Models and Algorithms for Genome Rearrangement with Positional Constraints

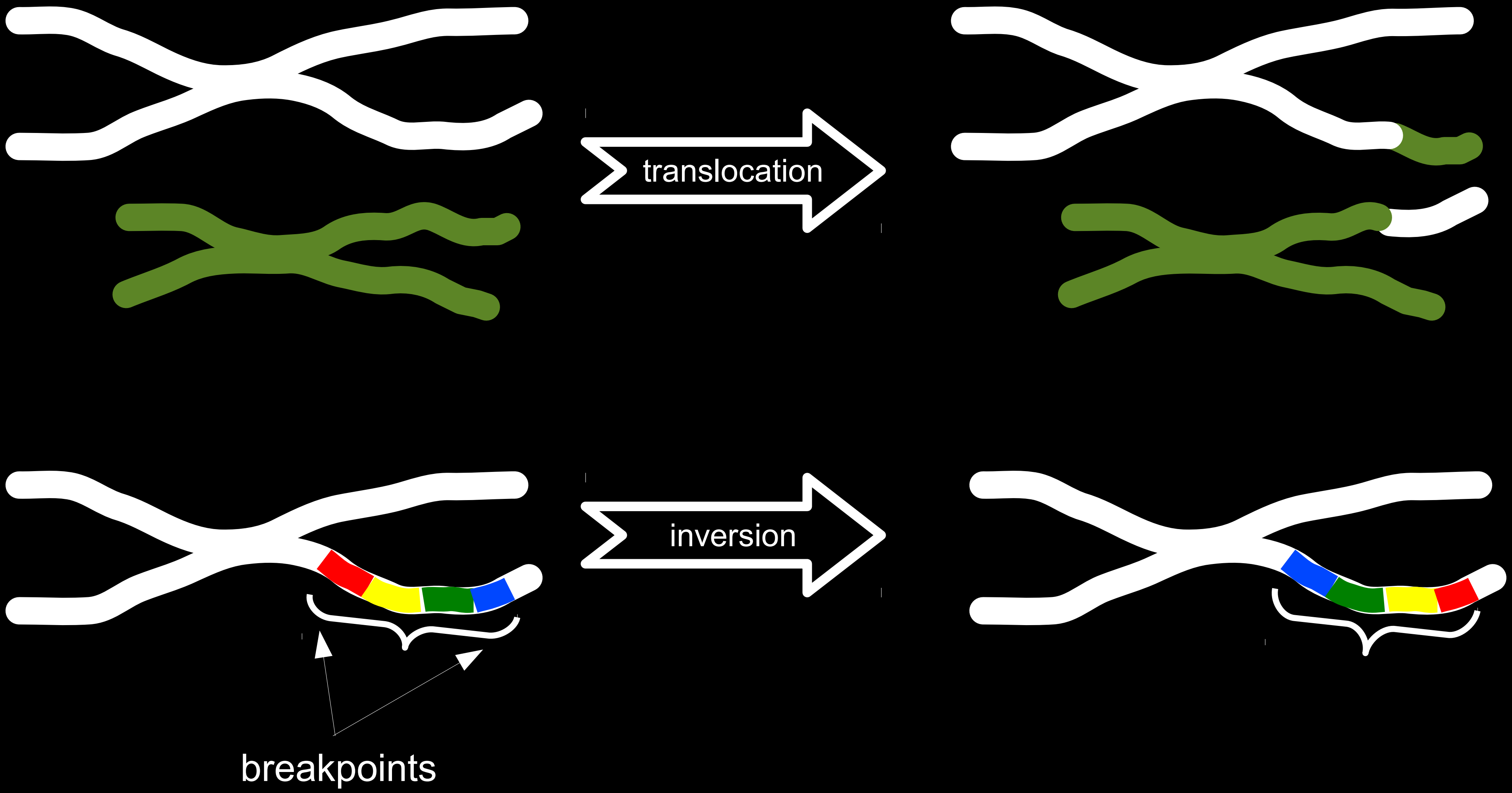
Krister Swenson

CNRS
LIRMM, Universite de Montpellier
France

Mathieu Blanchette

McGill University
Montreal, Canada

Genome Rearrangements



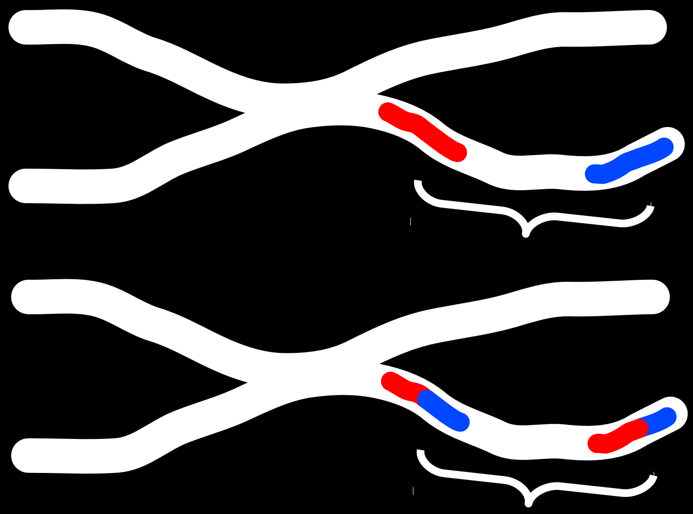
translocation

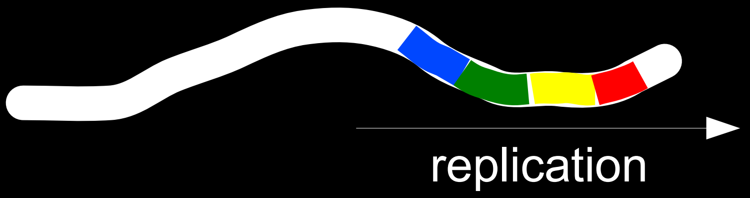
inversion

breakpoints

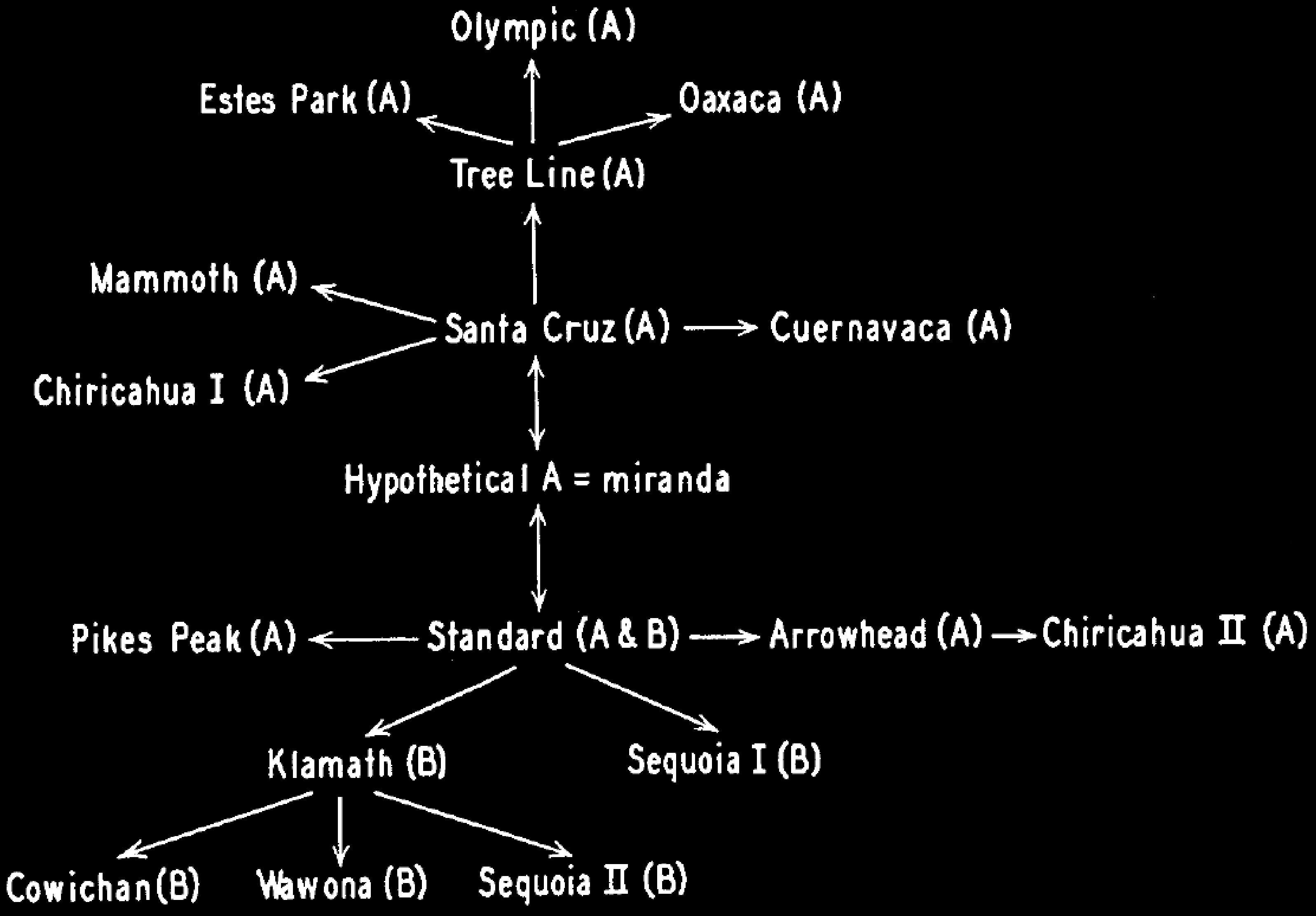
Consequences of Rearrangements

- Role in speciation
  - reproductive isolation
- Gene regulation
  - aberrant proteins
  - positional effects
- Disease
  - many cancers
  - hemophilia A
  - etc.

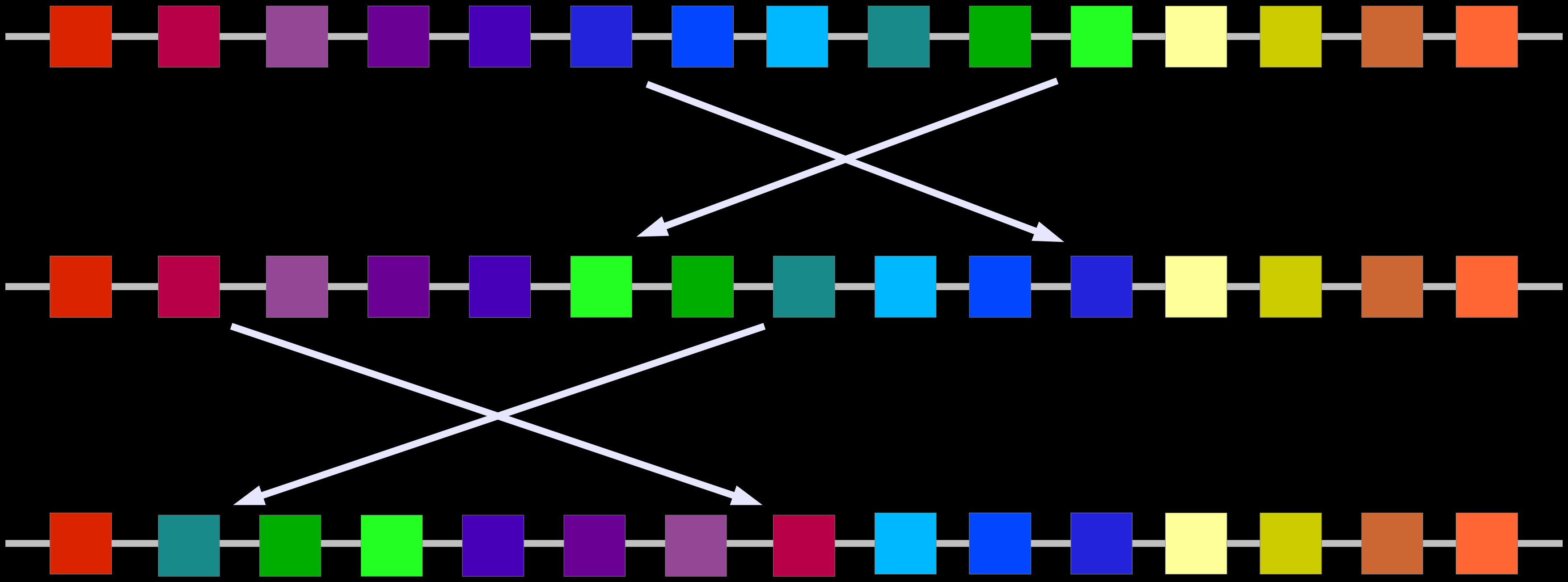




Phylogeny Reconstruction (~1930)

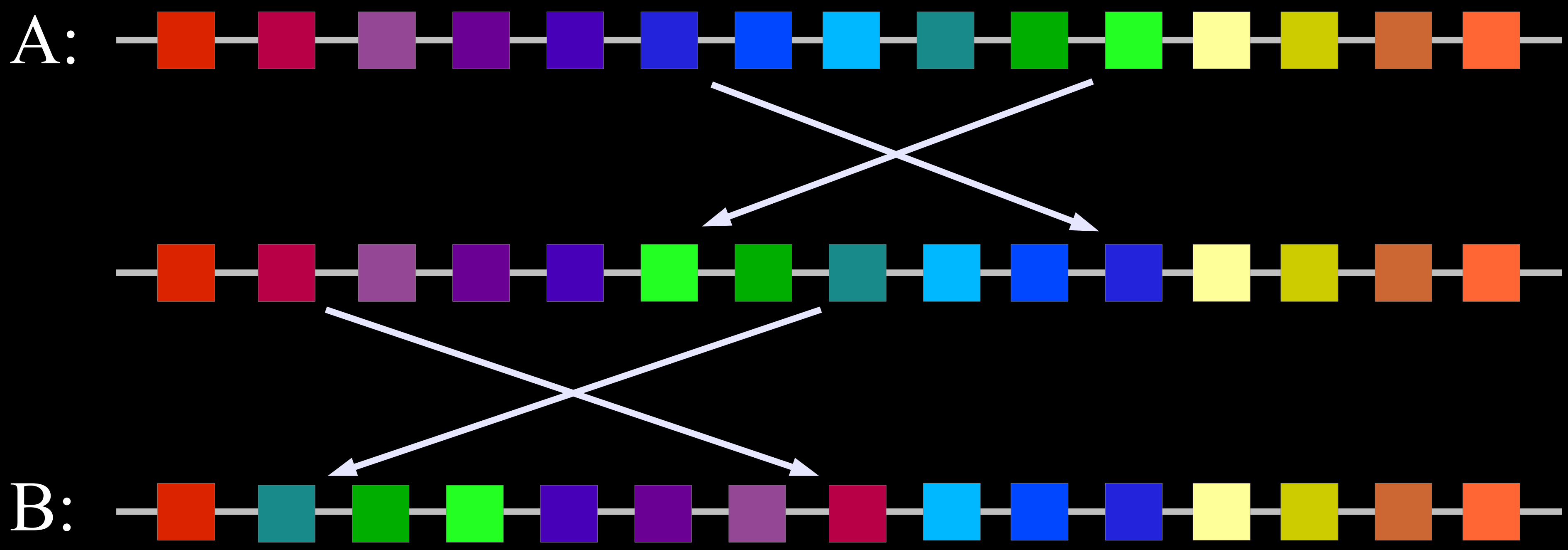


Rearrangement Scenario



Rearrangement Scenario

What is the distance between genome A and genome B?



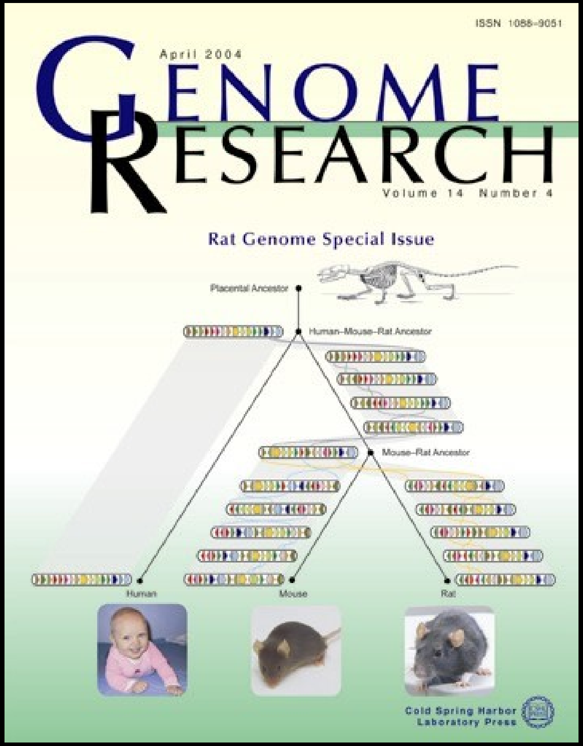
Whole Genome Analysis

- Pair-wise rearrangement distances
  - species tree reconstruction
  - gene homology inference

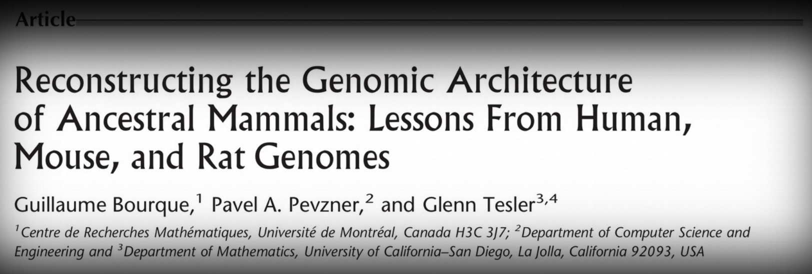
- Ancestral Reconstruction

Whole Genome Analysis

- Pair-wise rearrangement distances
  - species tree reconstruction
  - gene homology inference



- Ancestral Reconstruction



Article

Reconstructing the Genomic Architecture of Ancestral Mammals: Lessons From Human, Mouse, and Rat Genomes

Guillaume Bourque,[1] Pavel A. Pevzner,[2] and Glenn Tesler[3,4]

[1] Centre de Recherches Mathématiques, Université de Montréal, Canada H3C 3J7; [2] Department of Computer Science and Engineering and [3] Department of Mathematics, University of California–San Diego, La Jolla, California 92093, USA

Recent analysis of genome rearrangements in human and mouse genomes revealed evidence for more

Genome Research 2004

Chicken Special/Letter

Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages

Guillaume Bourque,[1,5] Evgeny M. Zdobnov,[2] Peer Bork,[2] Pavel A. Pevzner,[3] and Glenn Tesler[4]

[1] Genome Institute of Singapore, Singapore 138672, Republic of Singapore; [2] European Molecular Biology Laboratory, 69117 Heidelberg, Germany; [3] Department of Computer Science and Engineering, [4] Department of Mathematics, University of California, San Diego, La Jolla, California 92093, USA

Genome Research 2007

Resource

Breakpoint graphs and ancestral genome reconstructions

Max A. Alekseyev and Pavel A. Pevzner[1]

Department of Computer Science and Engineering, University of California at San Diego, La Jolla, California 92093-0404, USA

Recently completed whole-genome sequencing projects marked the transition from gene-based phylogenetic studies to phylogenomics analysis of entire genomes. We developed an algorithm MGRA for reconstructing ancestral genomes and used it to study the rearrangement history of seven mammalian genomes

Genome Research 2009

Whole Genome Analysis

“Initial sequencing and comparative analysis of the mouse genome”

- Nature 2002

“Genome sequence of the Brown Norway rat yields insights into mammalian evolution”

- Nature 2004

“Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution”

- Nature 2004

Conclusions:

- X chromosomes are scrambled in rodents but not humans (since common ancestor)
  - human X is the ancestral order
- rodent gene orders evolve faster (3x) than human and chicken lineages
- breakpoint reuse
- few translocations between human and chicken

Whole Genome Analysis

"Initial sequencing and comparative analysis of the mouse genome"

- Nature 2002

"Genome sequence of the Brown Norway rat yields insights into mammalian evolution"

- Nature 2004

"Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution"

- Nature 2004

all based solely on parsimony

Conclusions:

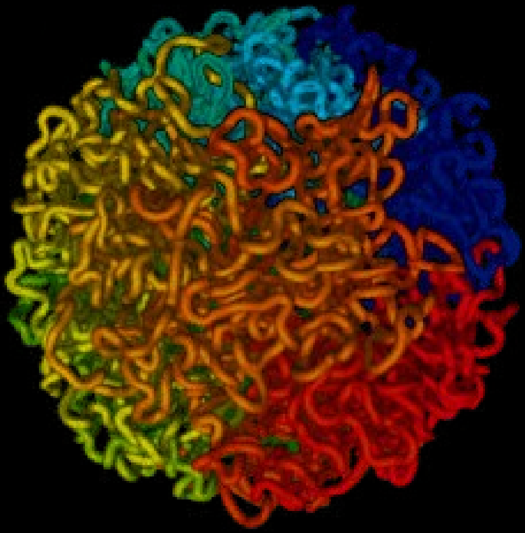
- X chromosomes are scrambled in rodents but not humans (since common ancestor)
  - human X is the ancestral order
- rodent gene orders evolve faster (3x) than human and chicken lineages
- breakpoint reuse
- few translocations between human and chicken

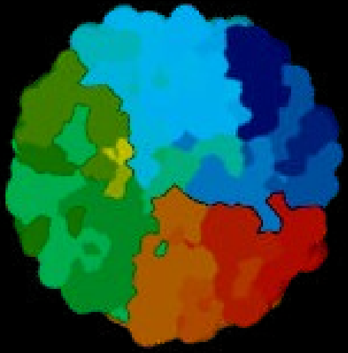
Addressing Limitations

- Limitation based on parsimony
  - uncertainty due to the LARGE search space

- Solution:

  introduce biological constraints





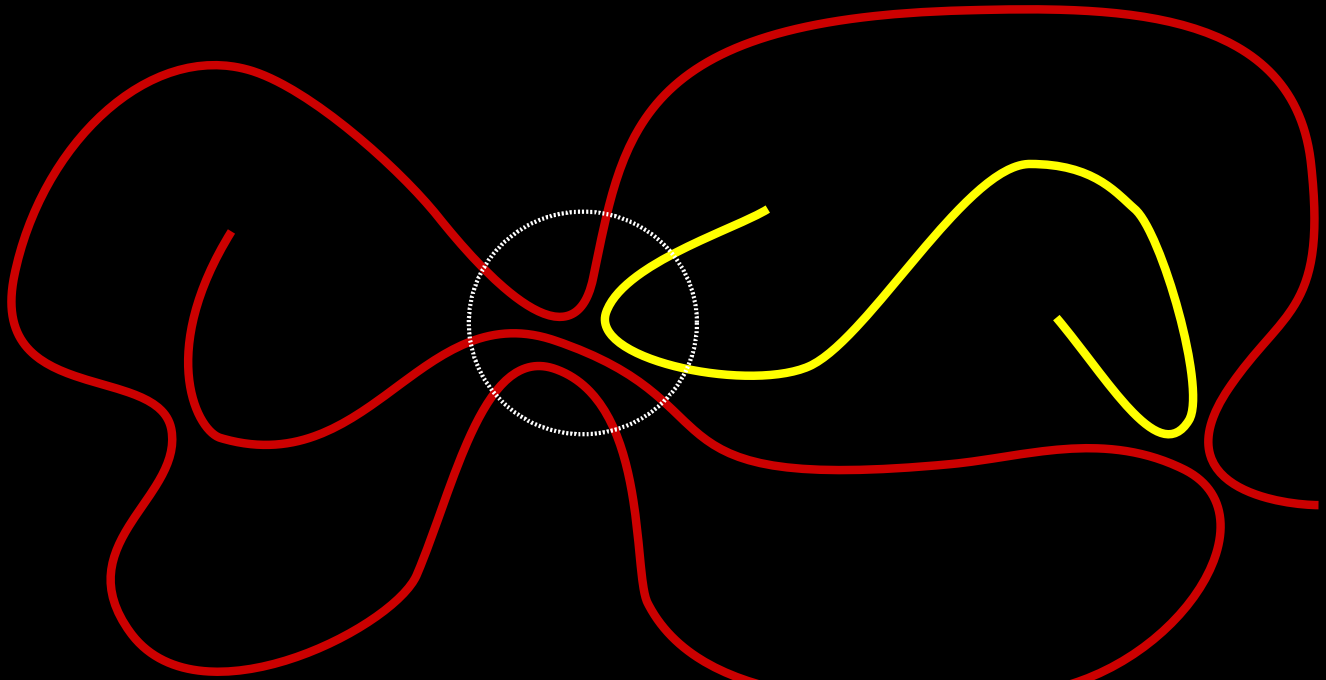
Lieberman-Aiden et al.

Hypothesis:

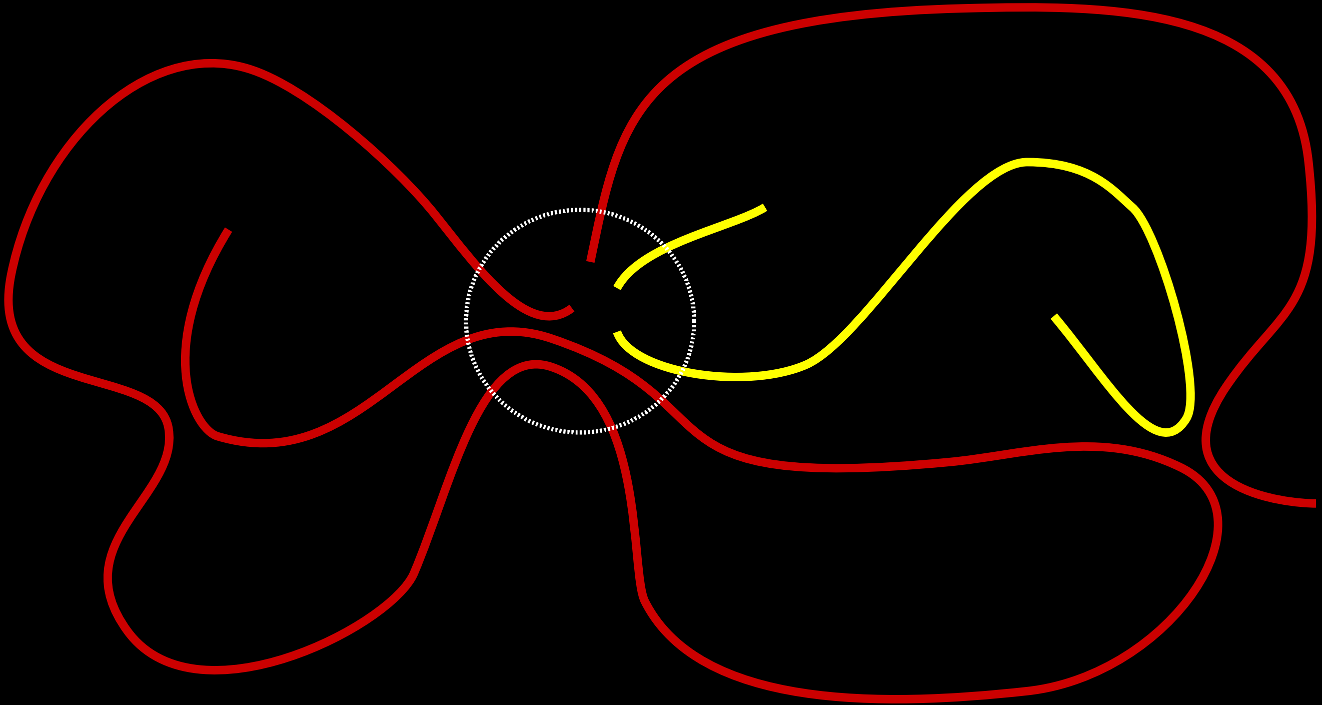
Rearrangement breakpoints are spatially close.

