Computational Genetics
Winter 2013
Lecture 10

Eleazar Eskin

University of California, Los Angeles

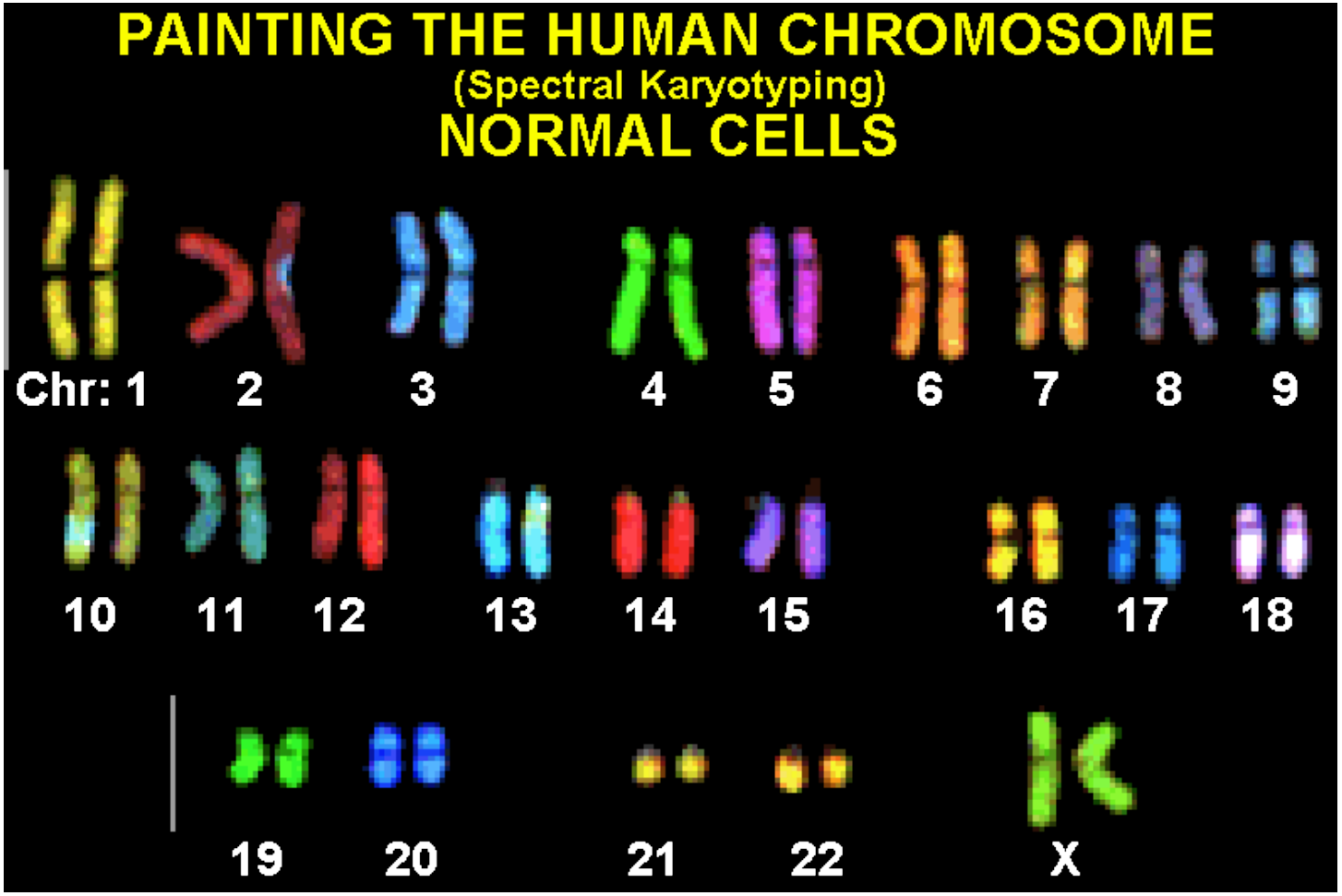
Pair End Sequencing

Lecture 10.

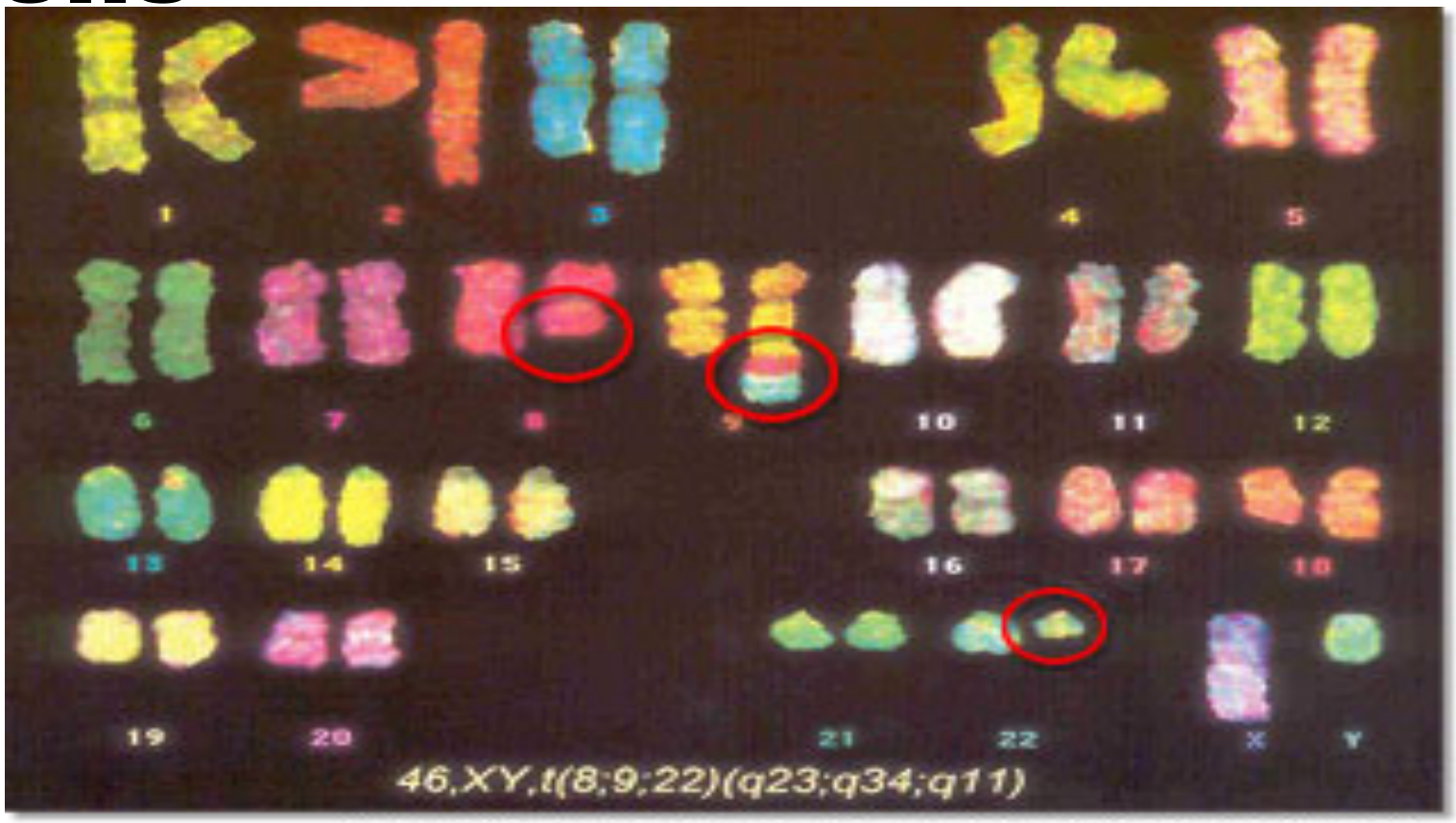
February 20th, 2013

(Slides from Ben Raphael)

Chromosome Painting: Normal Cells



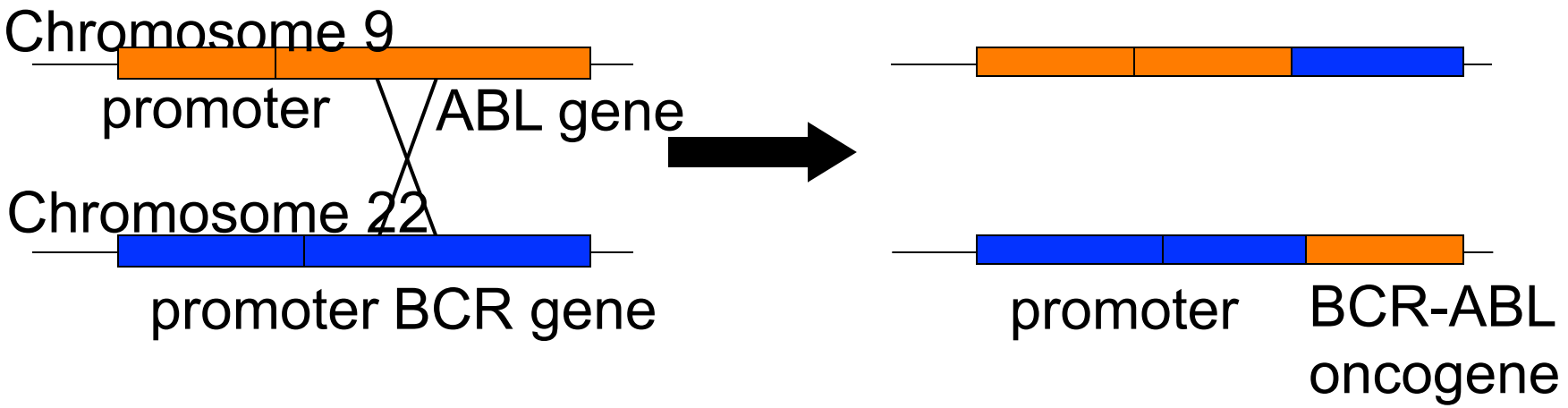
Chromosome Painting: Tumor Cells



Note: This karyotype was prepared using a FISH technique known as "chromosome painting". As well as having a translocation from chromosome 22, chromosome 9 also has translocated material from chromosome 8.

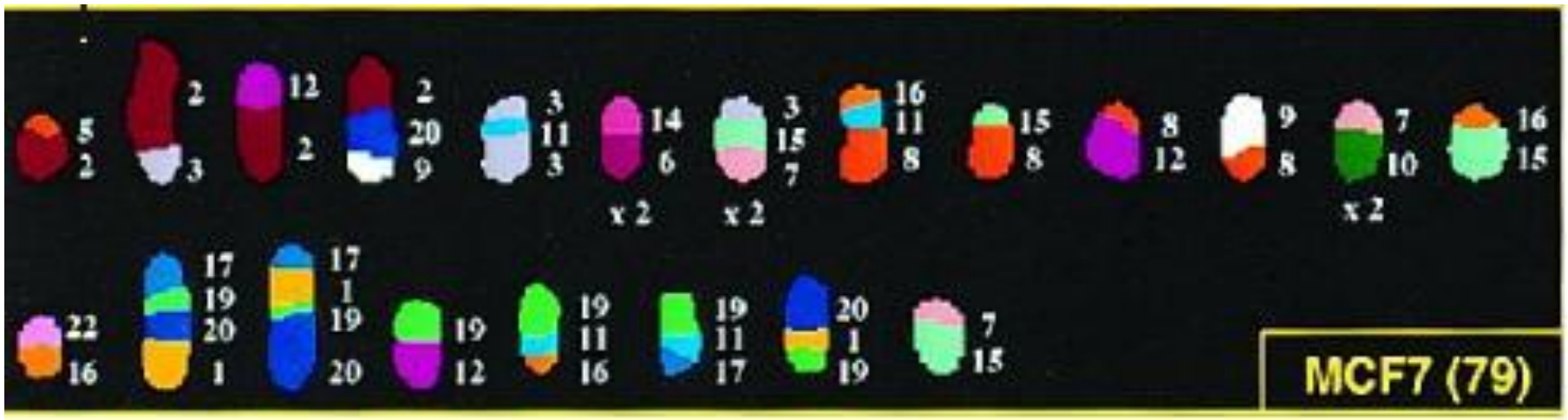
Rearrangements in Tumors

- Change gene structure and regulatory "wiring" of the genome.
- Create "bad" novel *fusion* genes and break "good" old genes.
- Example: translocation in leukemia.



- Gleevec™ (Novartis 2001) targets BCR-ABL oncogene.

Complex Tumor Genomes

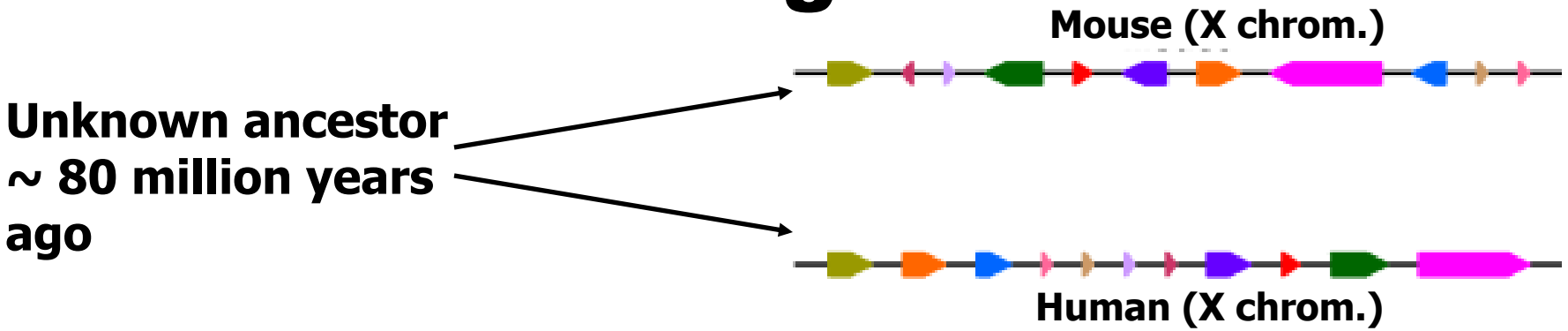


1) What are detailed architectures of tumor genomes?

