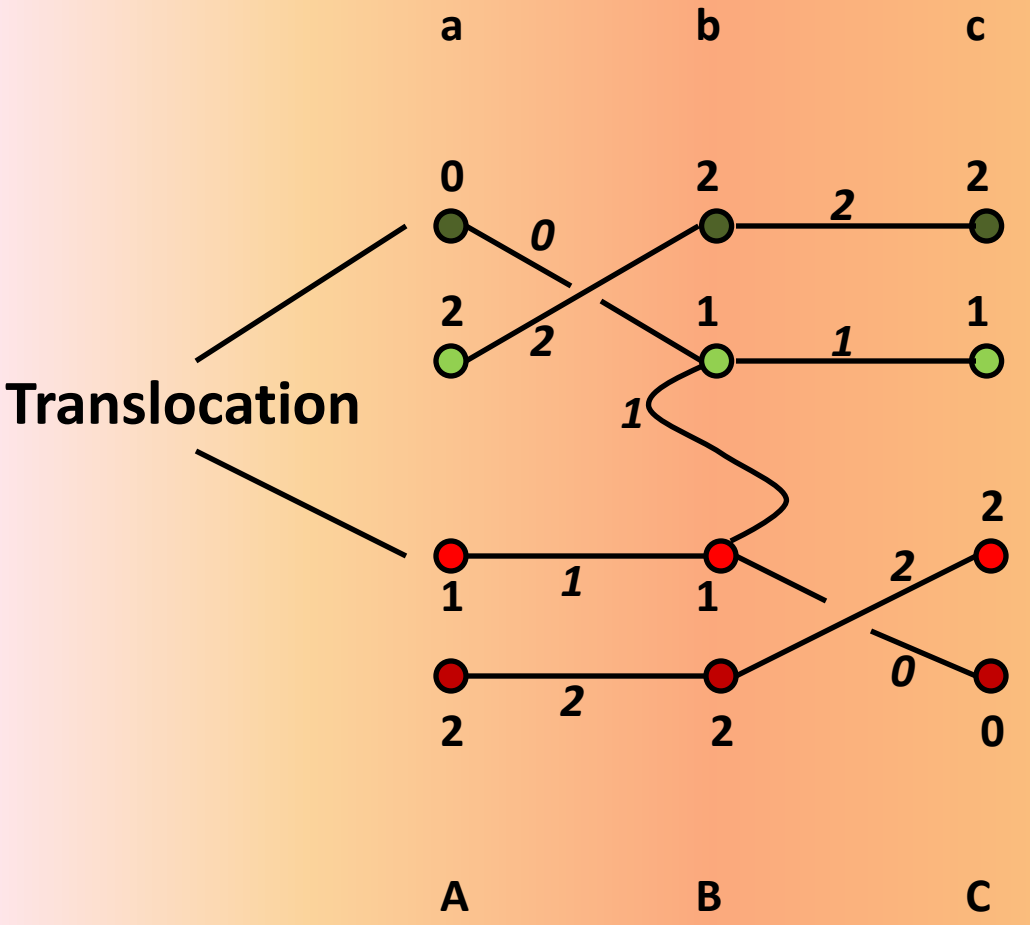
2 x abc

2 x ABC

Chromosomal Duplications



Unbalanced Translocation

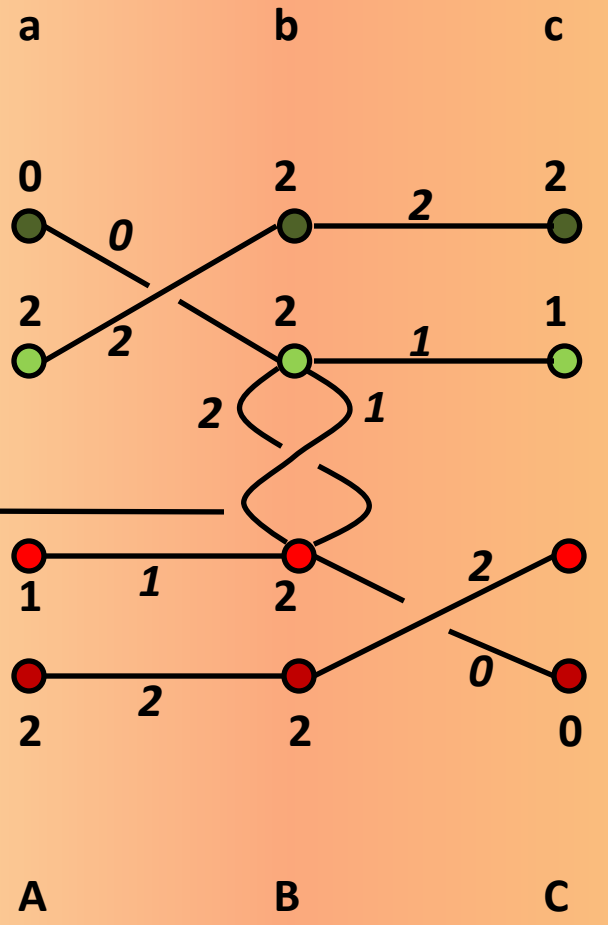


2 x abc

ABbc

2 x ABC

Tandem Duplication



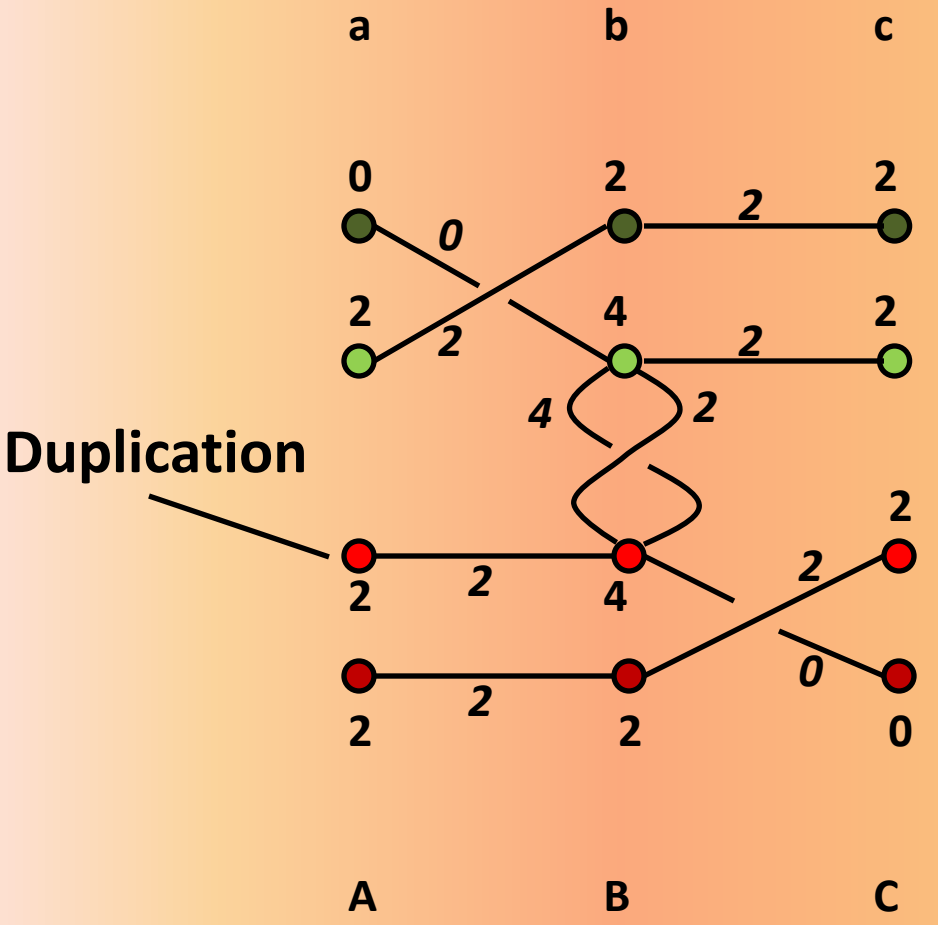
Duplication

2 x abc

ABbBbc

2 x ABC

Chromosomal Duplication



Duplication

Case II

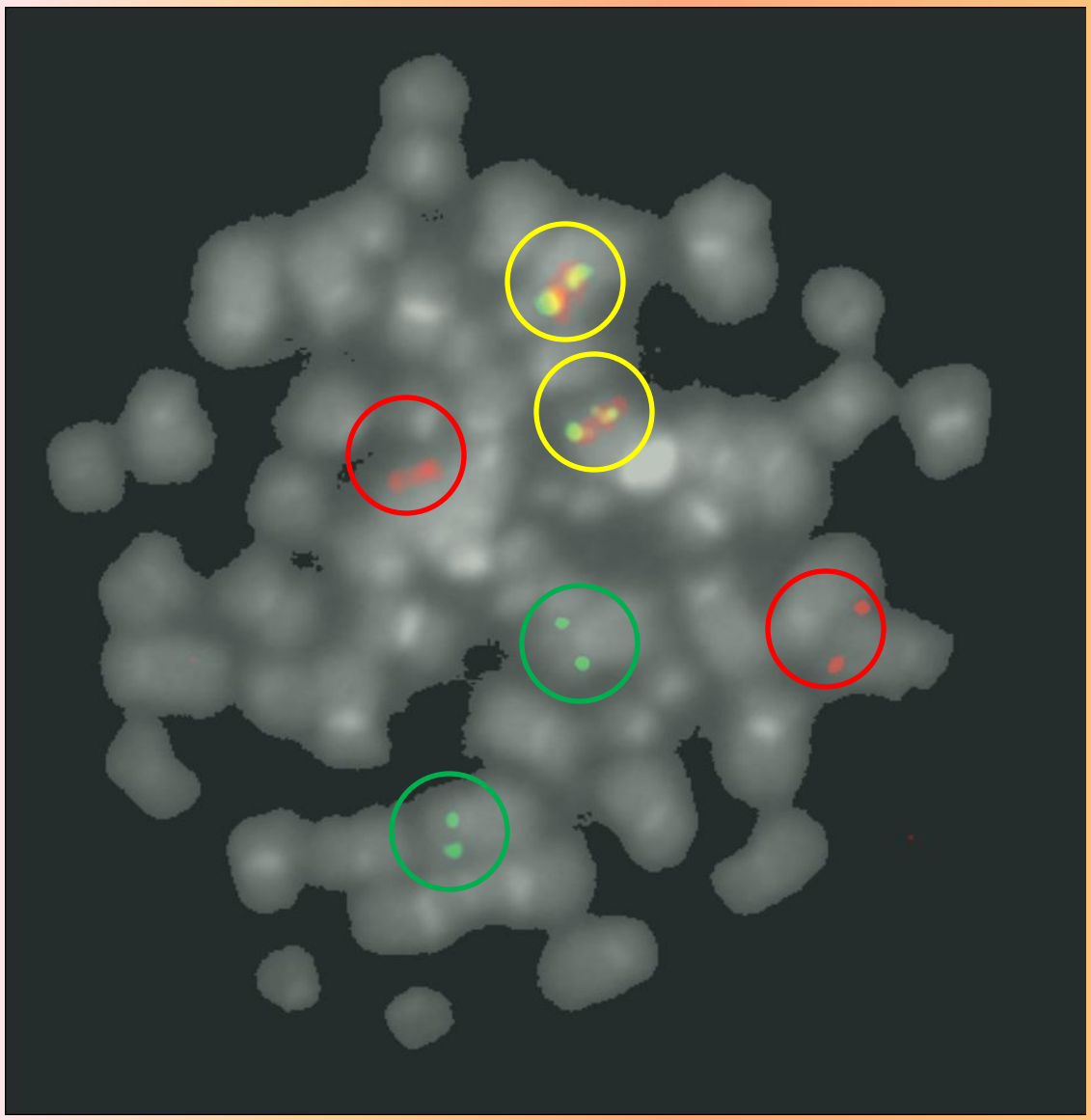