Véron, Lemaitre, Gautier, Lacroix and Sagot
"*Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny*"

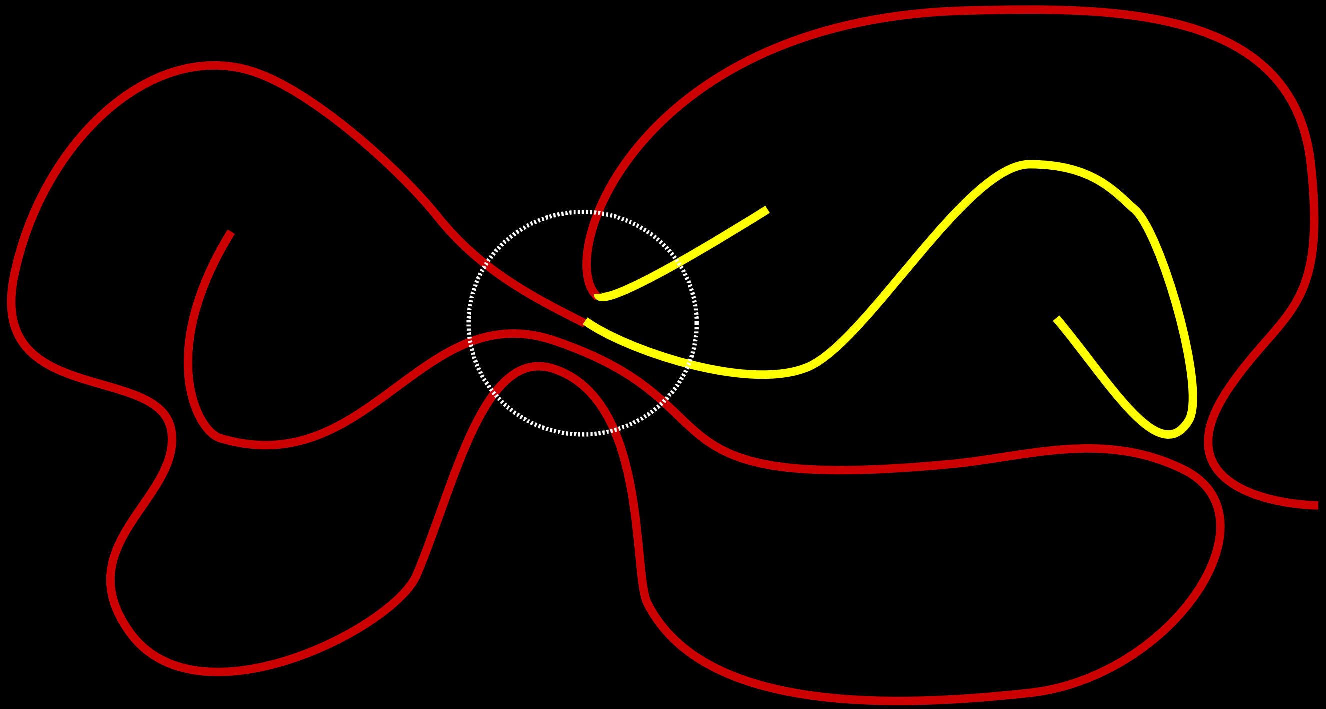
A “Local” Translocation



A “Local” Translocation

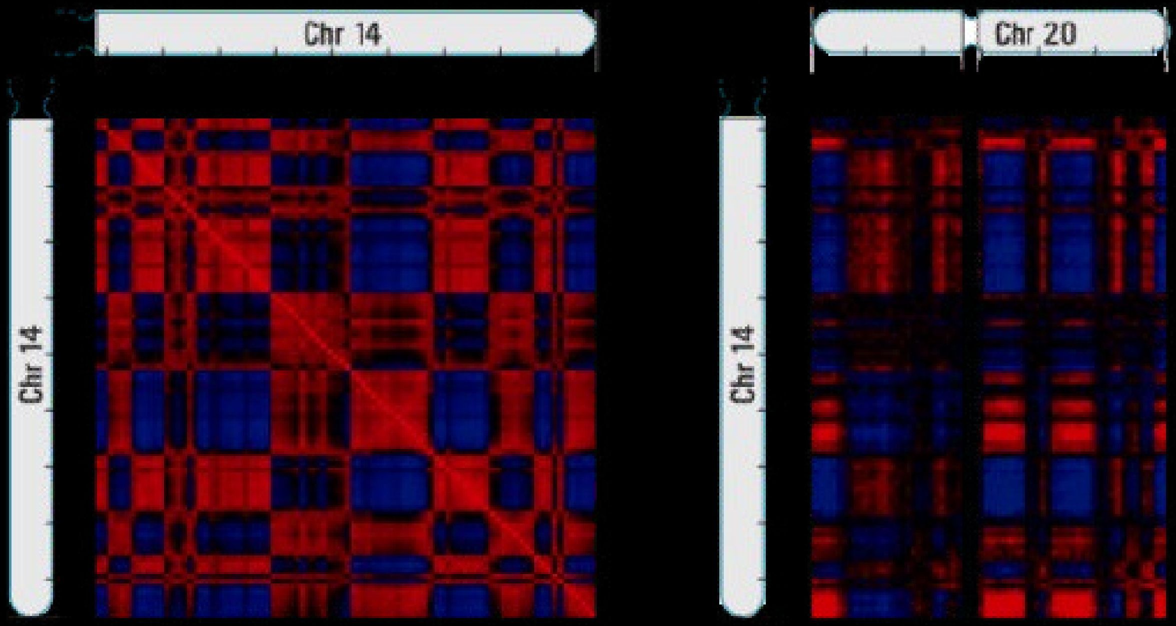


A “Local” Translocation

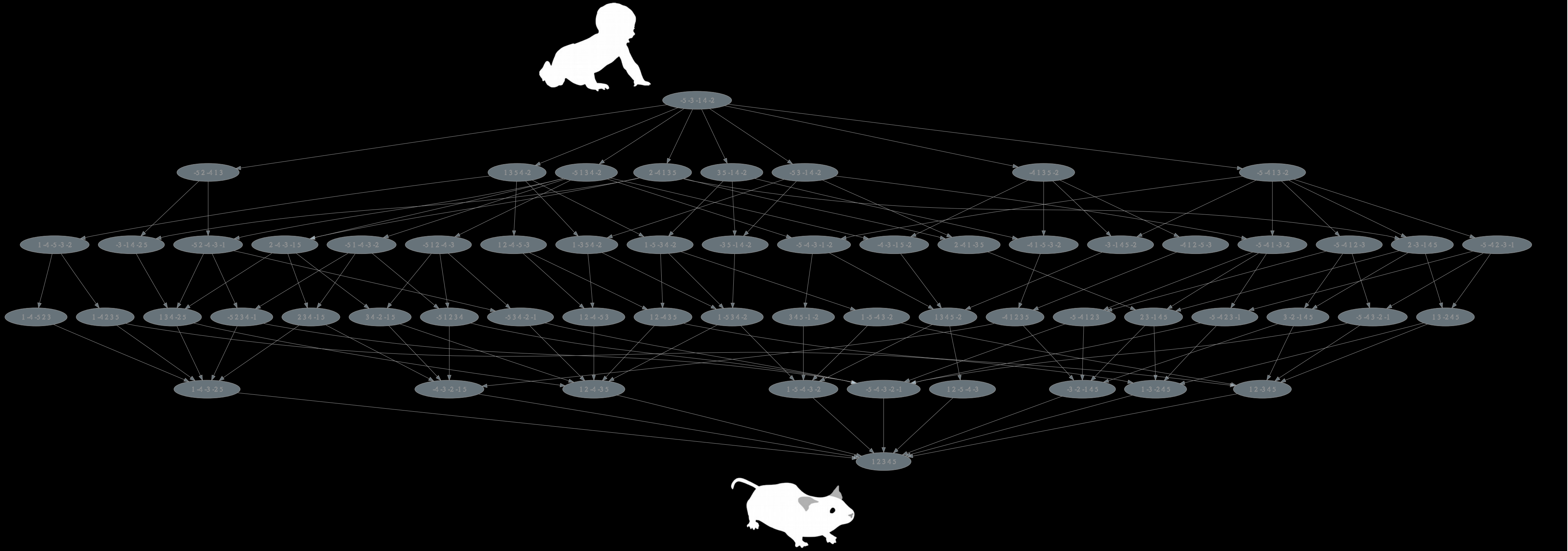


Hi-C Heatmaps

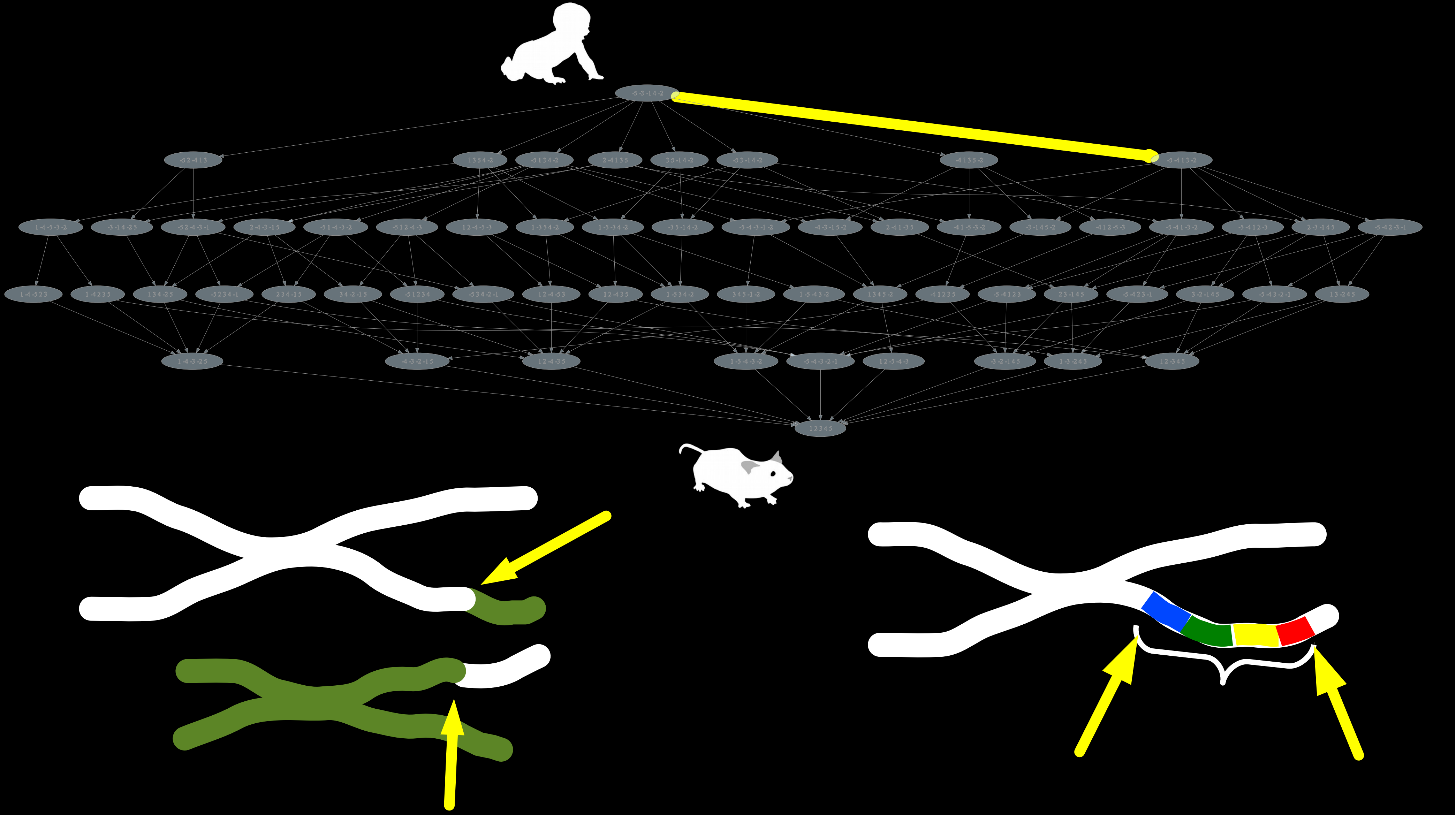
Each entry is proportional to spacial proximity.



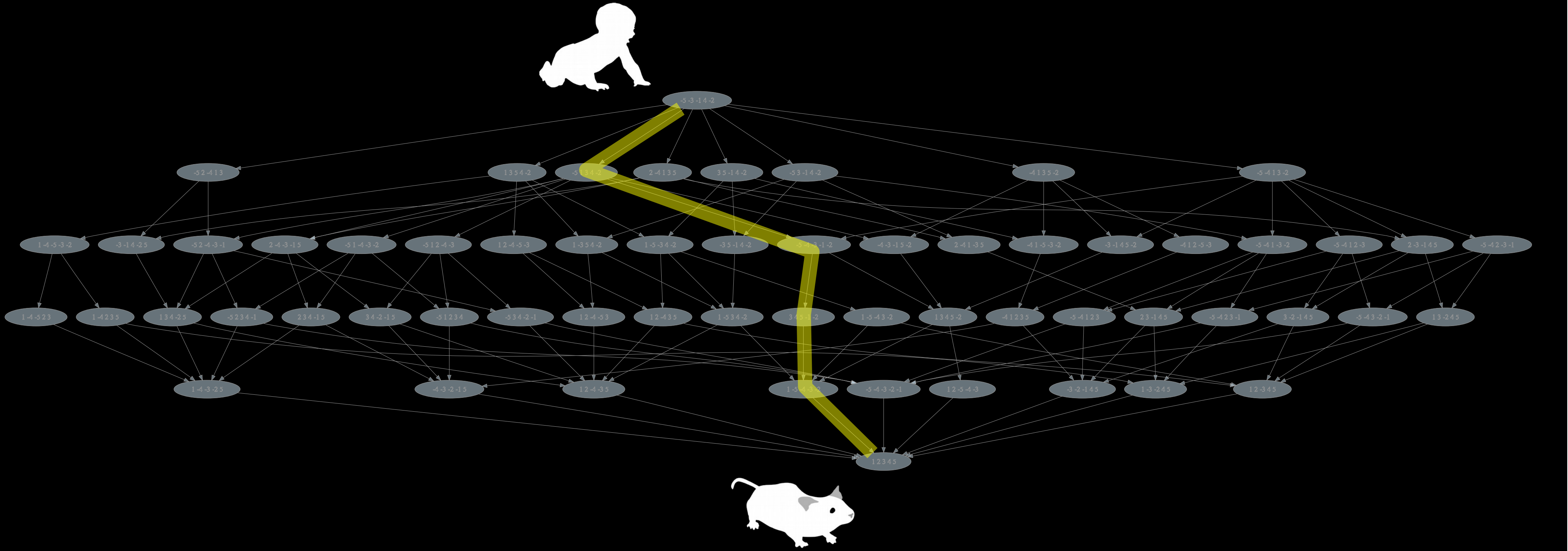
Evolutionary Context



Evolutionary Context



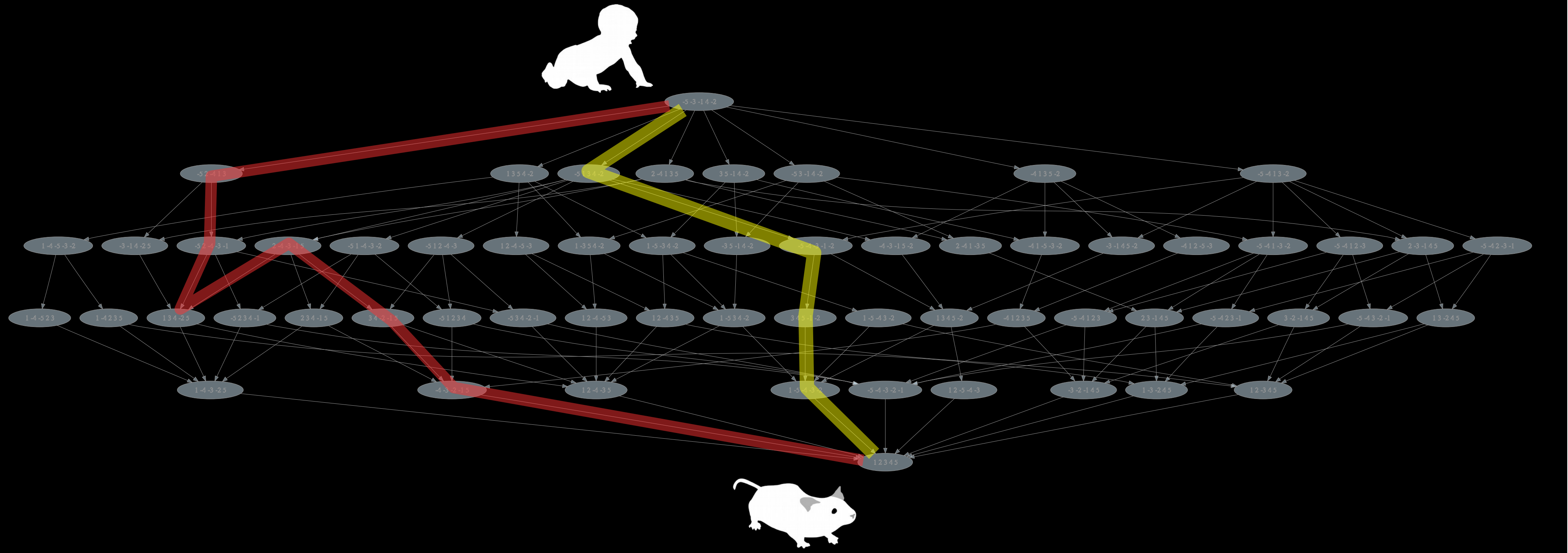
Evolutionary Context



Numerous *scenarios* between two genomes

- parsimonious

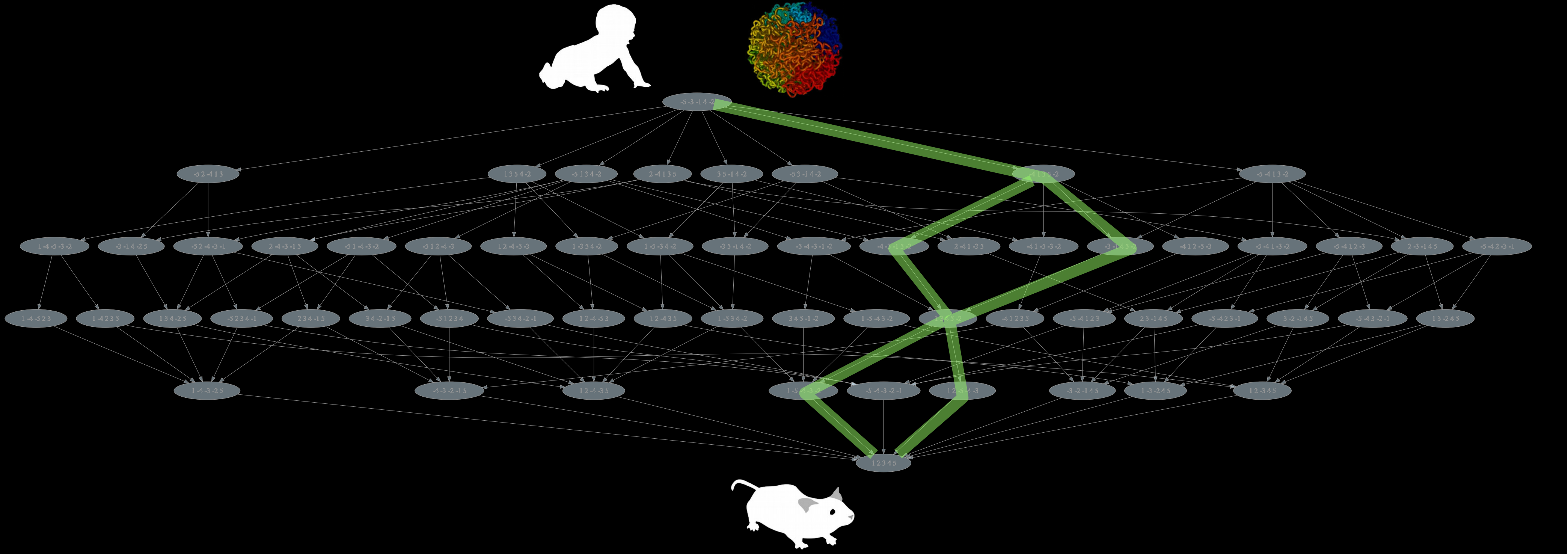
Evolutionary Context



Numerous *scenarios* between two genomes

- parsimonious / non-parsimonious

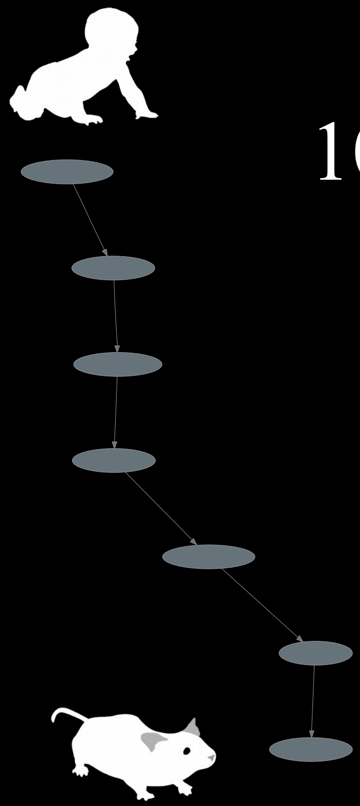
Evolutionary Context



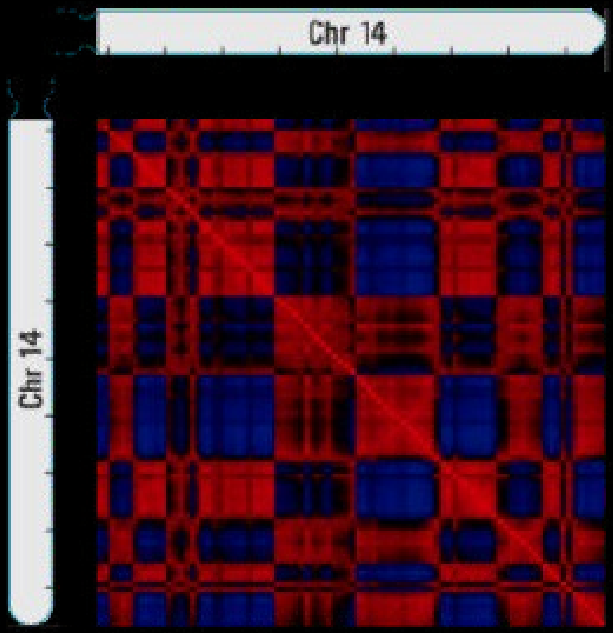
Numerous *scenarios* between two genomes