2) What rearrangements/duplications produce these architectures and what is the order of these events?

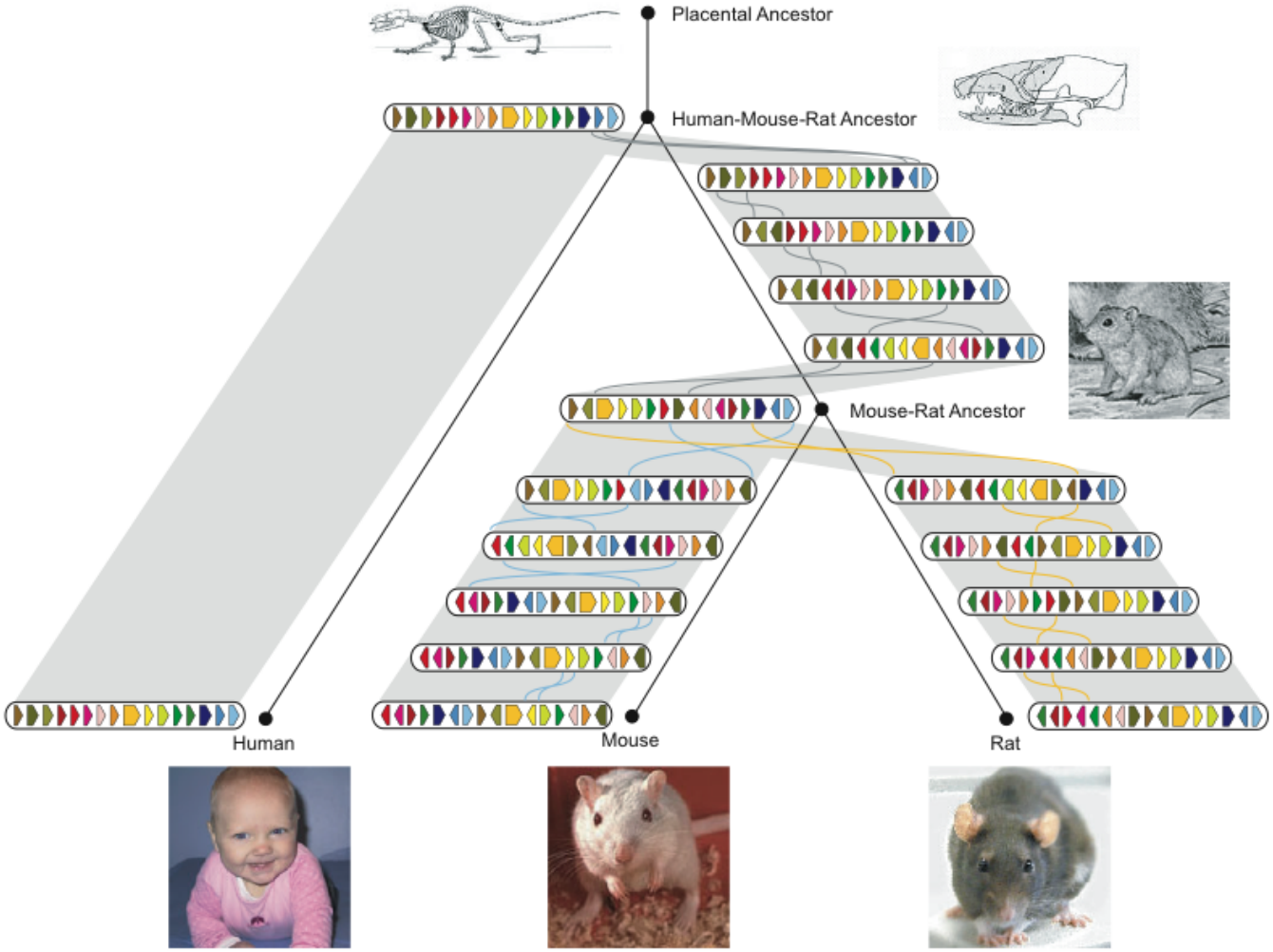
3) What are the novel fusion genes and old “broken” genes?

Genome rearrangements

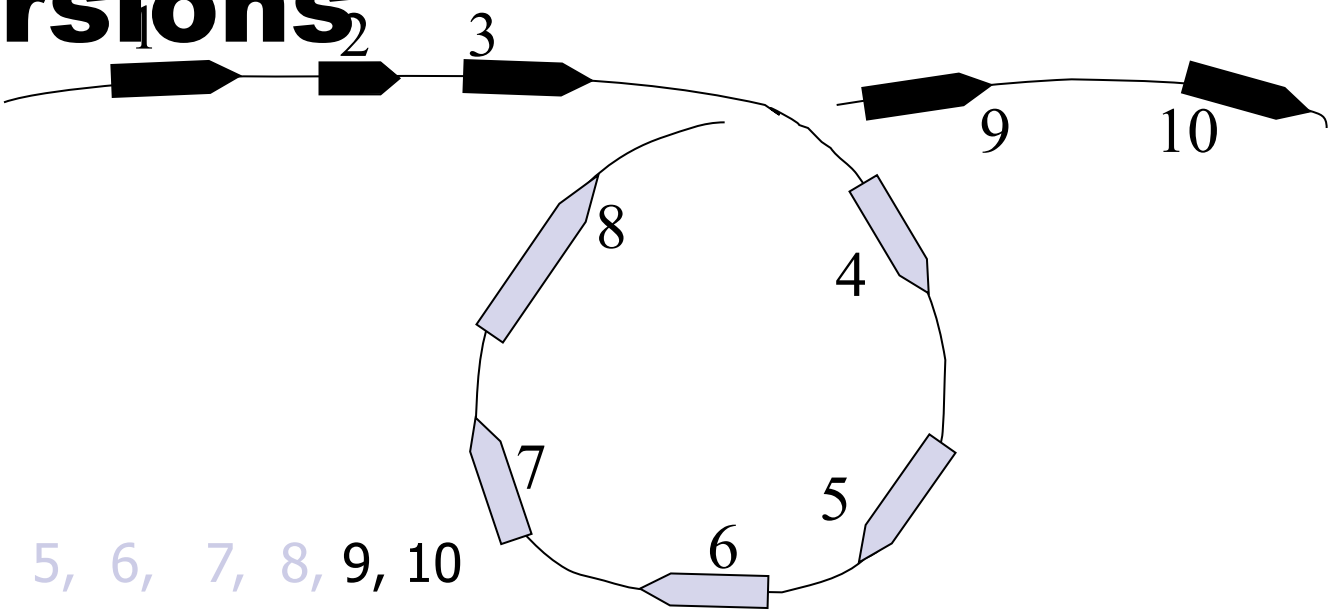


- What are the the “architectural blocks” forming the existing genomes and how to find them?
- What is the architecture of the ancestral genome?
- What is the evolutionary scenario for transforming one genome into the other?

History of Chromosome X



Inversions

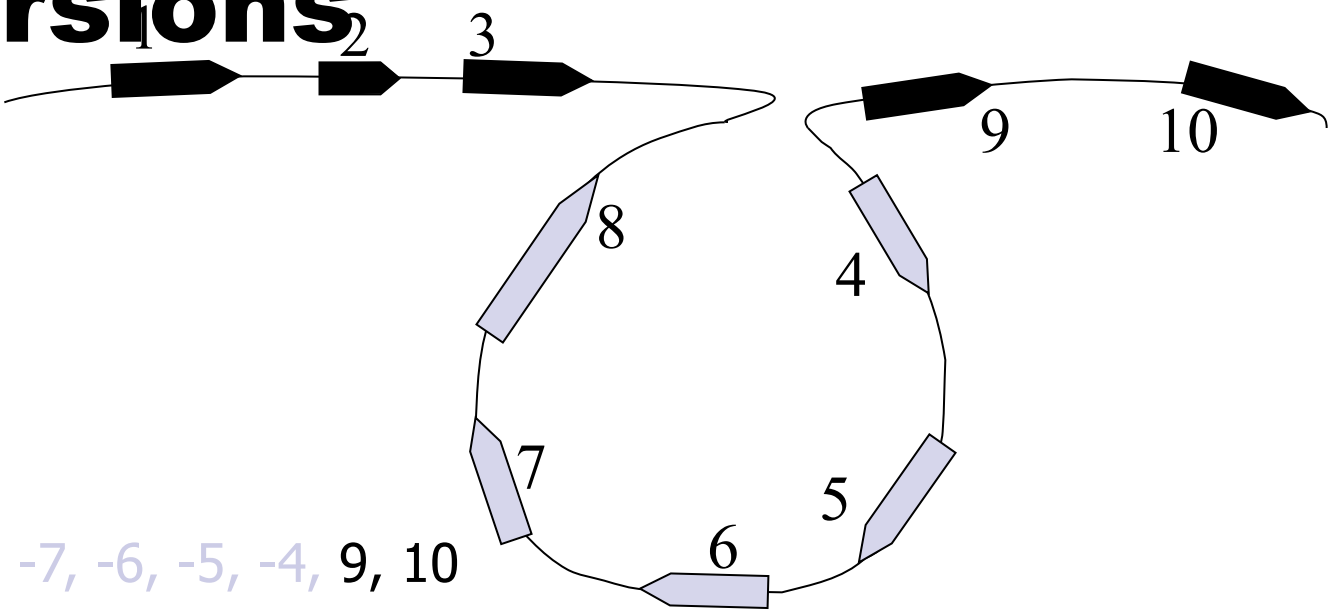


1, 2, 3, 4, 5, 6, 7, 8, 9, 10

- Blocks represent conserved genes.
- In the course of evolution or in a clinical context, blocks 1,2,…,10 could be misread as:

1, 2, 3, -8, -7, -6, -5, -4, 9, 10.

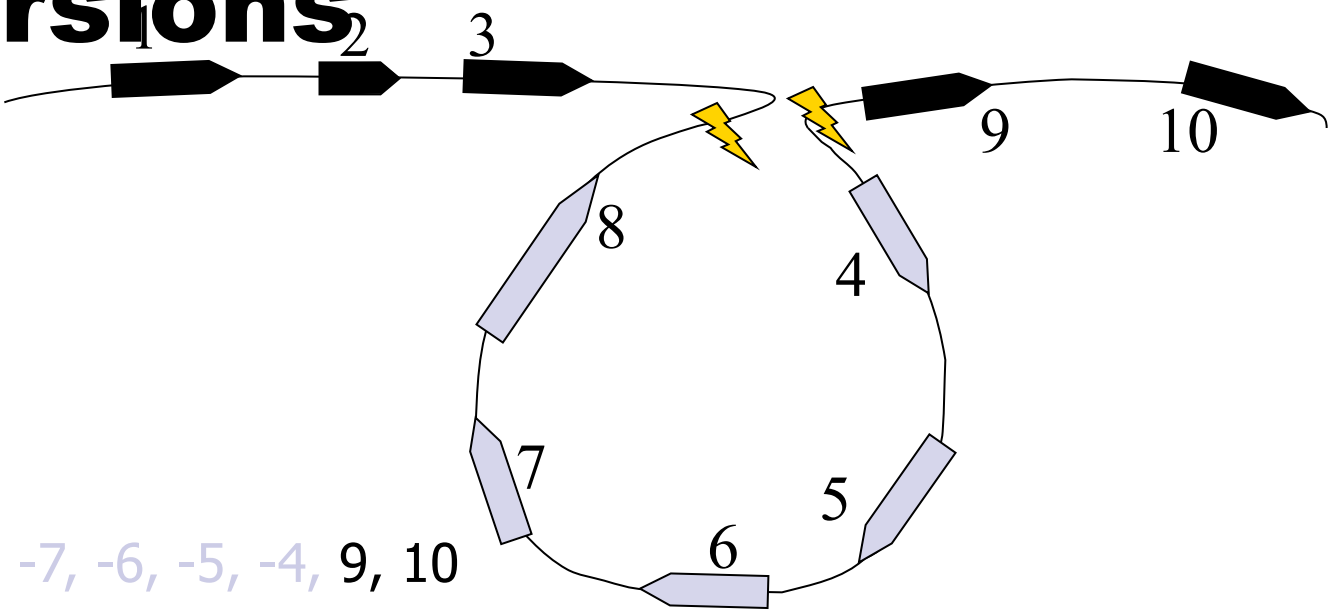
Inversions



1, 2, 3, -8, -7, -6, -5, -4, 9, 10

- Blocks represent conserved genes.
- In the course of evolution or in a clinical context, blocks 1,…,10 could be misread as 1, 2, 3, -8, -7, -6, -5, -4, 9, 10.
- **Evolution:** occurred one-two times every million years.
- **Cancer:** may occur every month.

Inversions

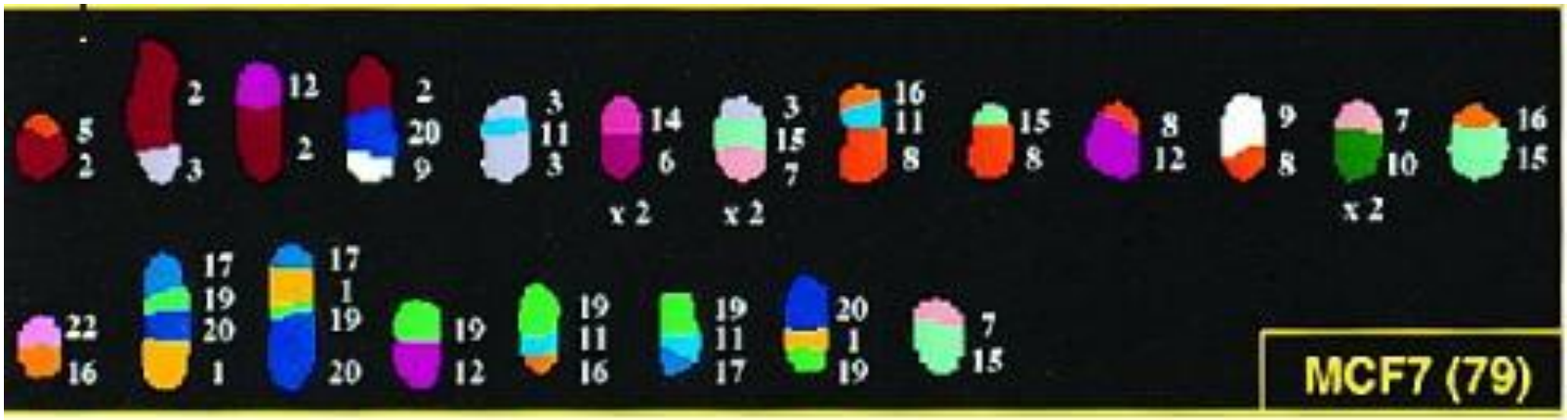


1, 2, 3, -8, -7, -6, -5, -4, 9, 10

The inversion introduced two *breakpoints* (disruptions in order).