2 x abc

2 x ABbBbc

2 x ABC

Unbalanced Translocation < Tandem Duplication < Chromosomal Duplication

FISH



b

B

Case II

2 x abc

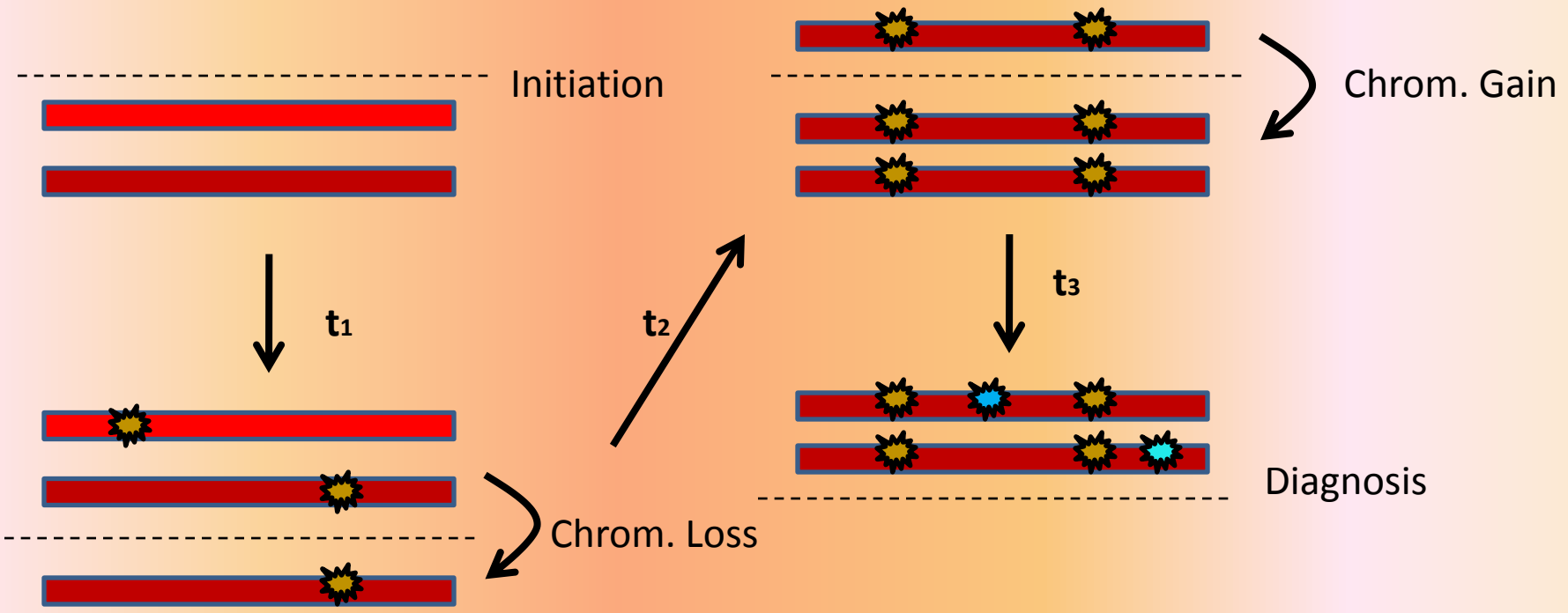
2 x ABbBbc

2 x ABC

Summary

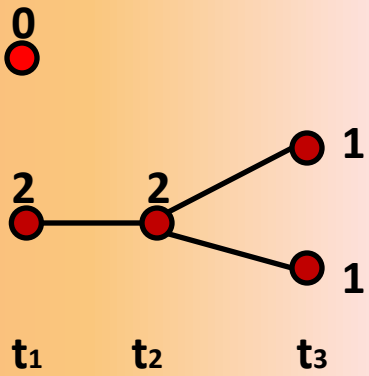
- Integer allelic copy number + aberrant paired reads =:
  - possible rearrangements
  - possible orders
  - possible chromosomes