- parsimonious / non-parsimonious
- spatially local

Sampling Scenarios



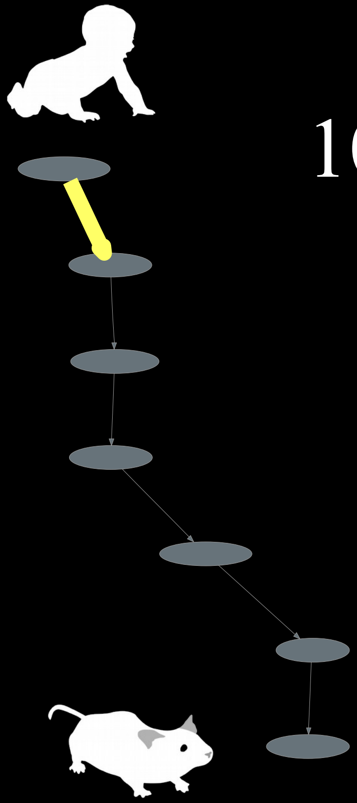
10,000 parsimonious scenario



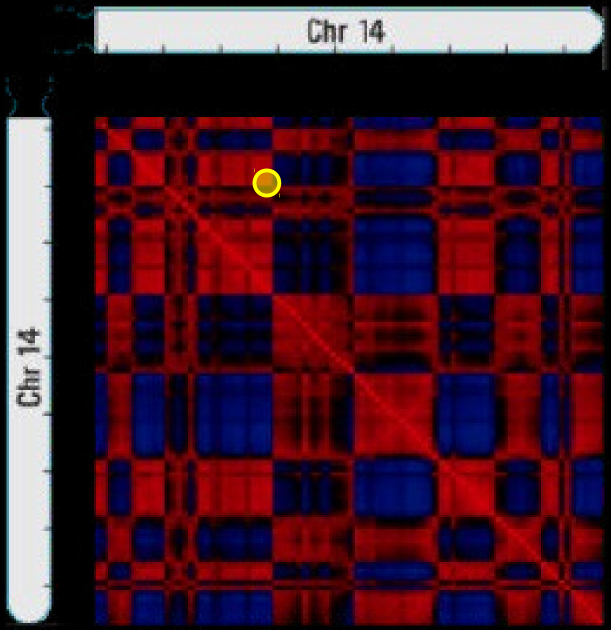
Chr 14

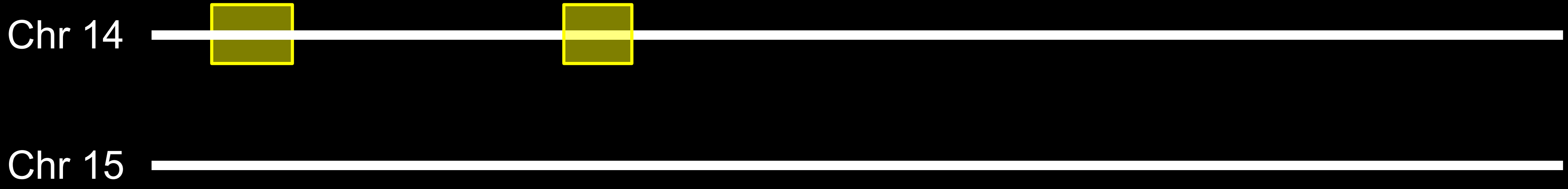
Chr 15

Sampling Scenarios

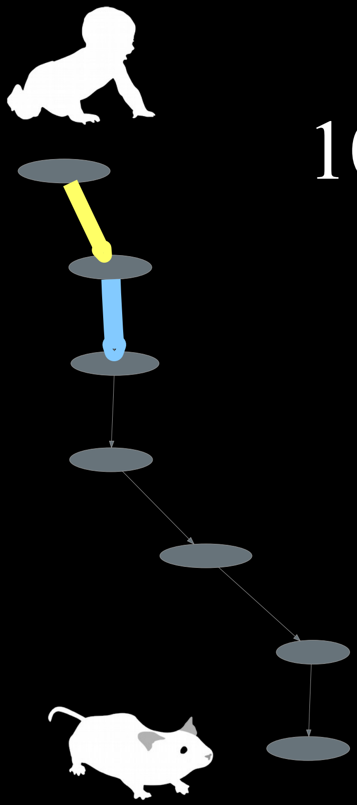


10,000 parsimonious scenario

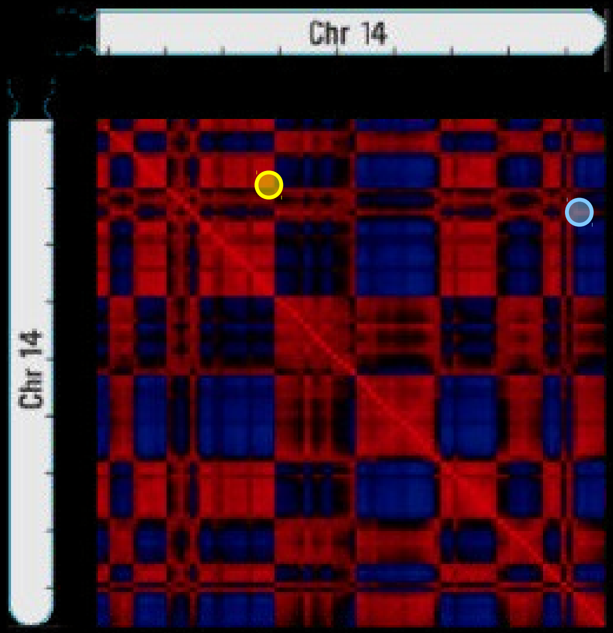


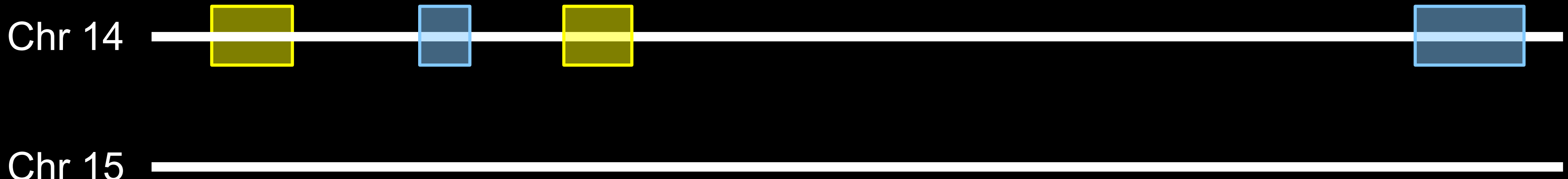


Sampling Scenarios

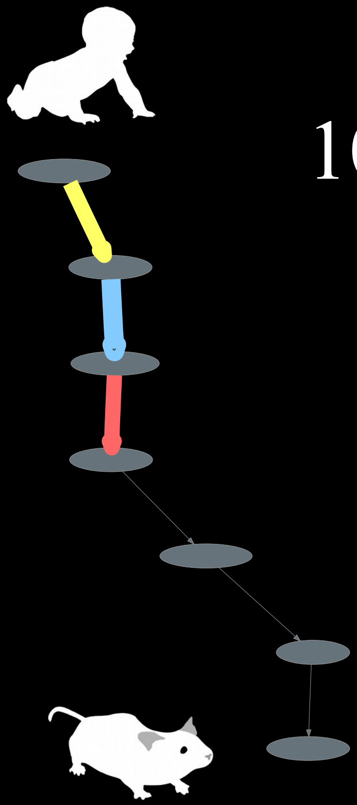


10,000 parsimonious scenario

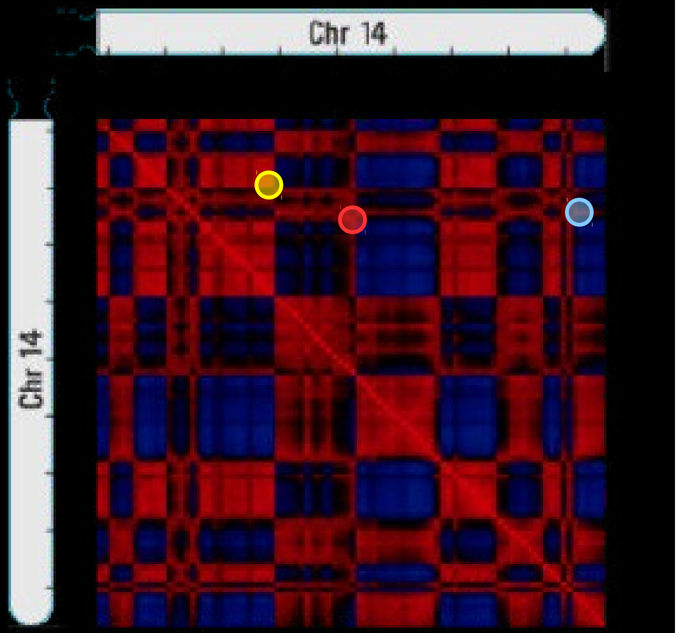


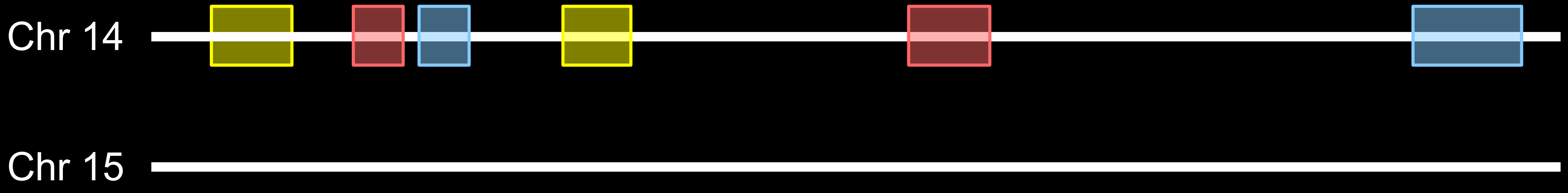


Sampling Scenarios

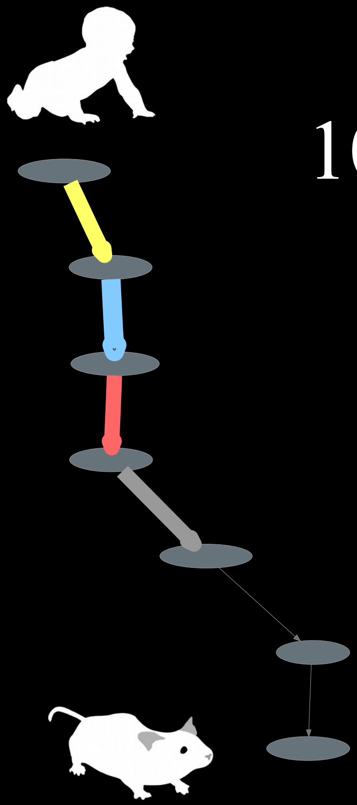


10,000 parsimonious scenario

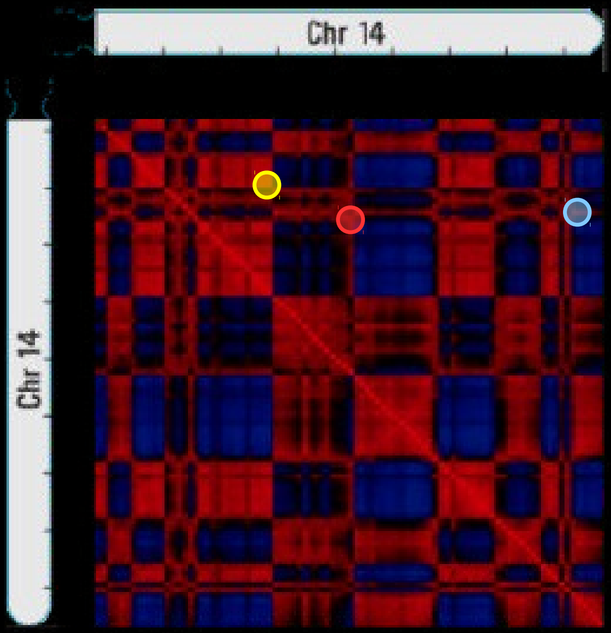


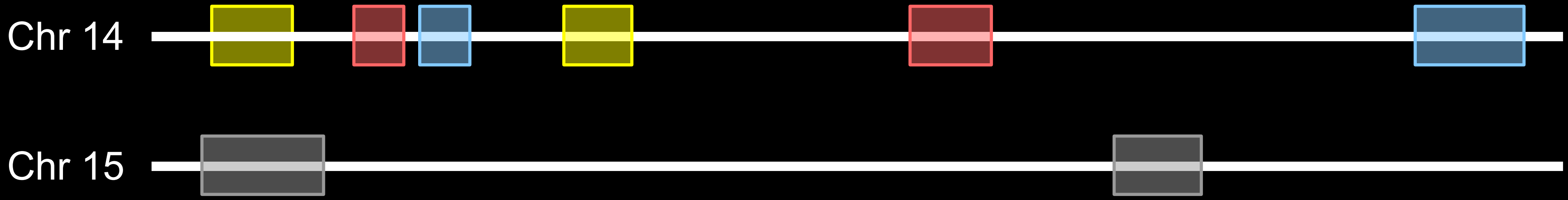


Sampling Scenarios

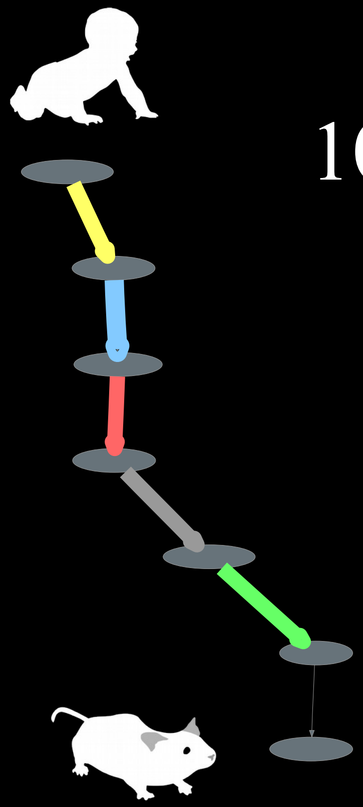


10,000 parsimonious scenario

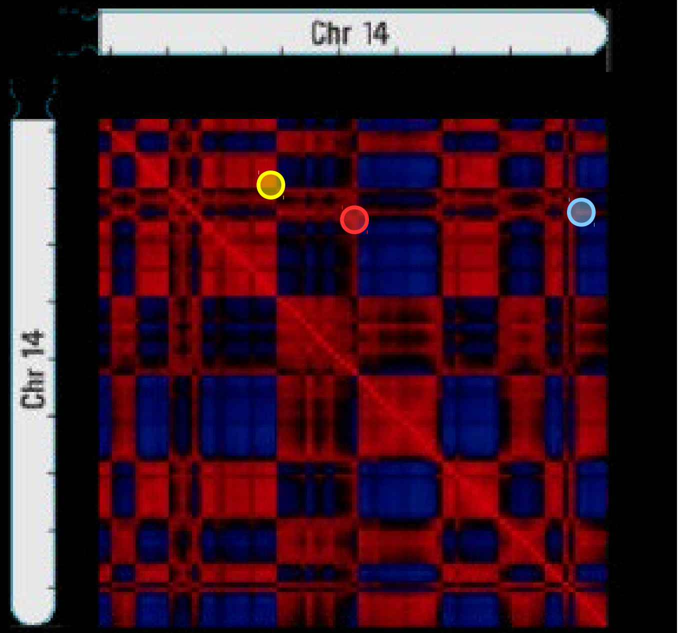


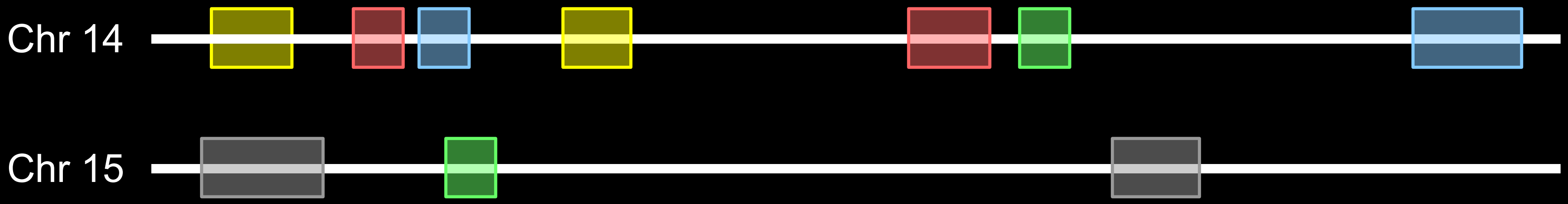


Sampling Scenarios



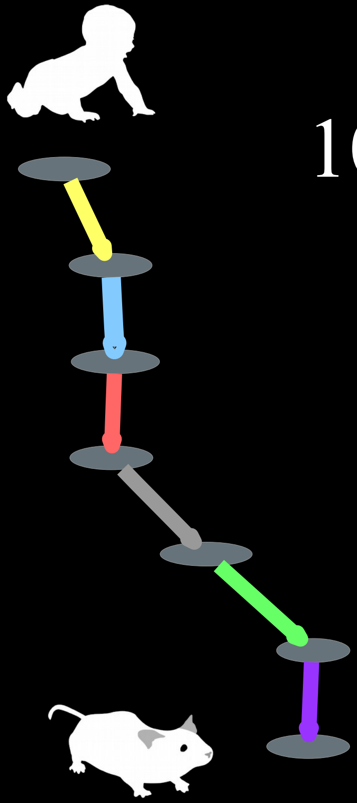
10,000 parsimonious scenario



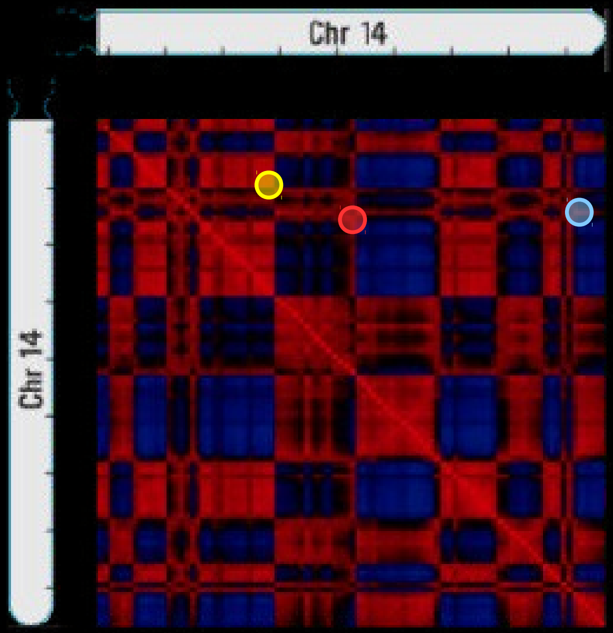

Chr 14

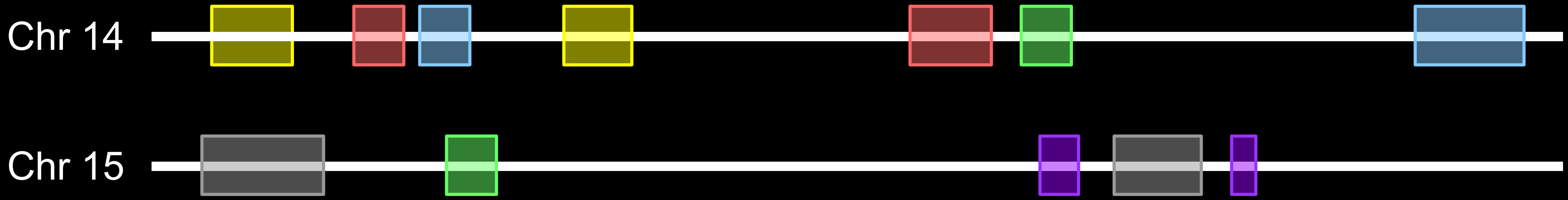
Chr 15

Sampling Scenarios

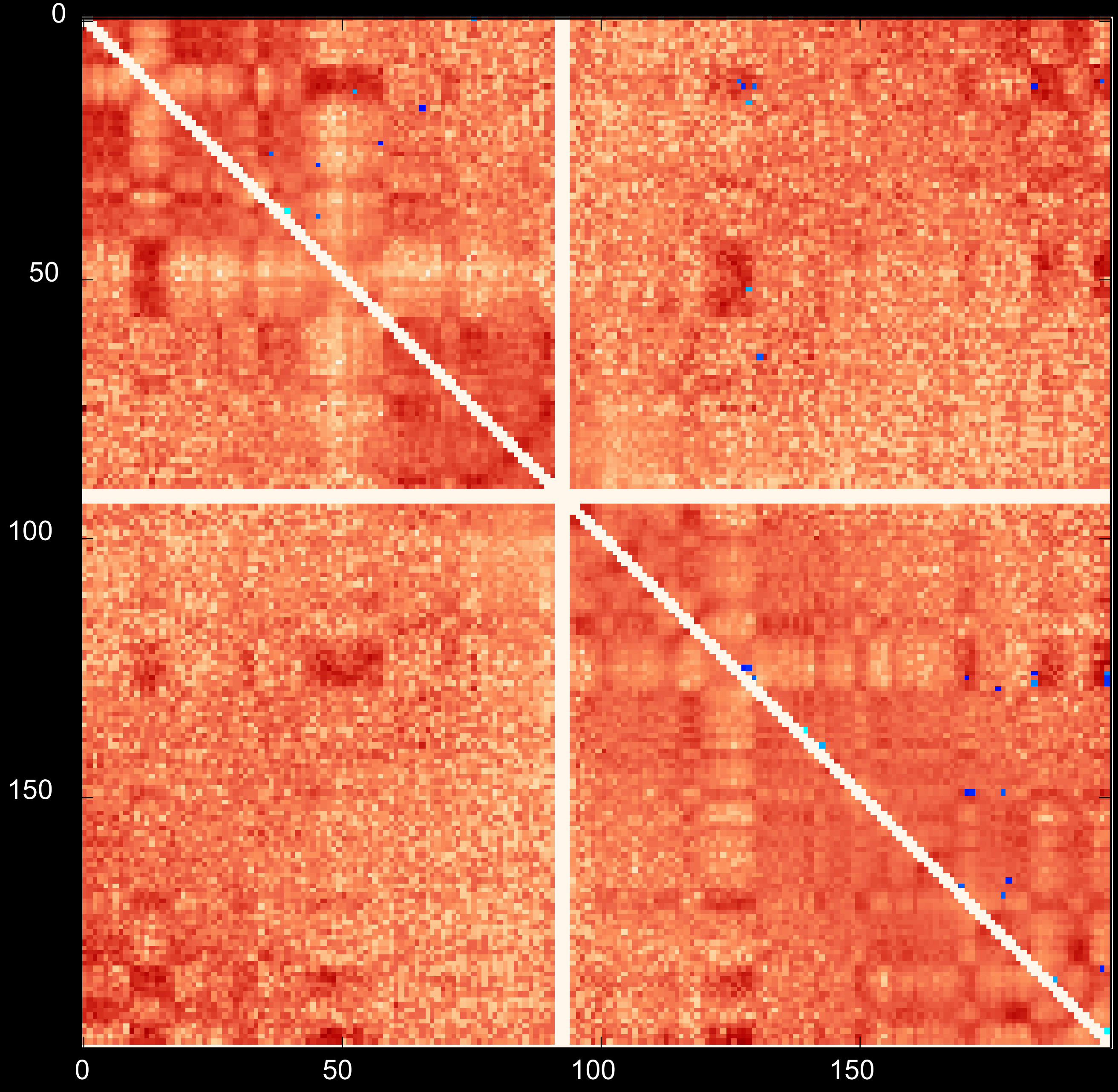


10,000 parsimonious scenario

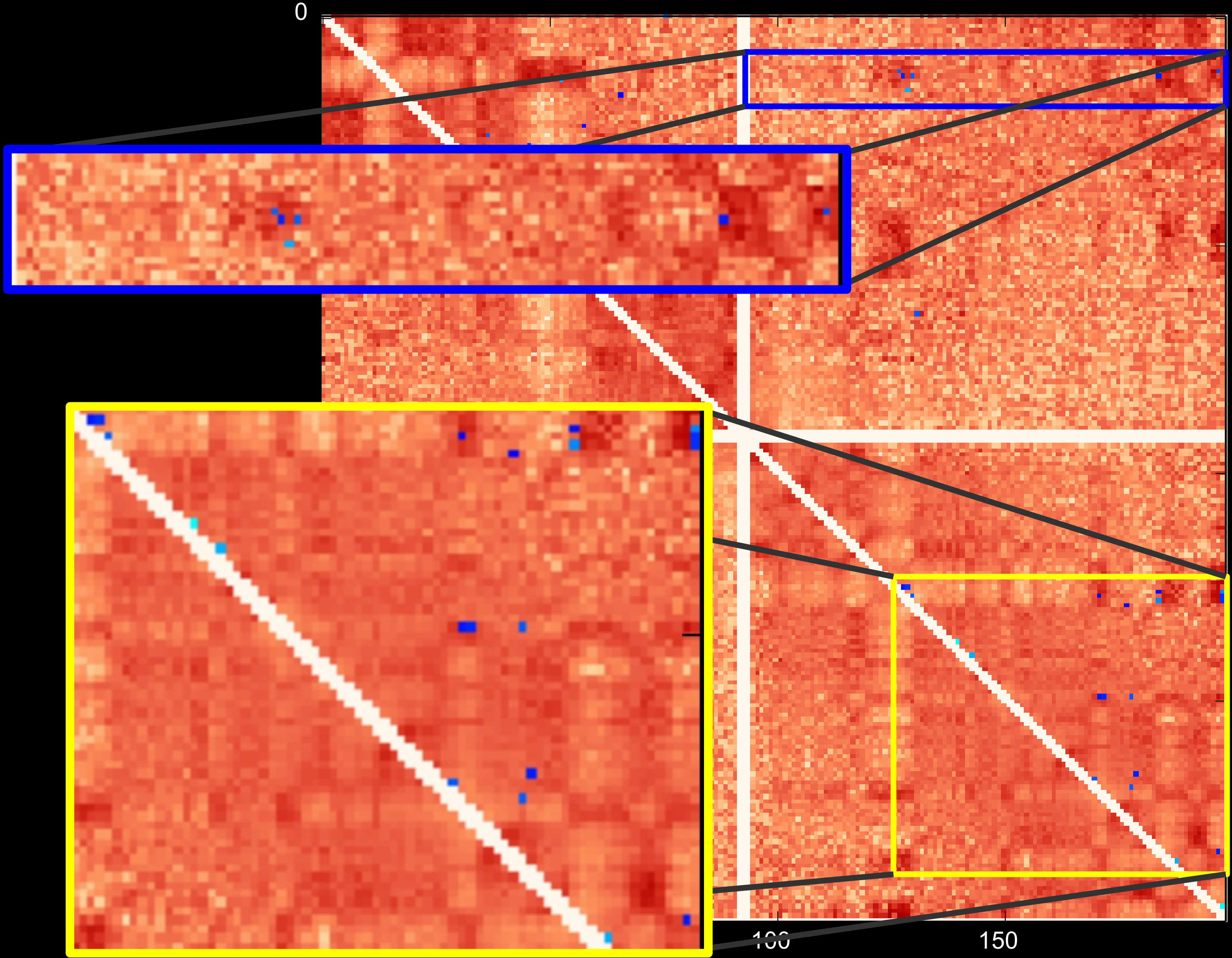




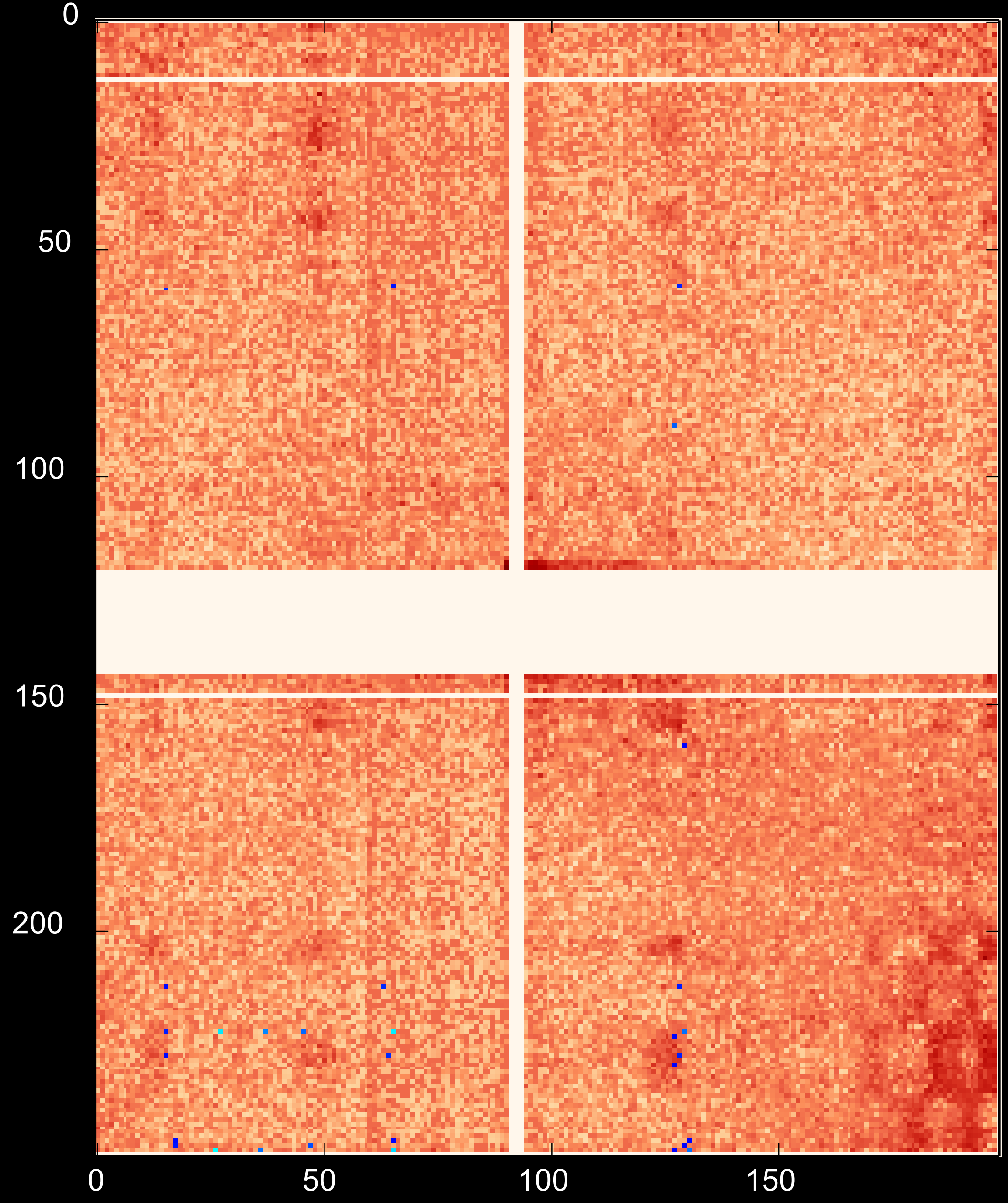
Sample from Chr 3



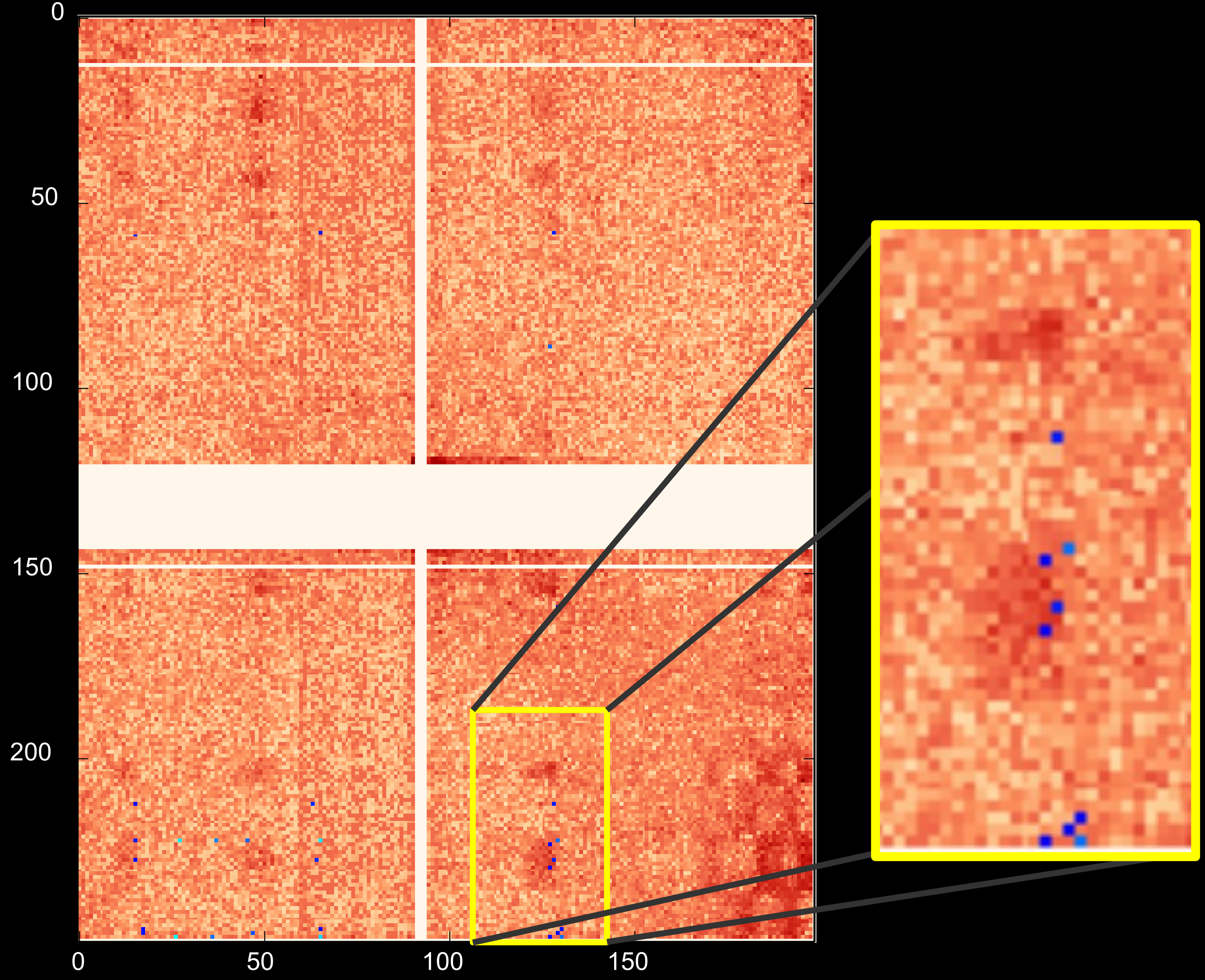
Sample from Chr 3



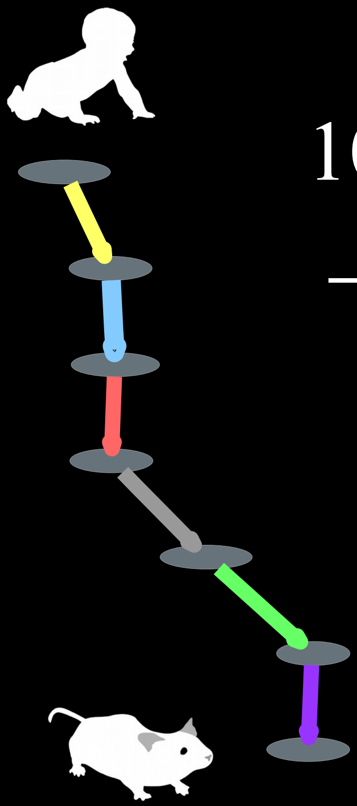
Sample from 1 vs. 3



Sample from 1 vs. 3

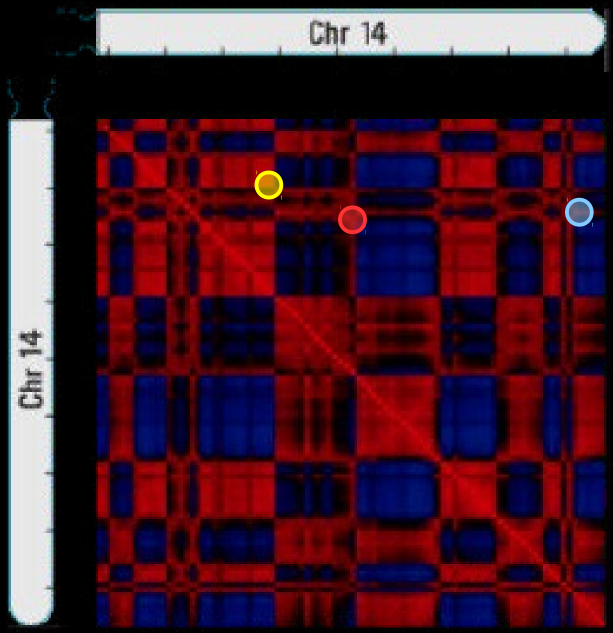


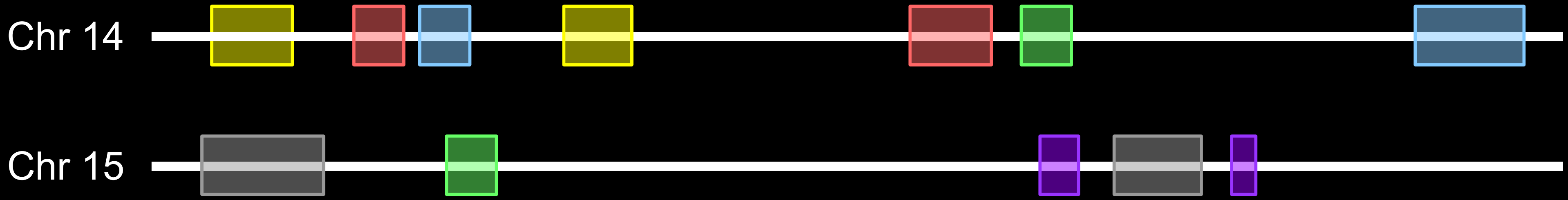
Sampling Scenarios



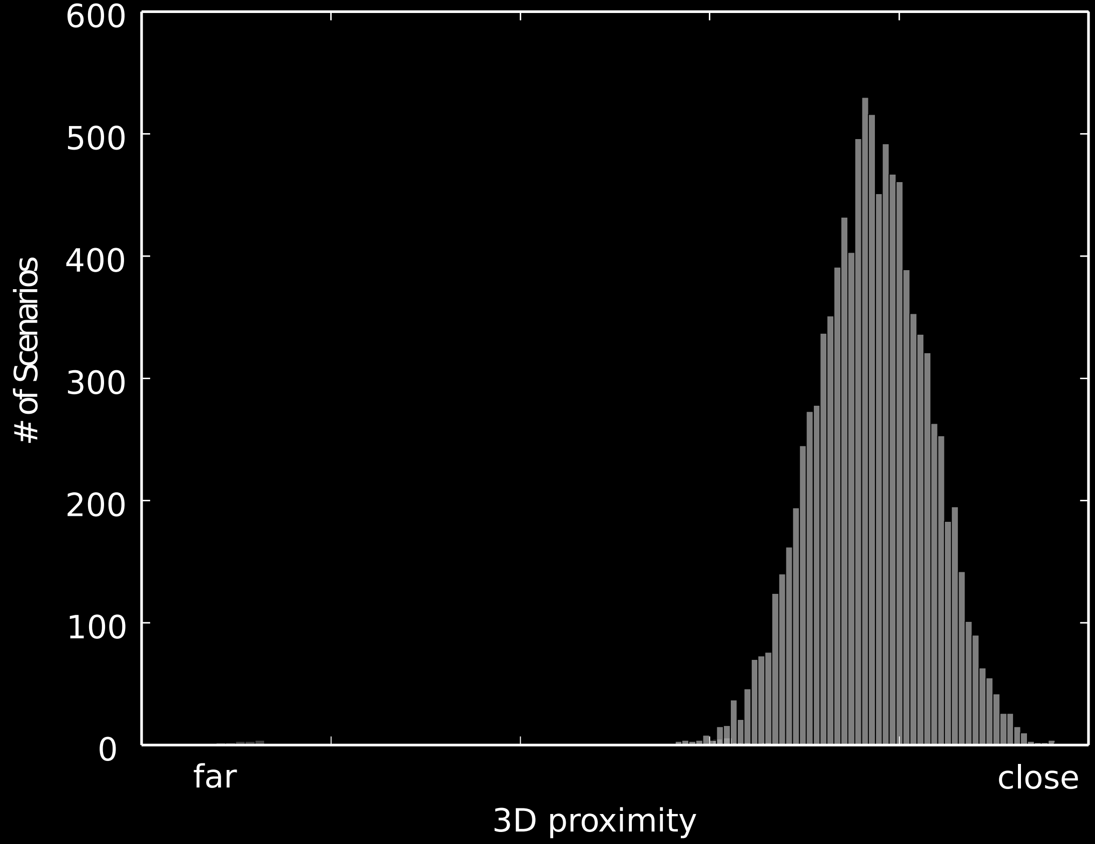
10,000 parsimonious scenario

- average over true breakpoints





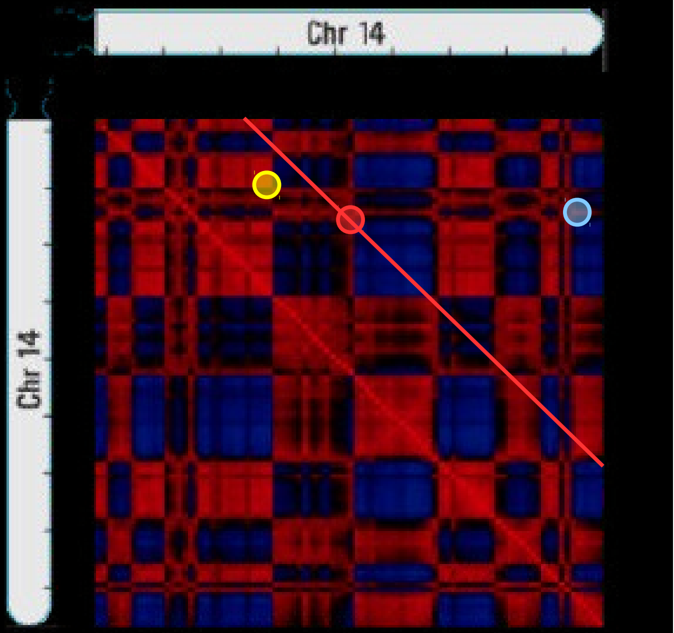