Measuring Structural Changes in Tumors: Cytogenetics

- Directly visualize (fluorescently) labeled chromosomes.
  - Chromosome banding, mFISH, SKY

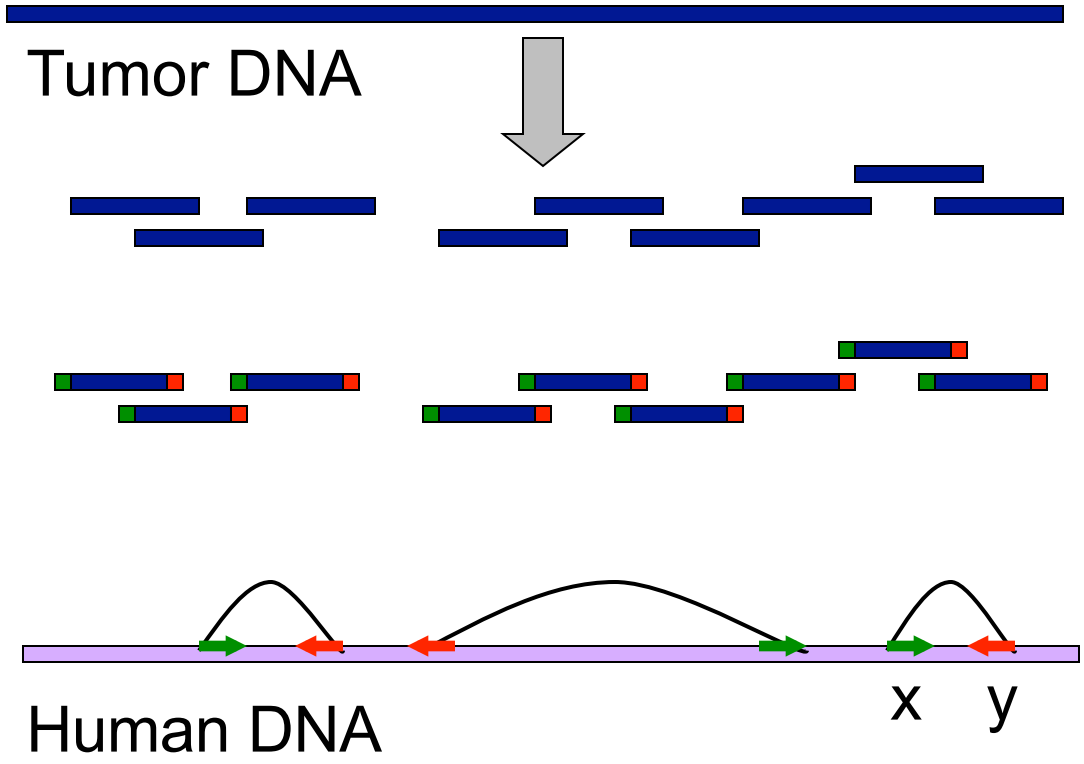


Weakness:

- Physical location of chromosomal junctions *not* revealed. Low resolution.
- No/little information about copy number changes.

Paired End Sequencing (PE)

C. Collins et al. (UCSF Cancer Center)



1) Pieces of tumor genome: clones (100-250kb).

2) Sequence ends of clones (500bp).

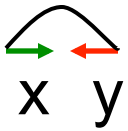
3) Map end sequences to human genome.

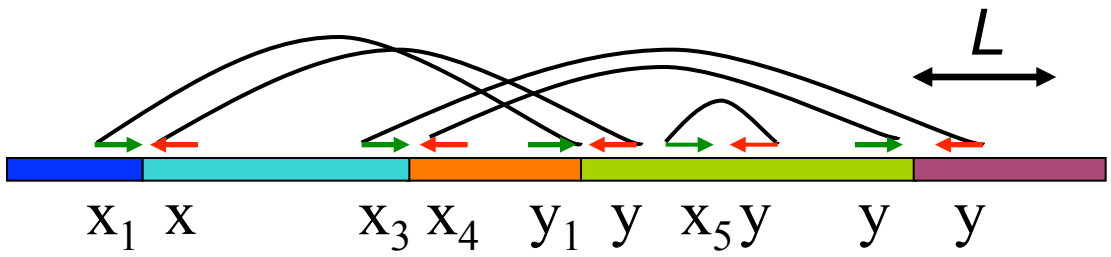
Each clone corresponds to pair of end sequences (***PE pair***) *(x,y).*

Typical Next Generation Sequencing read lengths are shorter.

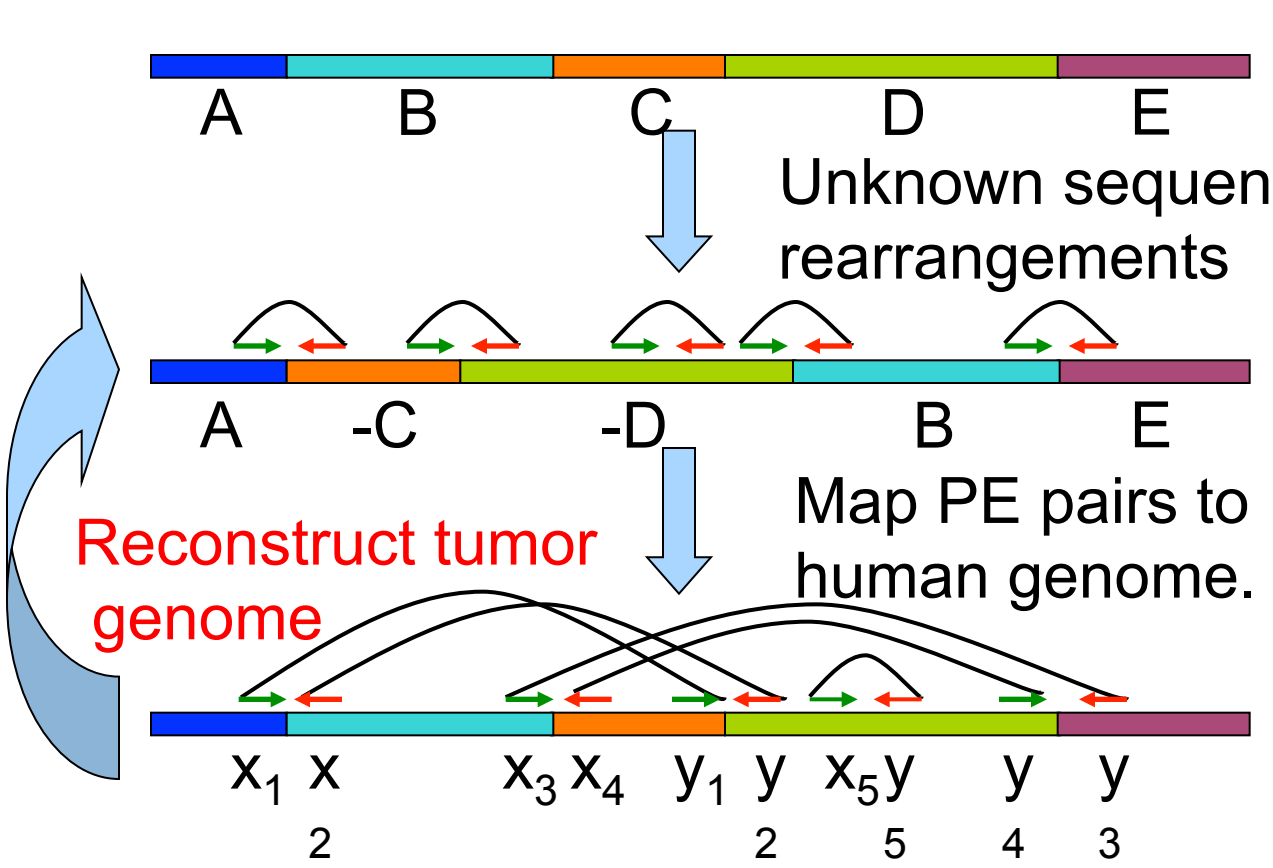
PE Pairs

- Order PE pair such that $x < y$.
- PE pair (x,y) is
  - ***valid*** if
    - x,y on same chromosome. and
    - $l \leq y - x \leq L$, min (max) size of clone.
    - x, y have opposite, convergent orientations
  - ***invalid,*** otherwise.
    - Results from rearrangement or experimental "noise".





Tumor Genome Reconstruction Puzzle

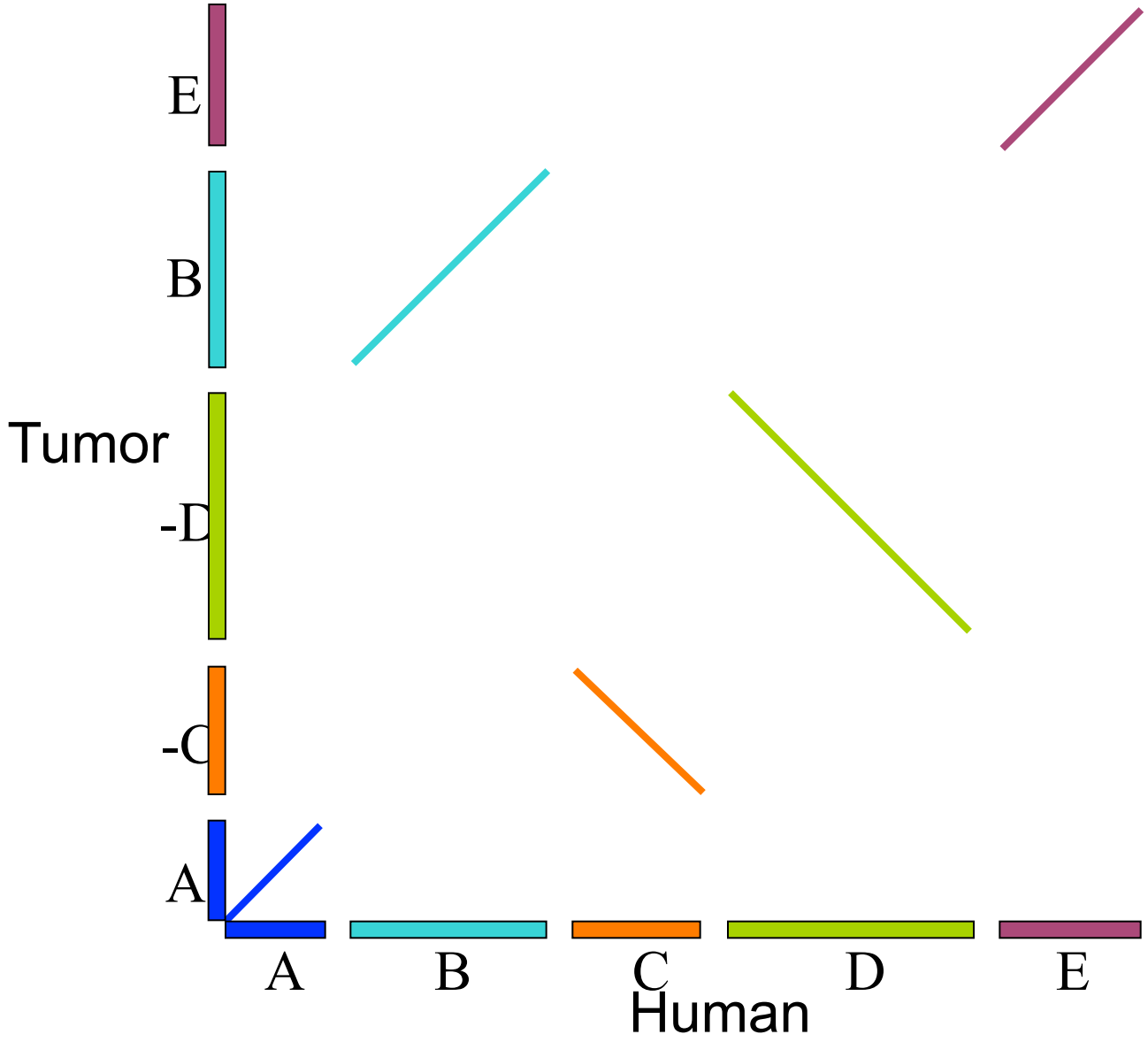


Human genome (known)

