

Chromosome communities in the human pangenome



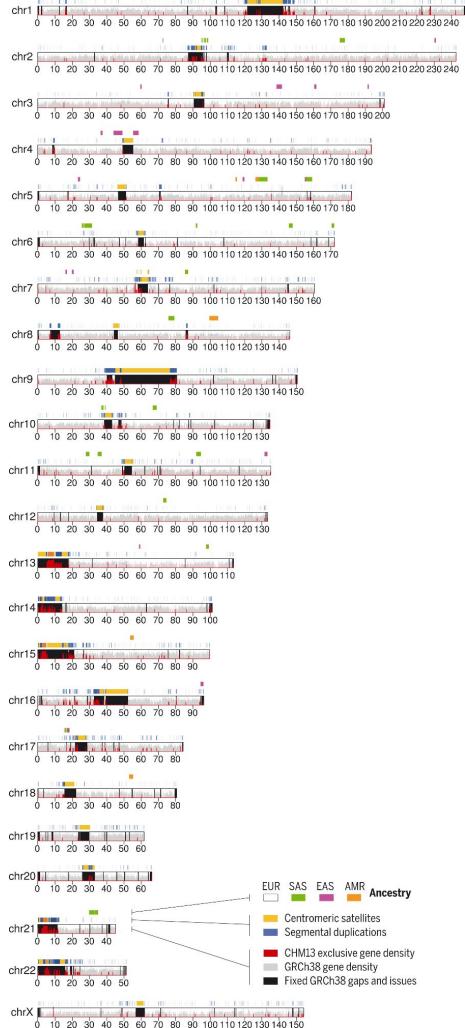
International Genome Graph Symposium 2022
Monte Verità, Ascona - 6 July 2022

Andrea Guaracino and Erik Garrison



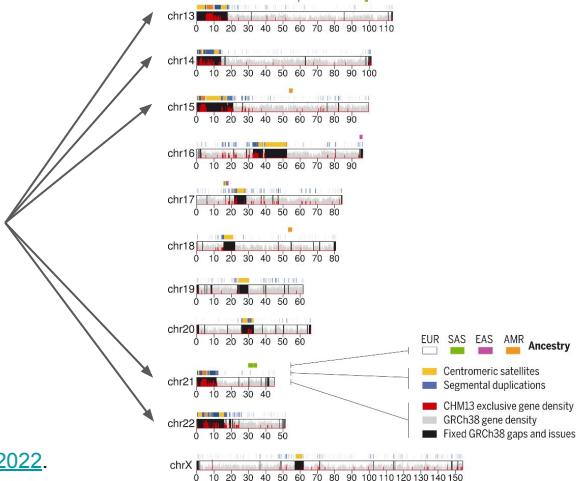
The complete sequence of a human genome

CHM13 filled 8% of the reference which was incomplete.

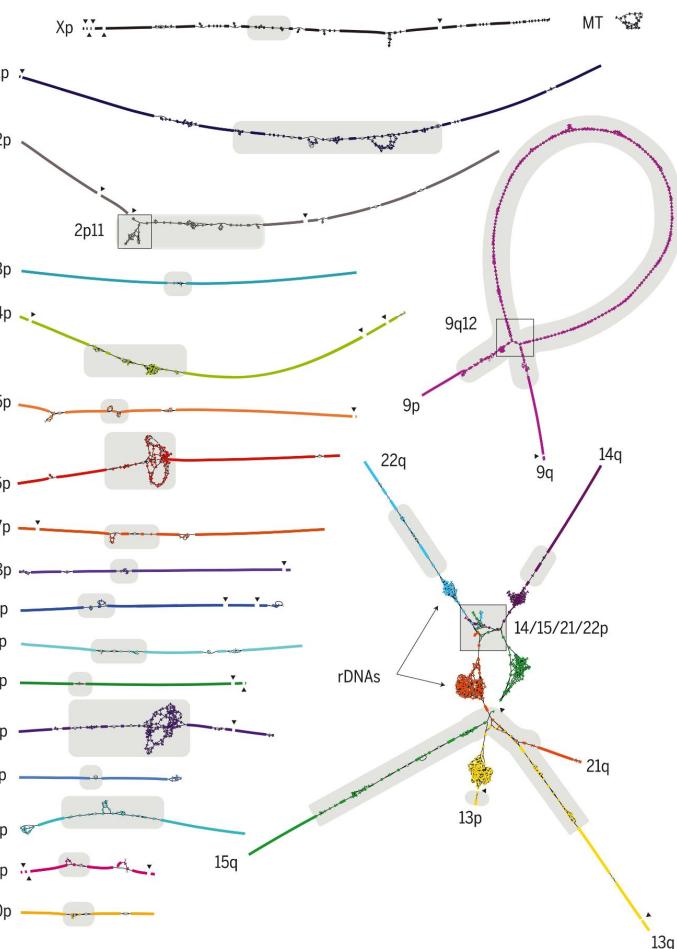


The complete sequence of a human genome

CHM13 filled 8% of the reference which was incomplete.

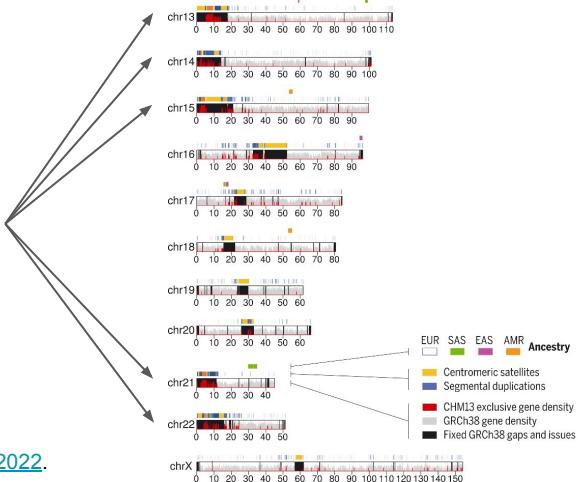


The acrocentric p-arms were assembled for the first time.

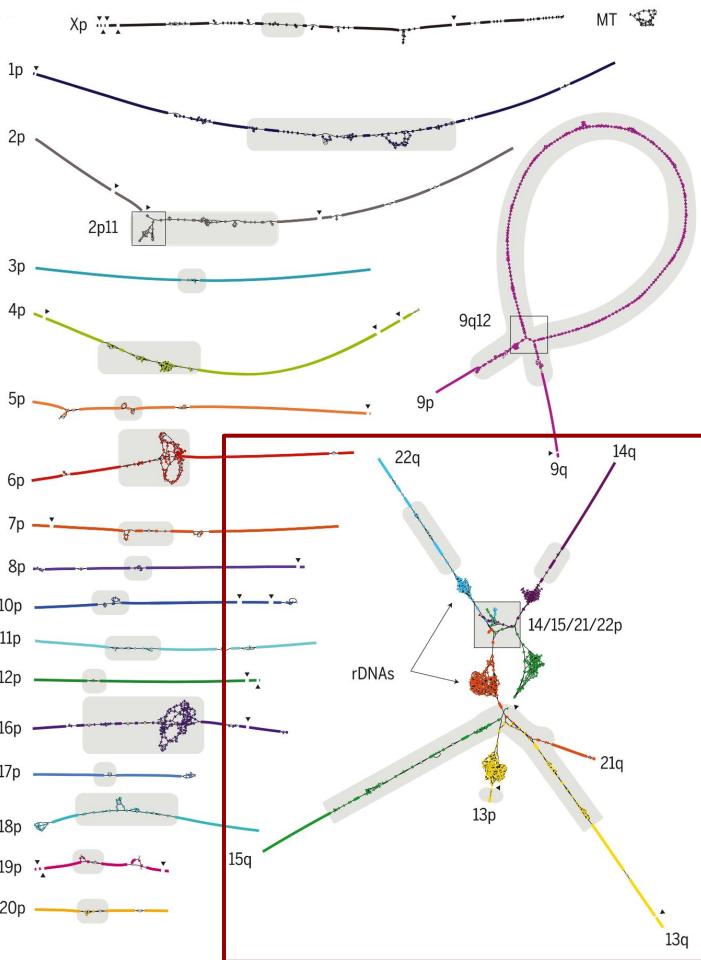


The complete sequence of a human genome

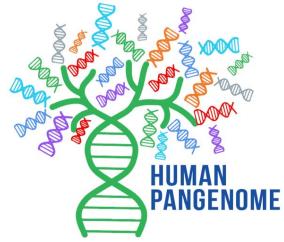
CHM13 filled 8% of the reference which was incomplete.