Human-Mouse Scenarios are Local

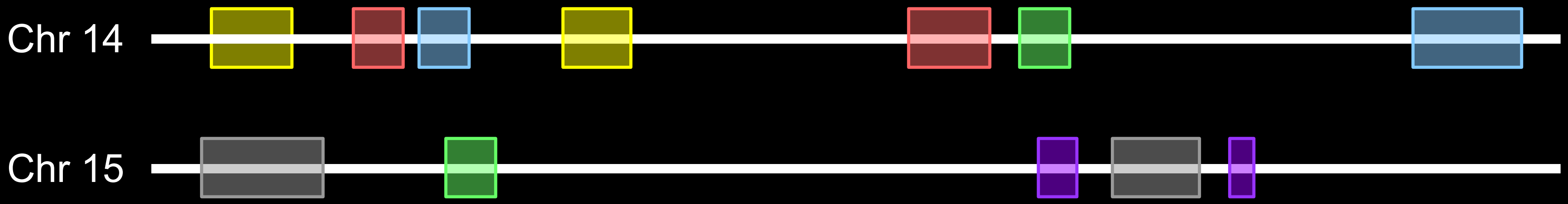


Sampling Scenarios

10,000 parsimonious scenario

- average over true breakpoints
- average over randomized breakpoints

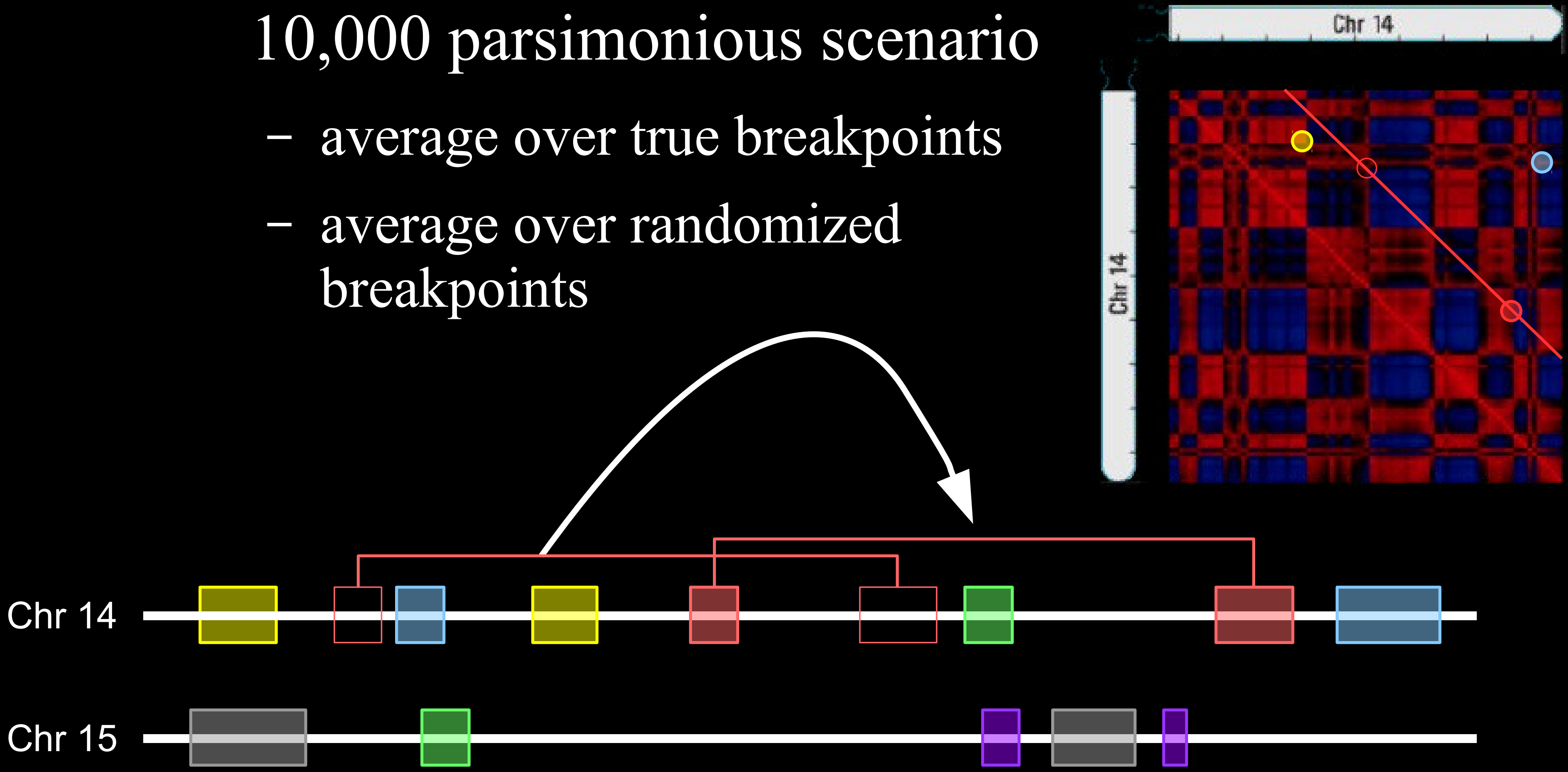




Sampling Scenarios

10,000 parsimonious scenario

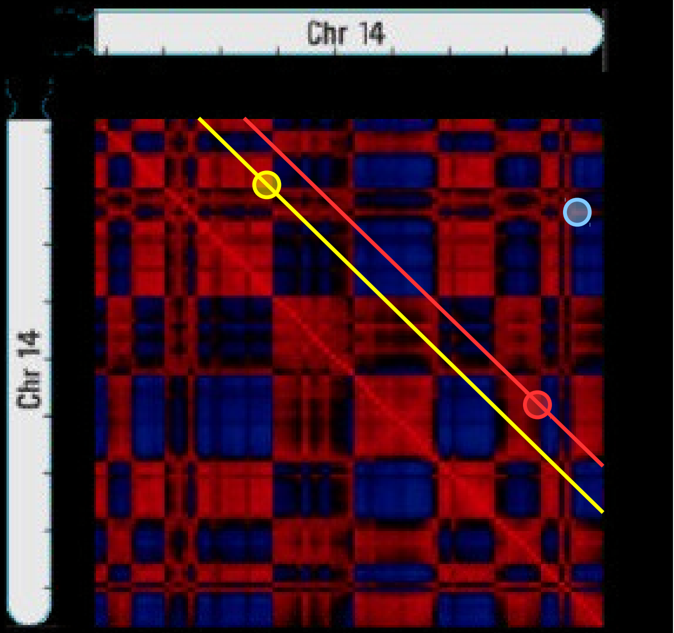
- average over true breakpoints
- average over randomized breakpoints

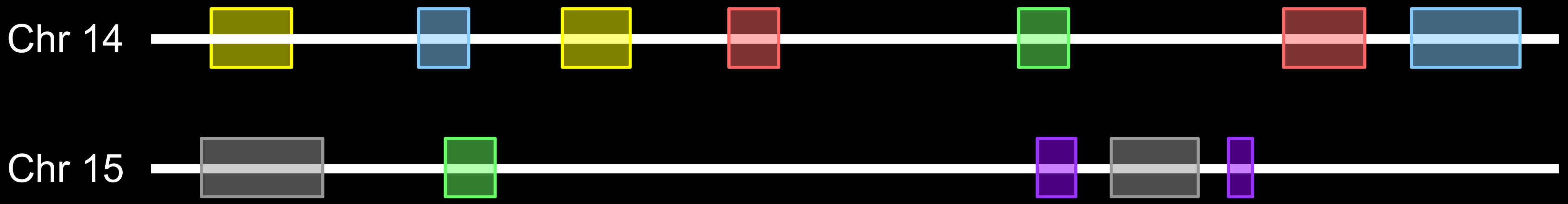


Sampling Scenarios

10,000 parsimonious scenario

- average over true breakpoints
- average over randomized breakpoints

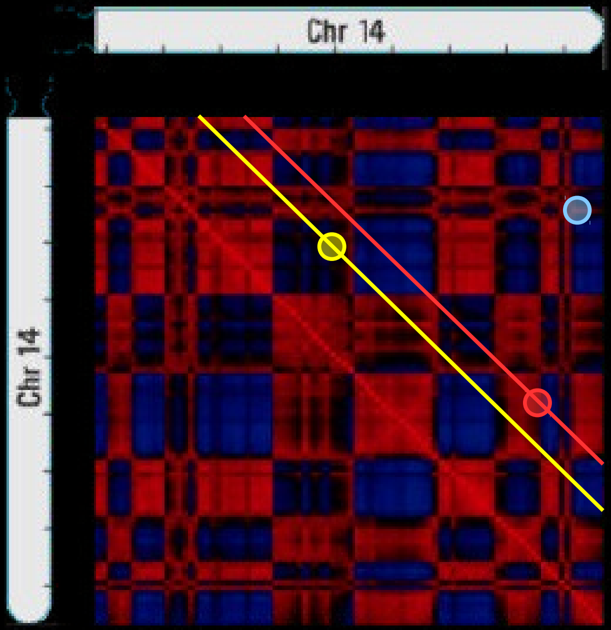


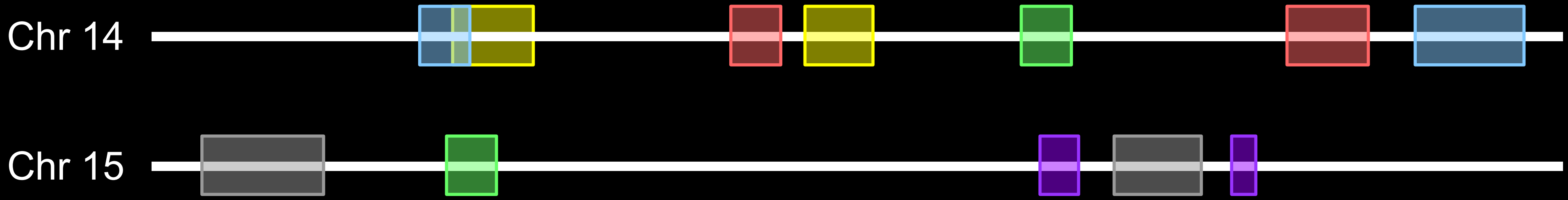


Sampling Scenarios

10,000 parsimonious scenario

- average over true breakpoints
- average over randomized breakpoints

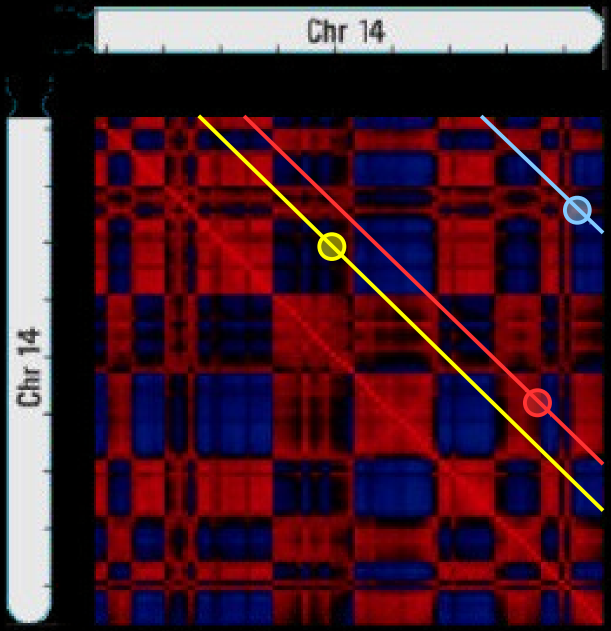


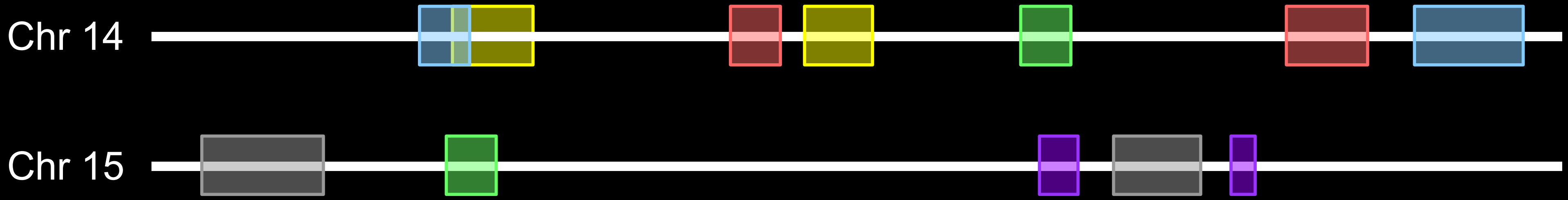


Sampling Scenarios

10,000 parsimonious scenario

- average over true breakpoints
- average over randomized breakpoints

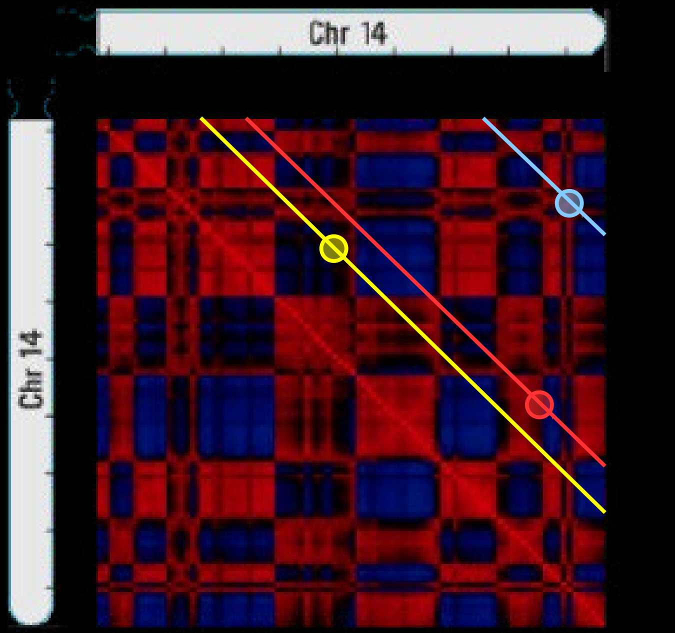


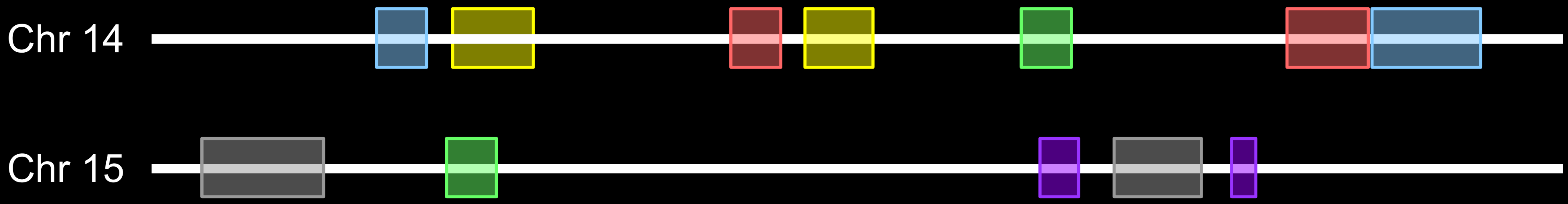


Sampling Scenarios

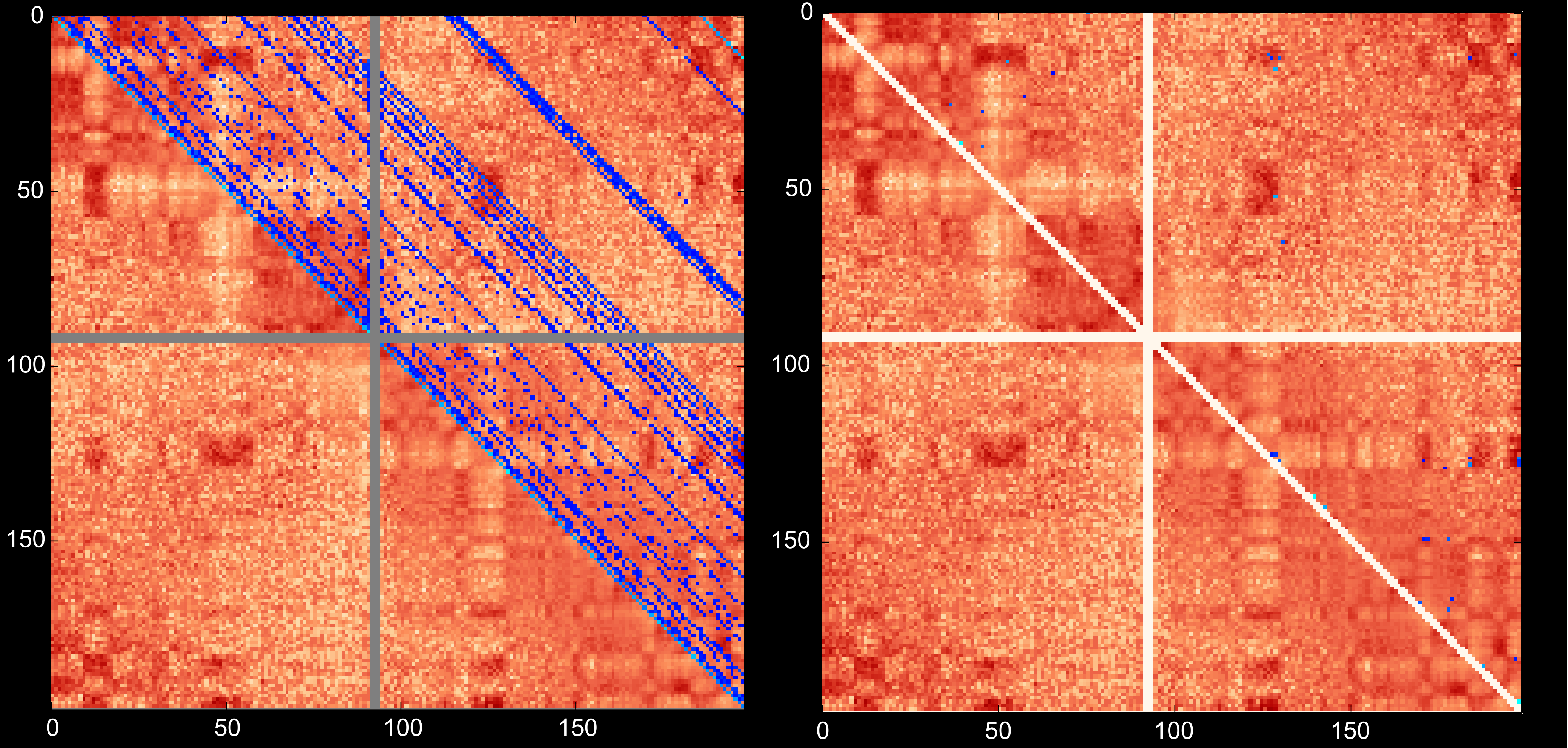
10,000 parsimonious scenario

- average over true breakpoints
- average over randomized breakpoints

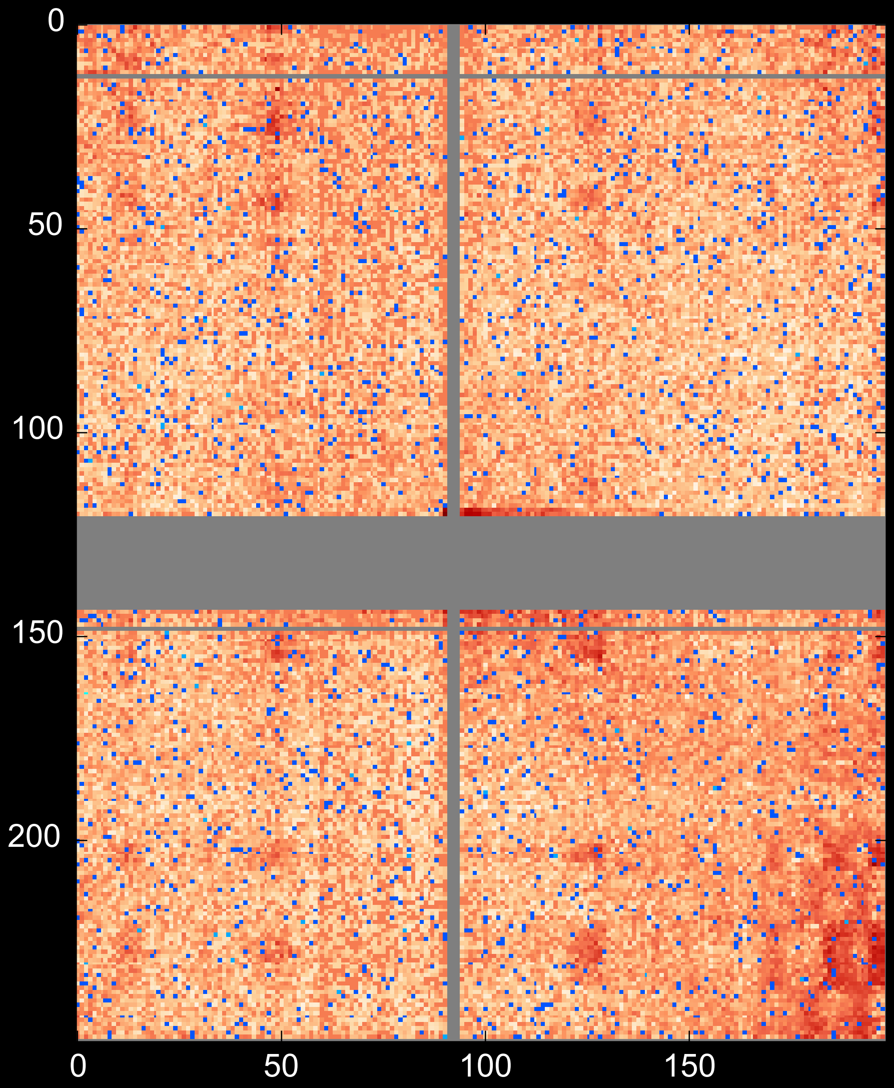


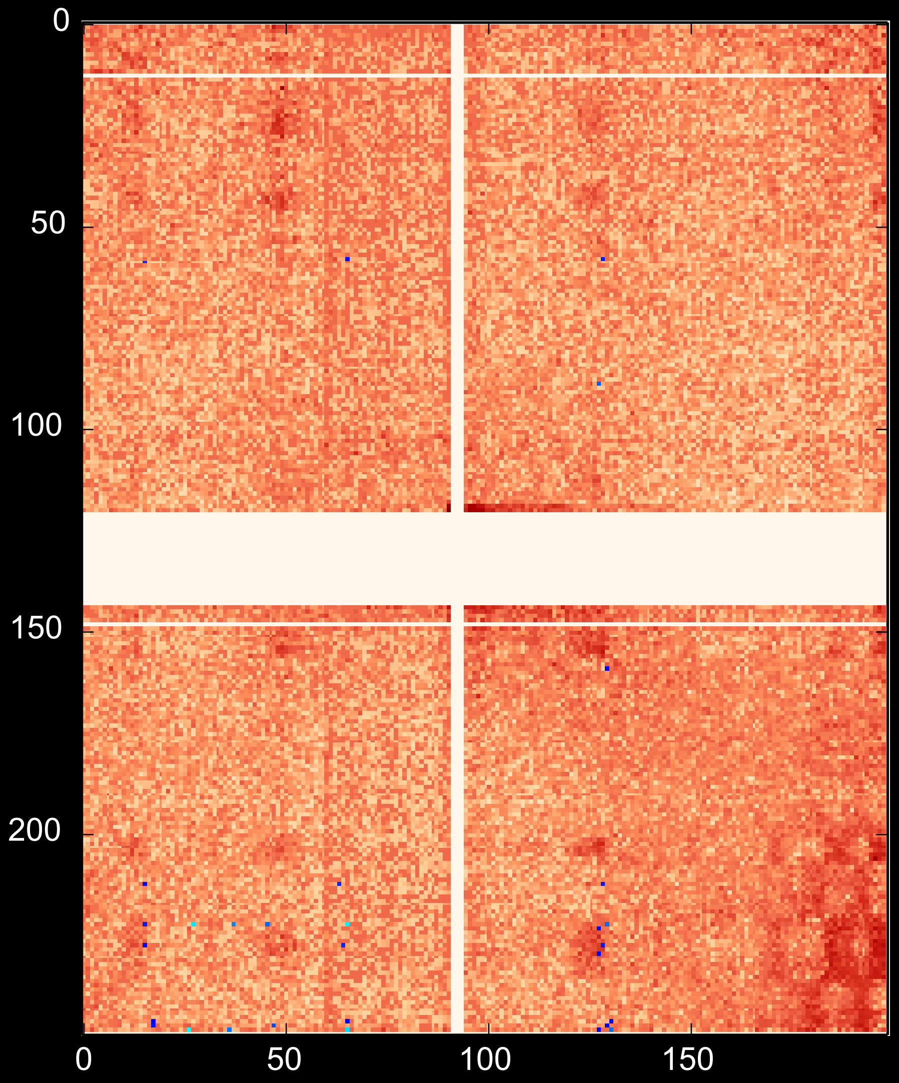


Sample from Chr 3

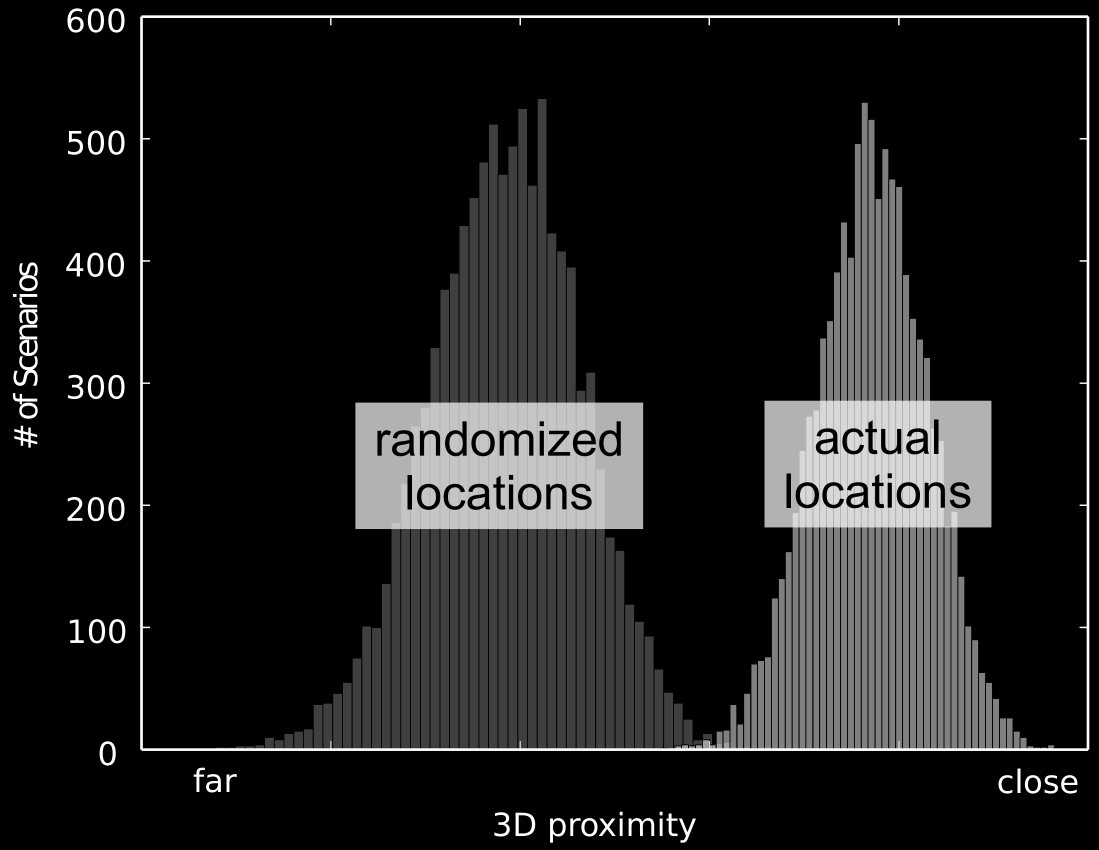


Sample from 1 vs. 3

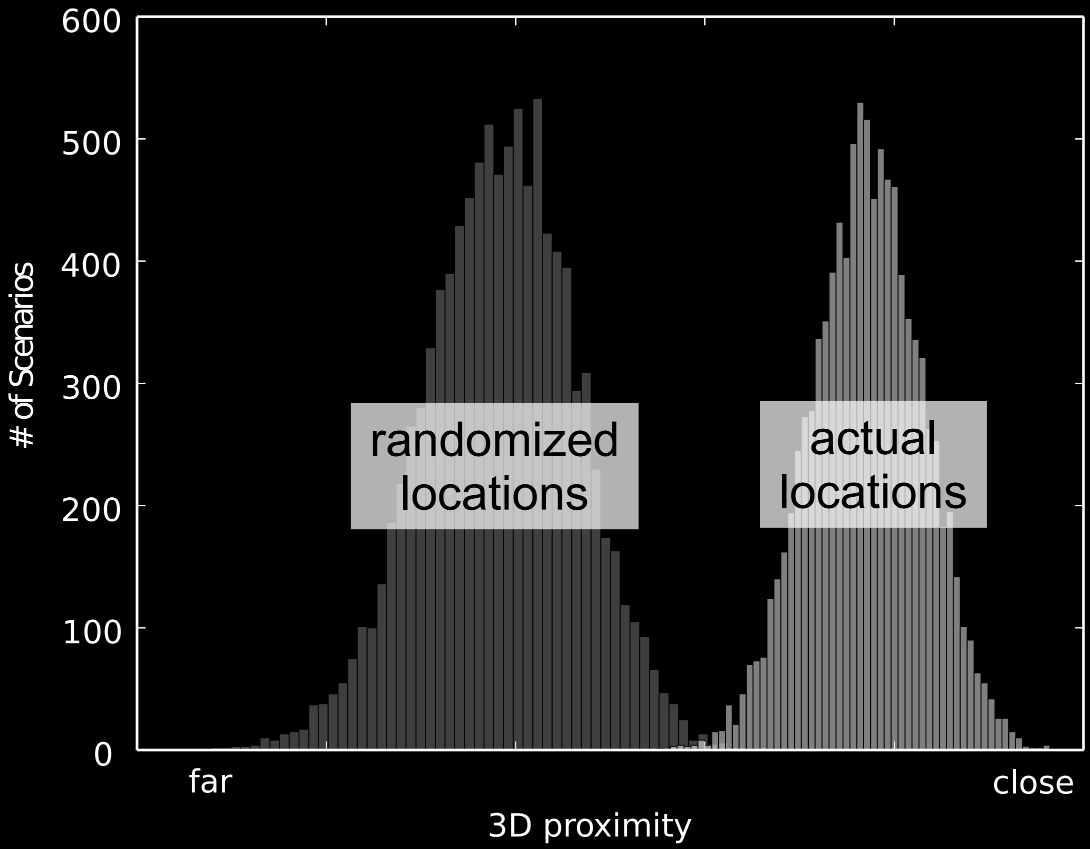




Human-Mouse Scenarios are Local

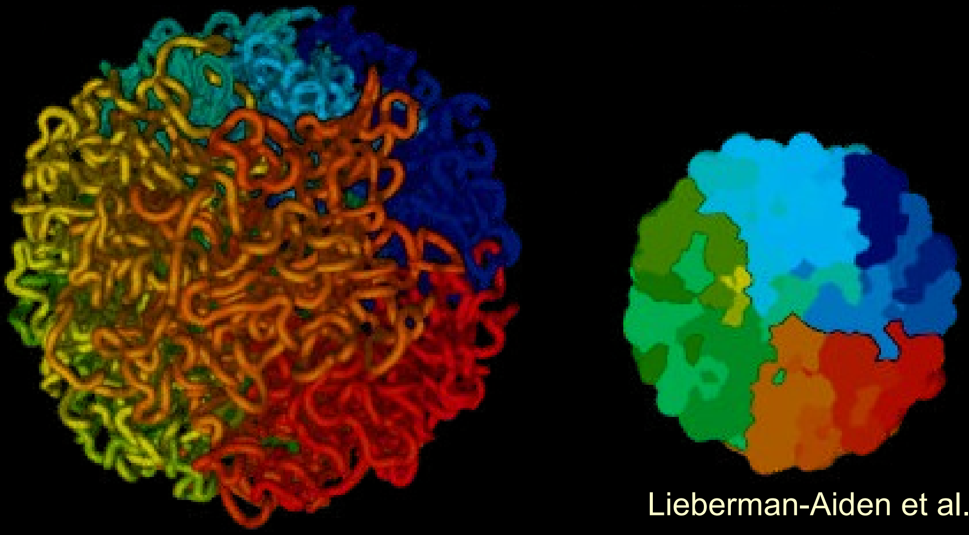


Human-Mouse Scenarios are Local



Intrachromosomal
and
interchromosomal
breakpoint pairs

- Evolutionarily conserved rearrangements are local
- Pattern exists despite only using human Hi-C



Lieberman-Aiden et al.