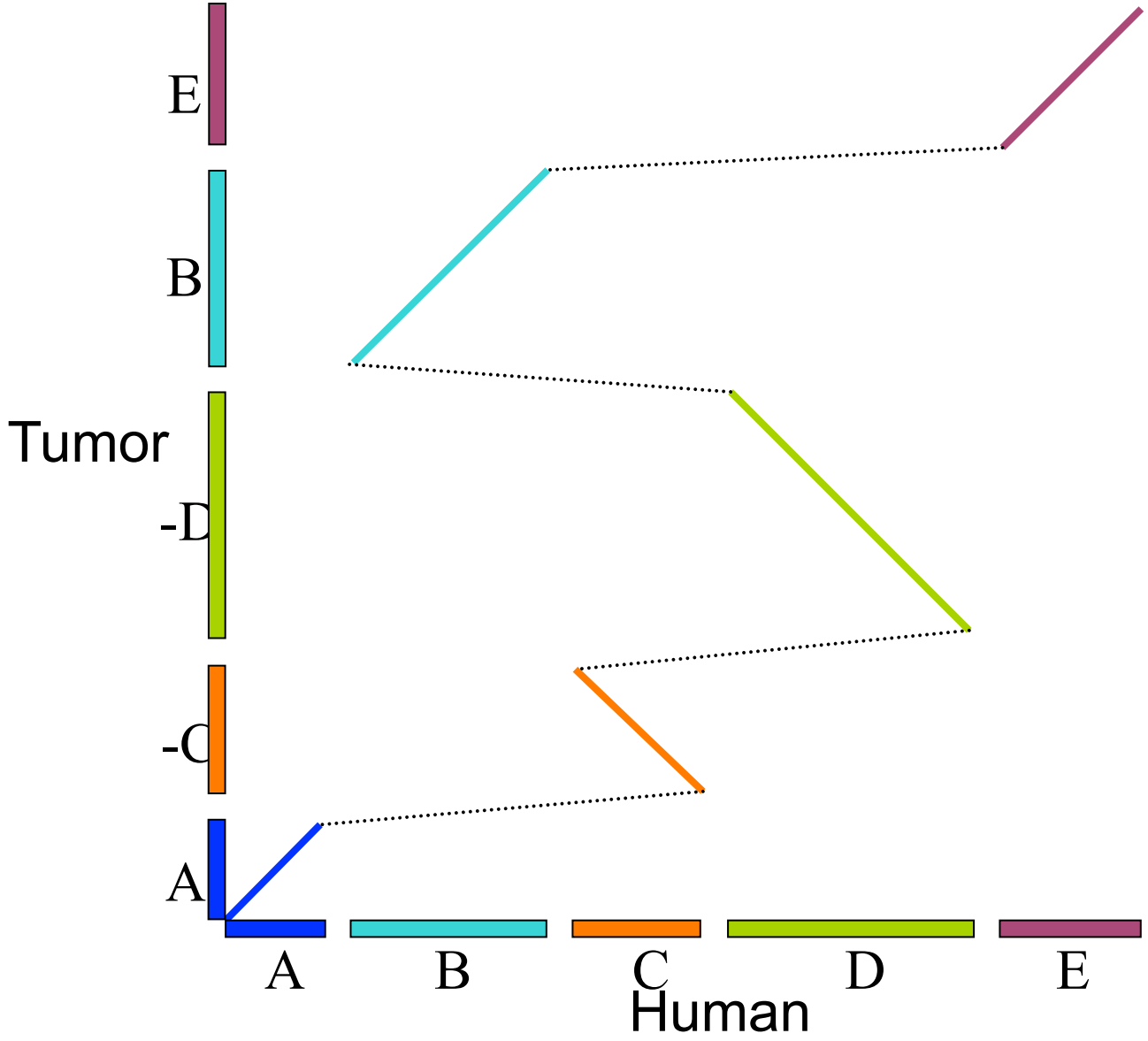
Tumor genome (unknown)

Location of PE pairs in human genome. (known)