The acrocentric p-arms were assembled for the first time.

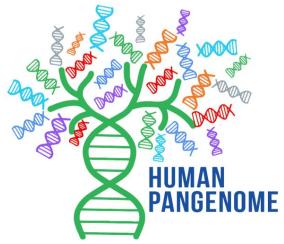


Community detection

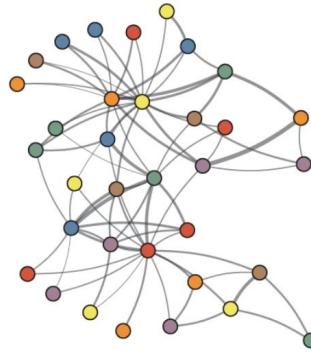


47 accurate and near-complete
haplotype-resolved human
genome assemblies

Community detection



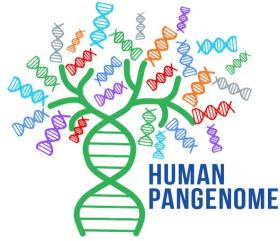
All-vs-all mapping



47 accurate and near-complete
haplotype-resolved human
genome assemblies

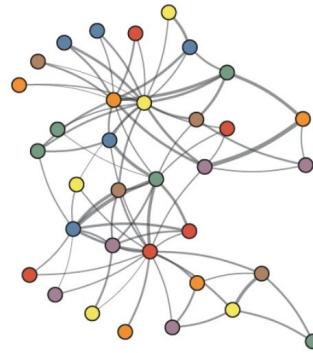
Mapping graph by keeping
only mappings between
contigs $\geq 1\text{Mbp}$

Community detection



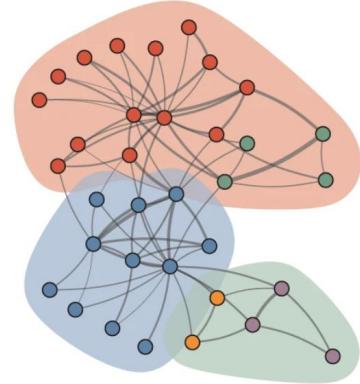
47 accurate and near-complete
haplotype-resolved human
genome assemblies

All-vs-all mapping



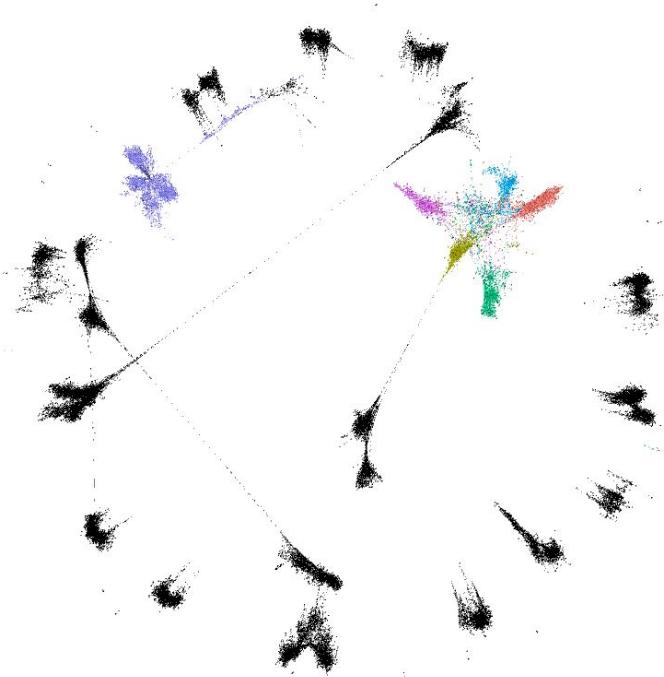
Mapping graph by keeping
only mappings between
contigs $\geq 1\text{Mbp}$

Community
detection

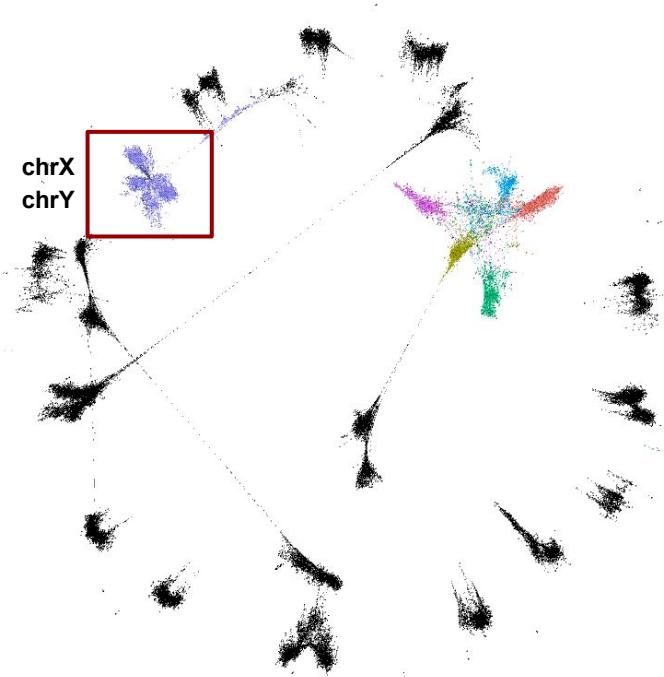


Community structure

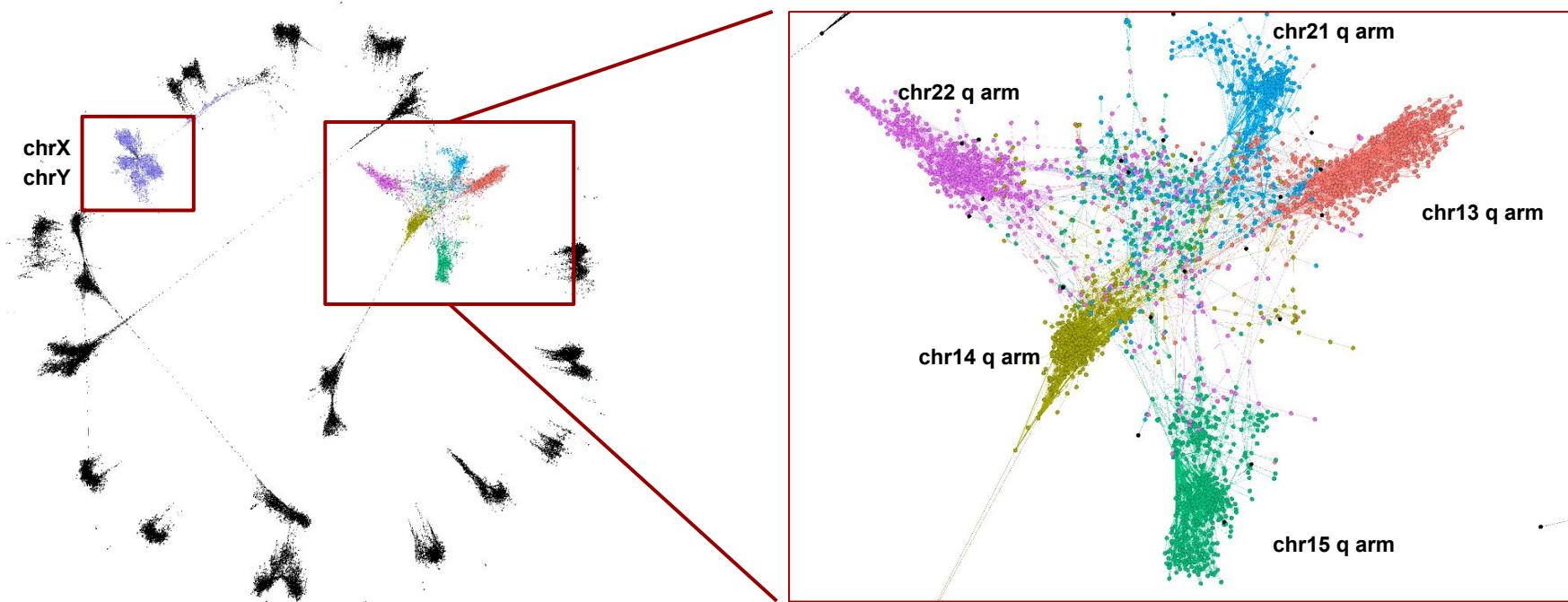
Community detection



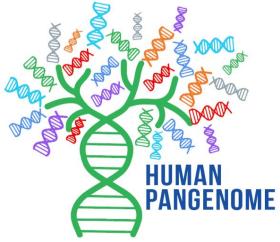
Community detection



Community detection



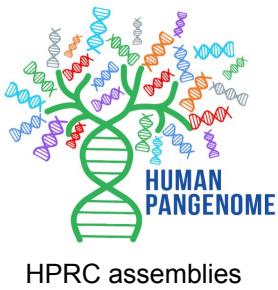
Workflow



HPRC assemblies

We decided to take a closer look, focusing
on the best assemblies in these regions.