Question:

How do we find scenarios that are spatially close?

Double Cut and Join

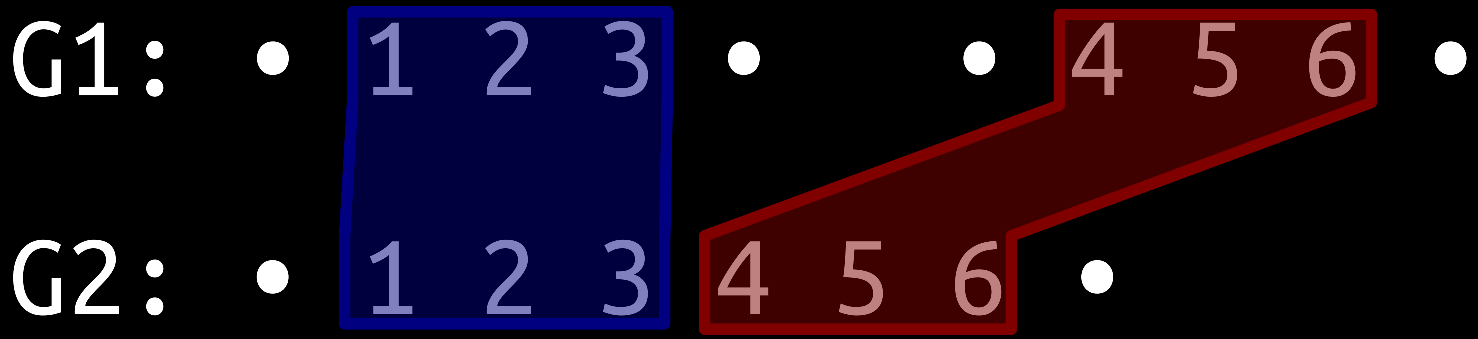
Cut 1 or 2 adjacencies

- IF 1 cut: create two telomeric adjacencies
- IF 2 cut: glue back 1 of 2 new ways

Double Cut and Join

Cut 1 or 2 adjacencies

- IF 1 cut: create two telomeric adjacencies
- IF 2 cut: glue back 1 of 2 new ways



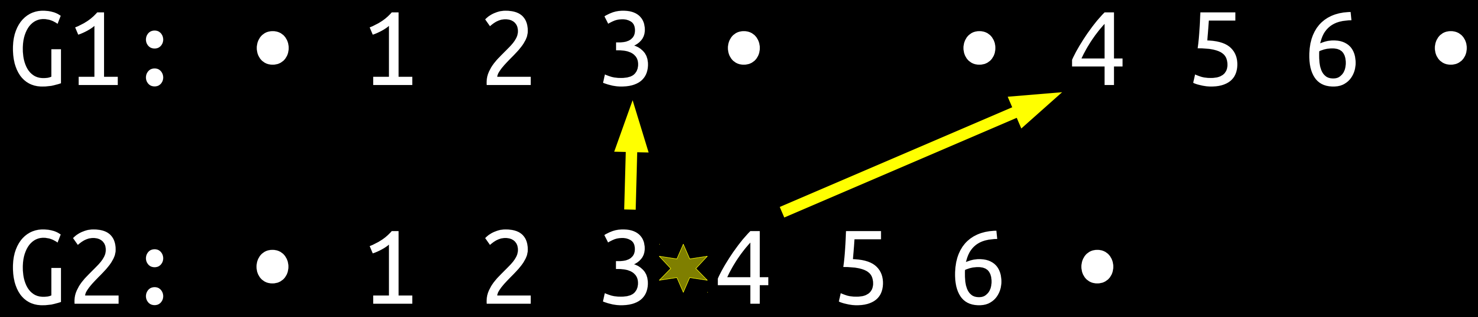
G1: • 1 2 3 • • 4 5 6 •

G2: • 1 2 3 4 5 6 •

Double Cut and Join

Cut 1 or 2 adjacencies

- IF **1 cut**: create two telomeric adjacencies
- IF 2 cut: glue back 1 of 2 new ways



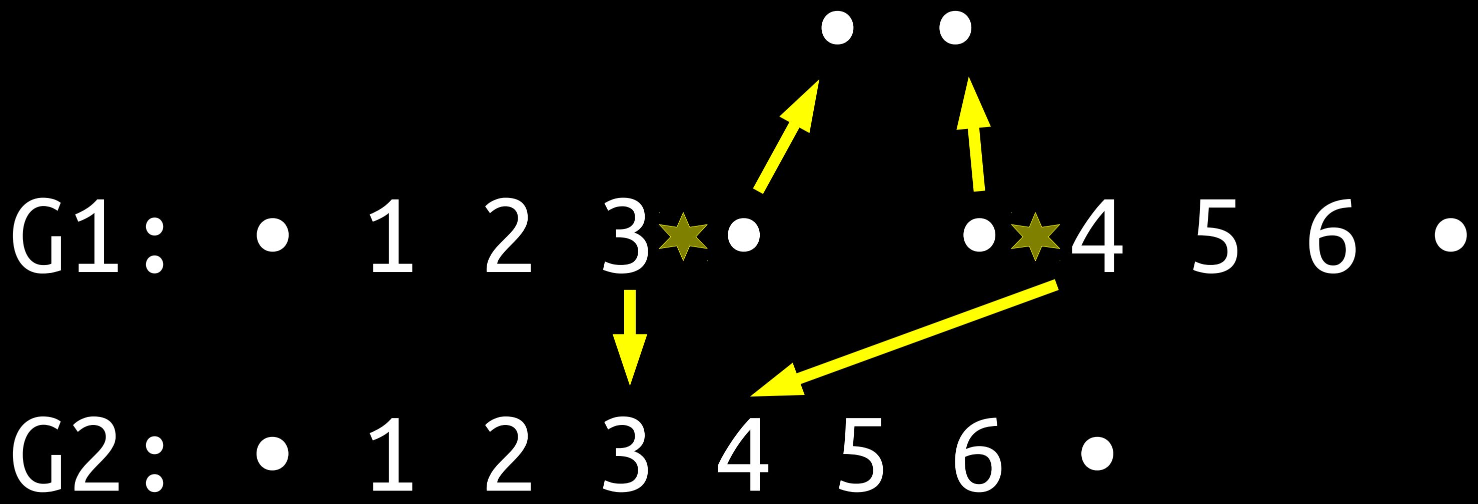
G1: • 1 2 3 • • 4 5 6 •

G2: • 1 2 3 ★ 4 5 6 •

Double Cut and Join

Cut 1 or 2 adjacencies

- IF 1 cut: create two telomeric adjacencies
- IF **2 cut**: glue back 1 of 2 new ways



• •

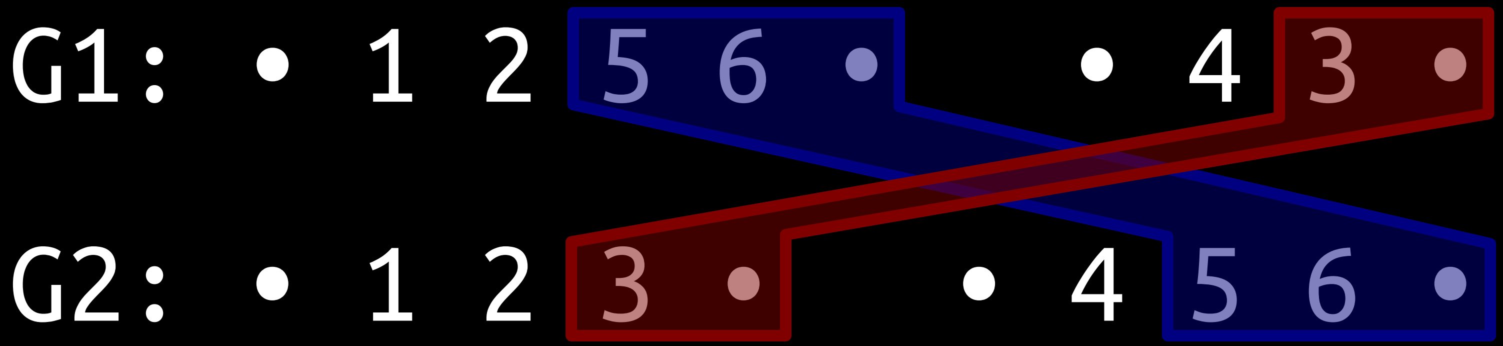
G1: • 1 2 3 • • 4 5 6 •

G2: • 1 2 3 4 5 6 •

Double Cut and Join

Cut 1 or 2 adjacencies

- IF 1 cut: create two telomeric adjacencies
- IF 2 cut: glue back 1 of 2 new ways



G1: • 1 2 5 6 •   • 4 3 •

G2: • 1 2 3 •   • 4 5 6 •

Double Cut and Join

Cut 1 or 2 adjacencies

- IF 1 cut: create two telomeric adjacencies
- IF **2 cut**: glue back 1 of 2 new ways

```
G1: • 1 2✶5 6 •    • 4✶3 •

G2: • 1 2 3 •   • 4 5 6 •
```

Double Cut and Join

Cut 1 or 2 adjacencies

- IF 1 cut: create two telomeric adjacencies
- IF **2 cut**: glue back 1 of 2 new ways

G1: 1 ★ $\underline{2 \qquad 3}$ ★

1 -3 ★ $\underline{-2}$ ★

G2: 1 -3 2

DCJ Scenarios

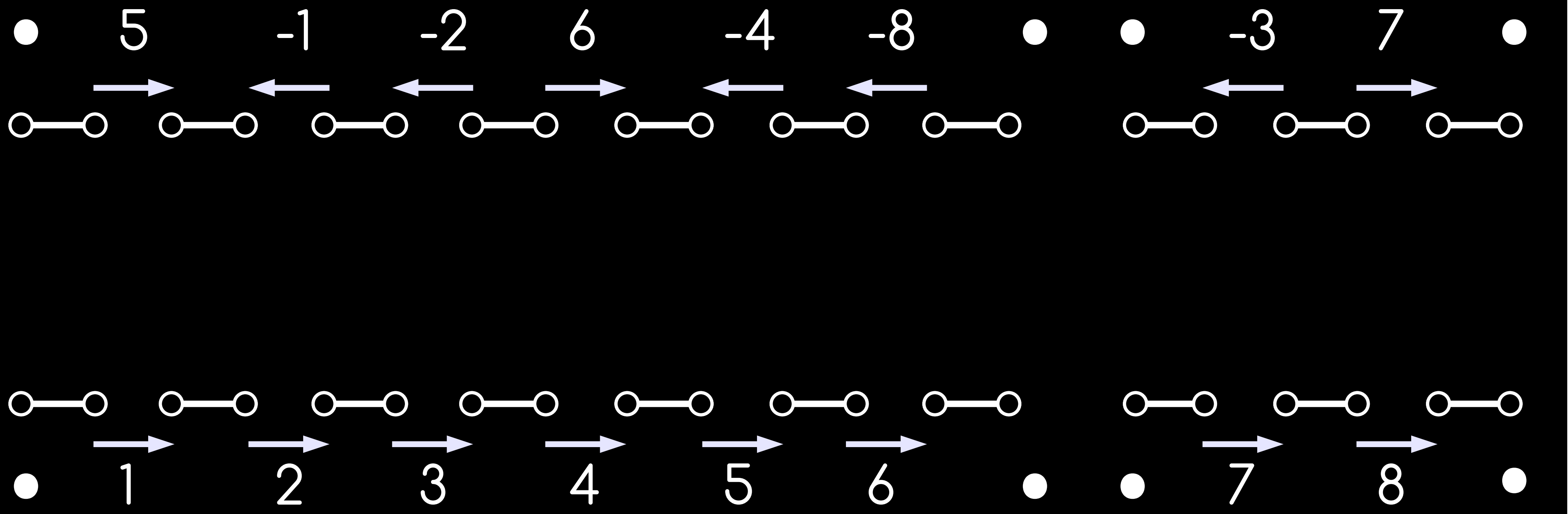
How do we find rearrangements?

● 5 -1 -2 6 -4 -8 ● ● -3 7 ●

● 1 2 3 4 5 6 ● ● 7 8 ●

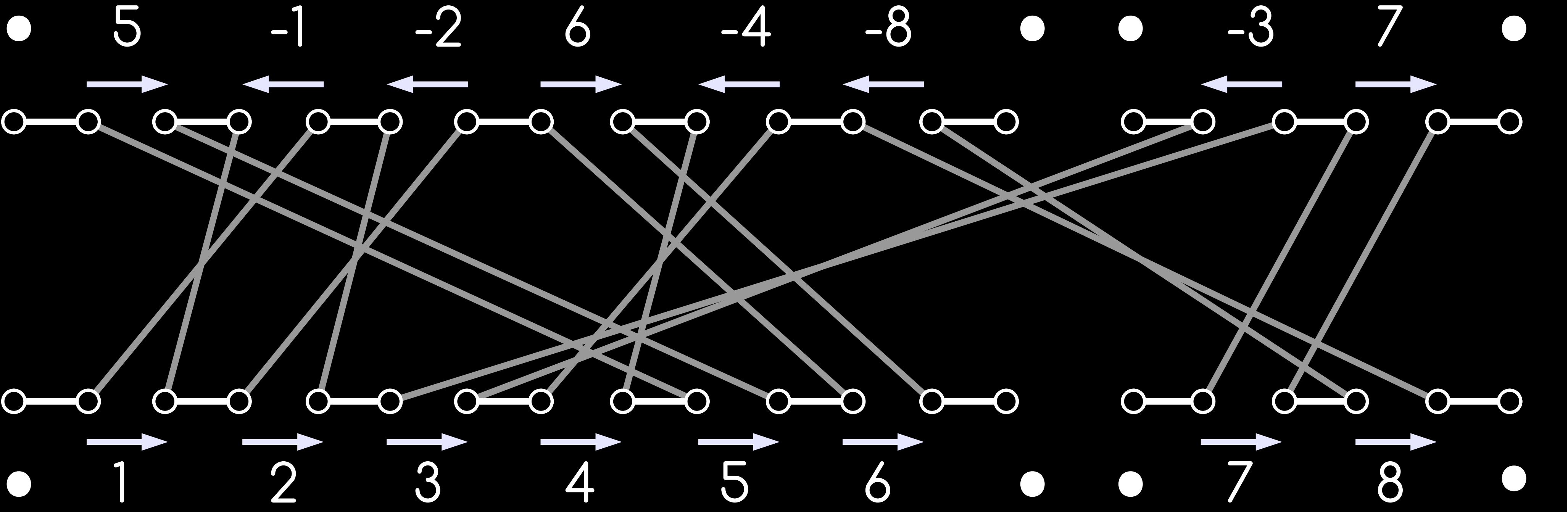
DCJ Scenarios

How do we find rearrangements?



DCJ Scenarios

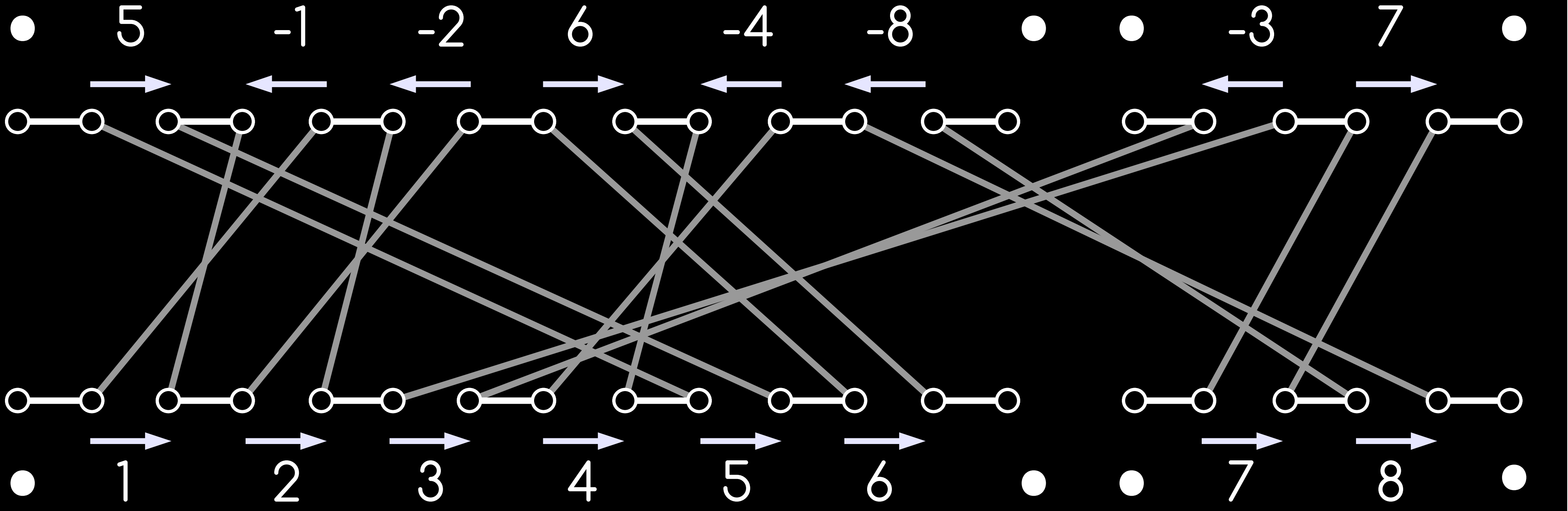
How do we find rearrangements?



DCJ Scenarios

How do we find rearrangements?

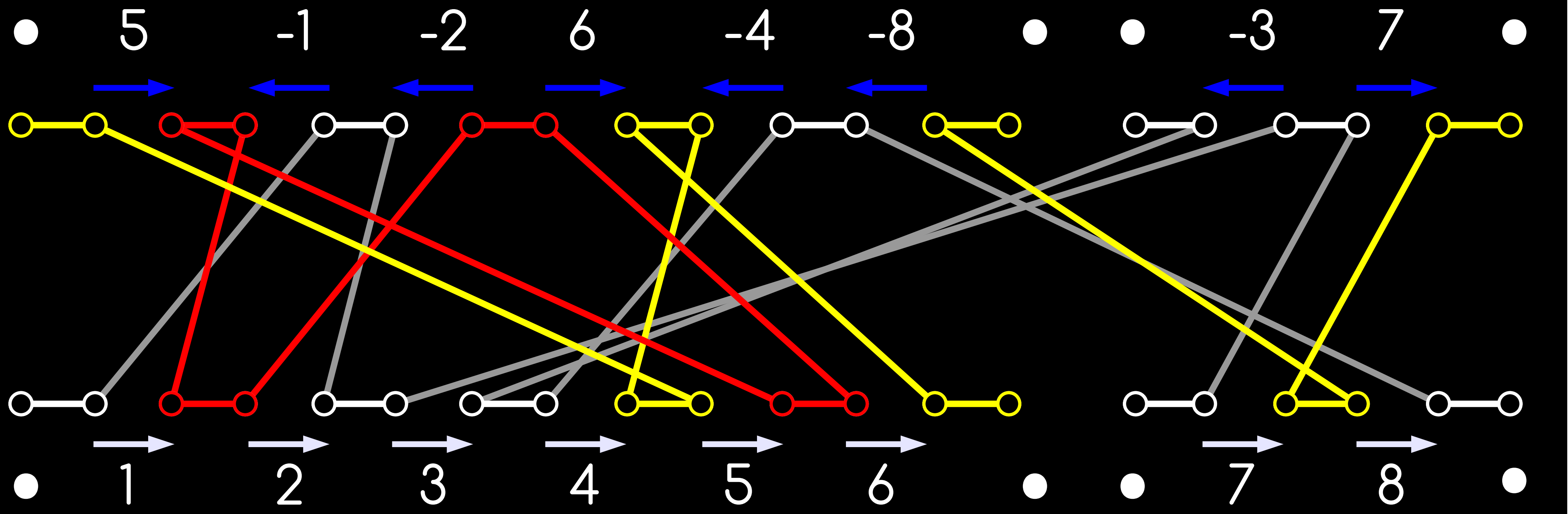
- d_{DCJ} (G1, G2) = N − (C + I/2)



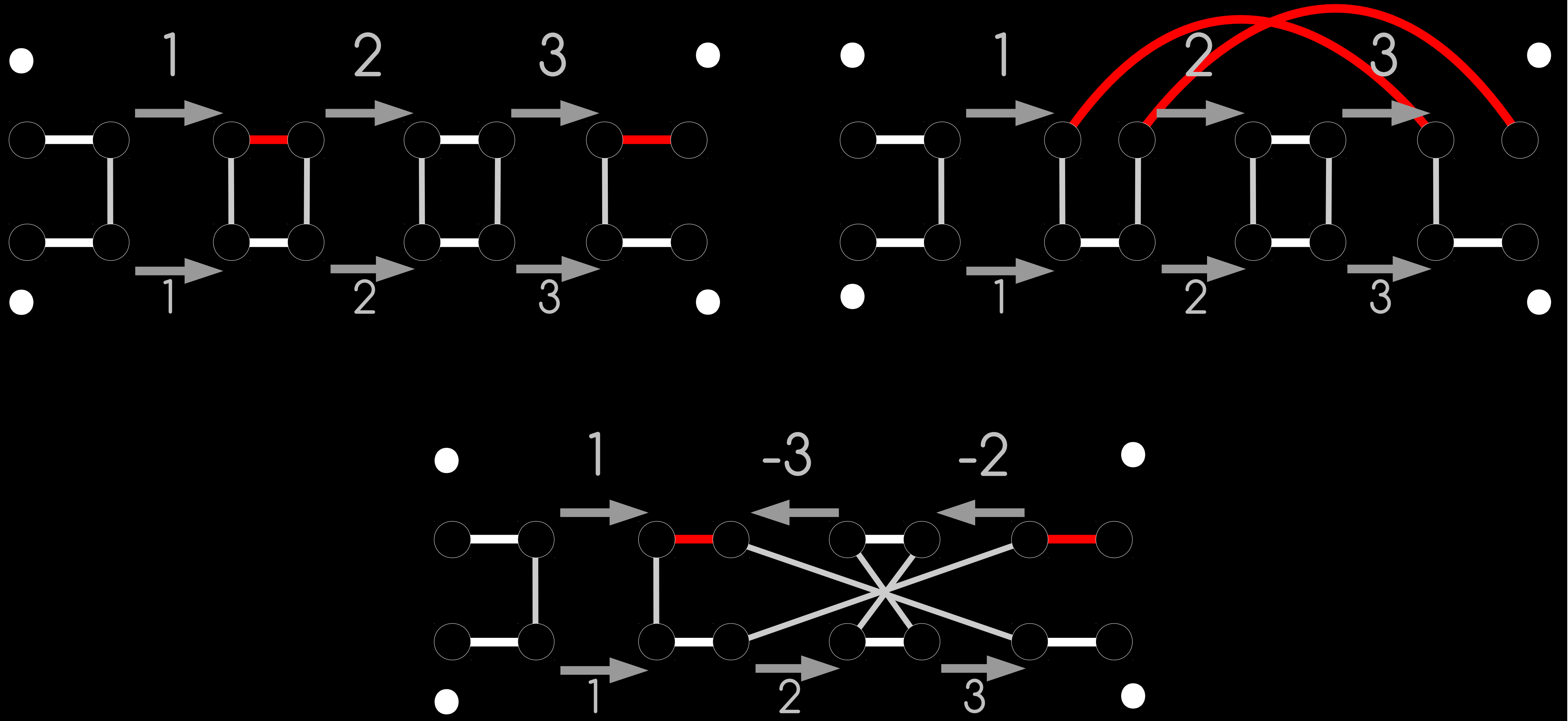
DCJ Scenarios

How do we find rearrangements?