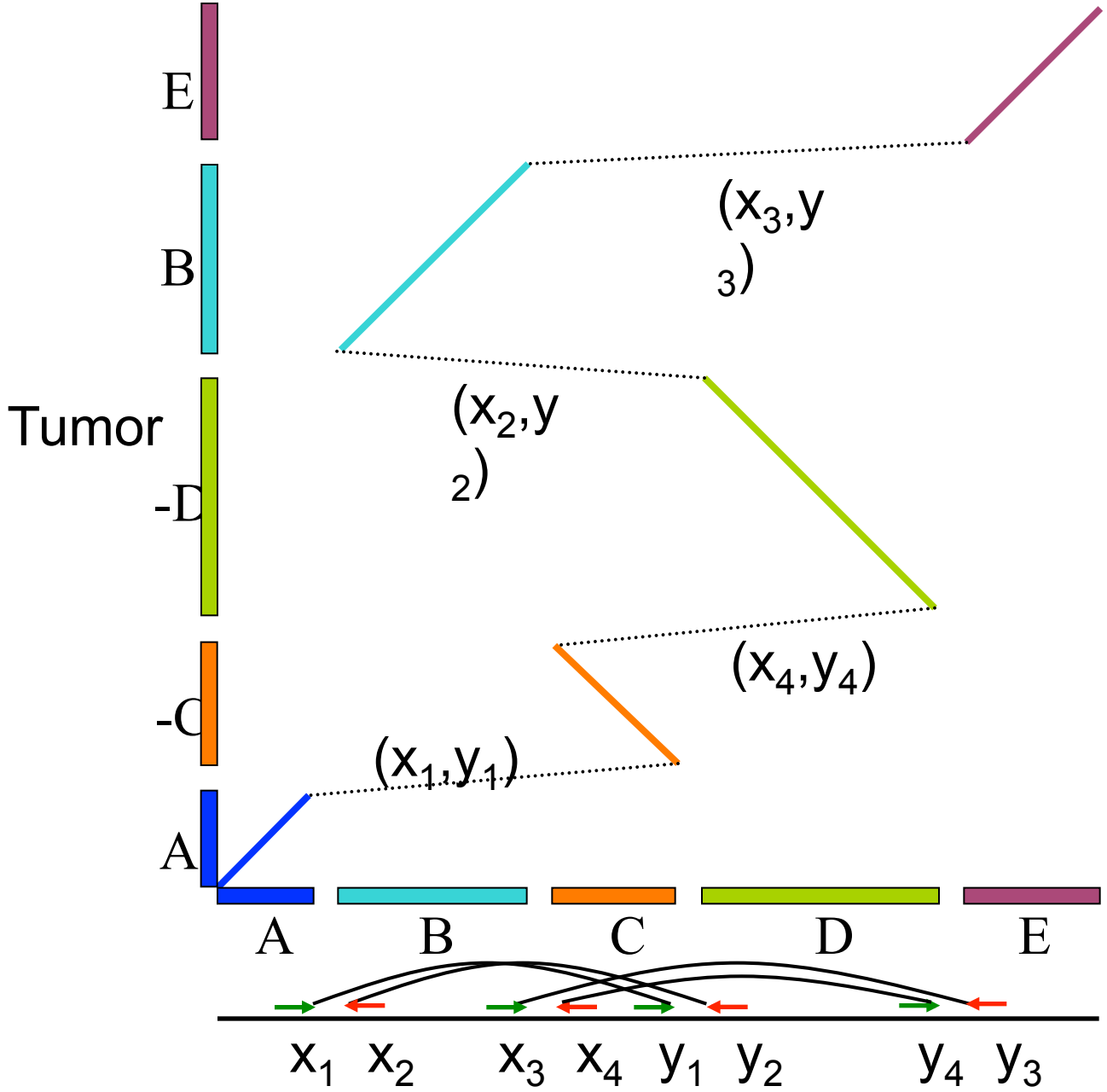
Tumor Genome Reconstruction



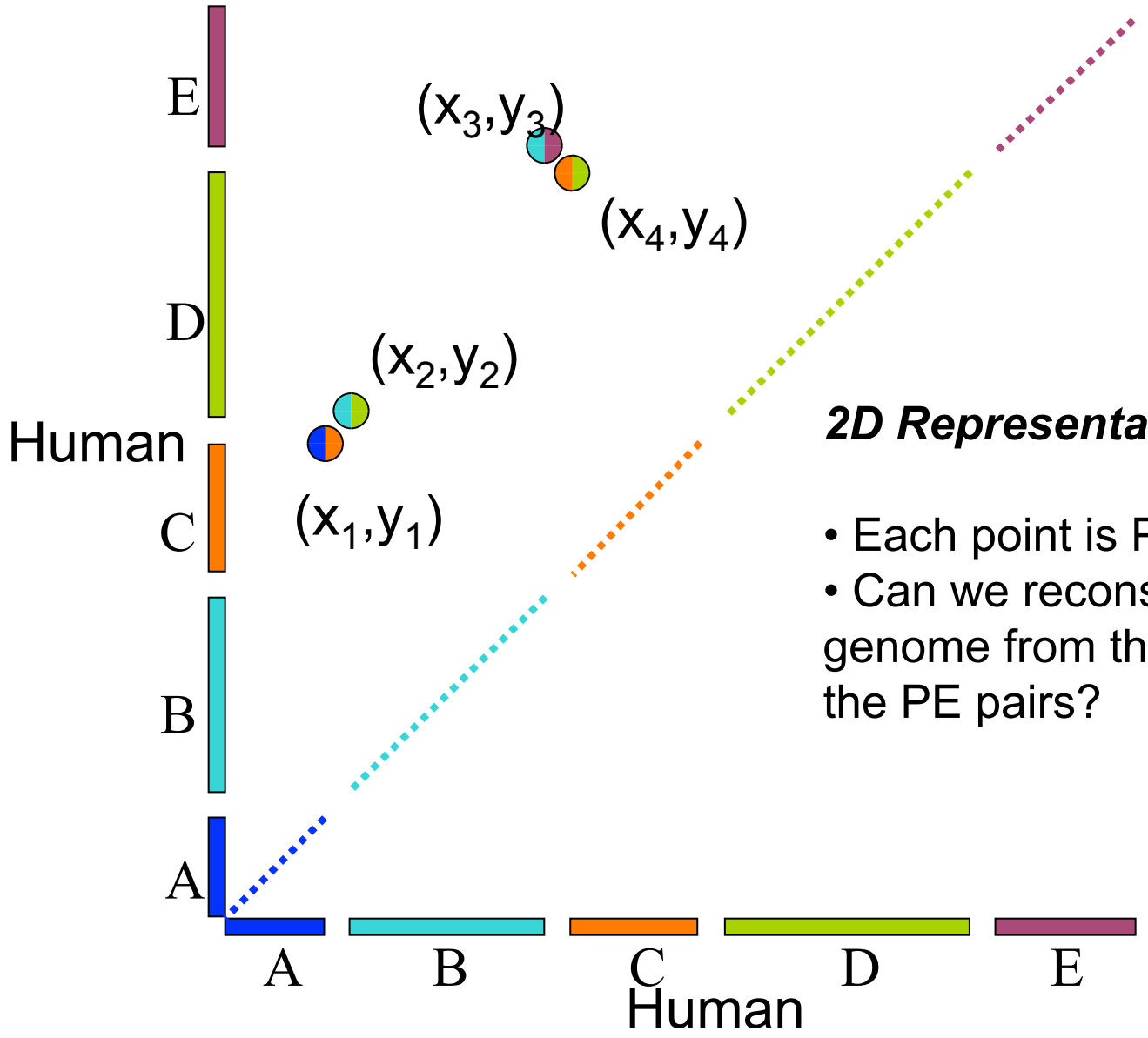
Tumor Genome Reconstruction



Tumor Genome Reconstruction

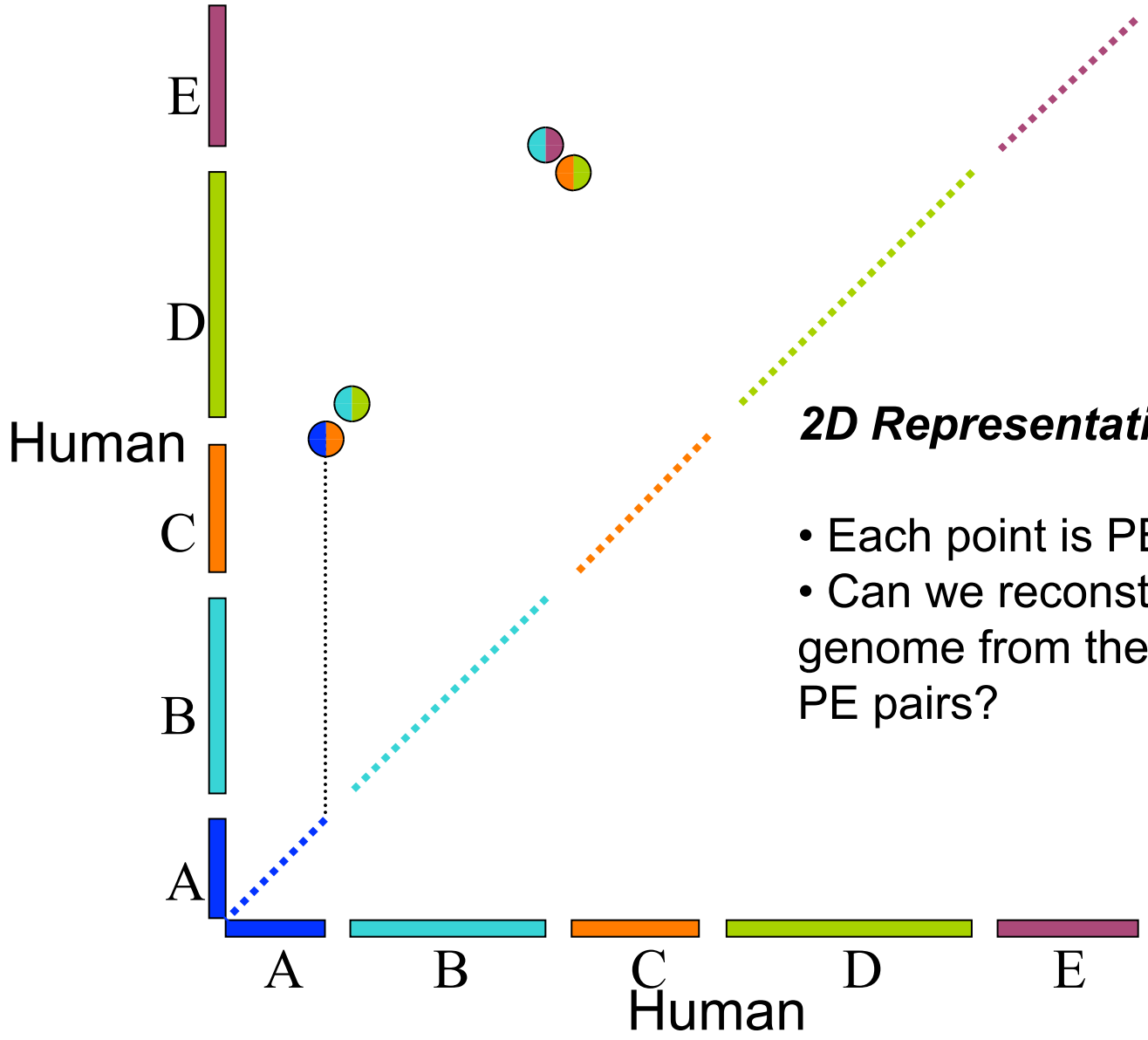


PE Plot



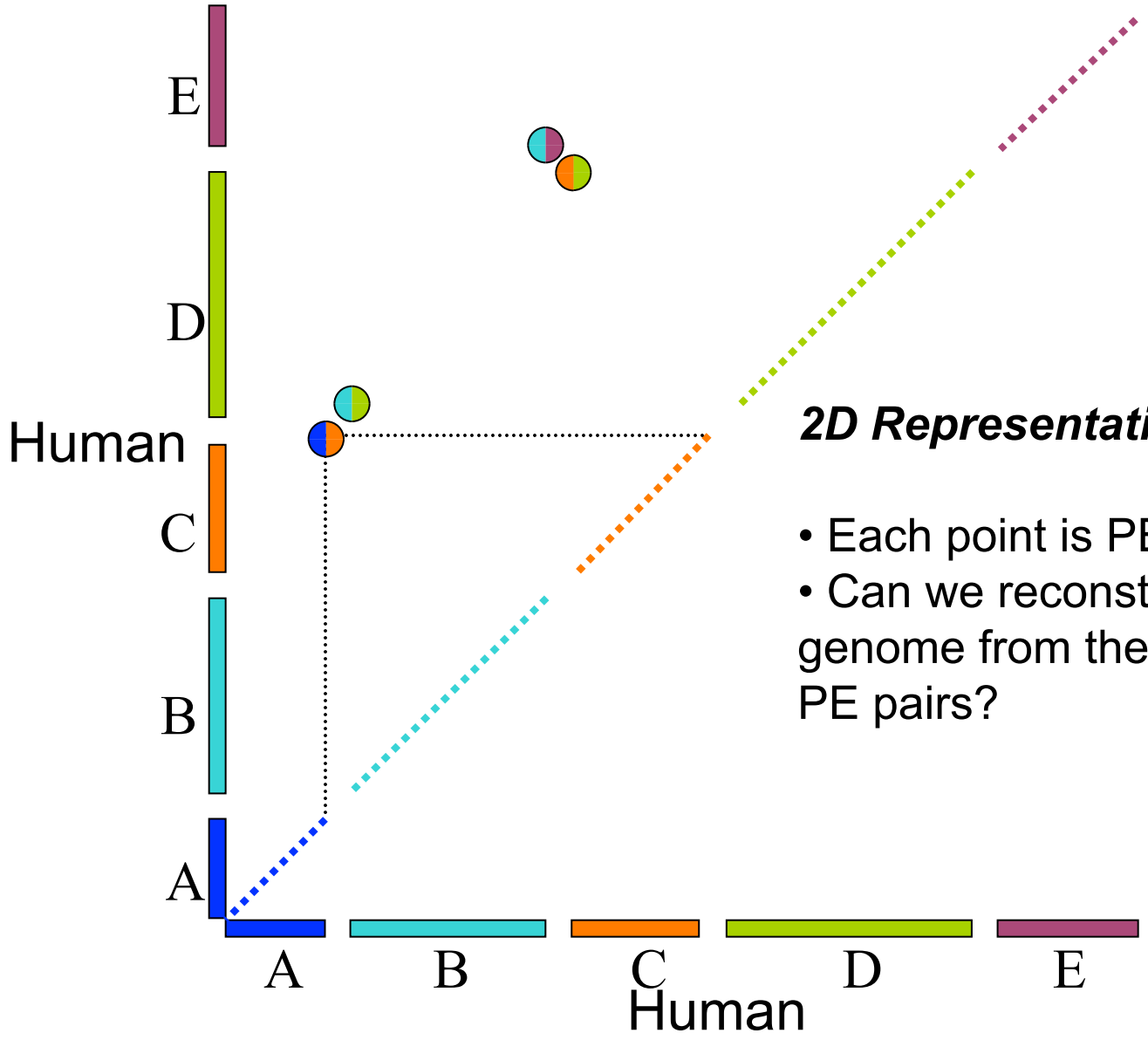
2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?



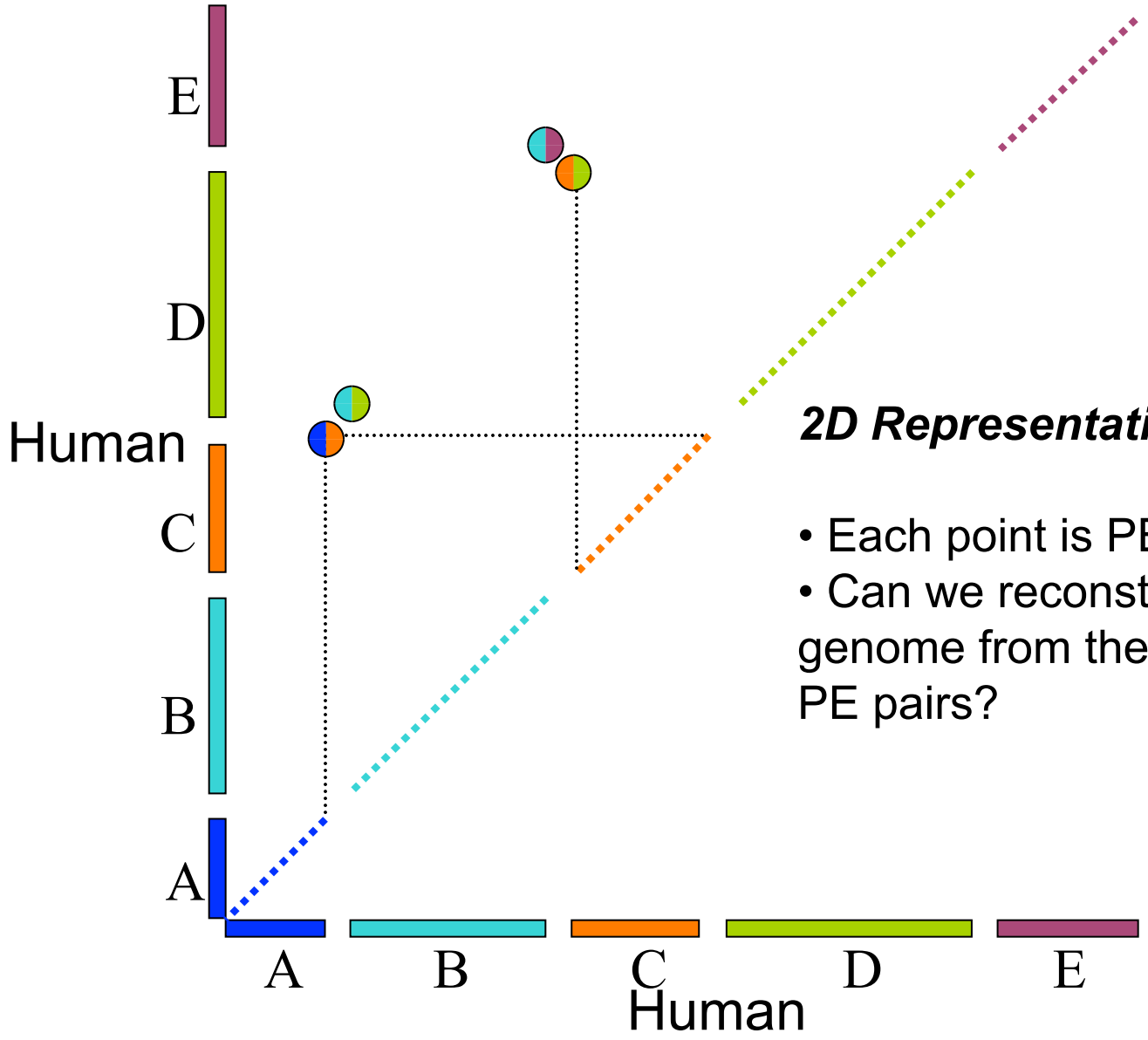
2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?



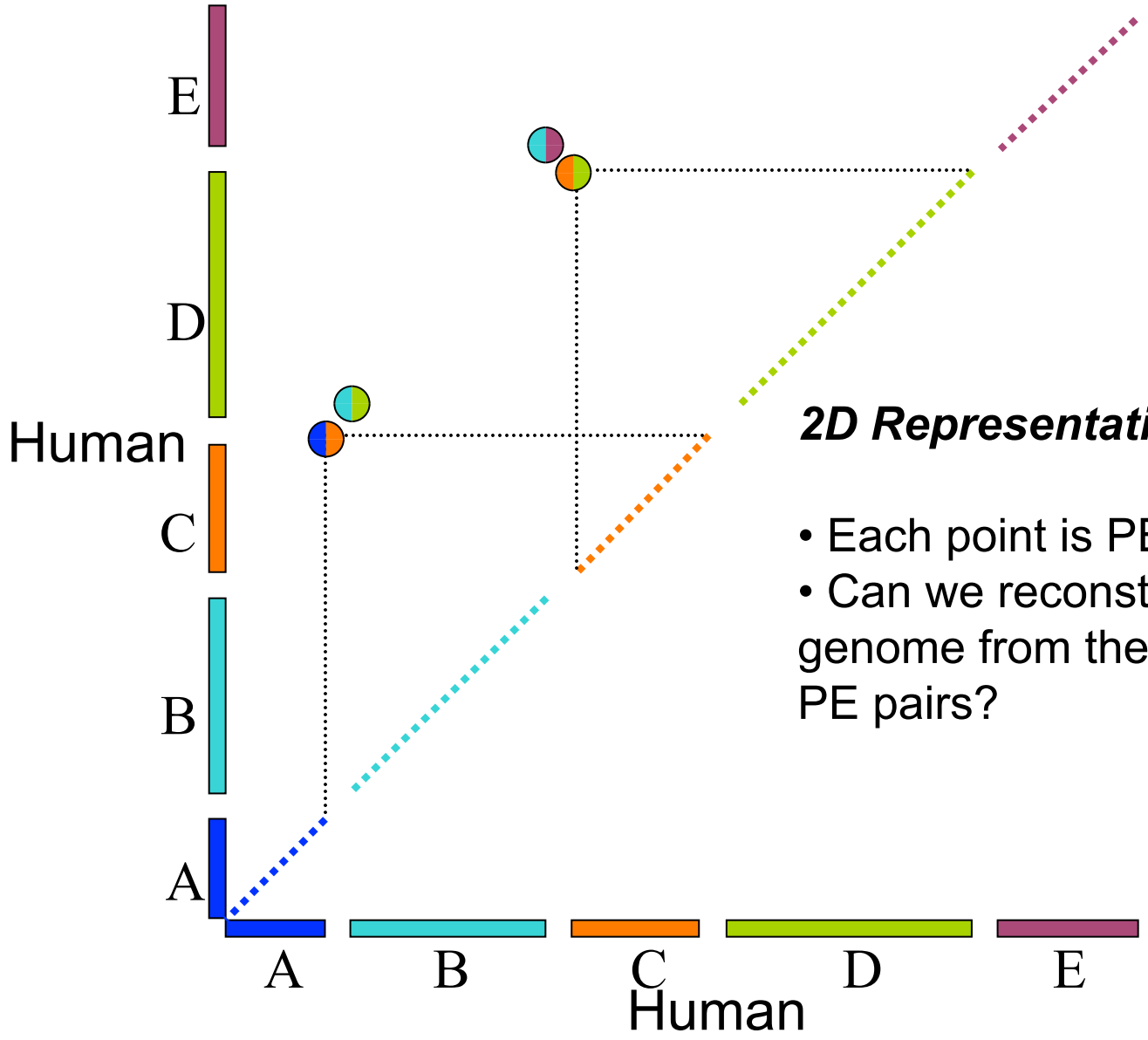
2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?