Workflow

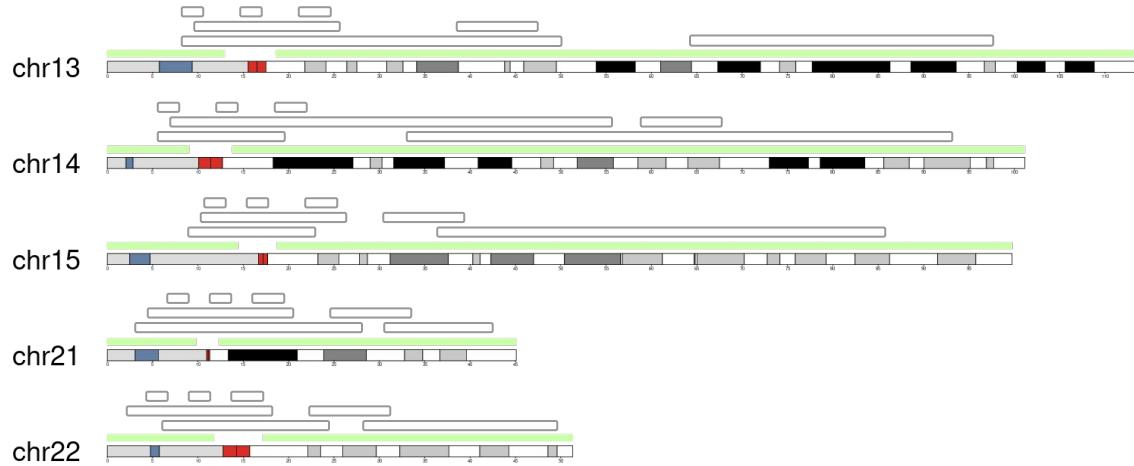


HPRC assemblies

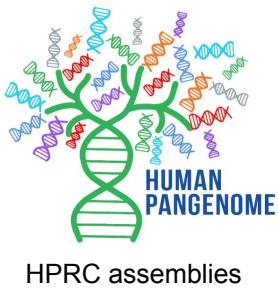


Mapping against
the whole CHM13

We decided to take a closer look, focusing
on the best assemblies in these regions.



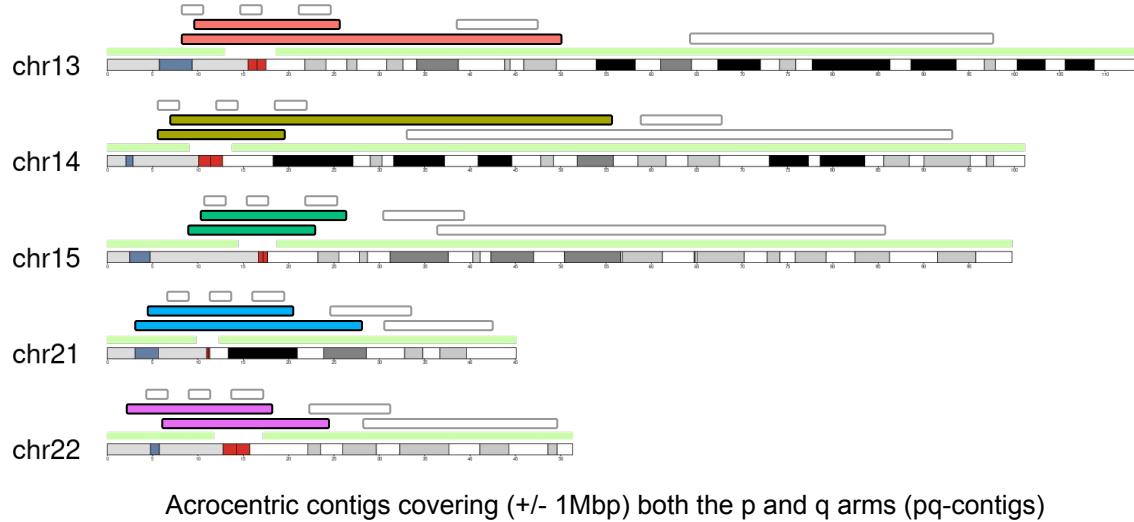
Workflow



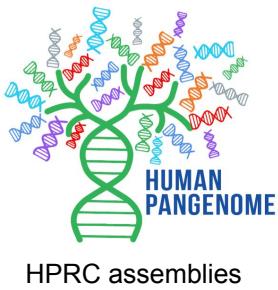
HPRC assemblies

Mapping against
the whole CHM13

We decided to take a closer look, focusing
on the best assemblies in these regions.



Workflow

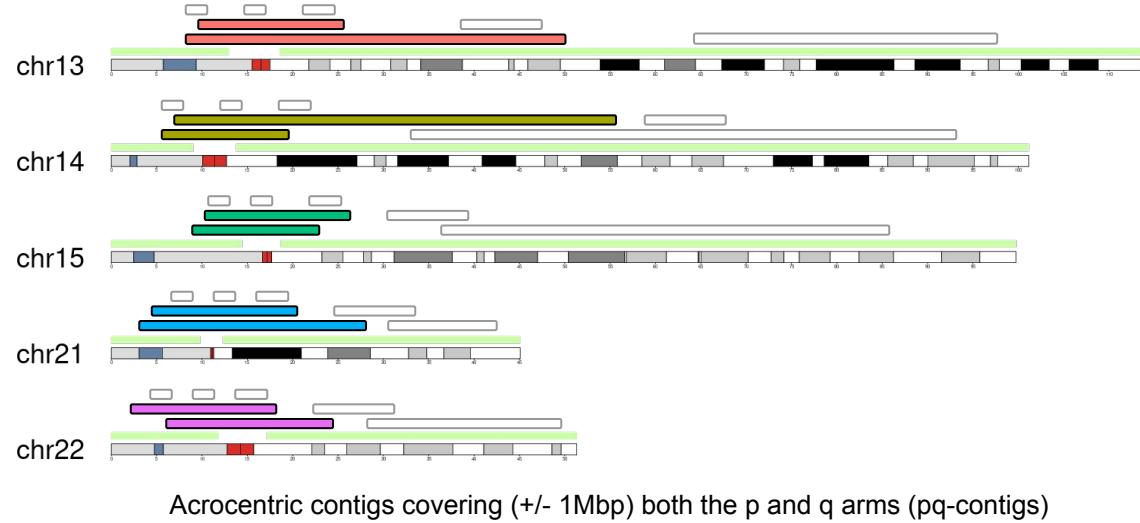


HPRC assemblies

We decided to take a closer look, focusing on the best assemblies in these regions.