- d_{DCJ} (G1, G2) = N − (C + I/2)

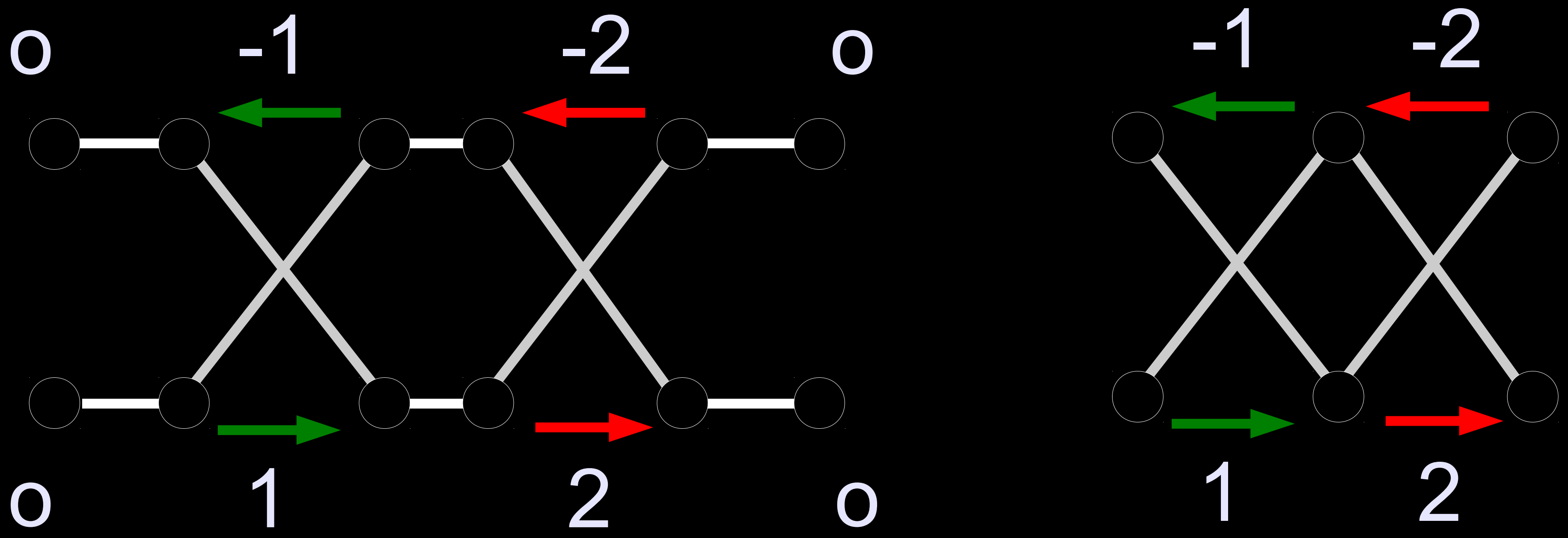


DCJ Moves



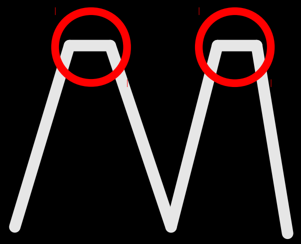
$$\boxed{d_{DCJ}\,(G1, G2) = N - (C + I/2)}$$

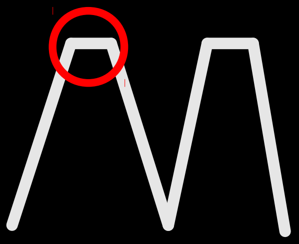
Shorthand



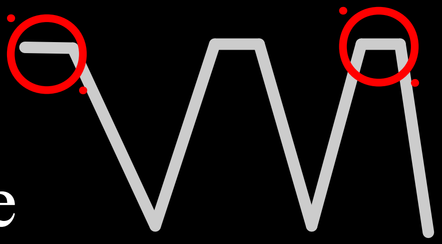
All DCJ Scenarios

- even path
  - extract cycle
  - path fission

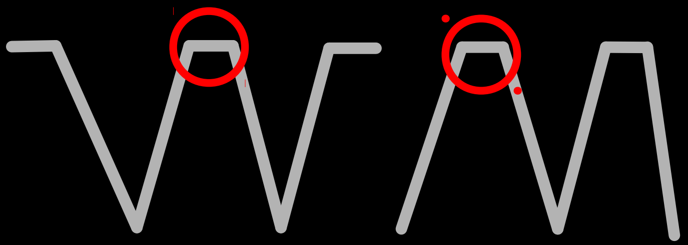




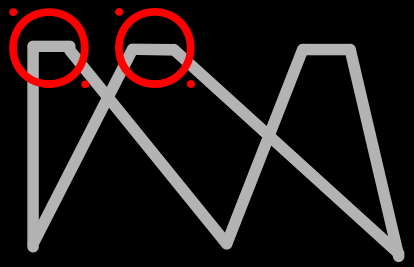
- odd path
  - extract cycle



- 2 even paths

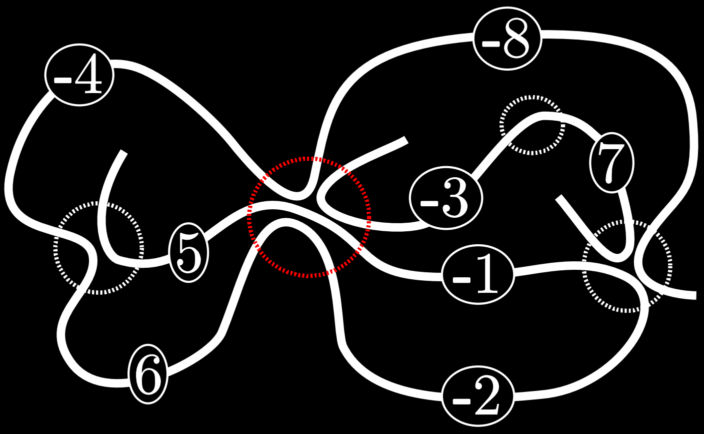


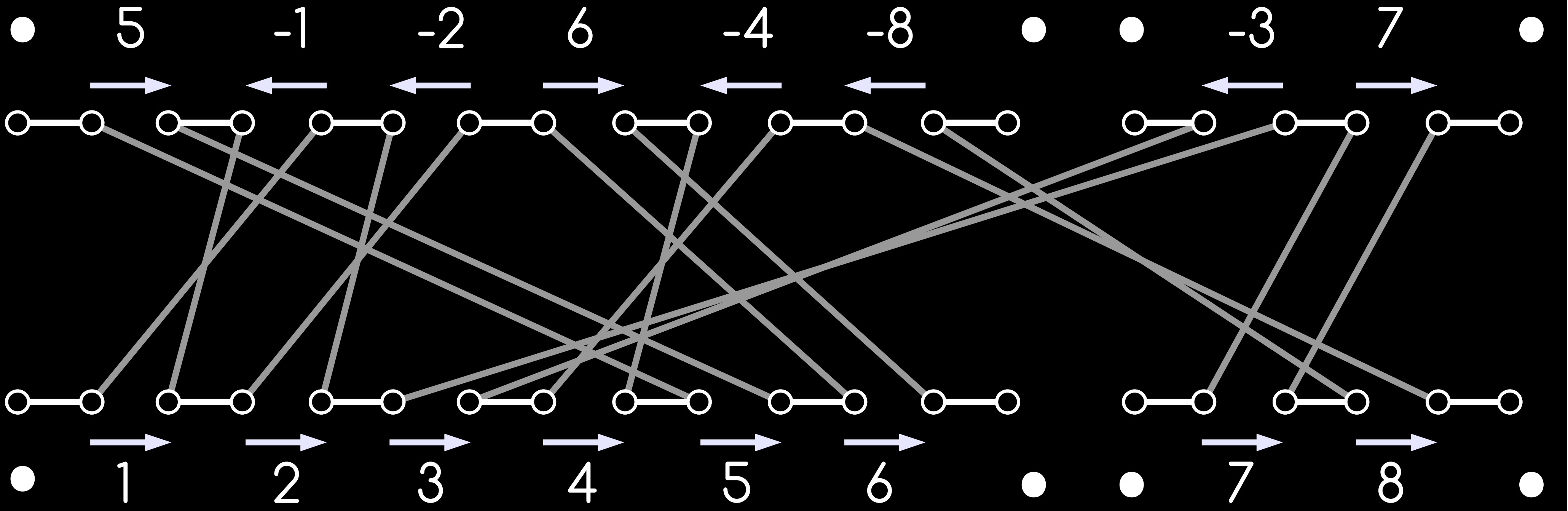
- cycle
  - split cycle



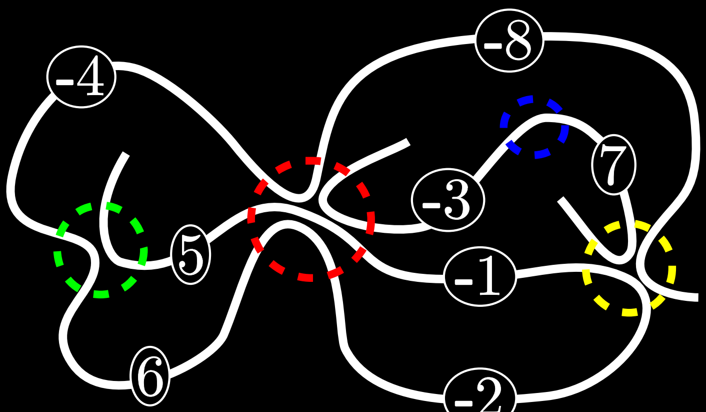
$$\boxed{d_{DCJ}(G1, G2) = N - (C + I/2)}$$

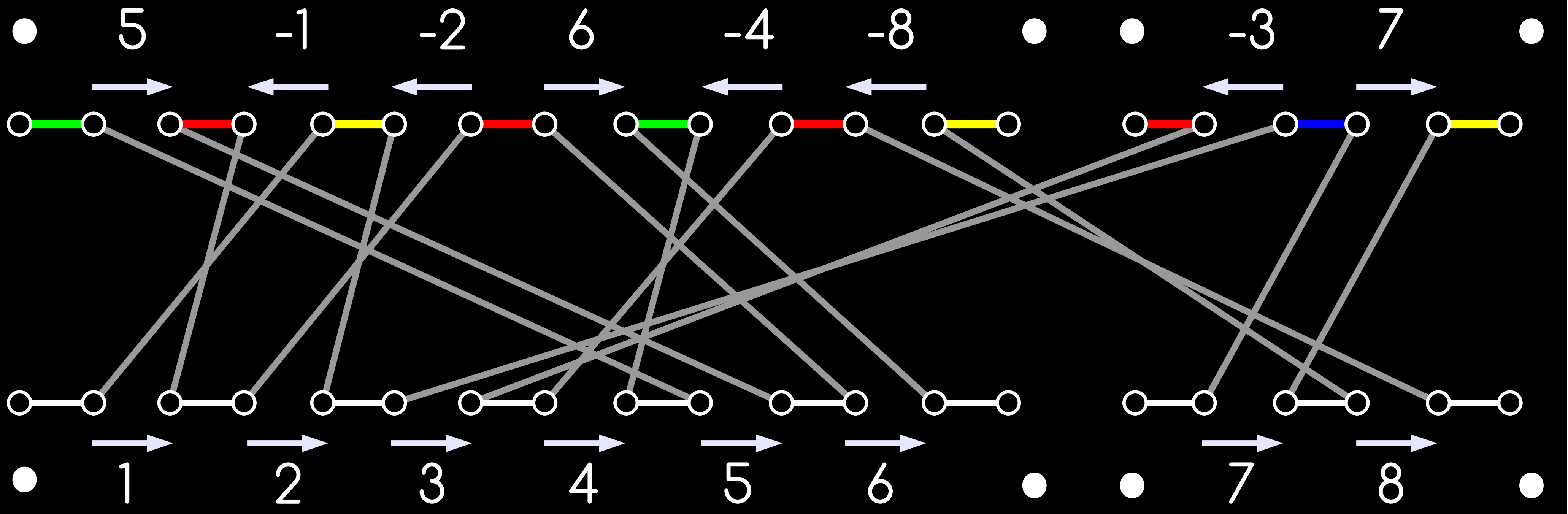
Local DCJ





Local DCJ

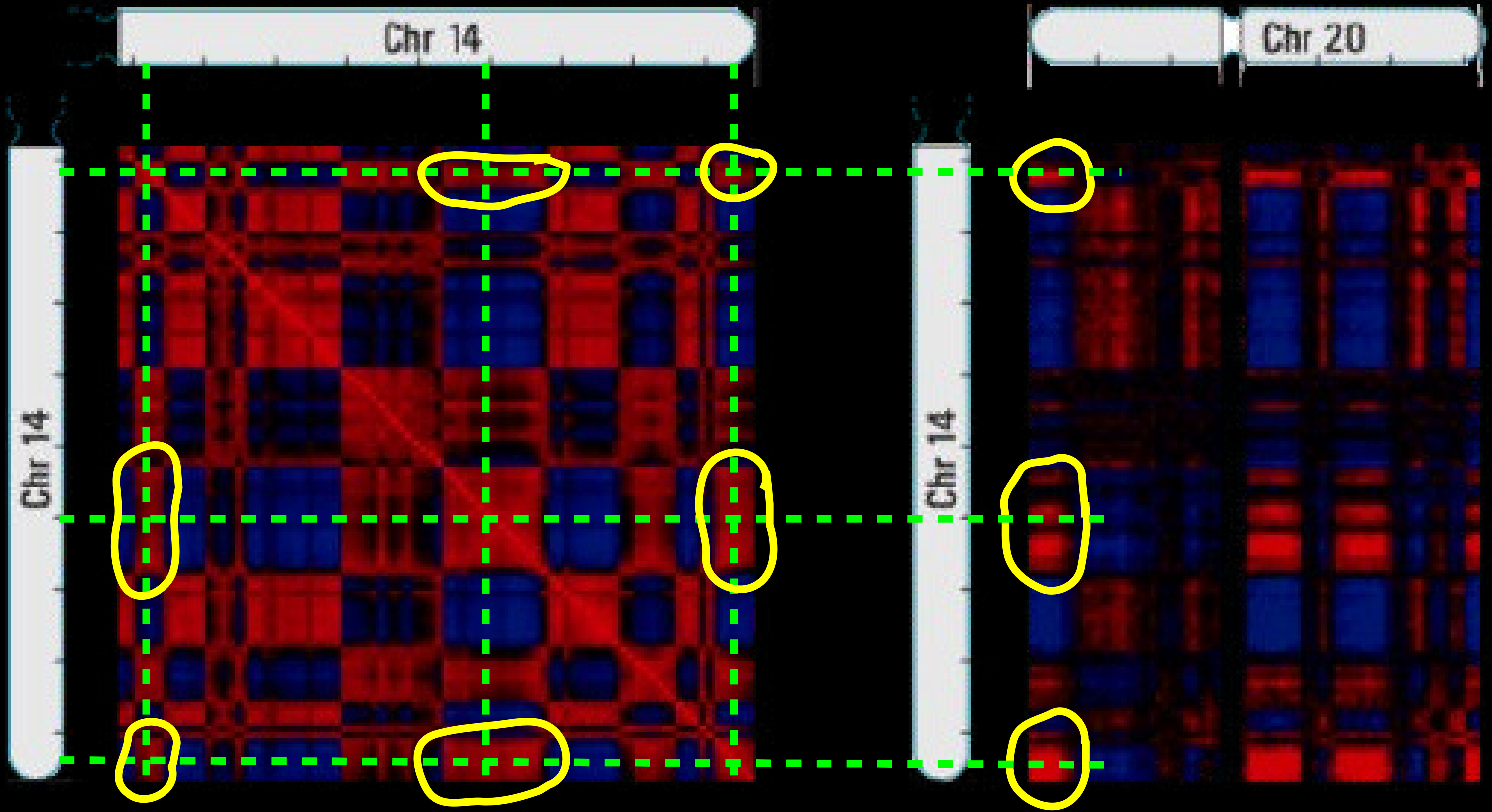




Hi-C Heatmaps

Heatmaps define a locality constraint.

- transitivity appears to hold



Non-locality

- Two problems...

  INPUT: two genomes with colored adjacencies

  - OUTPUT 1:
    scenario with minimum # of non-local moves

  - OUTPUT 2:
    a minimum length scenario, with a minimum # of non-local moves

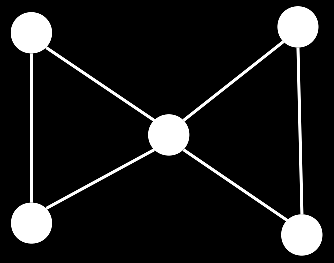
Non-locality

- Two problems...

  INPUT: two genomes with colored adjacencies

  - OUTPUT 1: NP-Hard
    scenario with minimum # of non-local moves
  - OUTPUT 2:
    a minimum length scenario, with a minimum # of non-local moves THIS PAPER Polynomial

Minimize Distant Rearrangements



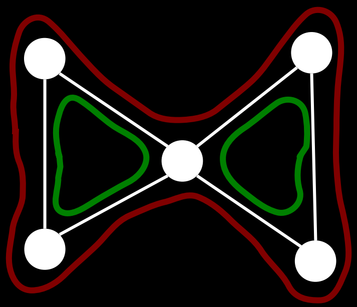
- NP-Hardness
  - Max Eulerian Cycle Decomposition

INPUT: Eulerian graph G = (V, E)

OUTPUT: partition of E into cycles

MEASURE: |E|

Minimize Distant Rearrangements



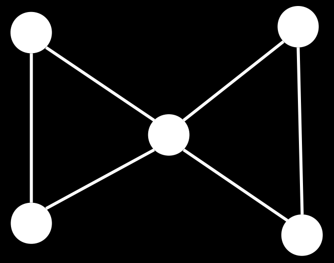
- NP-Hardness
  - Max Eulerian Cycle Decomposition

INPUT: Eulerian graph G = (V, E)

OUTPUT: partition of E into cycles

MEASURE: |E|

Minimize Distant Rearrangements

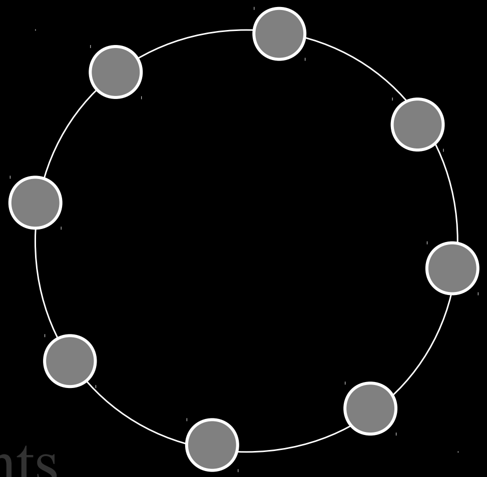


- NP-Hardness
  - Max Eulerian Cycle Decomposition

    INPUT: Eulerian graph G = (V, E)

    OUTPUT: partition of E into cycles

    MEASURE: |E|



- O(n^3) algorithm
  - Min Non-Crossing Colored Partition

    INPUT: ordered set of colored elements

    OUTPUT: non-crossing colored partition

    MEASURE: cardinality of the partition

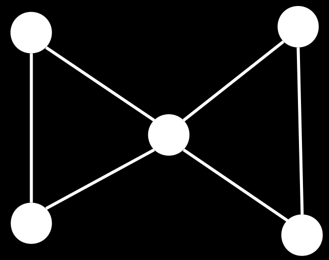
Minimize Distant Rearrangements

- NP-Hardness
  - Max Eulerian Cycle Decomposition

    INPUT: Eulerian graph G = (V, E)

    OUTPUT: partition of E into cycles

    MEASURE: |E|

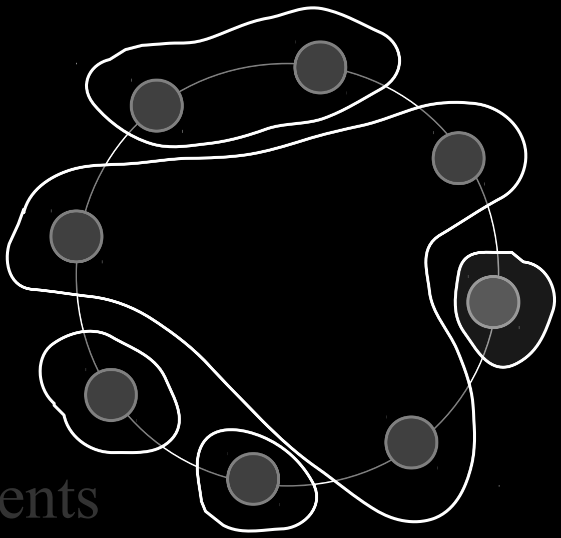


- O(n^3) algorithm
  - Min Non-Crossing Colored Partition

    INPUT: ordered set of colored elements

    OUTPUT: non-crossing colored partition

    MEASURE: cardinality of the partition



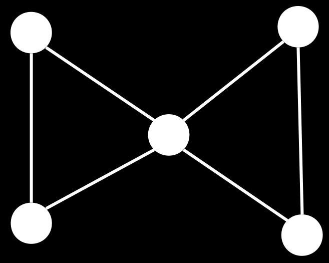
Minimize Distant Rearrangements

- NP-Hardness
  - Max Eulerian Cycle Decomposition

    INPUT: Eulerian graph G = (V, E)

    OUTPUT: partition of E into cycles

    MEASURE: |E|

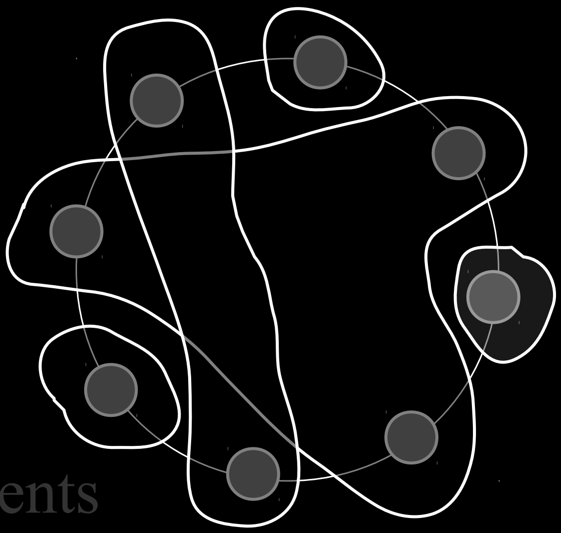


- O(n^3) algorithm
  - Min Non-Crossing Colored Partition

    INPUT: ordered set of colored elements

    OUTPUT: non-crossing colored partition

    MEASURE: cardinality of the partition



crossing!

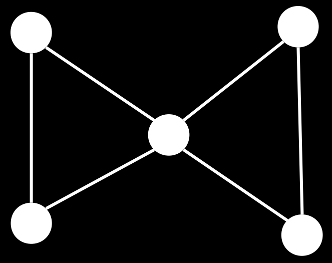
Minimize Distant Rearrangements

- NP-Hardness
  - Max Eulerian Cycle Decomposition

INPUT: Eulerian graph $G = (V, E)$

OUTPUT: partition of E into cycles

MEASURE: $|E|$

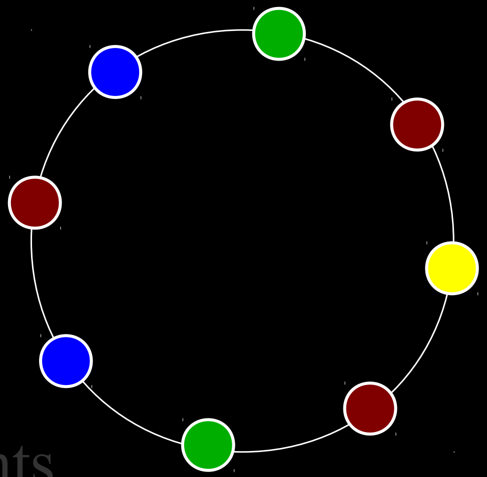


- $O(n^3)$ algorithm
  - Min Non-Crossing Colored Partition

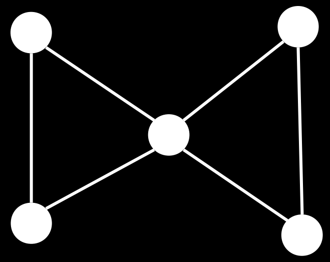
INPUT: ordered set of colored elements

OUTPUT: non-crossing colored partition

MEASURE: cardinality of the partition



Minimize Distant Rearrangements

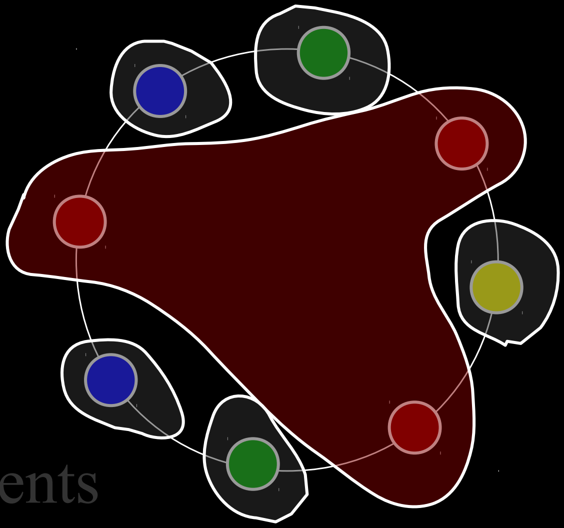


- NP-Hardness
  - Max Eulerian Cycle Decomposition

    INPUT: Eulerian graph $G = (V, E)$

    OUTPUT: partition of E into cycles

    MEASURE: $|E|$



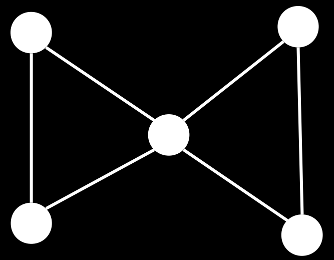
- $O(n^3)$ algorithm
  - Min Non-Crossing Colored Partition

    INPUT: ordered set of colored elements

    OUTPUT: non-crossing colored partition

    MEASURE: cardinality of the partition

Minimize Distant Rearrangements

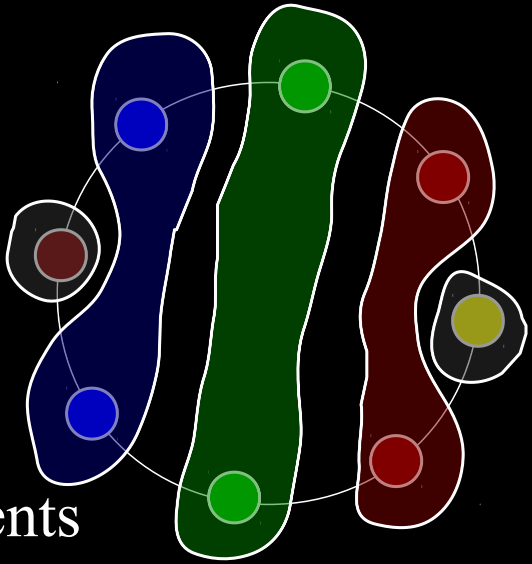


- NP-Hardness
  - Max Eulerian Cycle Decomposition

    INPUT: Eulerian graph G = (V, E)

    OUTPUT: partition of E into cycles

    MEASURE: |E|



- O(n^3) algorithm
  - Min Non-Crossing Colored Partition

    INPUT: ordered set of colored elements

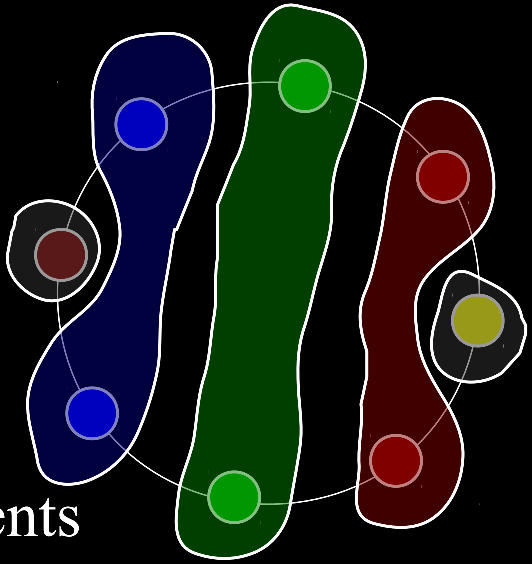
    OUTPUT: non-crossing colored partition

    MEASURE: cardinality of the partition

Minimize Distant Rearrangements

Generalization of Maximum
Independent Set on a Circle Graph

Solved by Dynamic Programming



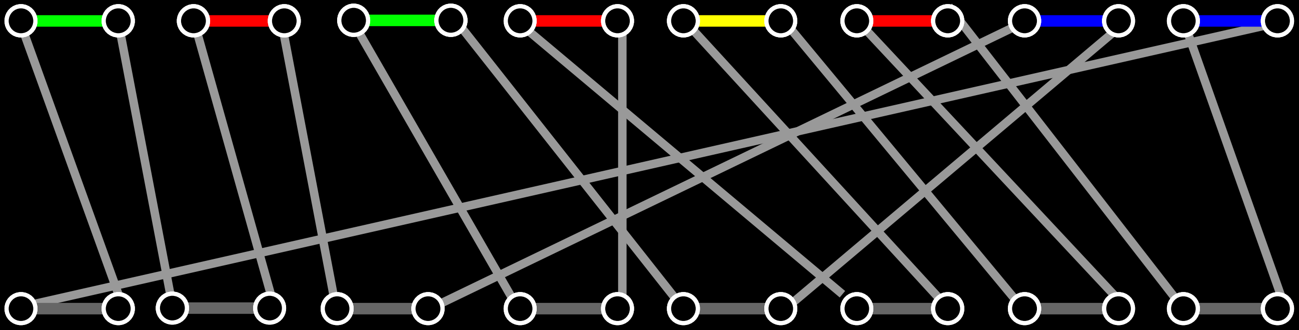
- $O(n^3)$ algorithm
  - Min Non-Crossing Colored Partition