- ‘When’ did they occur?

Single Nucleotide Mutations



Het ~ 1 ~

Hom ~ 2 ~

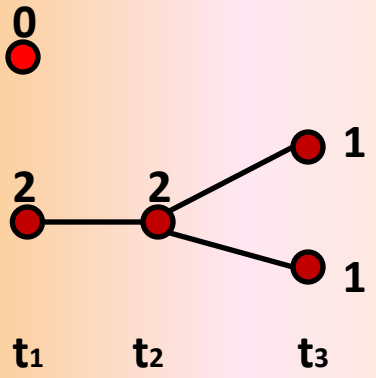


Timing Gain Events

Chromosome	**13 (92Mb)**	**17 (75Mb)**
Heterozygous (n_1)	**192**	**153**
Homozygous (n_2)	**331**	**127**
Chrom. Loss - t_1	**Unknown**	**Unknown**
Chrom. Gain - t_2	**77.6% (ci 74.3% - 80.6%)**	**62.4% (ci 57.0.3% - 67.8%)**
Diagnosis - t_3	**100%**	**100%**

$$n_1 \sim Poiss(L.2t_3)$$

$$n_2 \sim Poiss(L.(t_1 + t_2))$$



Timing Driver Mutations

Chromosome	**13 (110Mb)**	**17 (75Mb)**
Heterozygous	**192**	**153**
Homozygous	**331**	**127**
Chrom. Loss - t_1	**Unknown**	**Unknown**
Chrom. Gain - t_2	**77.6% (ci 74.3% - 80.6%)**	**62.4% (ci 57.0.3% - 67.8%)**
Diagnosis - t_3	**100%**	**100%**