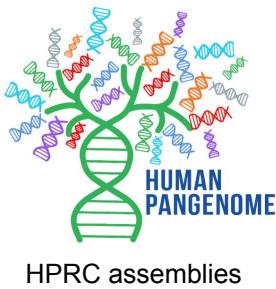
Mapping against
the whole CHM13



PanGenome Graph
Builder (PGGB)

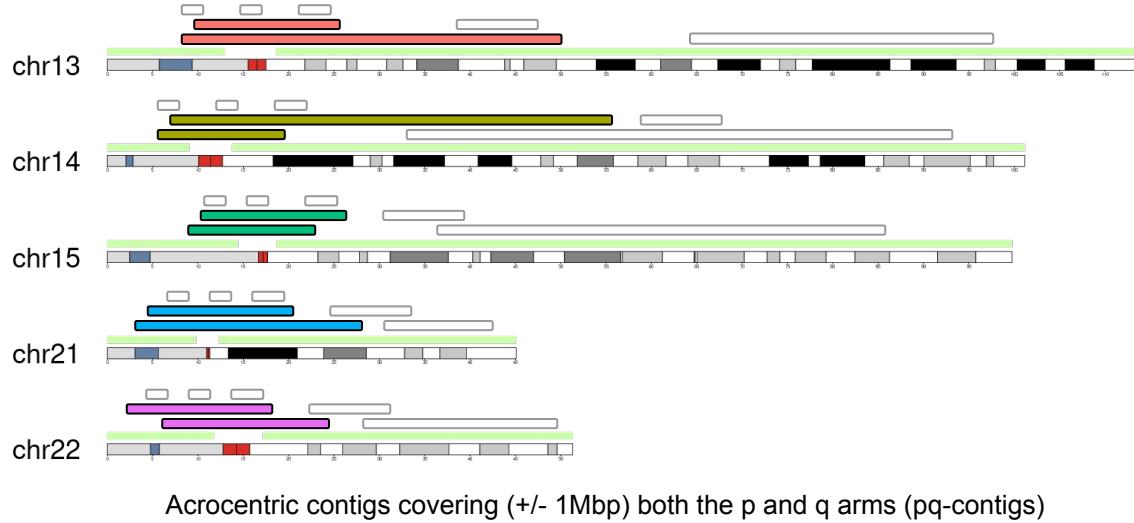
<https://github.com/pangenome/pggb>

Workflow

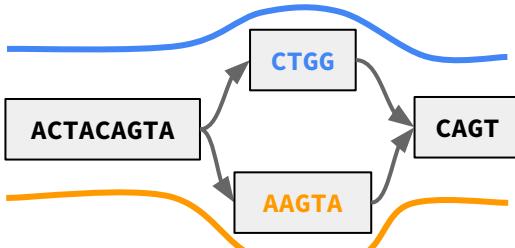


Mapping against
the whole CHM13

We decided to take a closer look, focusing
on the best assemblies in these regions.



- Genome 1: ACTACAGT**A**CT**G**GCA**T**
- Genome 2: ACTACAGTA**A**AGTA**C**AGT

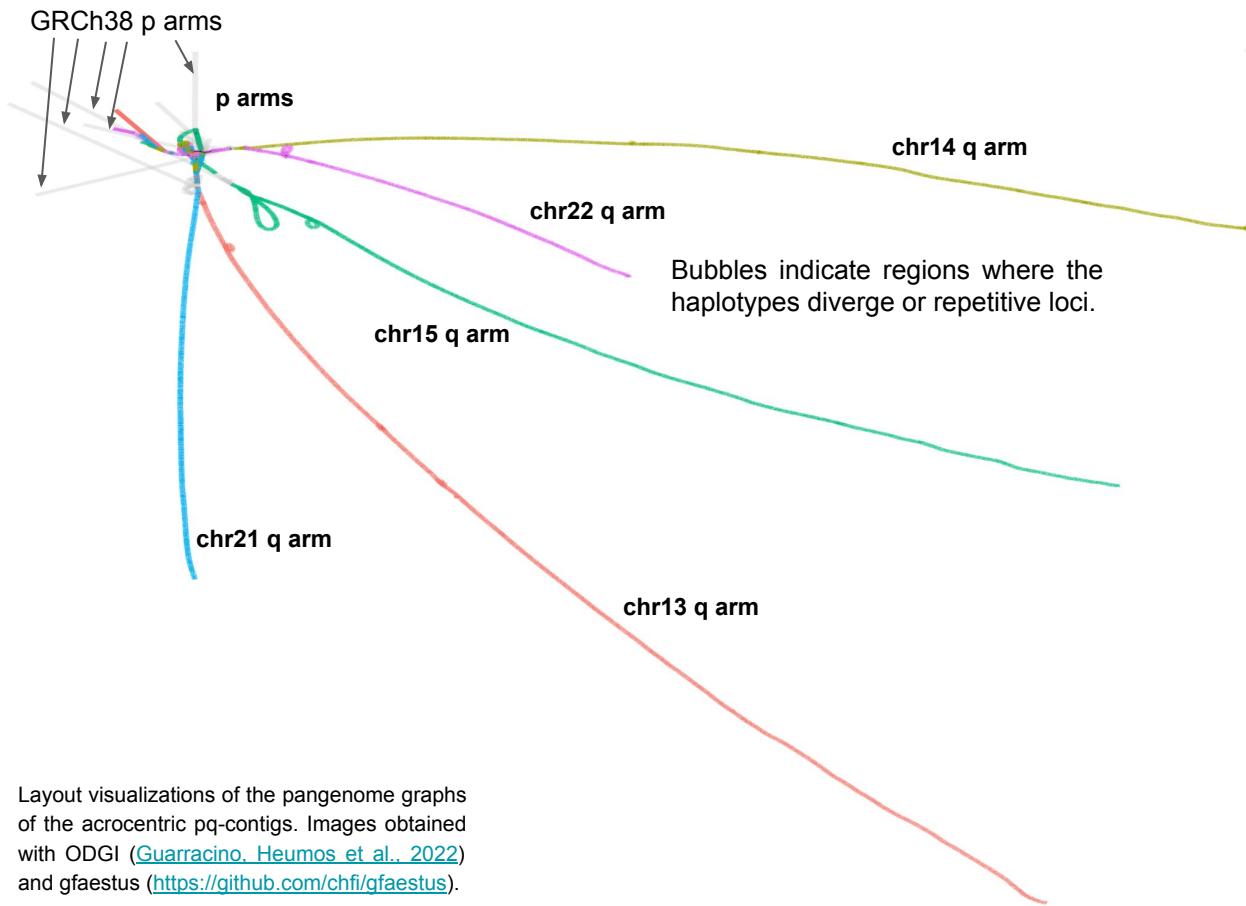


PanGenome Graph
Builder (PGGB)

<https://github.com/pangenome/pggb>

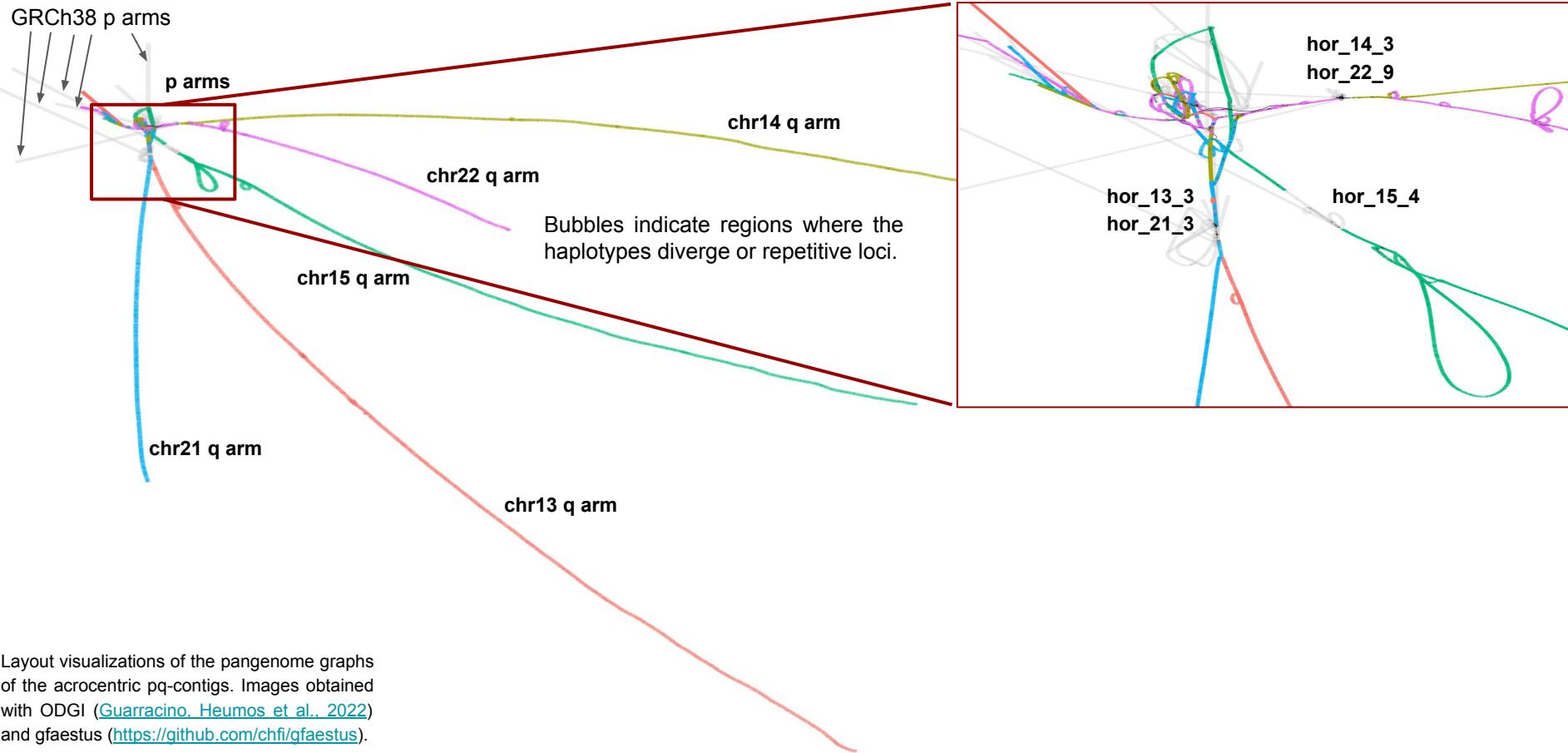
This provides an all-to-all alignment model that is non-redundant (one graph) and highly sensitive (transitive relationships are captured).