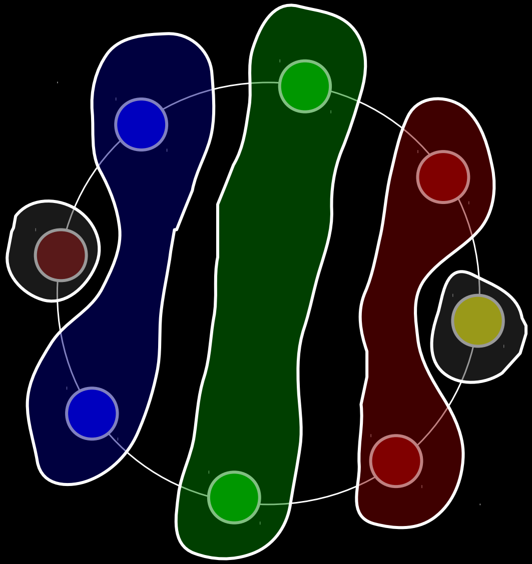
INPUT: ordered set of colored elements

OUTPUT: non-crossing colored partition

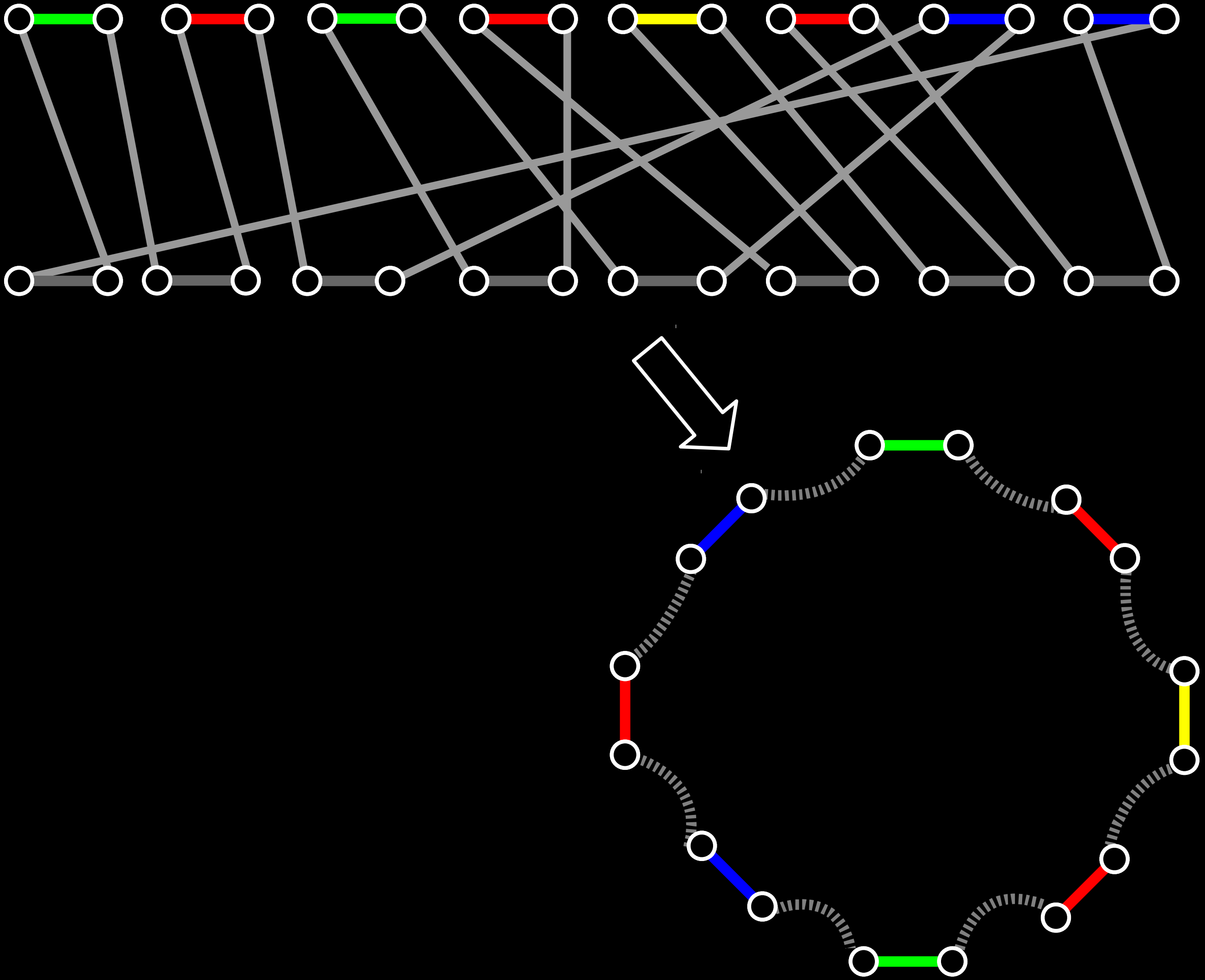
MEASURE: cardinality of the partition

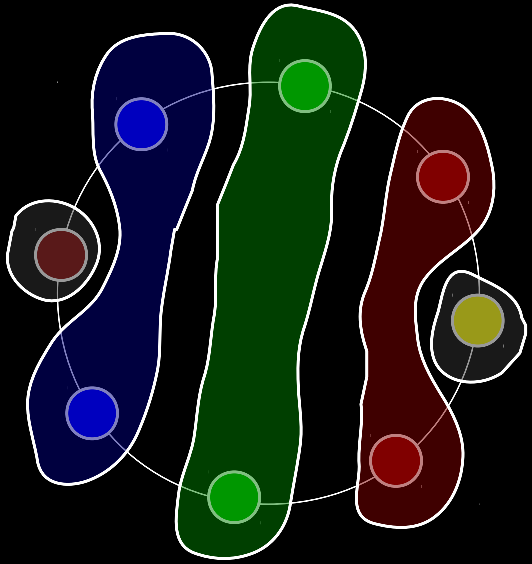
Sorting a Single Component



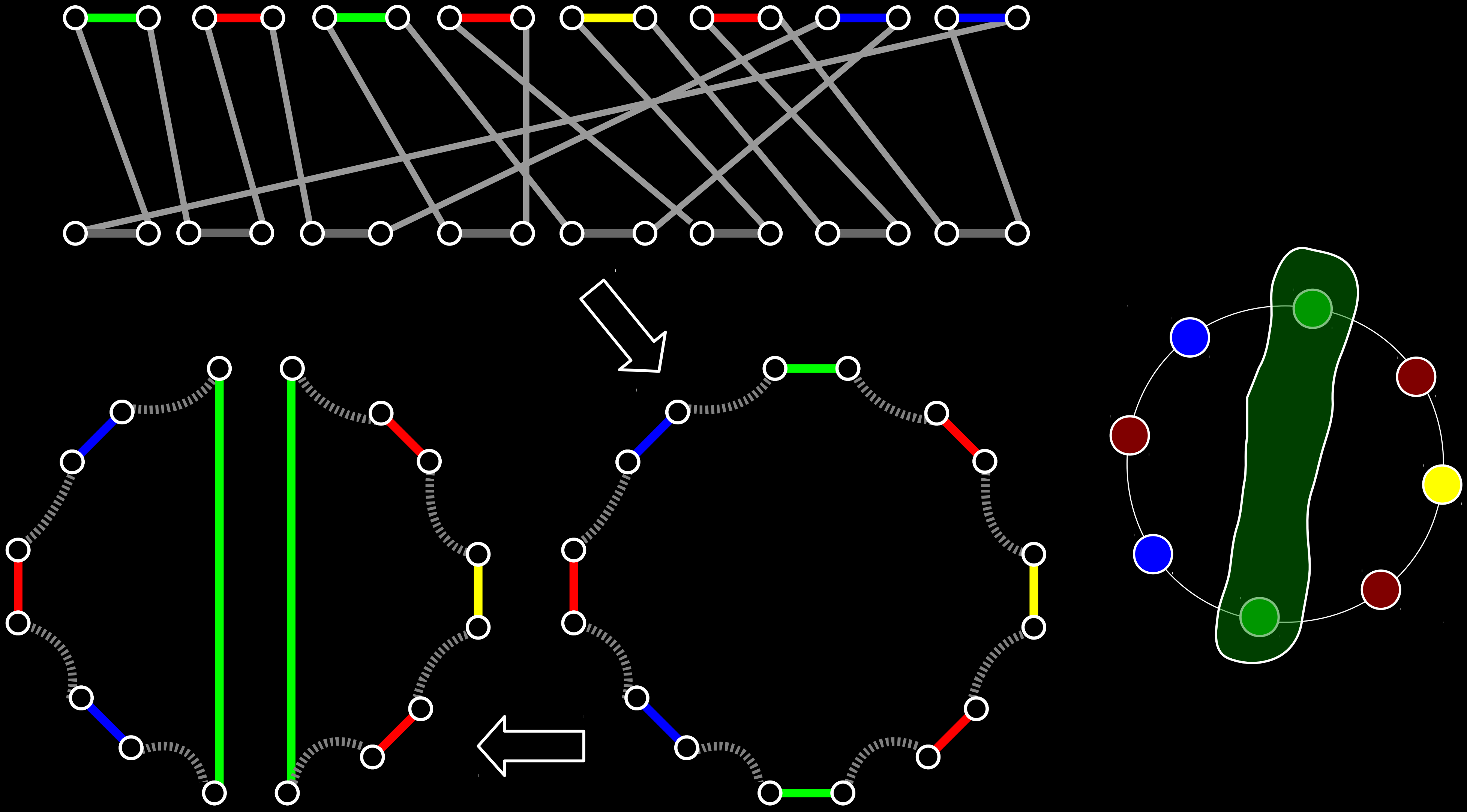


Sorting a Single Component



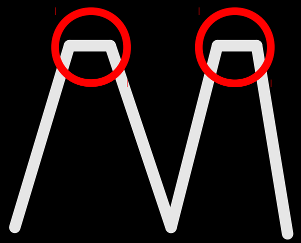


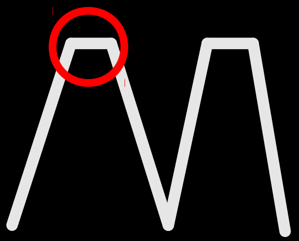
Sorting a Single Component



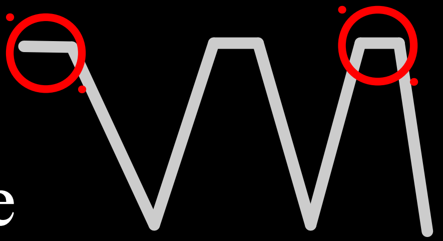
All DCJ Scenarios

- even path
  - extract cycle
  - path fission

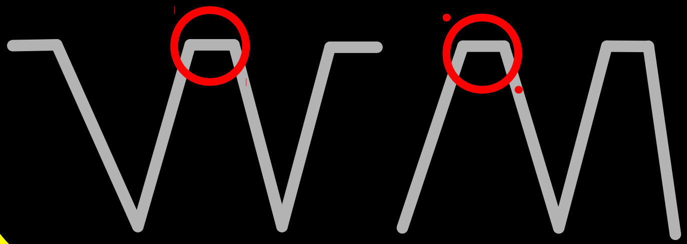




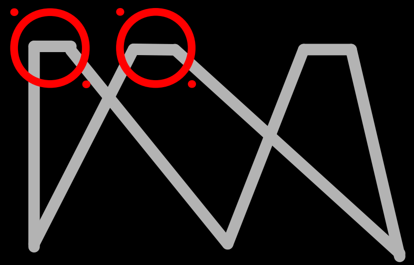
- odd path
  - extract cycle



- 2 even paths

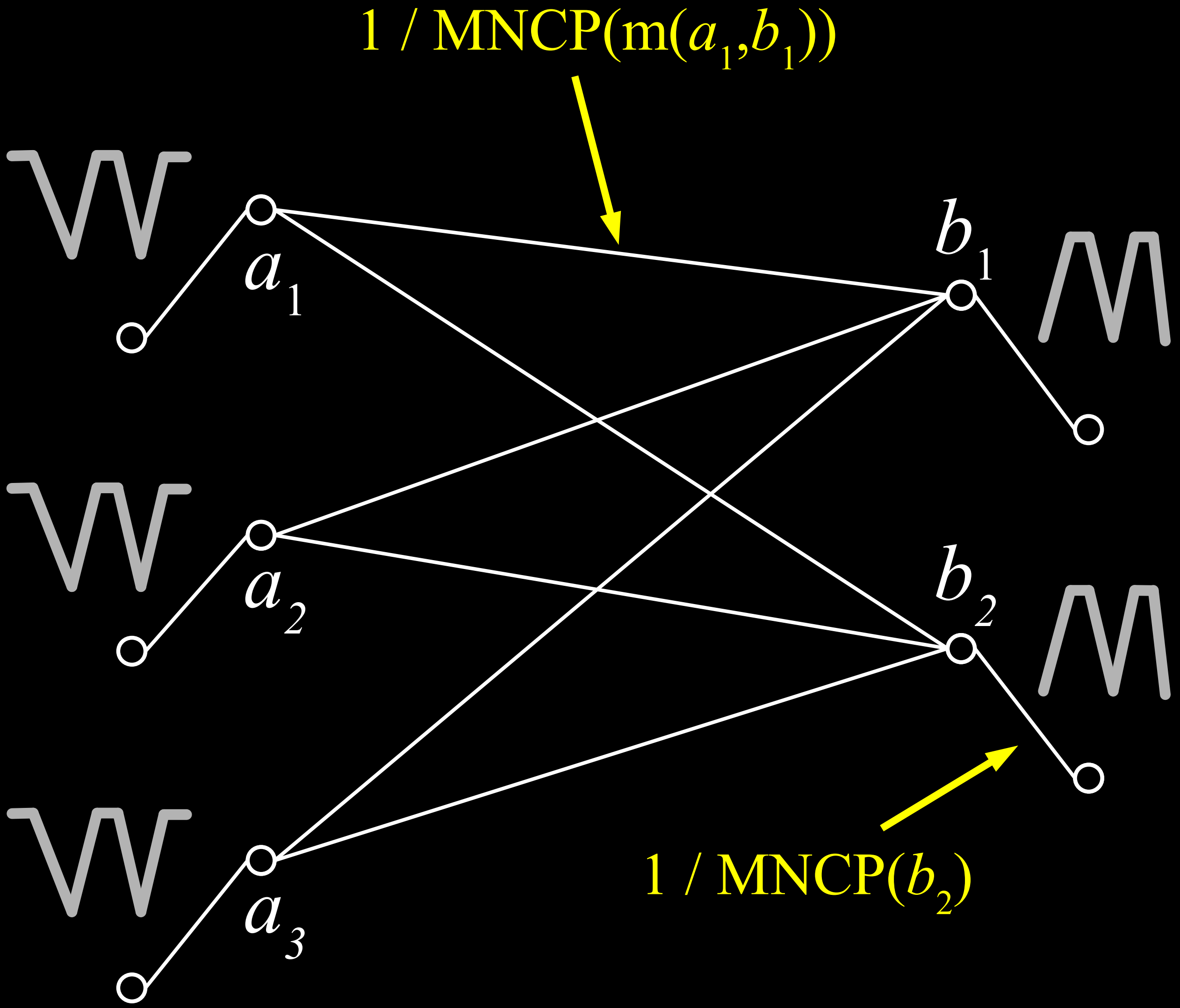


- cycle
  - split cycle



$$\boxed{d_{DCJ}(G1, G2) = N - (C + I/2)}$$

Multiple Even-Length Paths



Running Time

- O(n^3)
  - Min Non-Crossing Colored Partition
- O(N(W) N(M) n^3) ∈ O(n^5)
  - Labeling edges in the bipartite graph

Running Time

- O(n^3)
  - Min Non-Crossing Colored Partition
- O(N(W) N(M) n^3) ∈ ~~O(n^5)~~ $\boxed{O(n^4)}$
  - Labeling edges in the bipartite graph
- In practice N(W) N(M) is small!
  - 182 for human/mouse comparison

Future Work

- 3/2 approx to minimize # of non-local moves
- Other models of evolution
  - inversions
  - inversions/transpositions
- General weights
  - NP-Hard to minimize # of non-local
  - ?minimize # of non-local moves in *parsimonious*?
- 2-sided version of the problem
- Generalize to multiple species

Montpellier, France

Come to the mediterranean!

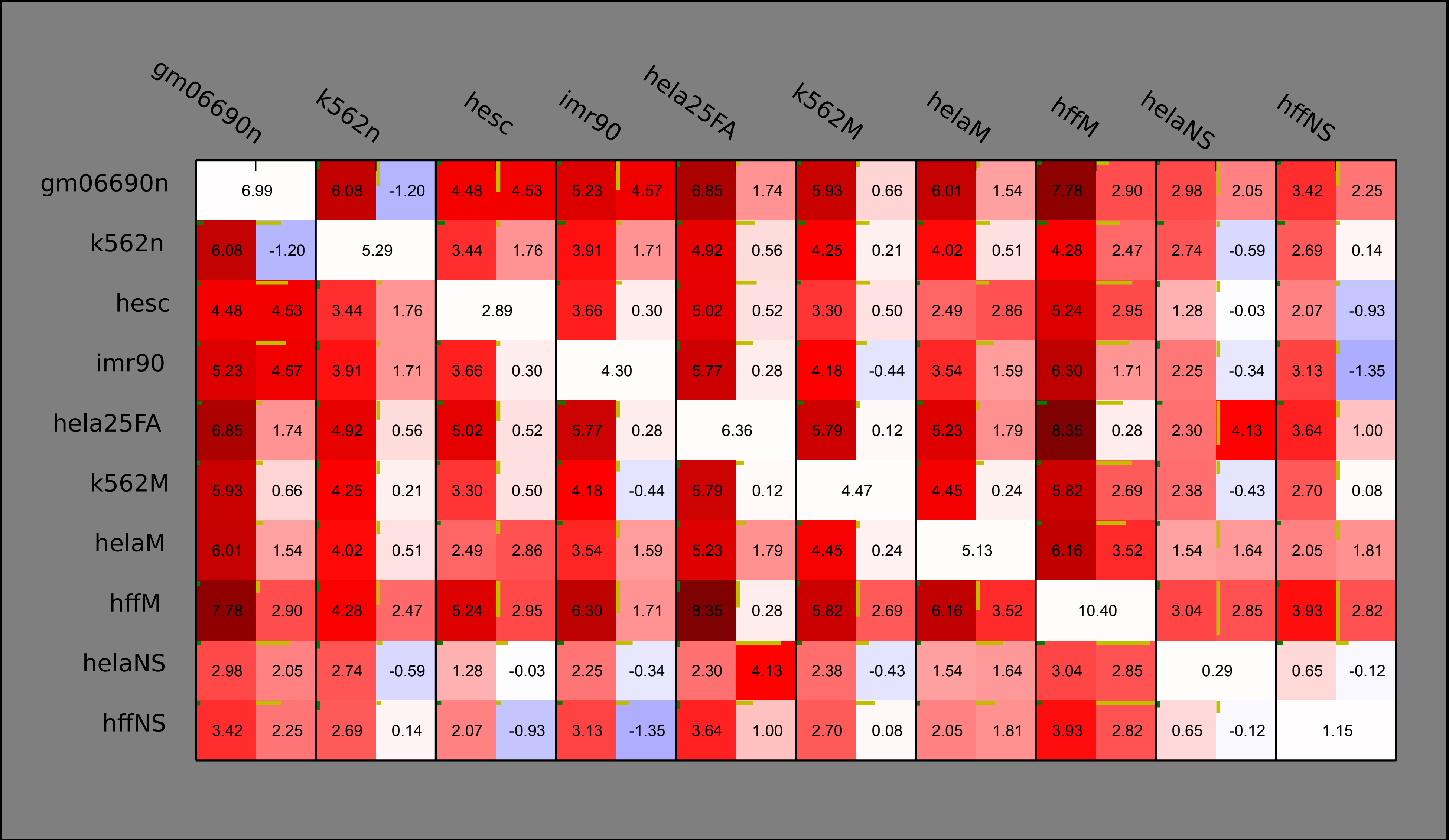
Kirkpatrick

2 post-doc fellowships

THE END

cnrs
dépasser les frontières

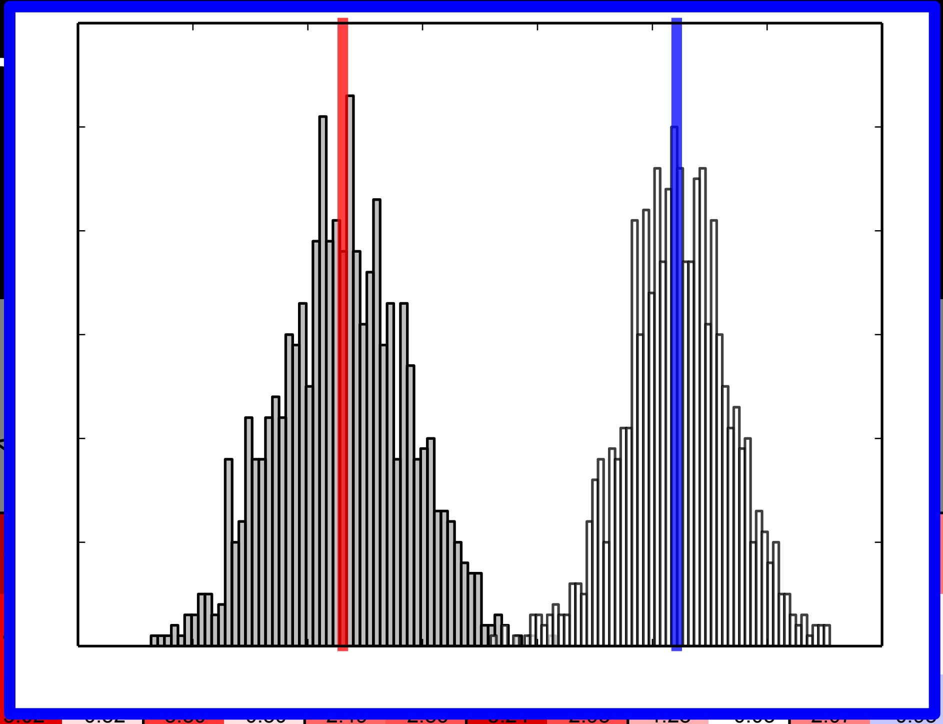
Cell Type Comparison - intra

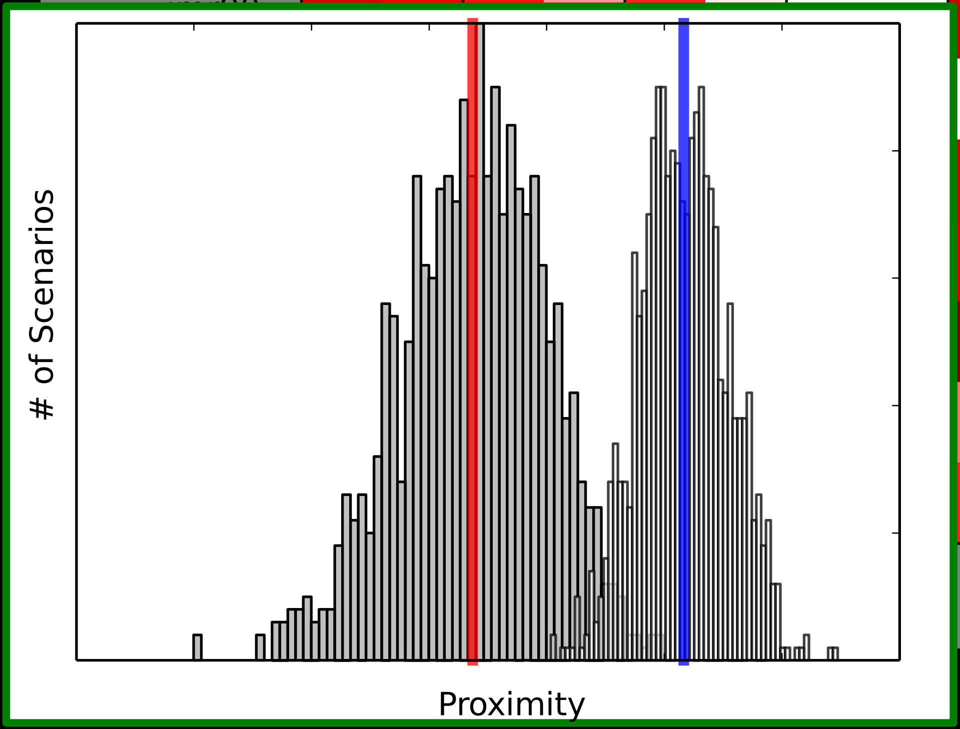


	gm06690n		k562n		hesc		imr90		hela25FA		k562M		helaM		hffM		helaNS		hffNS	
gm06690n	6.99		6.08	-1.20	4.48	4.53	5.23	4.57	6.85	1.74	5.93	0.66	6.01	1.54	7.78	2.90	2.98	2.05	3.42	2.25
k562n	6.08	-1.20	5.29		3.44	1.76	3.91	1.71	4.92	0.56	4.25	0.21	4.02	0.51	4.28	2.47	2.74	-0.59	2.69	0.14
hesc	4.48	4.53	3.44	1.76	2.89		3.66	0.30	5.02	0.52	3.30	0.50	2.49	2.86	5.24	2.95	1.28	-0.03	2.07	-0.93
imr90	5.23	4.57	3.91	1.71	3.66	0.30	4.30		5.77	0.28	4.18	-0.44	3.54	1.59	6.30	1.71	2.25	-0.34	3.13	-1.35
hela25FA	6.85	1.74	4.92	0.56	5.02	0.52	5.77	0.28	6.36		5.79	0.12	5.23	1.79	8.35	0.28	2.30	4.13	3.64	1.00
k562M	5.93	0.66	4.25	0.21	3.30	0.50	4.18	-0.44	5.79	0.12	4.47		4.45	0.24	5.82	2.69	2.38	-0.43	2.70	0.08
helaM	6.01	1.54	4.02	0.51	2.49	2.86	3.54	1.59	5.23	1.79	4.45	0.24	5.13		6.16	3.52	1.54	1.64	2.05	1.81
hffM	7.78	2.90	4.28	2.47	5.24	2.95	6.30	1.71	8.35	0.28	5.82	2.69	6.16	3.52	10.40		3.04	2.85	3.93	2.82
helaNS	2.98	2.05	2.74	-0.59	1.28	-0.03	2.25	-0.34	2.30	4.13	2.38	-0.43	1.54	1.64	3.04	2.85	0.29		0.65	-0.12
hffNS	3.42	2.25	2.69	0.14	2.07	-0.93	3.13	-1.35	3.64	1.00	2.70	0.08	2.05	1.81	3.93	2.82	0.65	-0.12	1.15	

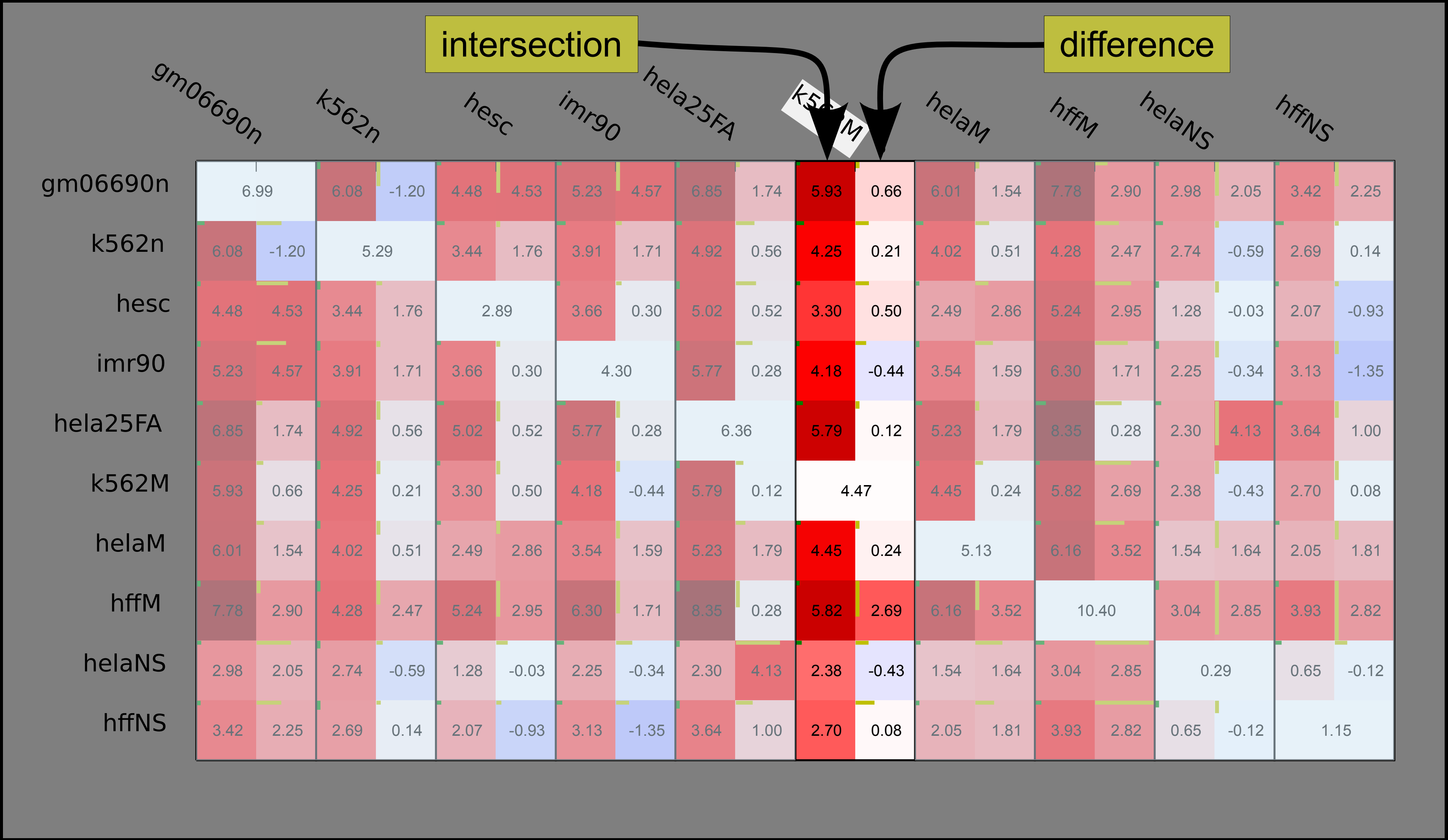
Cell Type Cor

	gm06690n	k562n	hesc	imr90	hela	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
gm06690n	6.99	6.08 -1.20	4.48 4.53	5.23 4.57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
k562n	6.08 -1.20	5.29	3.44 1.76	3.91 1.71	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
hesc	4.48 4.53	3.44 1.76	2.89	3.66 0.30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
imr90	[illegible]	[illegible]	[illegible]	[illegible]	5.77 0.28	4.18 -0.44	3.54 1.59	6.30 1.71	2.25 -0.34	3.13 -1.35
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	6.36	5.79 0.12	5.23 1.79	8.35 0.28	2.30 4.13	3.64 1.00
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5.79 0.12	4.47	4.45 0.24	5.82 2.69	2.38 -0.43	2.70 0.08
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5.23 1.79	4.45 0.24	5.13	6.16 3.52	1.54 1.64	2.05 1.81
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8.35 0.28	5.82 2.69	6.16 3.52	10.40	3.04 2.85	3.93 2.82
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.30 4.13	2.38 -0.43	1.54 1.64	3.04 2.85	0.29	0.65 -0.12
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.64 1.00	2.70 0.08	2.05 1.81	3.93 2.82	0.65 -0.12	1.15