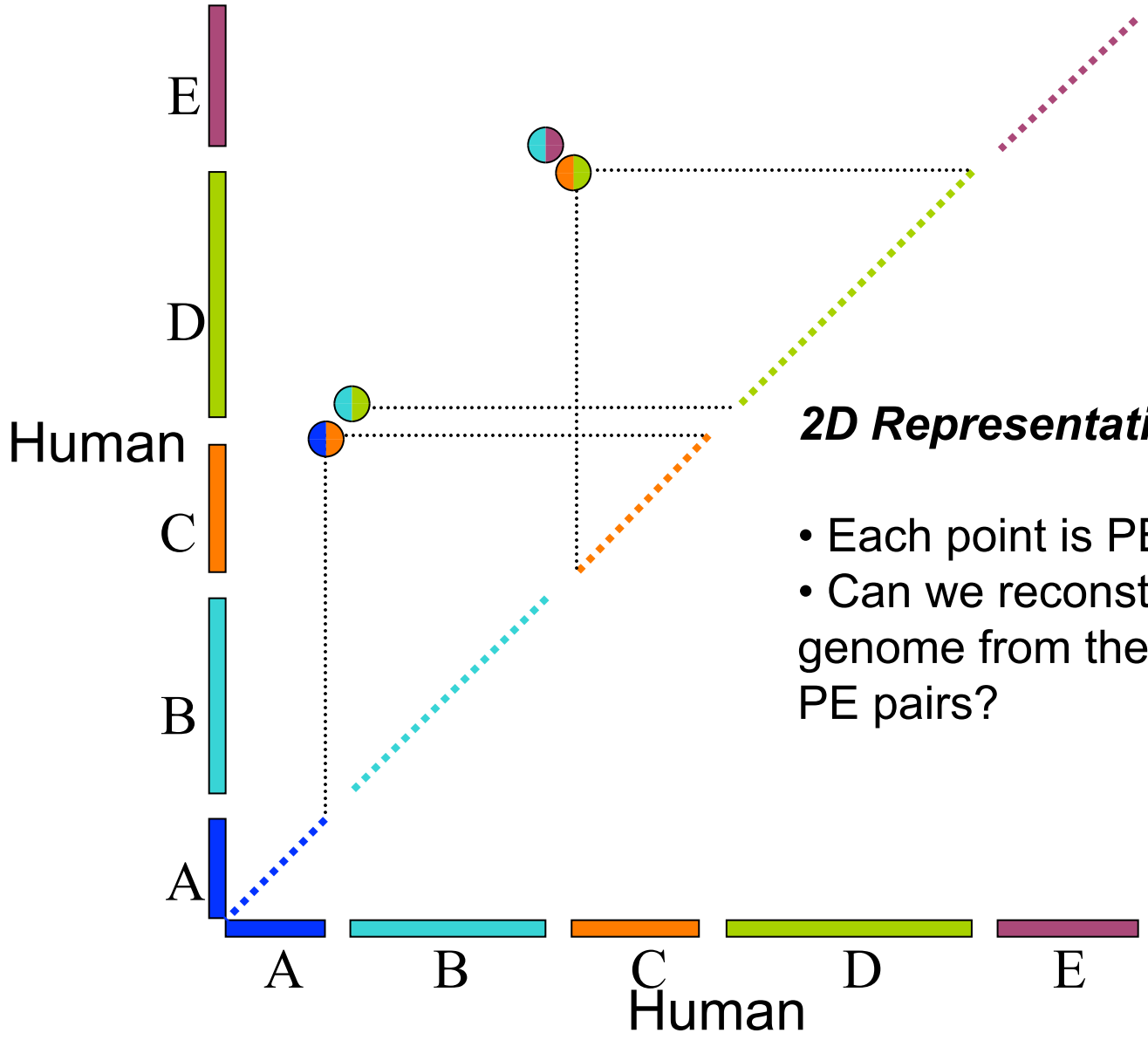
2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?



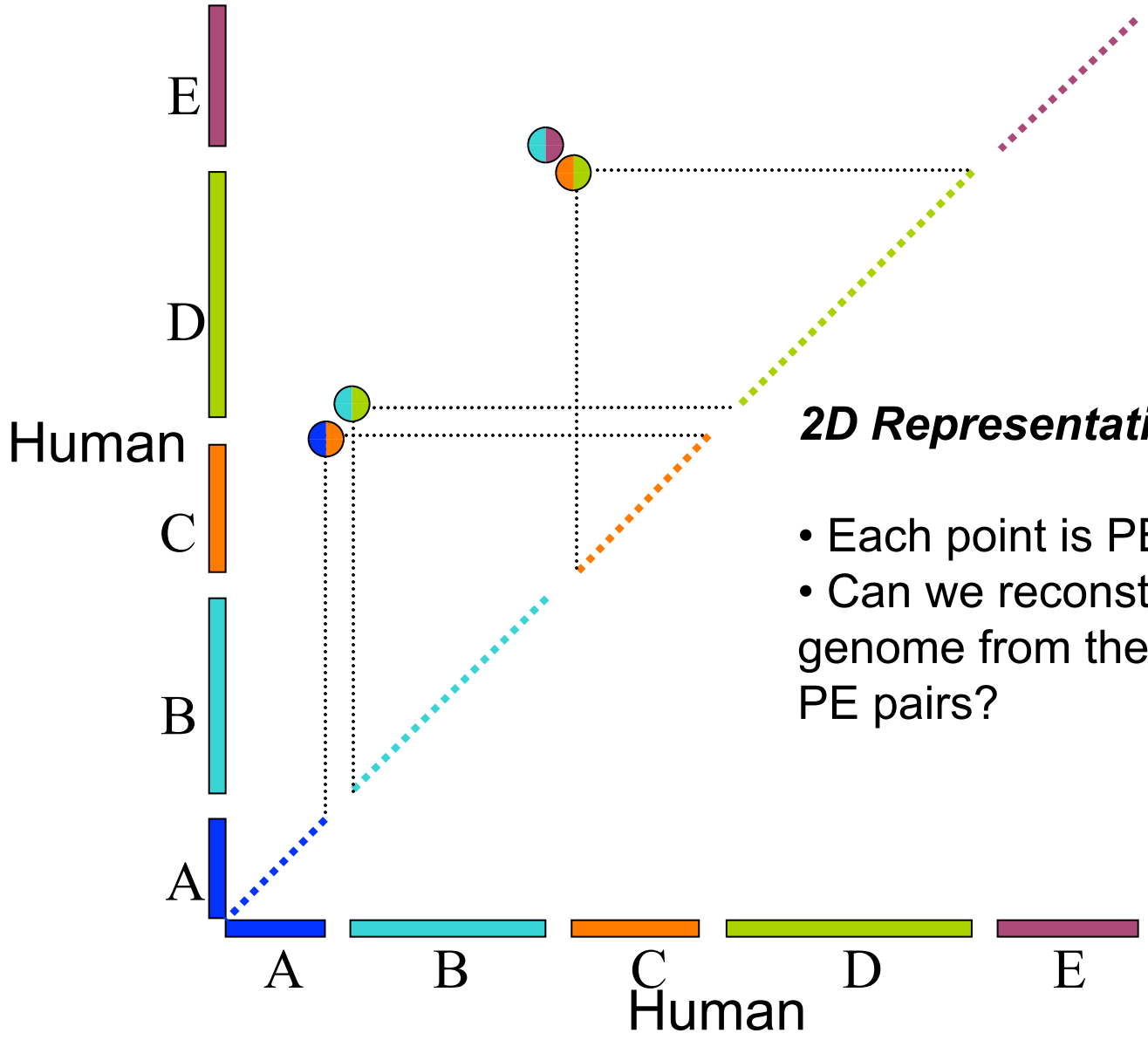
2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?



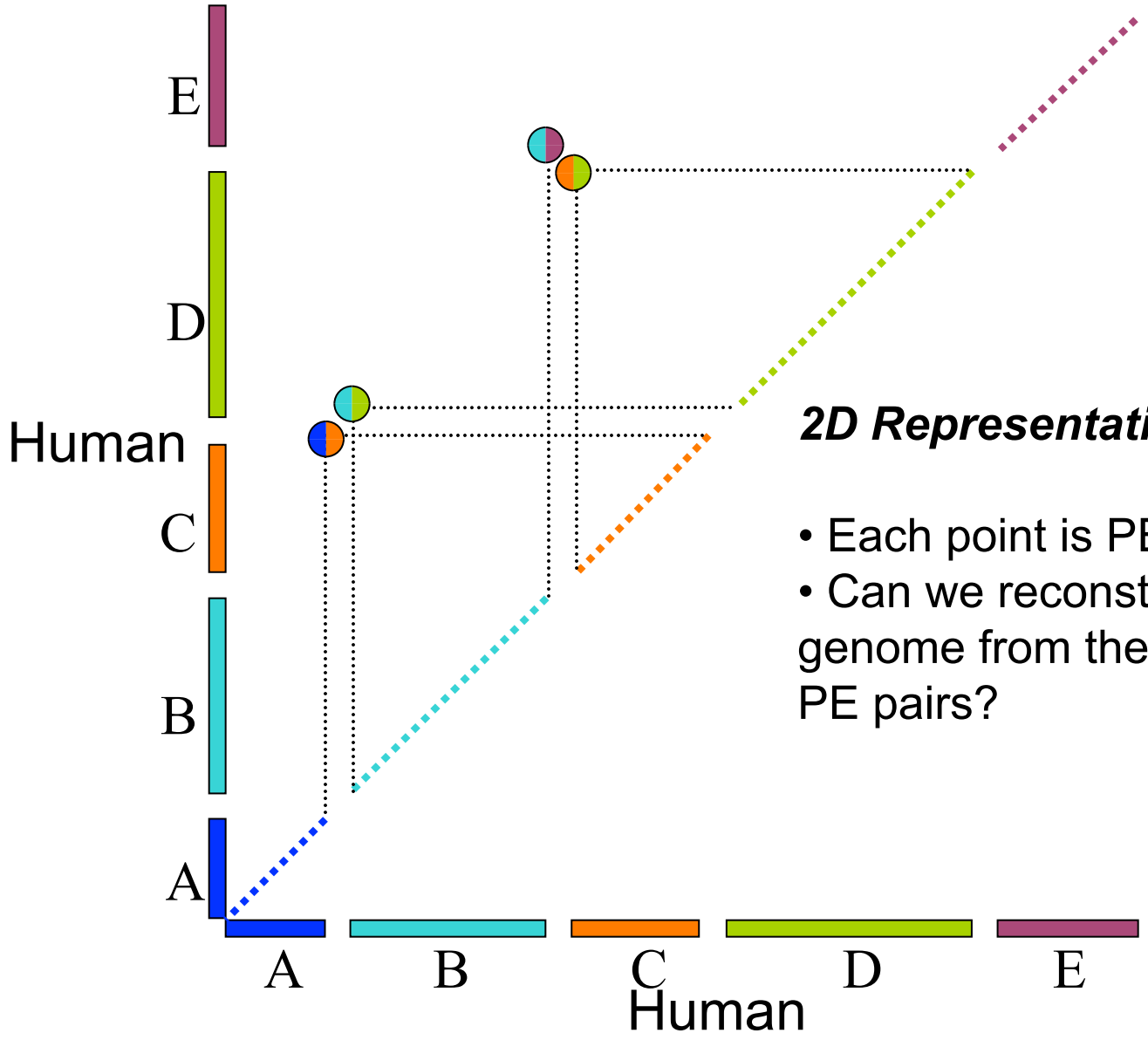
2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?



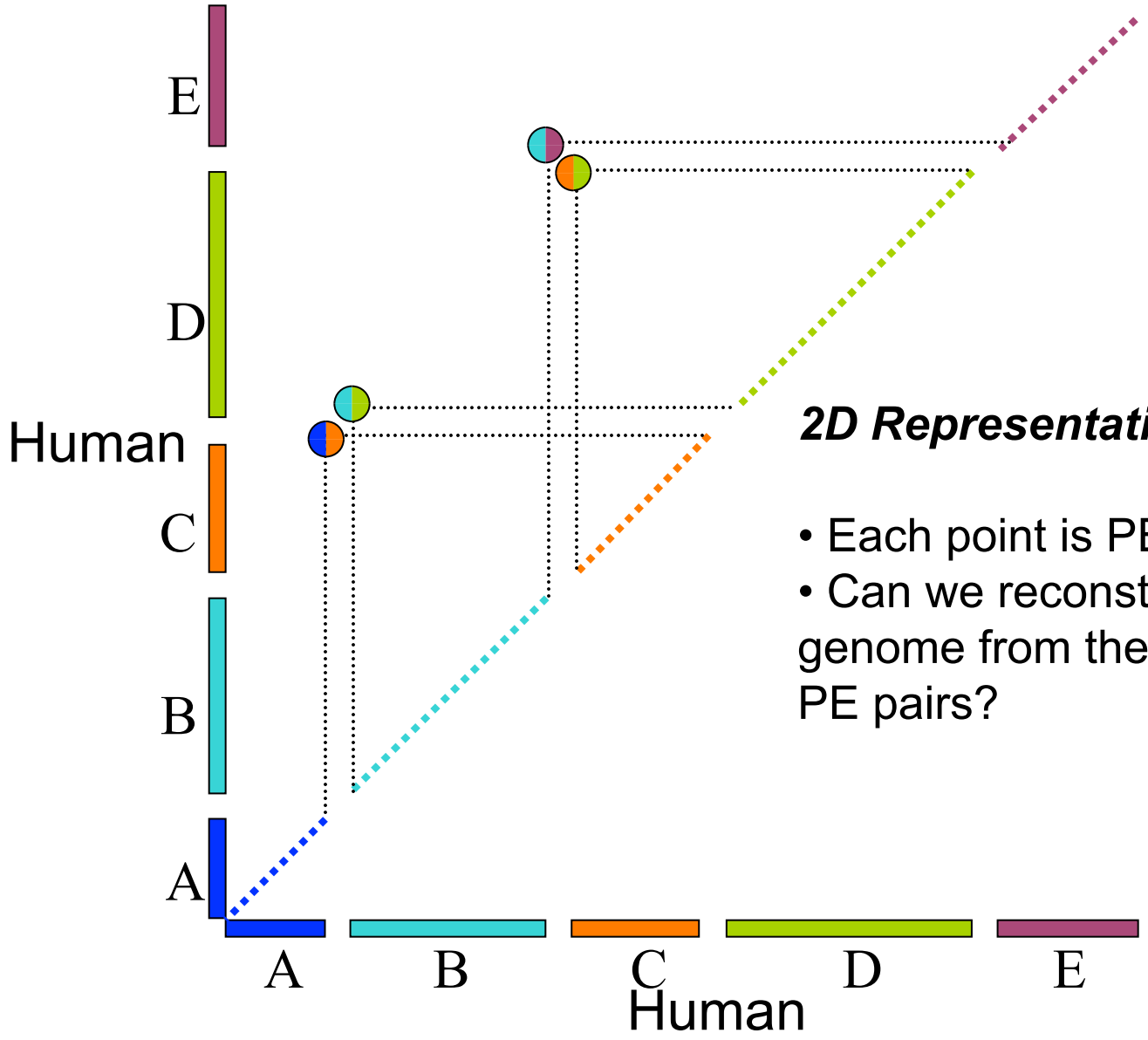
2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?