Acrocentric pq-contig pangenome graph

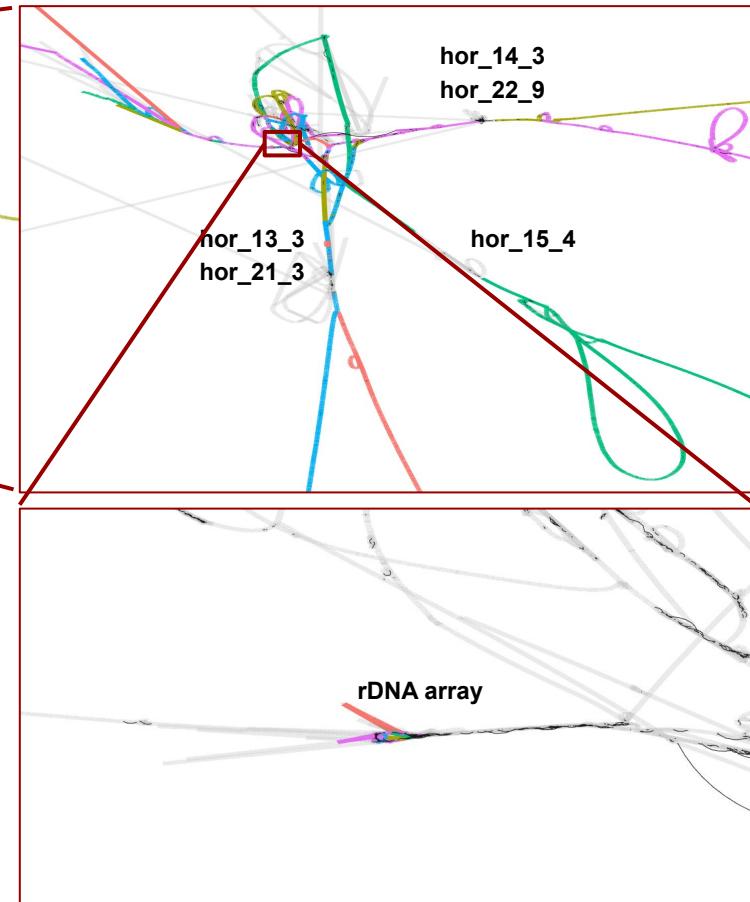
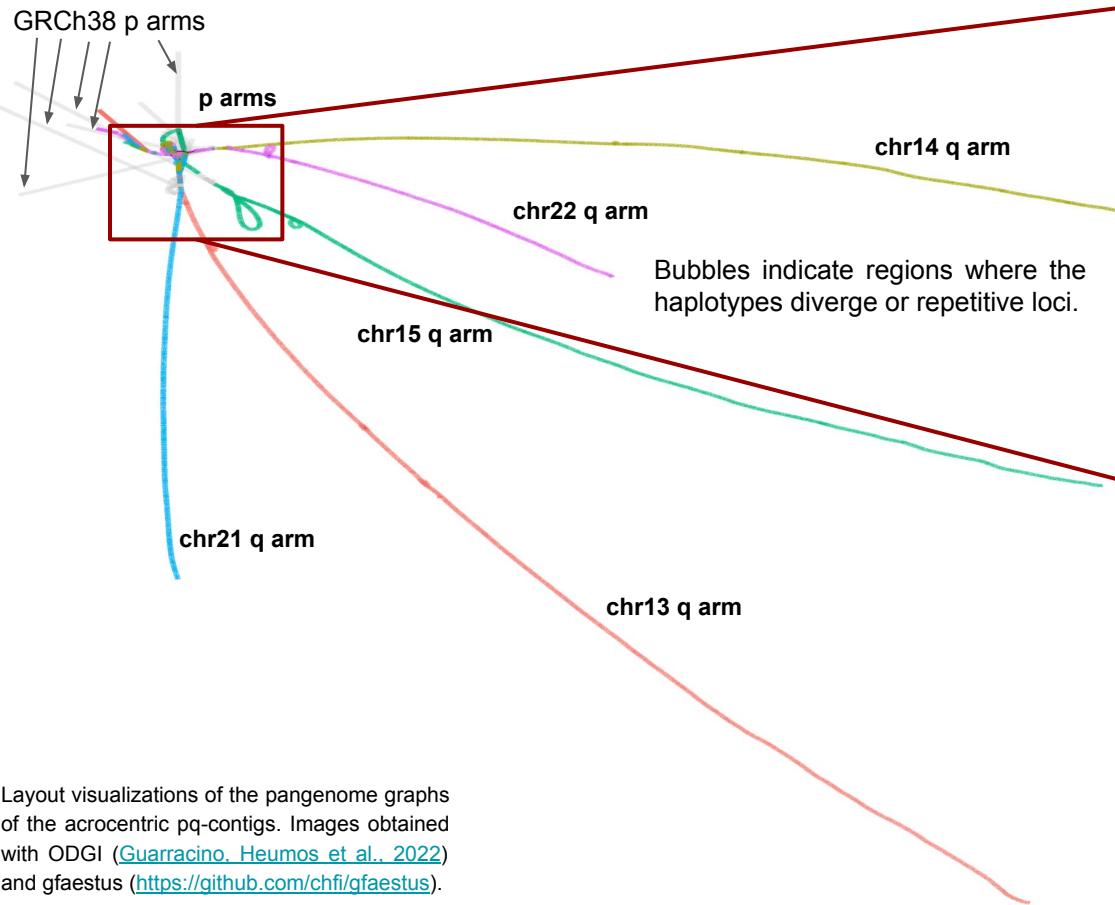


The pq-contig set produces the same graph structure we observed in the full HPRC acrocentric graph.

Acrocentric pq-contig pangenome graph

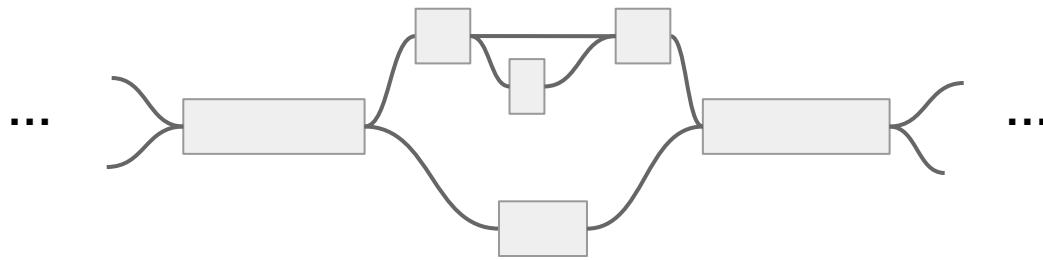


Acrocentric pq-contig pangenome graph



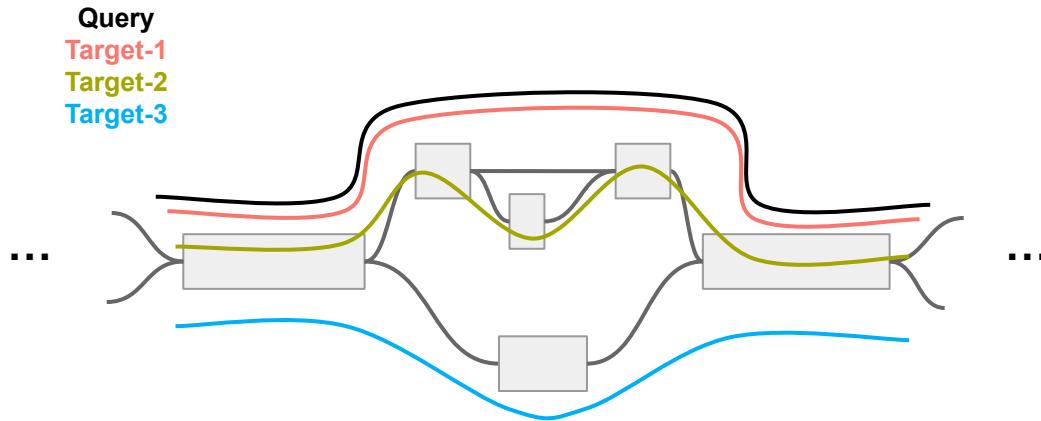
Pangenome untangling

Untangling extracts pairwise alignments from variation graphs.