Cell Type Comparison - intra

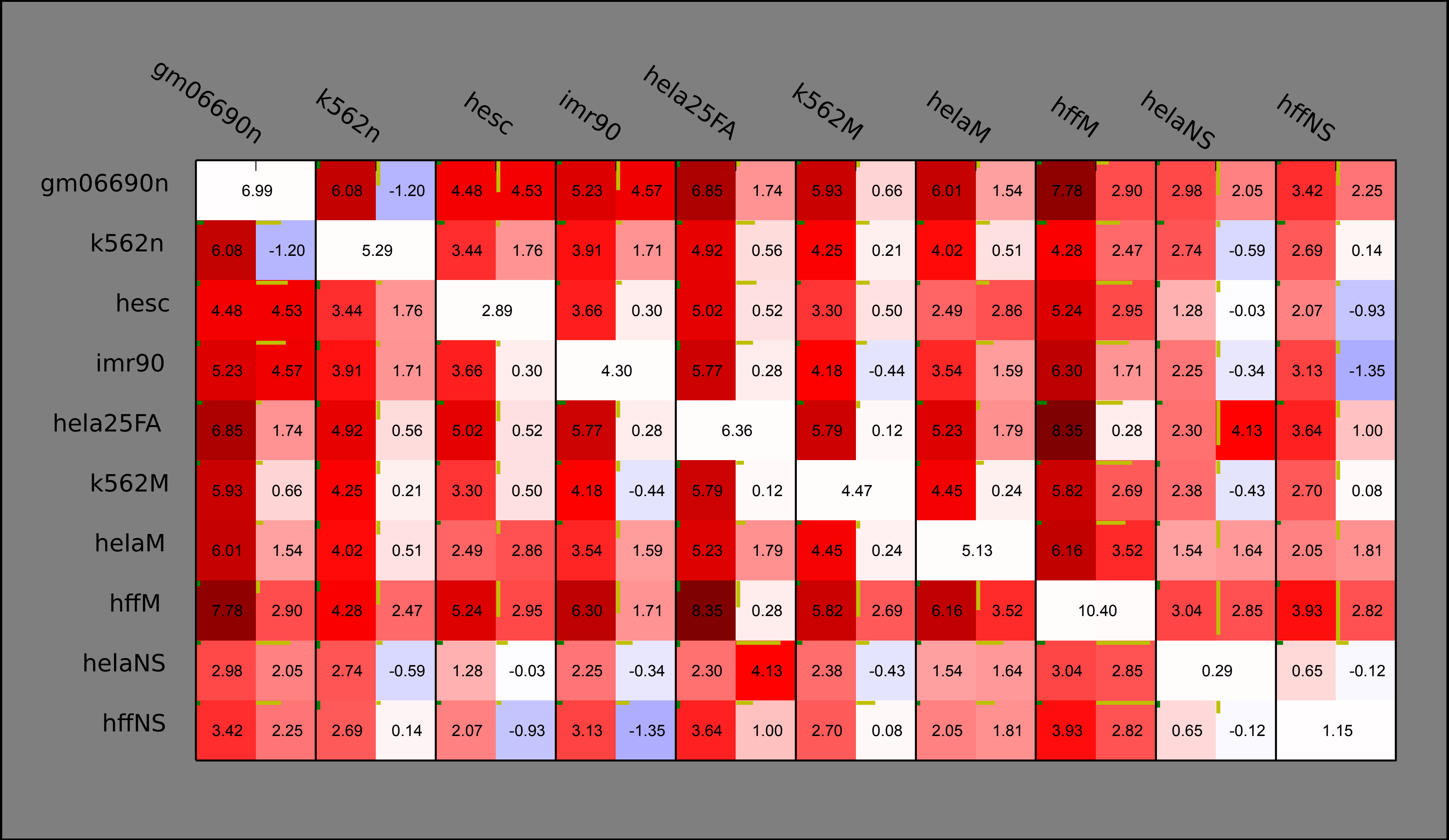


intersection

difference

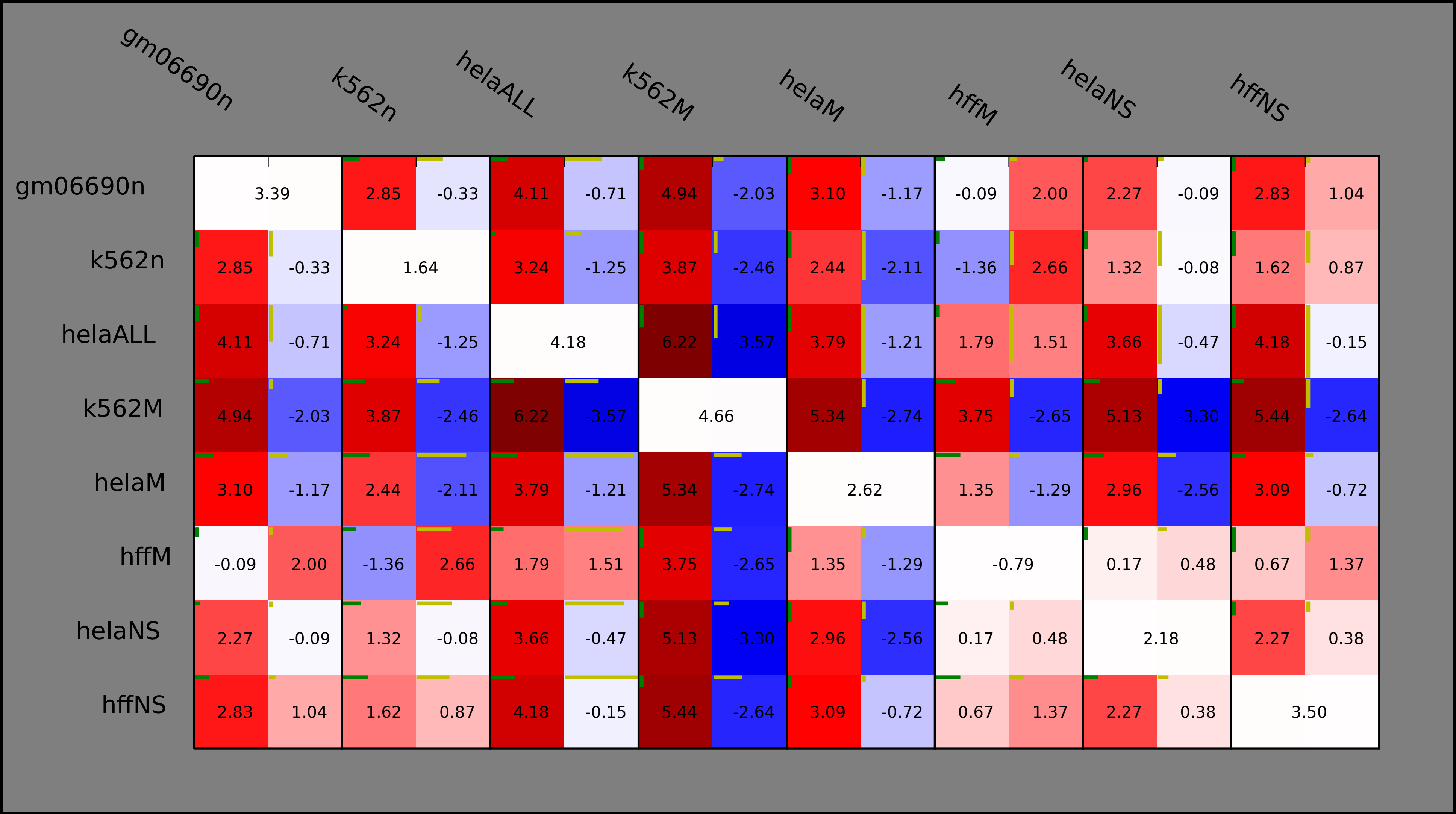
	gm06690n		k562n		hesc		imr90		hela25FA		k562M		helaM		hffM		helaNS		hffNS	
gm06690n	6.99		6.08	-1.20	4.48	4.53	5.23	4.57	6.85	1.74	5.93	0.66	6.01	1.54	7.78	2.90	2.98	2.05	3.42	2.25
k562n	6.08	-1.20	5.29		3.44	1.76	3.91	1.71	4.92	0.56	4.25	0.21	4.02	0.51	4.28	2.47	2.74	-0.59	2.69	0.14
hesc	4.48	4.53	3.44	1.76	2.89		3.66	0.30	5.02	0.52	3.30	0.50	2.49	2.86	5.24	2.95	1.28	-0.03	2.07	-0.93
imr90	5.23	4.57	3.91	1.71	3.66	0.30	4.30		5.77	0.28	4.18	-0.44	3.54	1.59	6.30	1.71	2.25	-0.34	3.13	-1.35
hela25FA	6.85	1.74	4.92	0.56	5.02	0.52	5.77	0.28	6.36		5.79	0.12	5.23	1.79	8.35	0.28	2.30	4.13	3.64	1.00
k562M	5.93	0.66	4.25	0.21	3.30	0.50	4.18	-0.44	5.79	0.12	4.47		4.45	0.24	5.82	2.69	2.38	-0.43	2.70	0.08
helaM	6.01	1.54	4.02	0.51	2.49	2.86	3.54	1.59	5.23	1.79	4.45	0.24	5.13		6.16	3.52	1.54	1.64	2.05	1.81
hffM	7.78	2.90	4.28	2.47	5.24	2.95	6.30	1.71	8.35	0.28	5.82	2.69	6.16	3.52	10.40		3.04	2.85	3.93	2.82
helaNS	2.98	2.05	2.74	-0.59	1.28	-0.03	2.25	-0.34	2.30	4.13	2.38	-0.43	1.54	1.64	3.04	2.85	0.29		0.65	-0.12
hffNS	3.42	2.25	2.69	0.14	2.07	-0.93	3.13	-1.35	3.64	1.00	2.70	0.08	2.05	1.81	3.93	2.82	0.65	-0.12	1.15	

Cell Type Comparison - intra



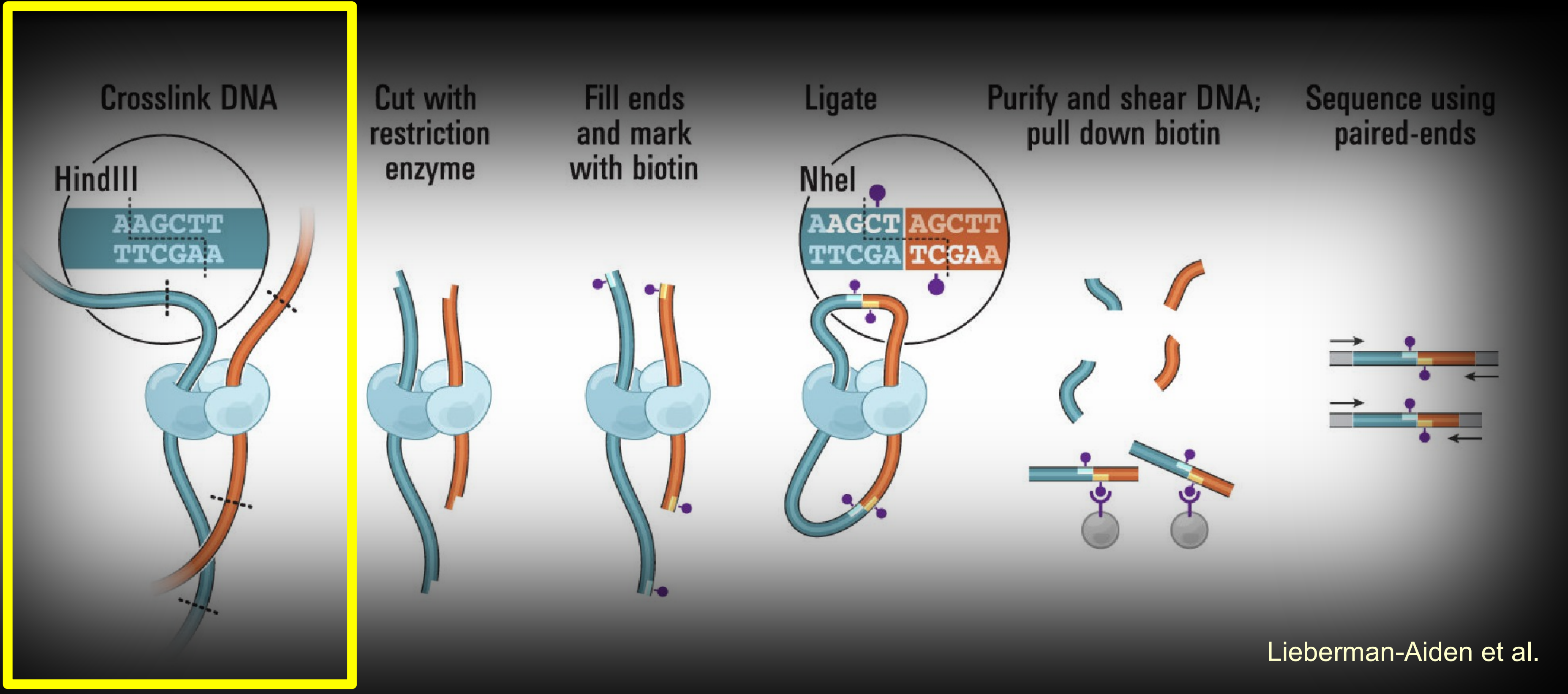
	gm06690n		k562n		hesc		imr90		hela25FA		k562M		helaM		hffM		helaNS		hffNS	
gm06690n	6.99		6.08	-1.20	4.48	4.53	5.23	4.57	6.85	1.74	5.93	0.66	6.01	1.54	7.78	2.90	2.98	2.05	3.42	2.25
k562n	6.08	-1.20	5.29		3.44	1.76	3.91	1.71	4.92	0.56	4.25	0.21	4.02	0.51	4.28	2.47	2.74	-0.59	2.69	0.14
hesc	4.48	4.53	3.44	1.76	2.89		3.66	0.30	5.02	0.52	3.30	0.50	2.49	2.86	5.24	2.95	1.28	-0.03	2.07	-0.93
imr90	5.23	4.57	3.91	1.71	3.66	0.30	4.30		5.77	0.28	4.18	-0.44	3.54	1.59	6.30	1.71	2.25	-0.34	3.13	-1.35
hela25FA	6.85	1.74	4.92	0.56	5.02	0.52	5.77	0.28	6.36		5.79	0.12	5.23	1.79	8.35	0.28	2.30	4.13	3.64	1.00
k562M	5.93	0.66	4.25	0.21	3.30	0.50	4.18	-0.44	5.79	0.12	4.47		4.45	0.24	5.82	2.69	2.38	-0.43	2.70	0.08
helaM	6.01	1.54	4.02	0.51	2.49	2.86	3.54	1.59	5.23	1.79	4.45	0.24	5.13		6.16	3.52	1.54	1.64	2.05	1.81
hffM	7.78	2.90	4.28	2.47	5.24	2.95	6.30	1.71	8.35	0.28	5.82	2.69	6.16	3.52	10.40		3.04	2.85	3.93	2.82
helaNS	2.98	2.05	2.74	-0.59	1.28	-0.03	2.25	-0.34	2.30	4.13	2.38	-0.43	1.54	1.64	3.04	2.85	0.29		0.65	-0.12
hffNS	3.42	2.25	2.69	0.14	2.07	-0.93	3.13	-1.35	3.64	1.00	2.70	0.08	2.05	1.81	3.93	2.82	0.65	-0.12	1.15	

Cell Type Comparison - inter



	gm06690n		k562n		helaALL		k562M		helaM		hffM		helaNS		hffNS	
gm06690n	3.39		2.85	-0.33	4.11	-0.71	4.94	-2.03	3.10	-1.17	-0.09	2.00	2.27	-0.09	2.83	1.04
k562n	2.85	-0.33	1.64		3.24	-1.25	3.87	-2.46	2.44	-2.11	-1.36	2.66	1.32	-0.08	1.62	0.87
helaALL	4.11	-0.71	3.24	-1.25	4.18		6.22	-3.57	3.79	-1.21	1.79	1.51	3.66	-0.47	4.18	-0.15
k562M	4.94	-2.03	3.87	-2.46	6.22	-3.57	4.66		5.34	-2.74	3.75	-2.65	5.13	-3.30	5.44	-2.64
helaM	3.10	-1.17	2.44	-2.11	3.79	-1.21	5.34	-2.74	2.62		1.35	-1.29	2.96	-2.56	3.09	-0.72
hffM	-0.09	2.00	-1.36	2.66	1.79	1.51	3.75	-2.65	1.35	-1.29	-0.79		0.17	0.48	0.67	1.37
helaNS	2.27	-0.09	1.32	-0.08	3.66	-0.47	5.13	-3.30	2.96	-2.56	0.17	0.48	2.18		2.27	0.38
hffNS	2.83	1.04	1.62	0.87	4.18	-0.15	5.44	-2.64	3.09	-0.72	0.67	1.37	2.27	0.38	3.50	

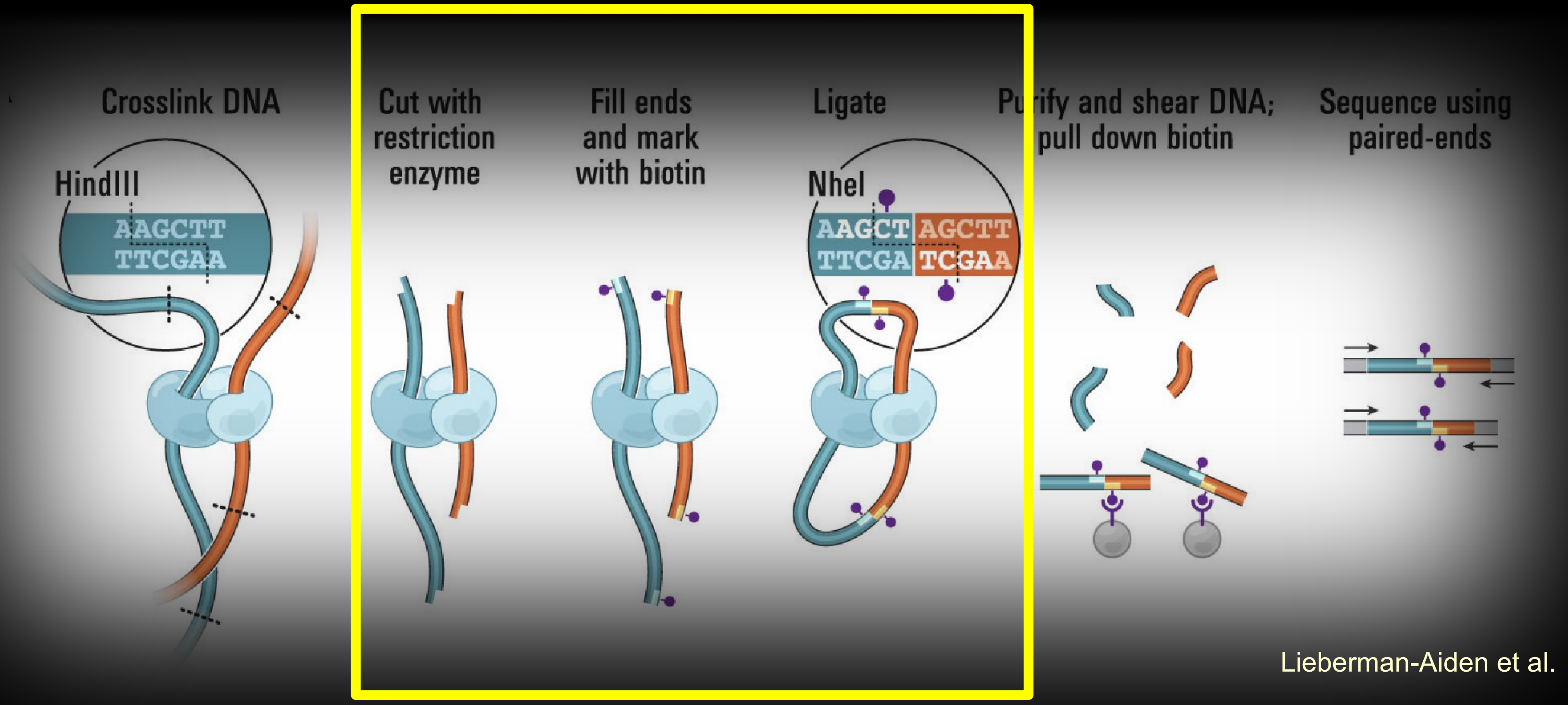
Data: Hi-C


Crosslink DNA
HindIII
AAGCTT
TTCGAA
Cut with restriction enzyme
Fill ends and mark with biotin
Ligate
NheI
AAGCT AGCTT
TTCGA TCGAA
Purify and shear DNA; pull down biotin
Sequence using paired-ends

Lieberman-Aiden et al.

1) crosslink

Data: Hi-C

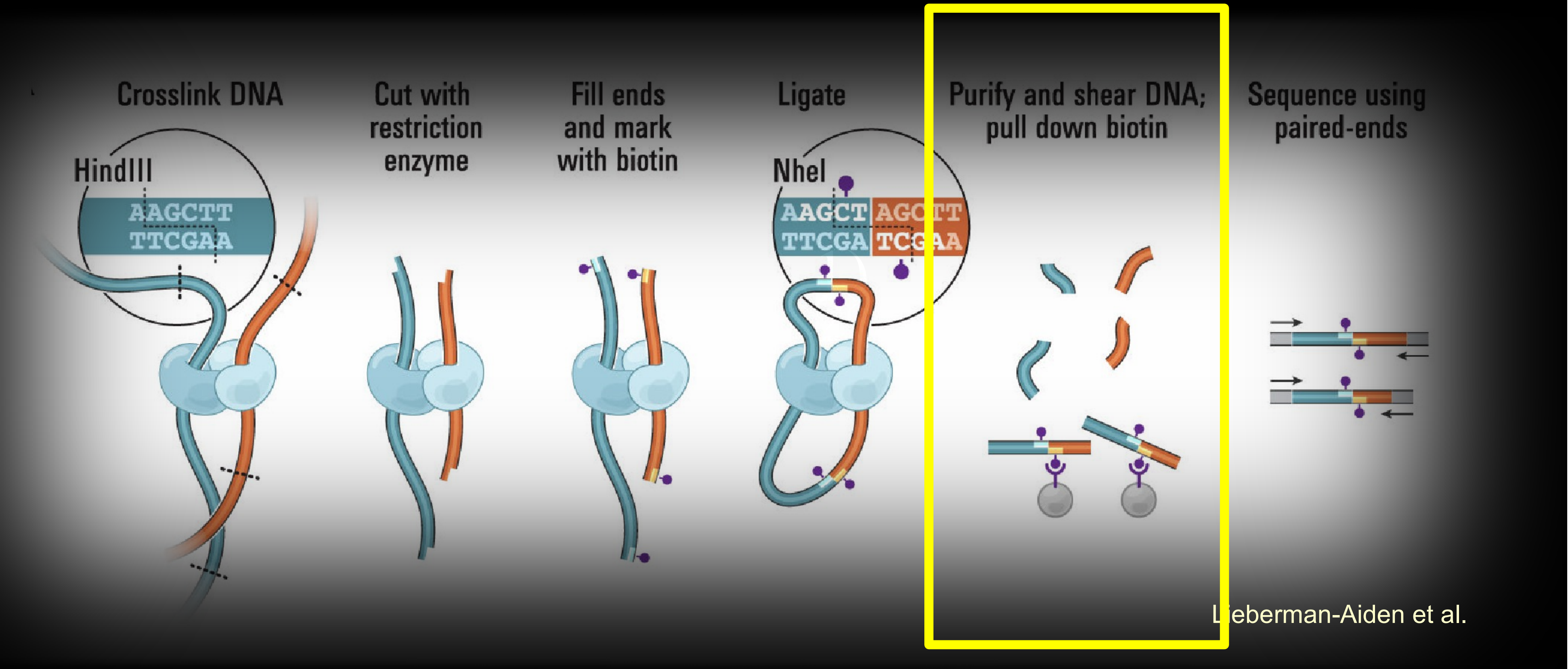

Crosslink DNA
HindIII
AAGCTT
TTCGAA
Cut with restriction enzyme
Fill ends and mark with biotin
Ligate
NheI
AAGCT AGCTT
TTCGA TCGAA
Purify and shear DNA; pull down biotin
Sequence using paired-ends

Lieberman-Aiden et al.

1) crosslink

2) ligate

Data: Hi-C


Crosslink DNA
HindIII
AAGCTT
TTCGAA
Cut with restriction enzyme
Fill ends and mark with biotin
Ligate
NheI
AAGCT AGCTT
TTCGA TCGAA
Purify and shear DNA; pull down biotin
Sequence using paired-ends

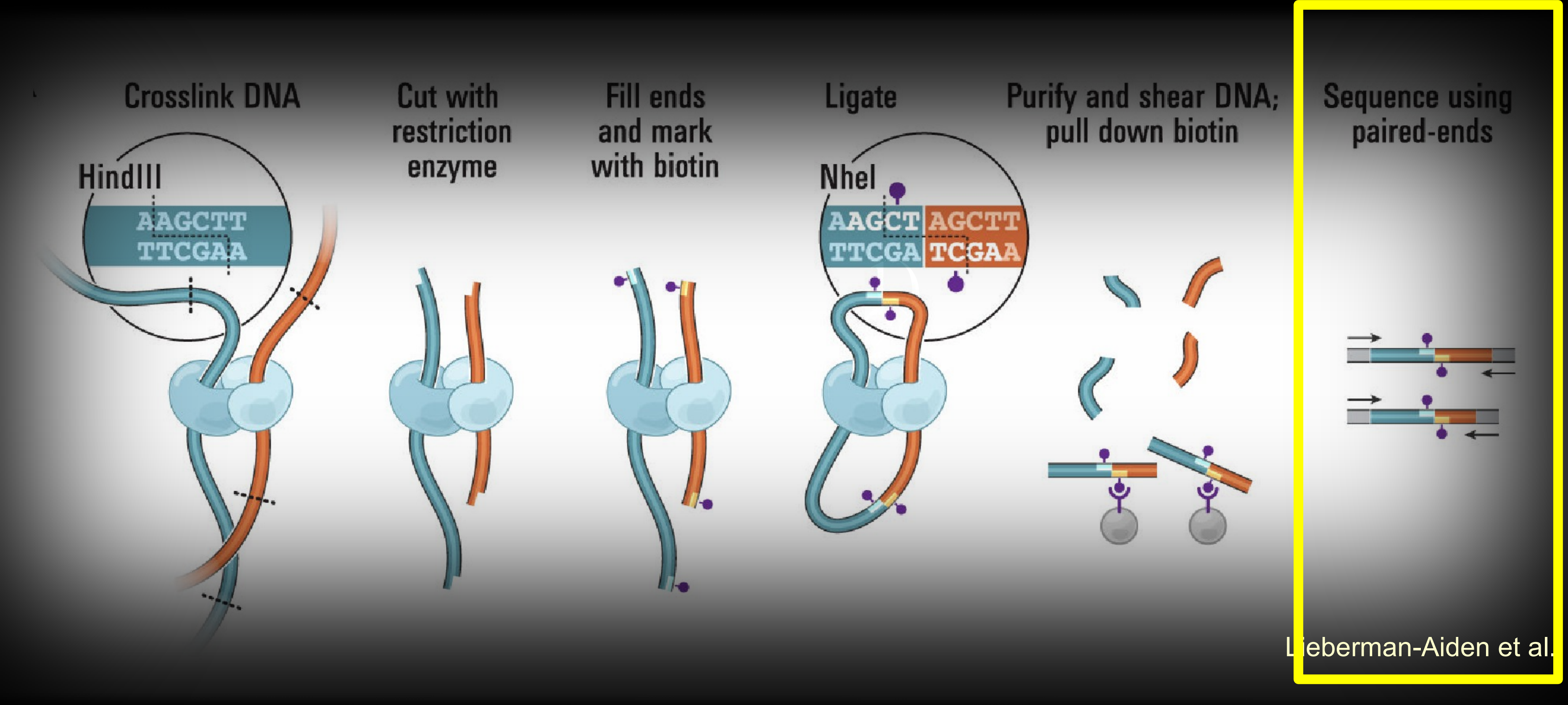
Lieberman-Aiden et al.

1) crosslink

2) ligate

3) shear

Data: Hi-C


Crosslink DNA
HindIII
AAGCTT
TTCGAA
Cut with restriction enzyme
Fill ends and mark with biotin
Ligate
NheI
AAGCT AGCTT
TTCGA TCGAA
Purify and shear DNA; pull down biotin
Sequence using paired-ends

Lieberman-Aiden et al.

1) crosslink

2) ligate

3) shear

4) sequence

Cell Lines

- 3 different labs
- 10 different experiments
  - 3 metaphase cell lines
  - 6 types of cells

Cell Lines

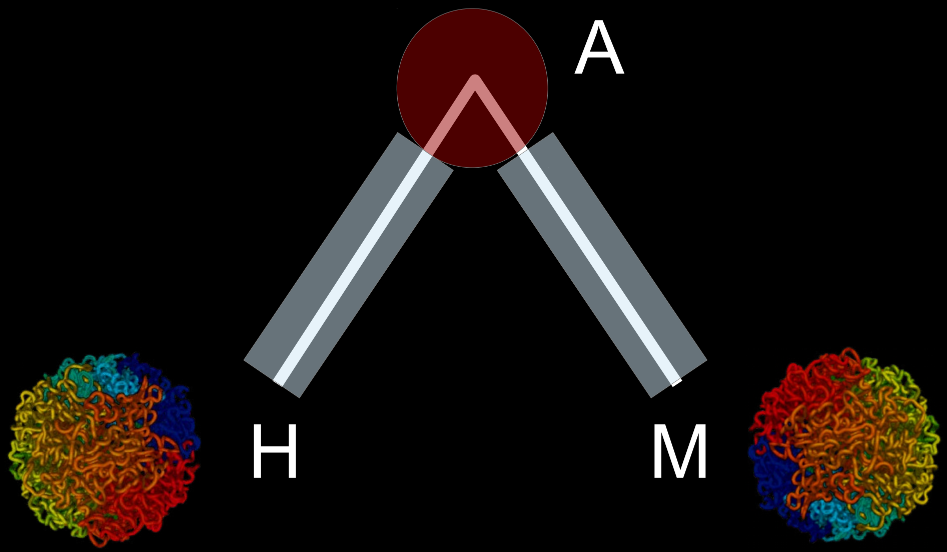
- 3 different labs
- 10 different experiments
  - 3 metaphase cell lines
  - 6 types of cells

We see significant similarities between all of them!

→ selection on breakpoints?

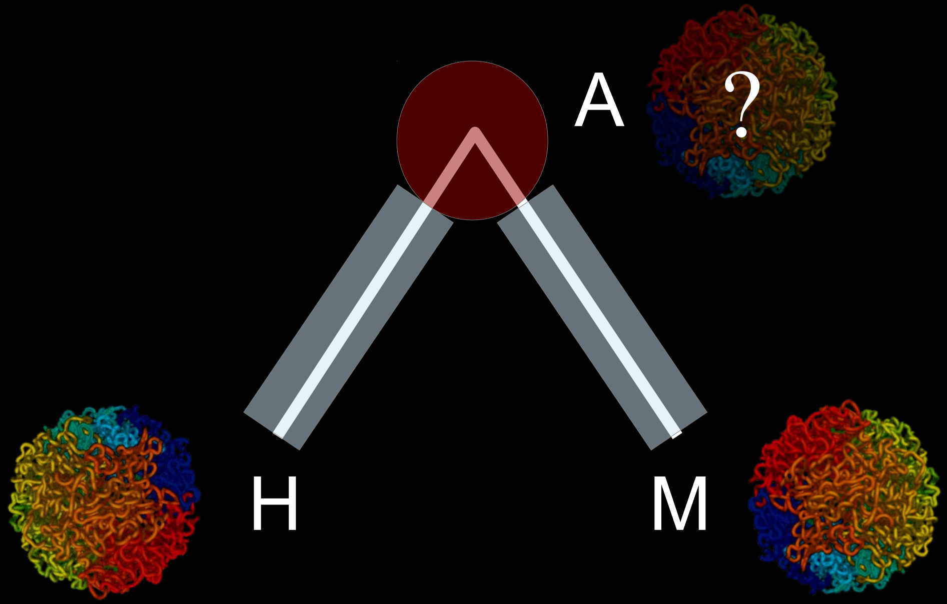
Directions

- Search for “local” scenarios.
- Use Mouse AND Human data.
- Place rearrangements on path to ancestor.

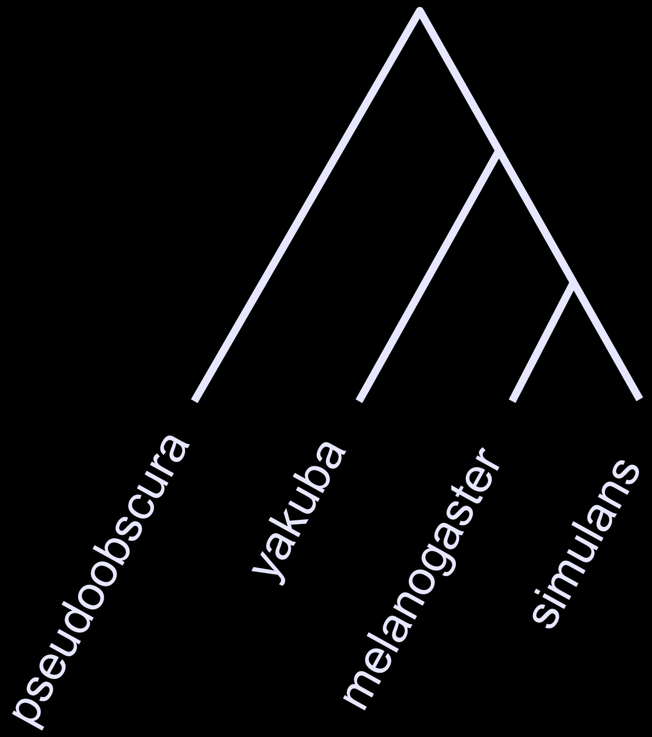


Directions

- Search for “local” scenarios.
- Use Mouse AND Human data.
- Place rearrangements on path to ancestor.



Directions



Drosophila!

(Giacomo Cavalli Lab)