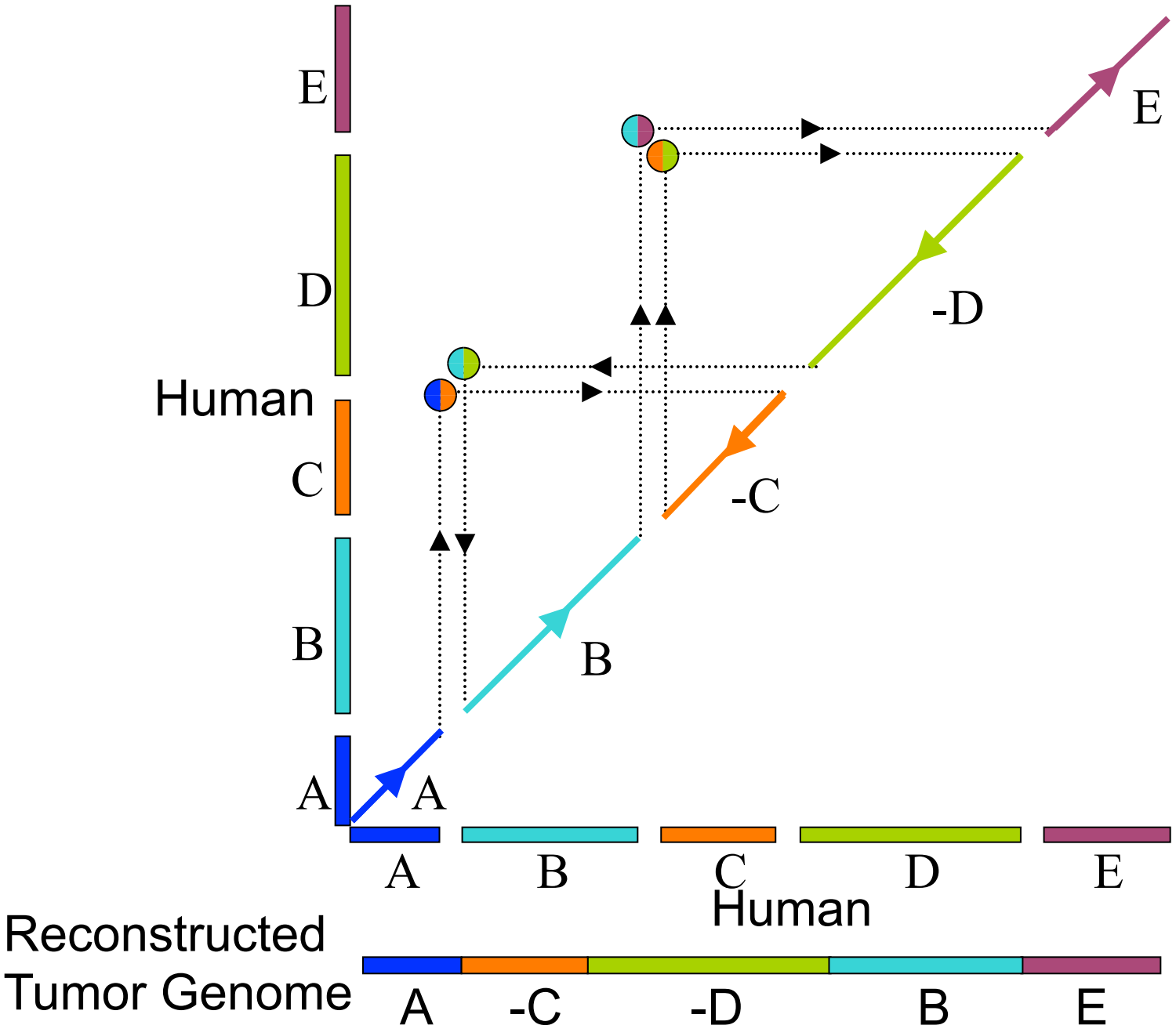
2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?

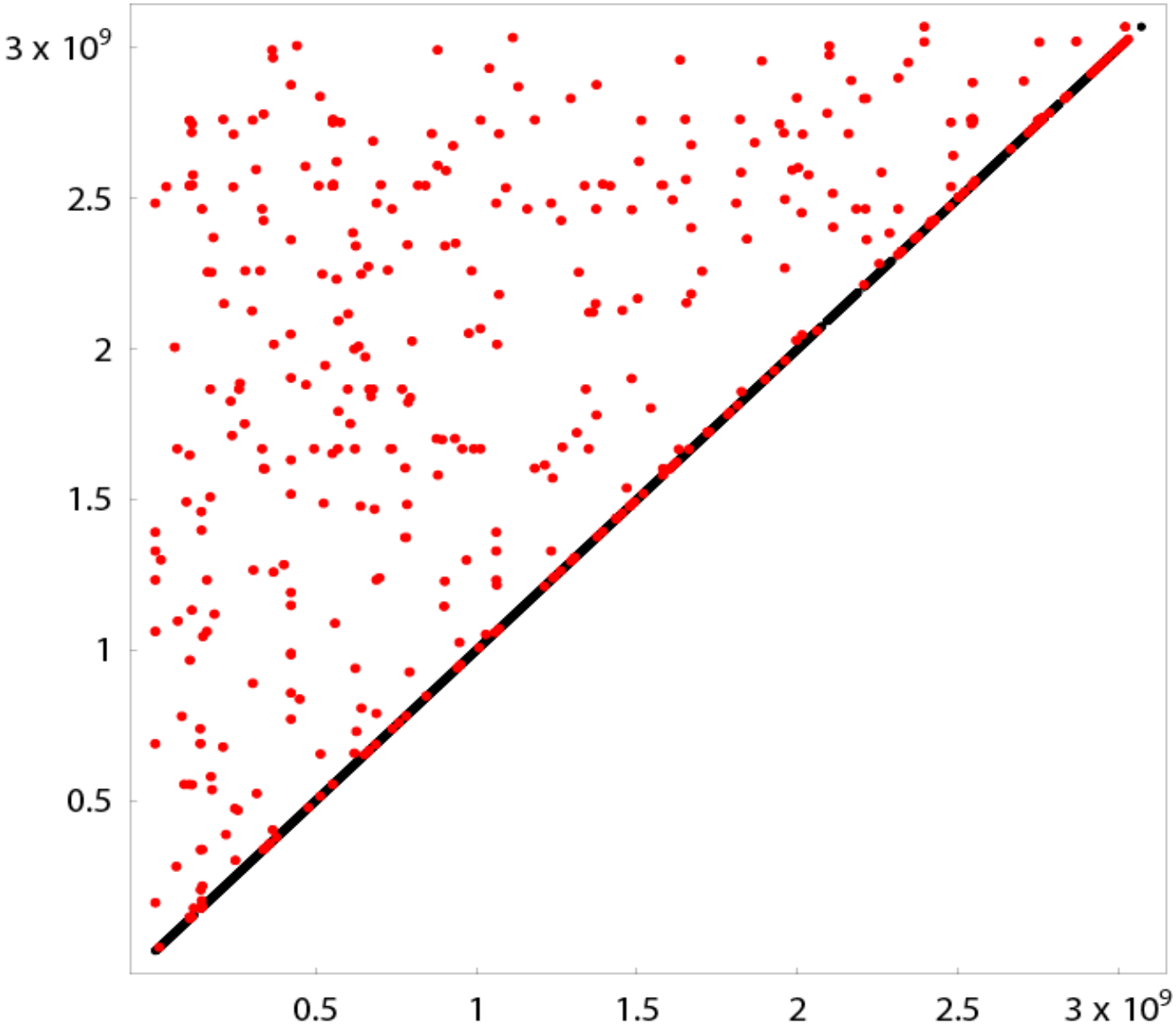


2D Representation of PE Data

- Each point is PE pair.
- Can we reconstruct the tumor genome from the positions of the PE pairs?


E
D
Human
C
B
A
A
B
C
D
E
-D
-C
B
E
A
B
C
D
E
Human
Reconstructed
Tumor Genome
A
-C
-D
B
E

Real data noisy and incomplete!

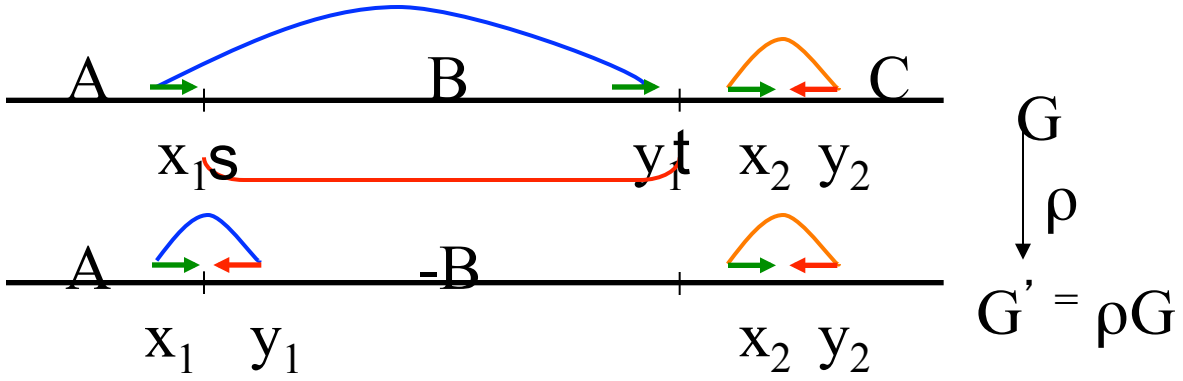


Reconstruction of Tumor History

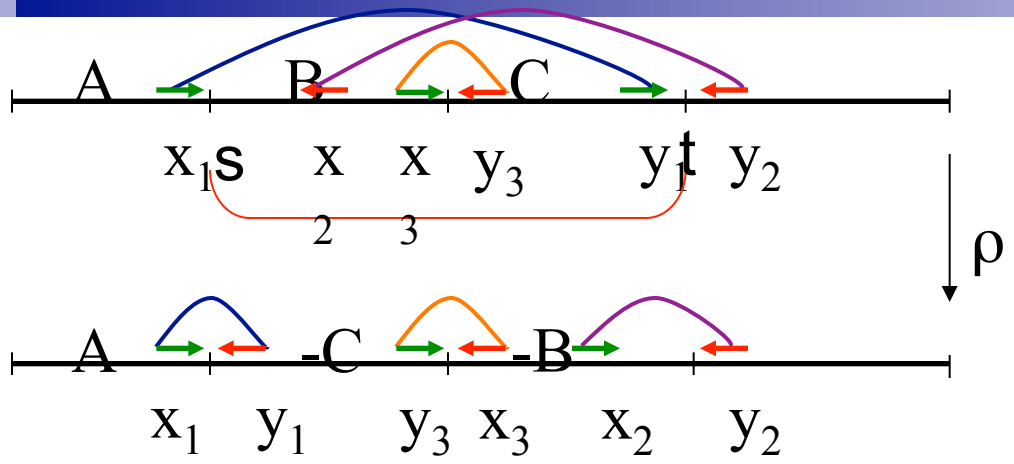
- Use knowledge of known rearrangement mechanisms
  - **e.g. inversions, translocations, etc.**
- Find *simplest* explanation for data, given these mechanisms.
- Motivation: *Sorting by Reversals*

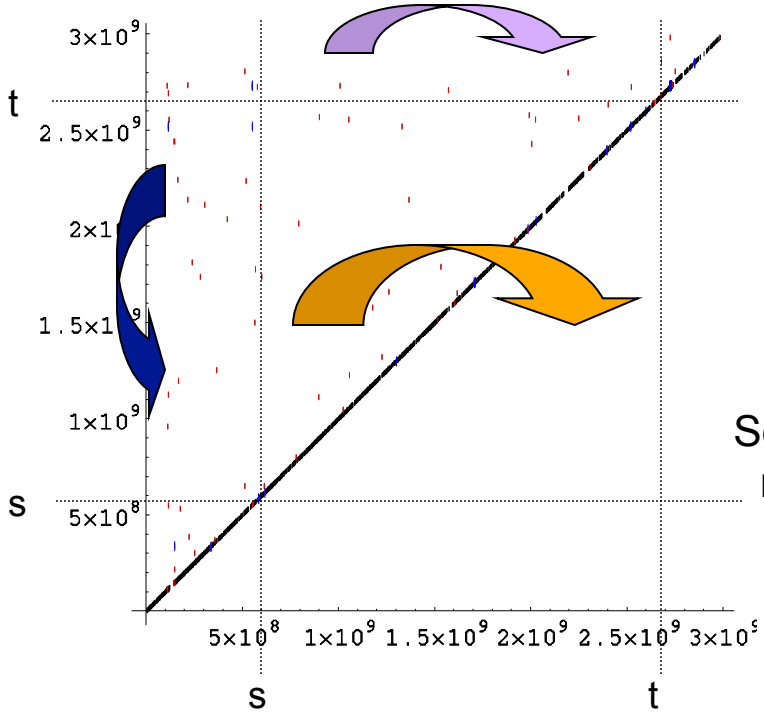
PE Sorting Problem

- **G** = [0,*M*], unichromosomal genome.
- Reversal $\rho_{s,t}(x) = \begin{cases} x, & \text{if } x < s \text{ or } x > t, \\ \boldsymbol{t - (x - s)}, & \textbf{otherwise.} \end{cases}$

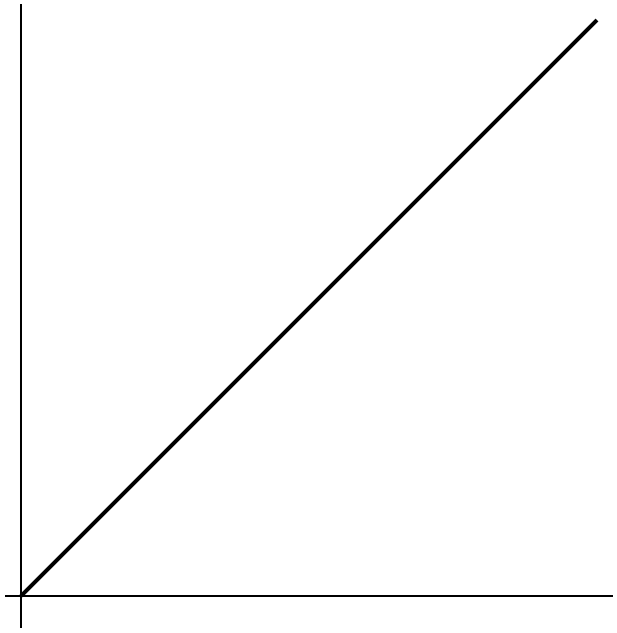


Given: PE pairs $(x_1, y_1), \ldots, (x_n, y_n)$
Find: Minimum number of reversals $\rho_{s1,t1}, \ldots, \rho_{sn,\,tn}$ such that if $\rho = \rho_{s1,t1} \cdots \rho_{sn,\,tn}$ then $(\rho\, x_1, \rho\, y_1), \ldots, (\rho\, x_n, \rho\, y_n)$ are valid PE pairs.