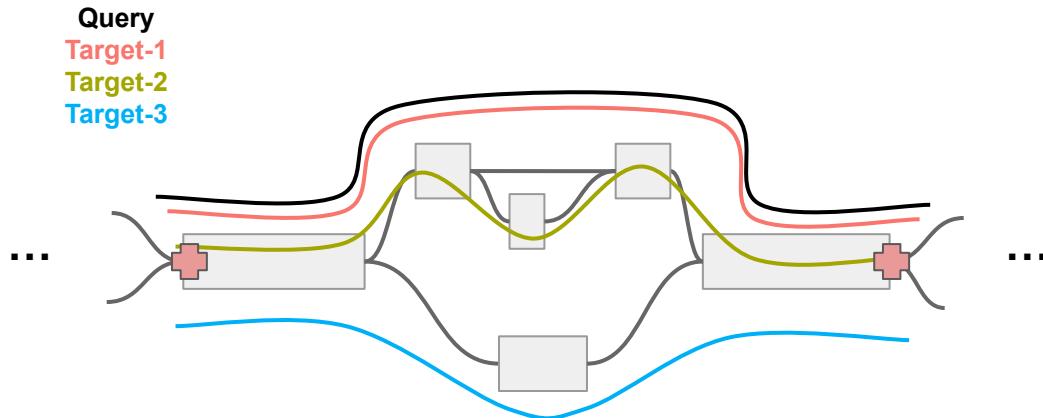
Pangenome untangling

Untangling extracts pairwise alignments from variation graphs.



Pangenome untangling

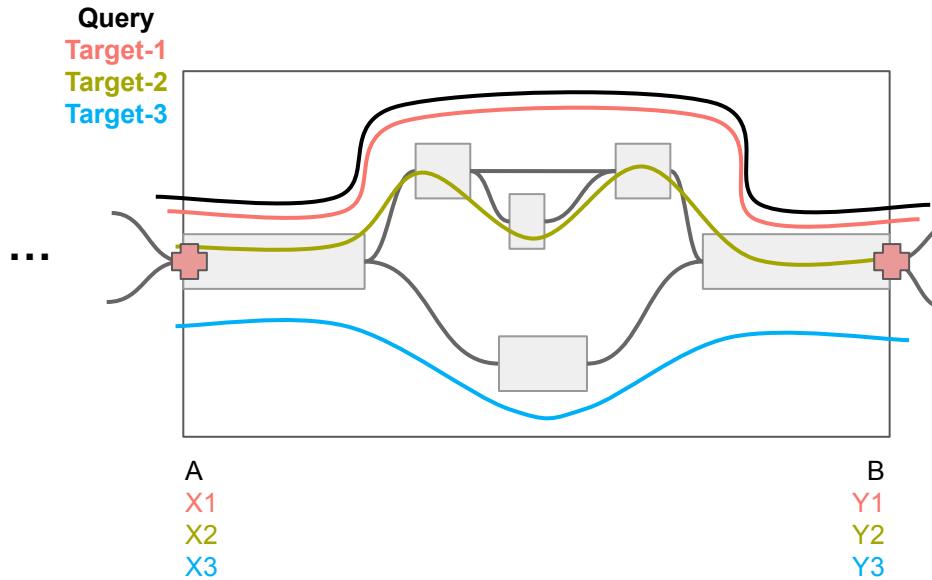
Untangling extracts pairwise alignments from variation graphs.



Identify cut points in
the graph

Pangenome untangling

Untangling extracts pairwise alignments from variation graphs.

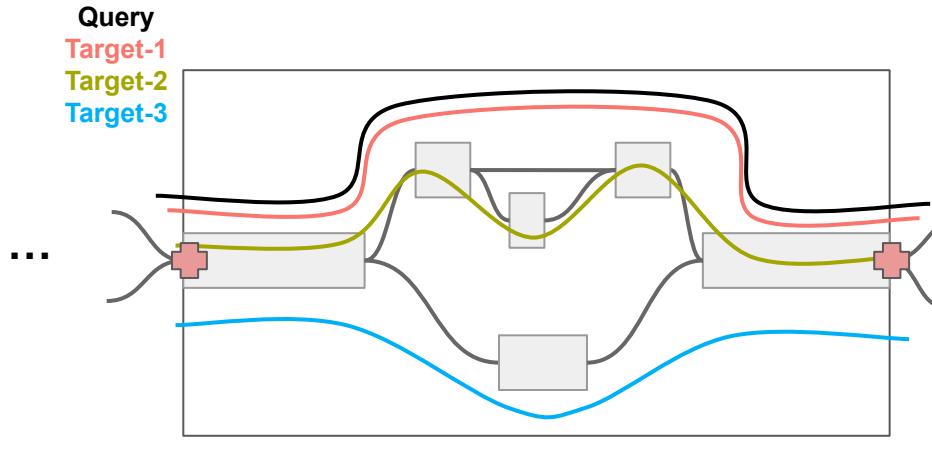


Identify cut points in
the graph

Define segment
boundaries

Pangenome untangling

Untangling extracts pairwise alignments from variation graphs.



query	start	end	target	start	end	jaccard	rank
Query	A	B	Target-1	X1	Y1	1	1
Query	A	B	Target-2	X2	Y2	0.95	2
Query	A	B	Target-3	X3	Y3	0.7	3

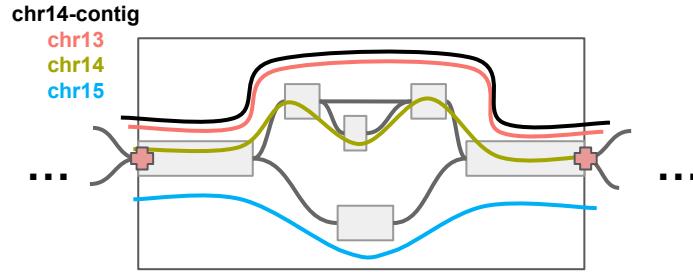
Identify cut points in the graph

Define segment boundaries

Compare segments

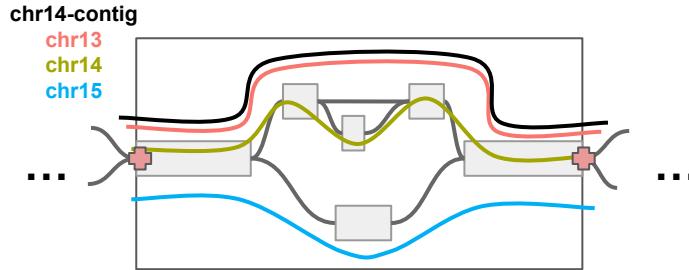
with jaccard in node (allele) space

Pangenome untangling: grounding



query	start	end	target	start	end	jaccard	rank
chr14-contig	A	B	chr13	X1	Y1	1	1
chr14-contig	A	B	chr14	X2	Y2	0.95	2
chr14-contig	A	B	chr15	X3	Y3	0.7	3

Pangenome untangling: grounding

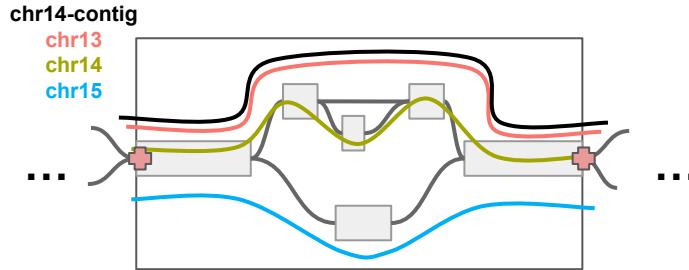


query	start	end	target	start	end	jaccard	rank
chr14-contig	A	B	chr13	X1	Y1	1	1
chr14-contig	A	B	chr14	X2	Y2	0.95	2
chr14-contig	A	B	chr15	X3	Y3	0.7	3

query	start	end	target	start	end	jaccard	rank
chr14-contig	A	B	chr14	X2	Y2	0.95	1