P53

17

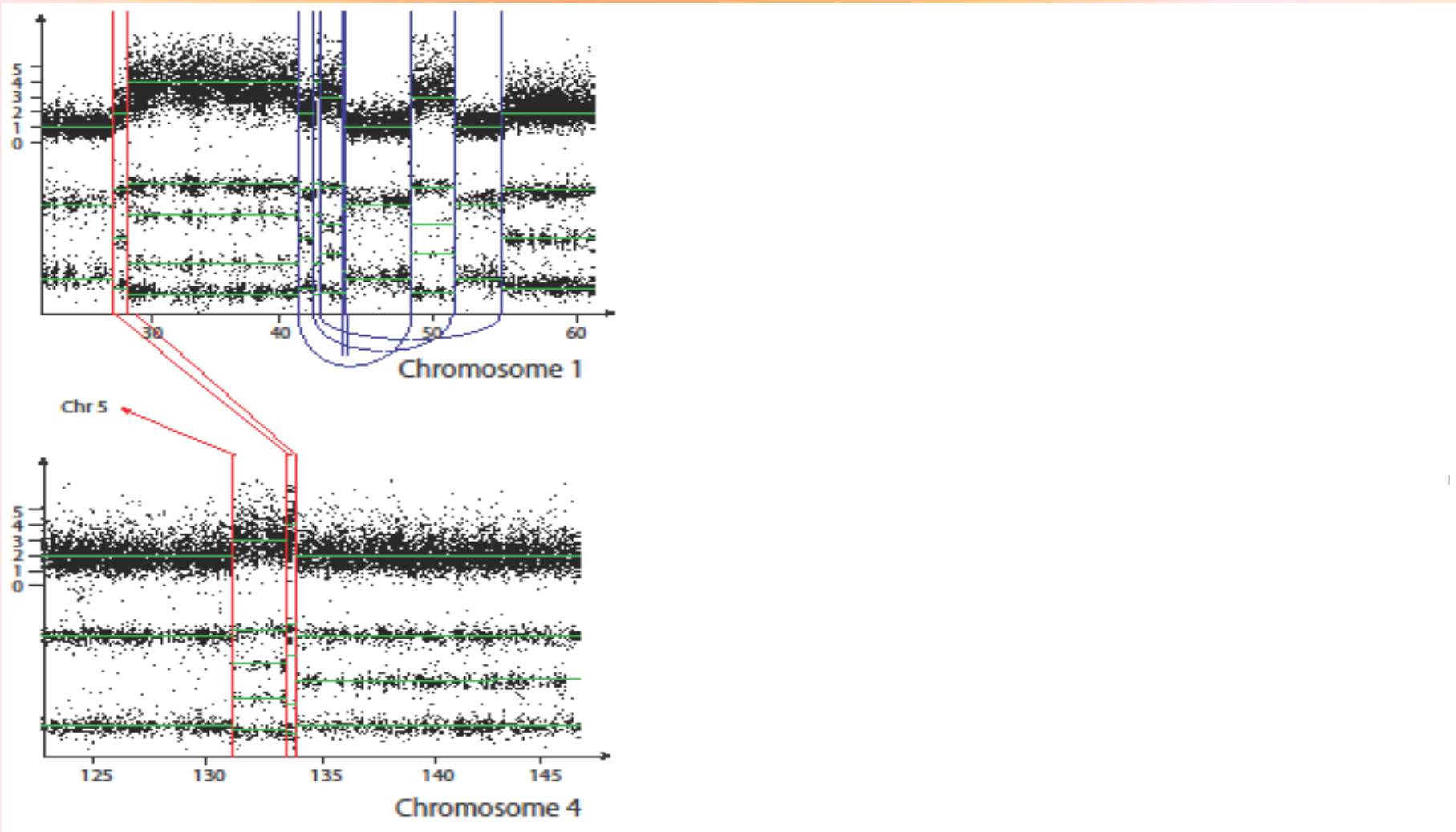
Mut < 62.4%

RB1

13

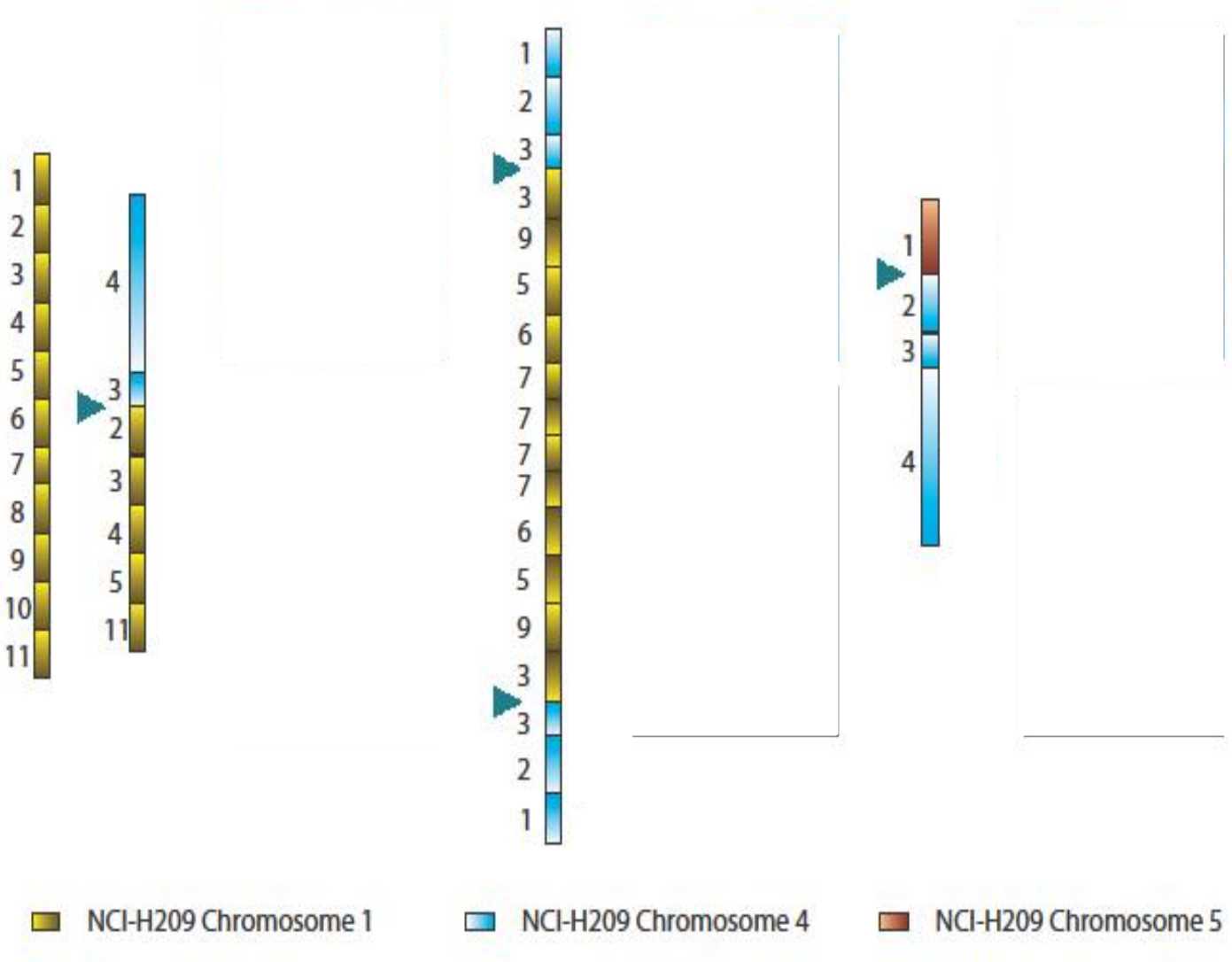
Mut < 77.6%

Complex Example NCI-H209

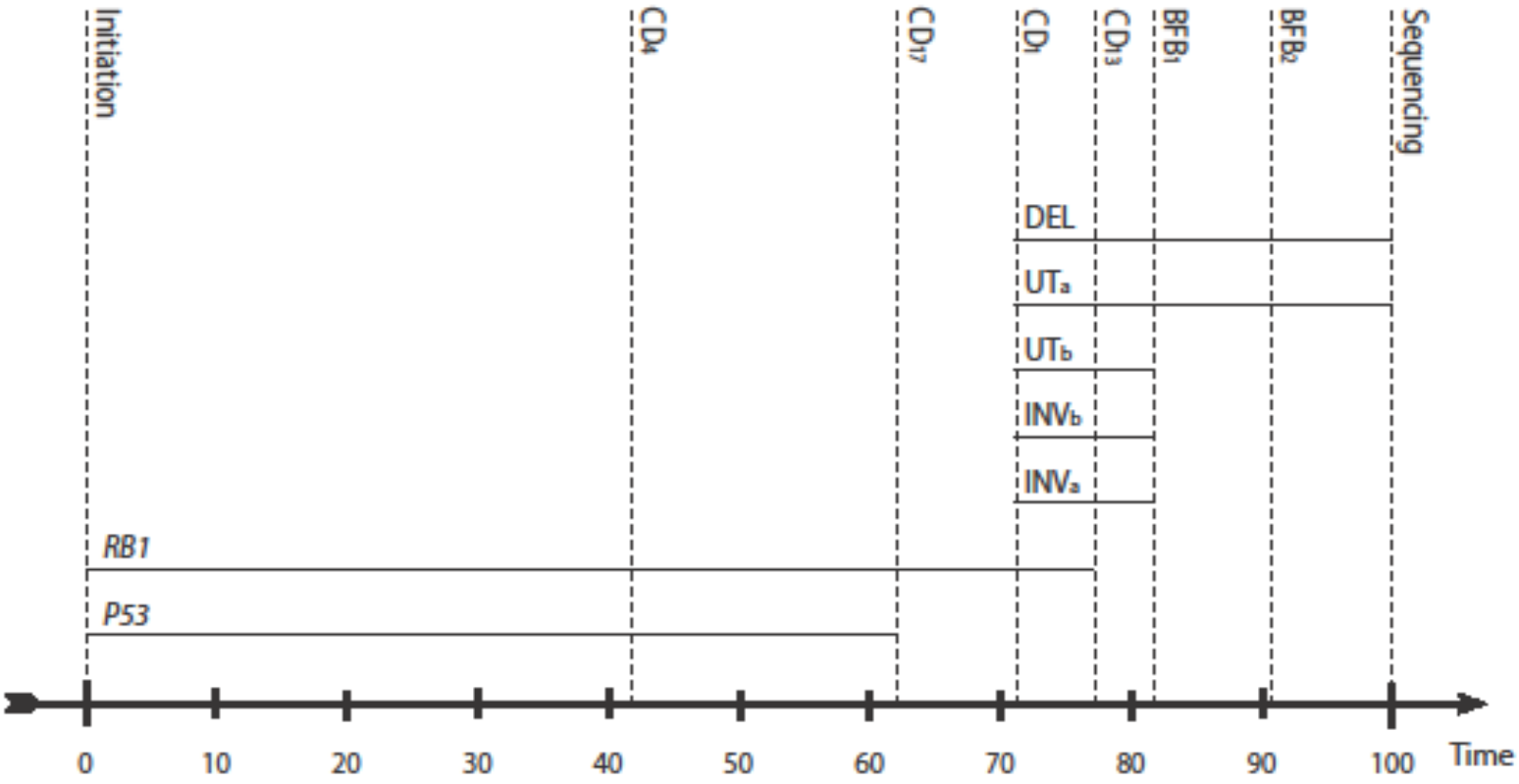


$\{CD_1, CD_4\}<\{UT_a, UT_b\}, CD_1<\{DEL, INV_a, INV_b\}, \{UT_b, INV_a, INV_b\}<BFB_b<BFB_a$

NCI-H209 FISH



Complex Example NCI-H209 – Time Line



Conclusions

- Inputs:
  - Aberrant paired read data
  - Allelic Integral Copy Number
  - Single Nucleotide Somatic Mutation Multiplicity Counts
- Outputs:
  - Chromosomal Makeup
  - Rearrangement Types
  - Possible Orders
  - Timings for Copy Number Increases
  - Inference upon timing of cancer causing mutations
- GRAFT:
  - Genomic Rearrangements Assembly For Tumours
  - www.sanger.ac.uk/genetics/cgp/software/GRAFT

Acknowledgements

Erin Pleasance
Peter Campbell
Mike Statton
Andy Futreal
Dave Beare
Dave Jones
Paul Edwards
Scott Newman
Fengtang Yang
Nigel Carter