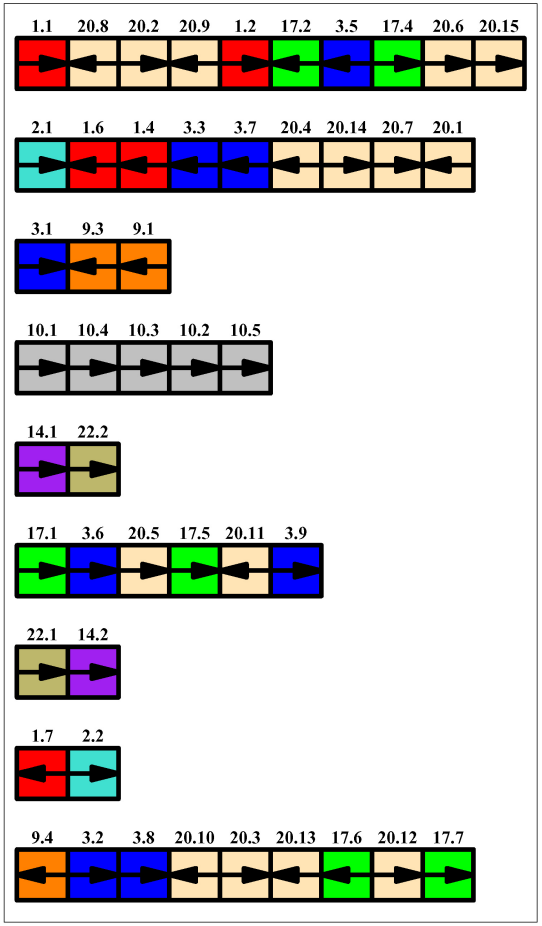




Sequence of reversals.



All ES pairs valid.

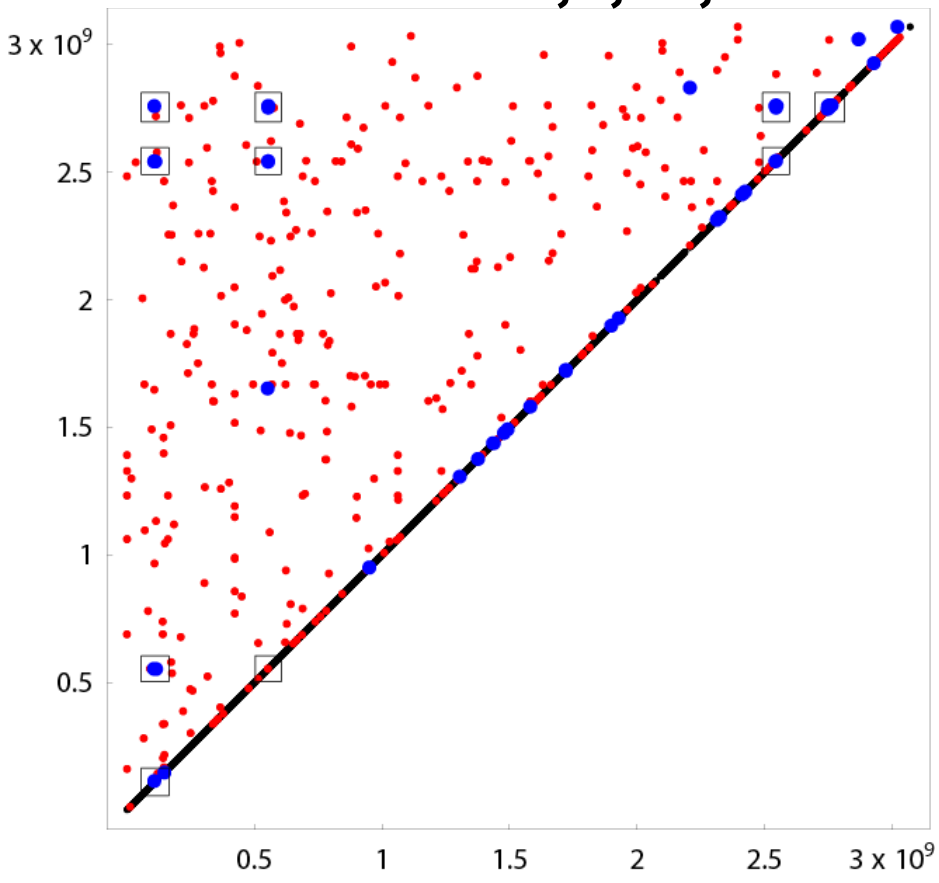


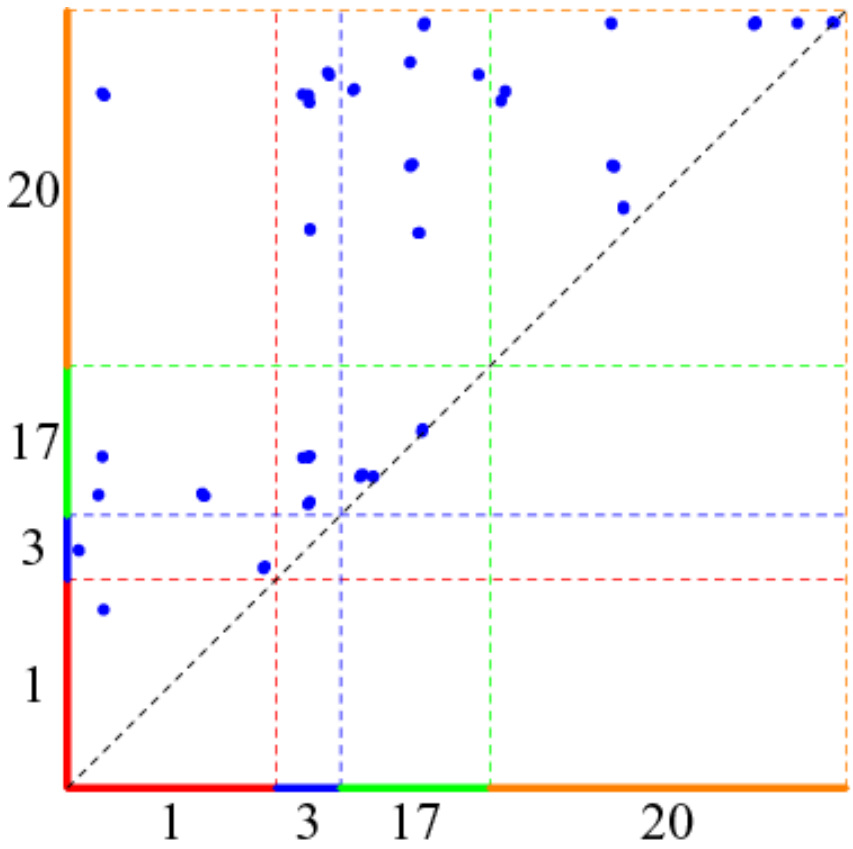
Human chromosomes → MCF7 chromosomes

5 inversions
15 translocations
Raphael et al.

Complications with MCF7:

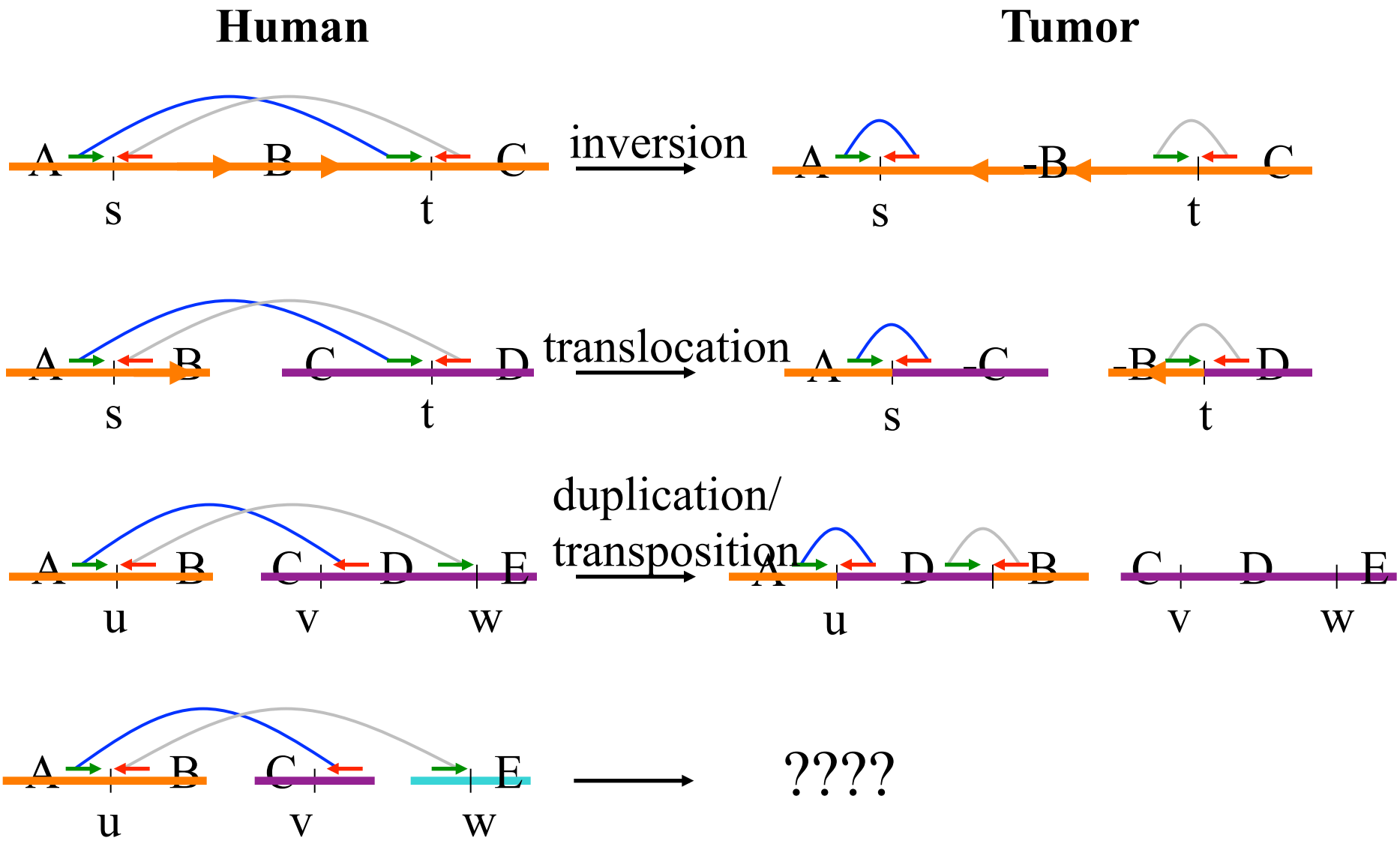
Chromosomes 1,3,17, 20



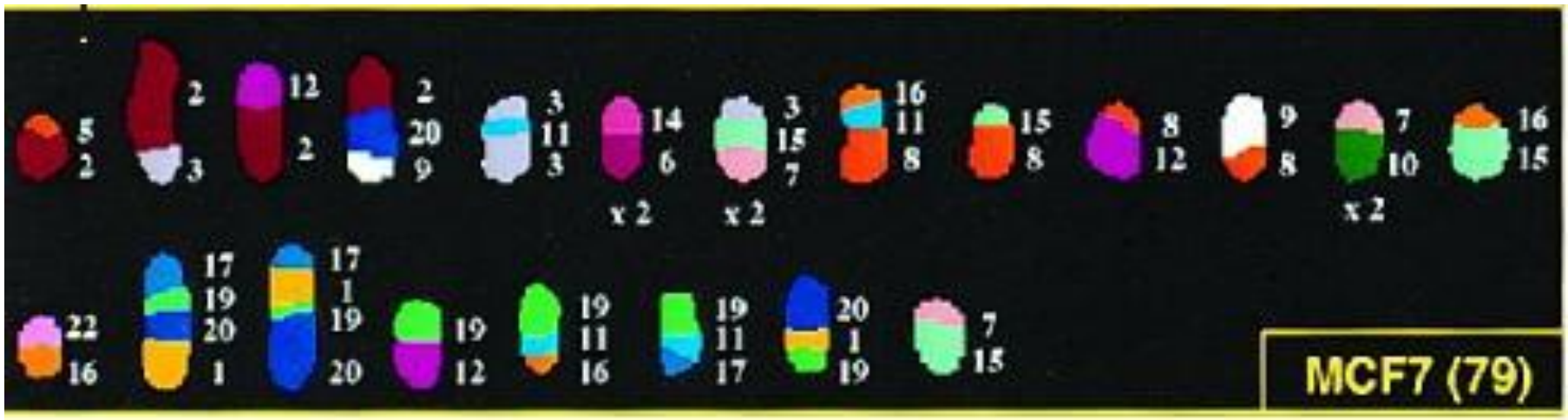


33/70 clusters
Total length: 31Mb

Rearrangement Signatures



Complex Tumor Genomes



Structure of Duplications in Tumors?

- Duplicated segments may co-localize
  (Guan et al. *Nat.Gen.*1994)

Human genome

Tumor genome

- Mechanisms not well understood.

Tumor Amplisomes