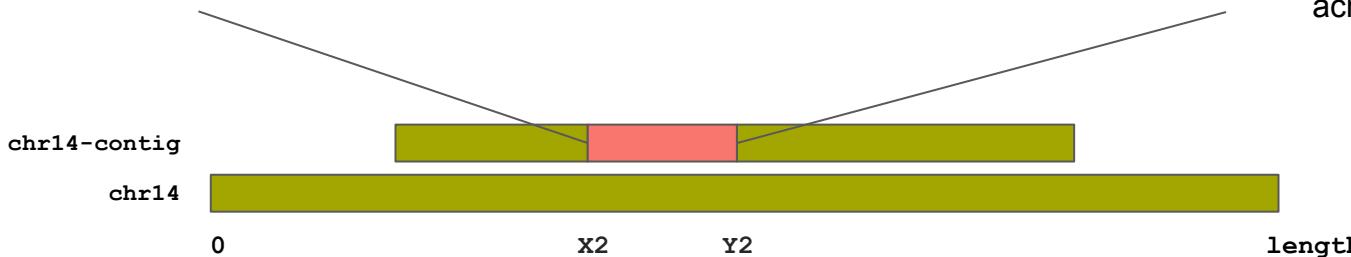
For each query segment, we find its best mapping against a specific acrocentric chromosome.

Pangenome untangling: grounding



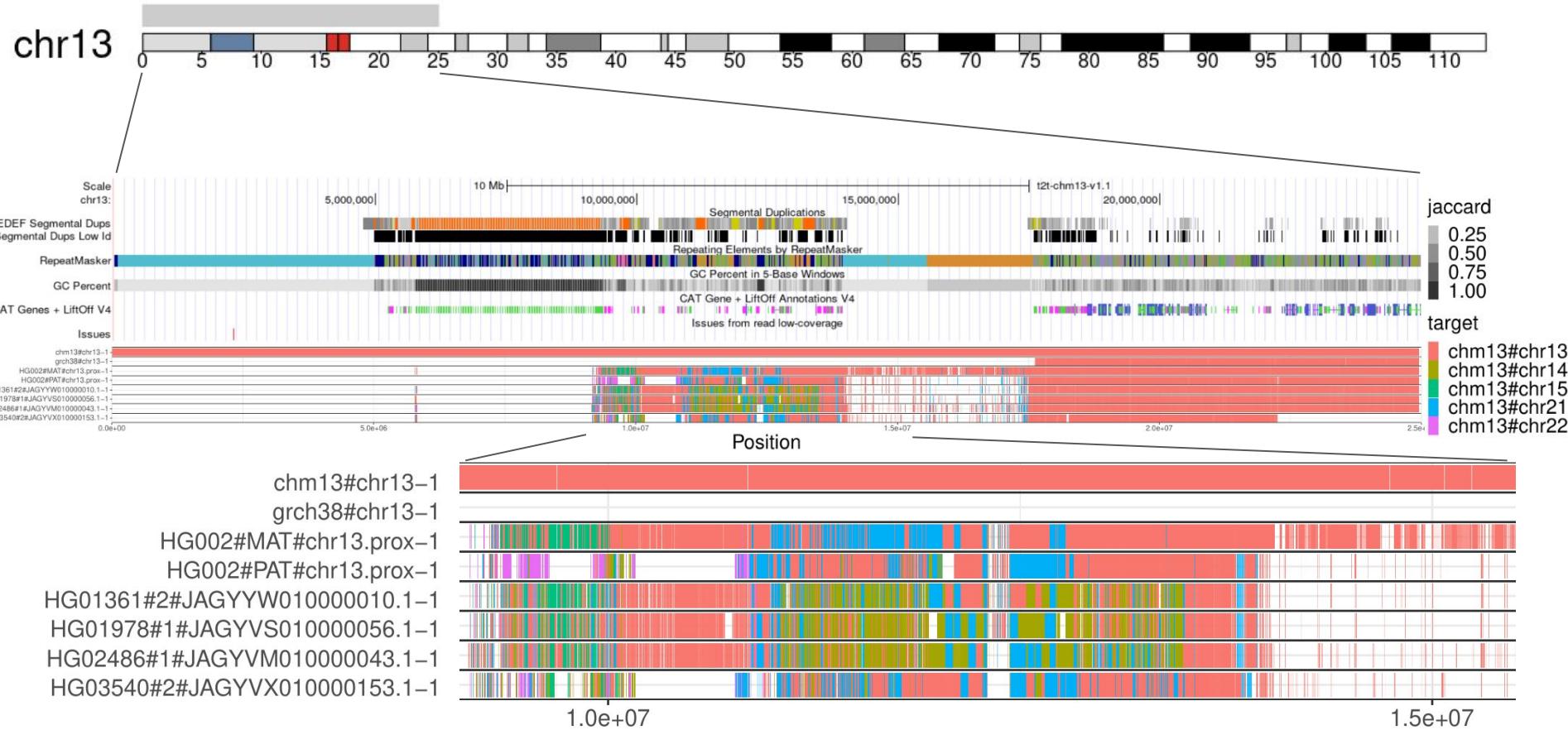
query	start	end	target	start	end	jaccard	rank
chr14-contig	A	B	chr13	X1	Y1	1	1
chr14-contig	A	B	chr14	X2	Y2	0.95	2
chr14-contig	A	B	chr15	X3	Y3	0.7	3

query	start	end	target	start	end	jaccard	rank
chr14-contig	A	B	chr14	X2	Y2	0.95	1

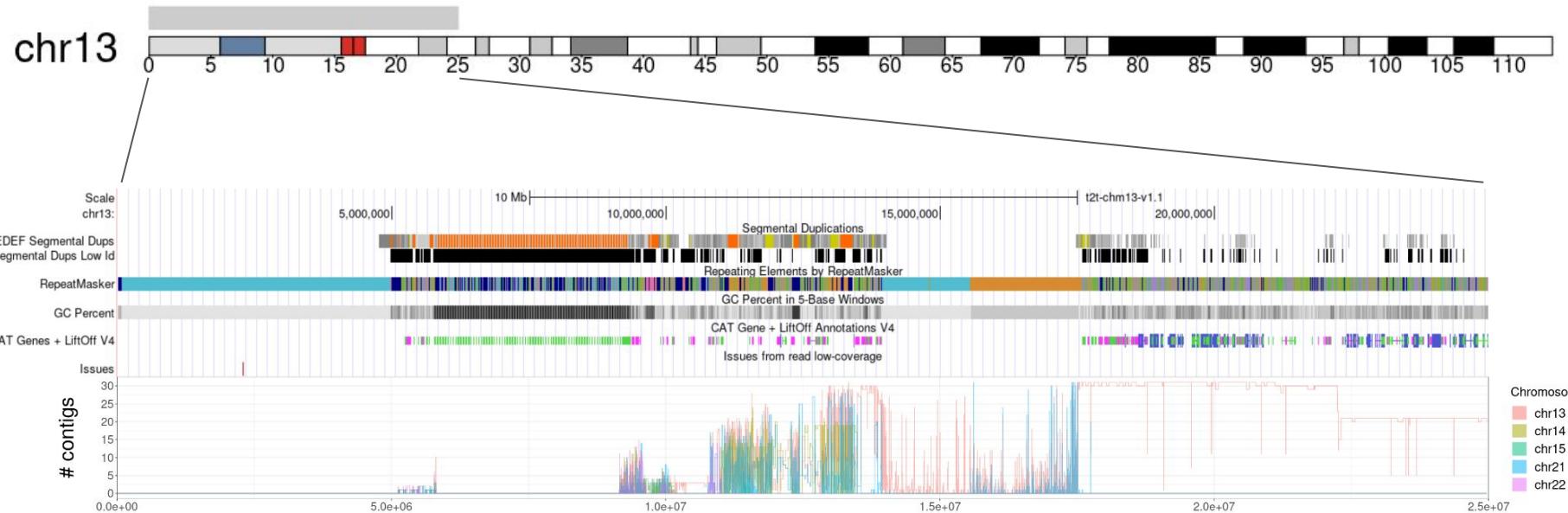


For each query segment, we find its best mapping against a specific acrocentric chromosome.

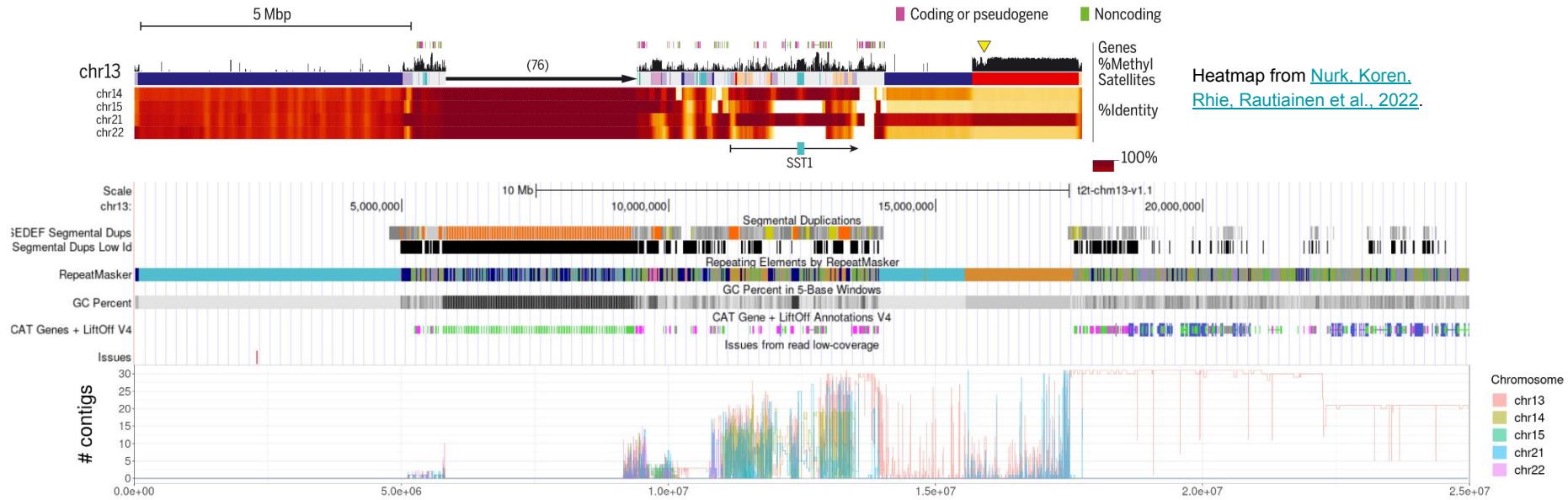
Untangling results - By contig - chr13



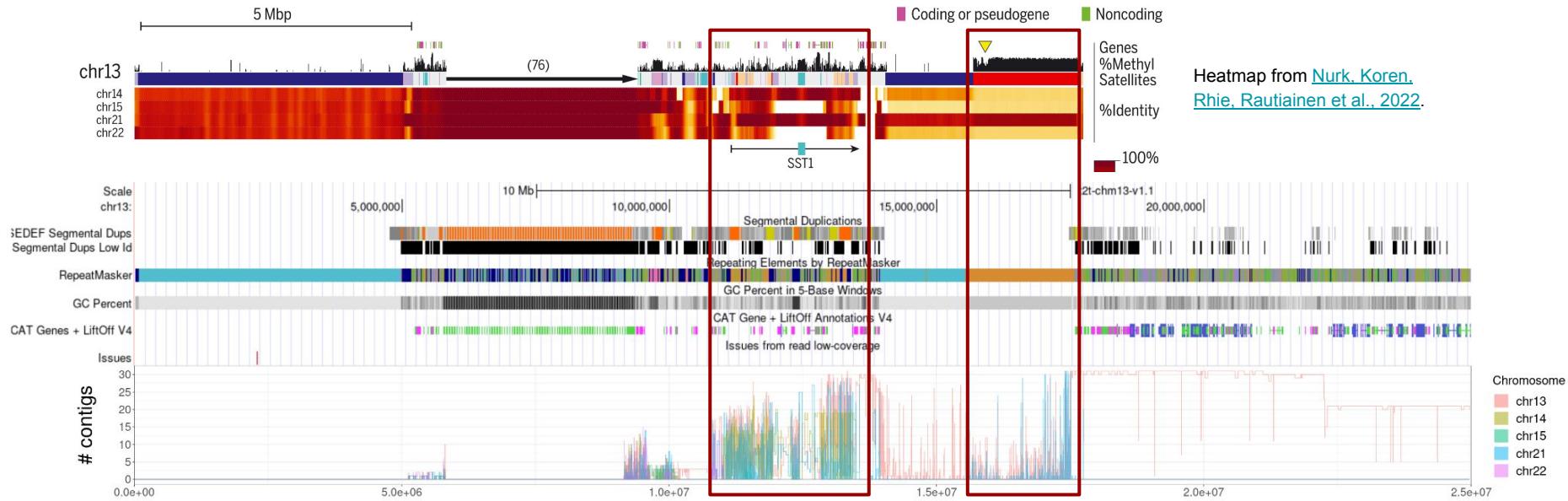
Untangling results - Aggregated - chr13



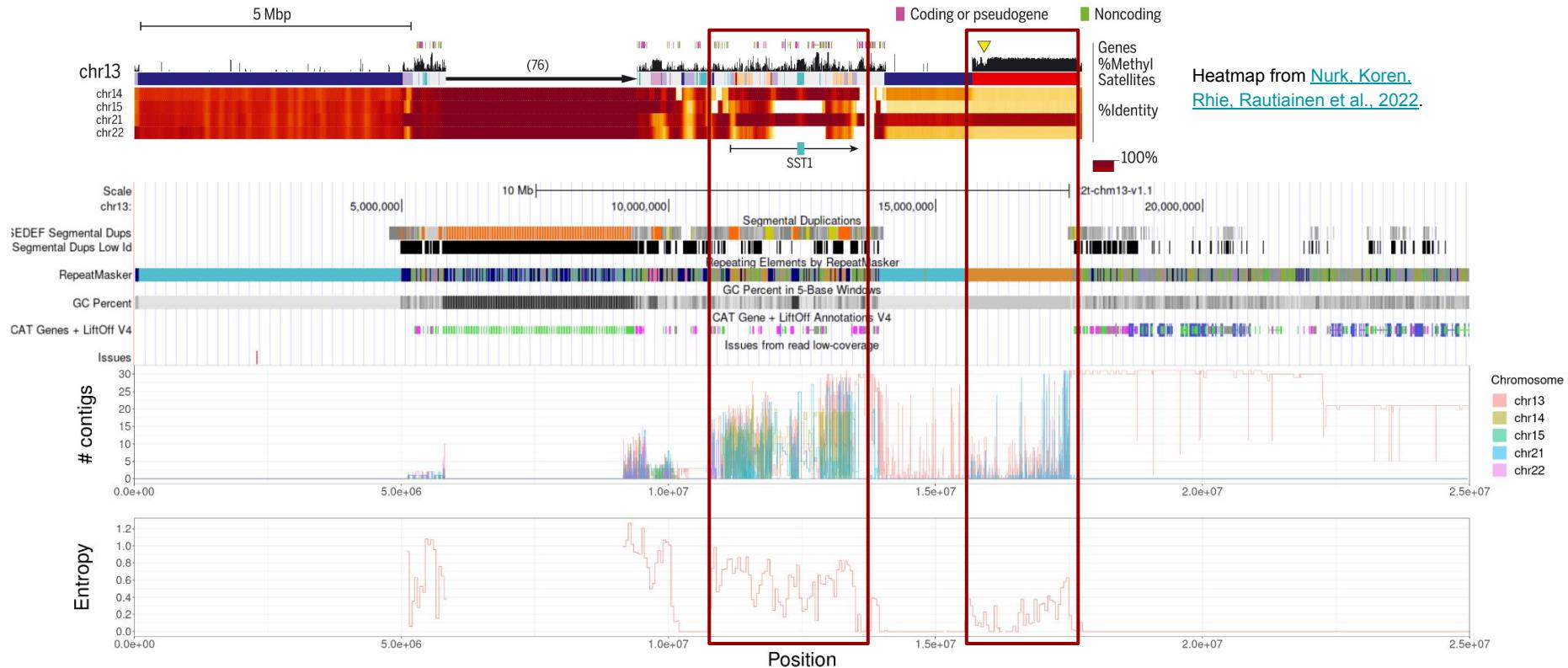
Untangling results - Aggregated - chr13



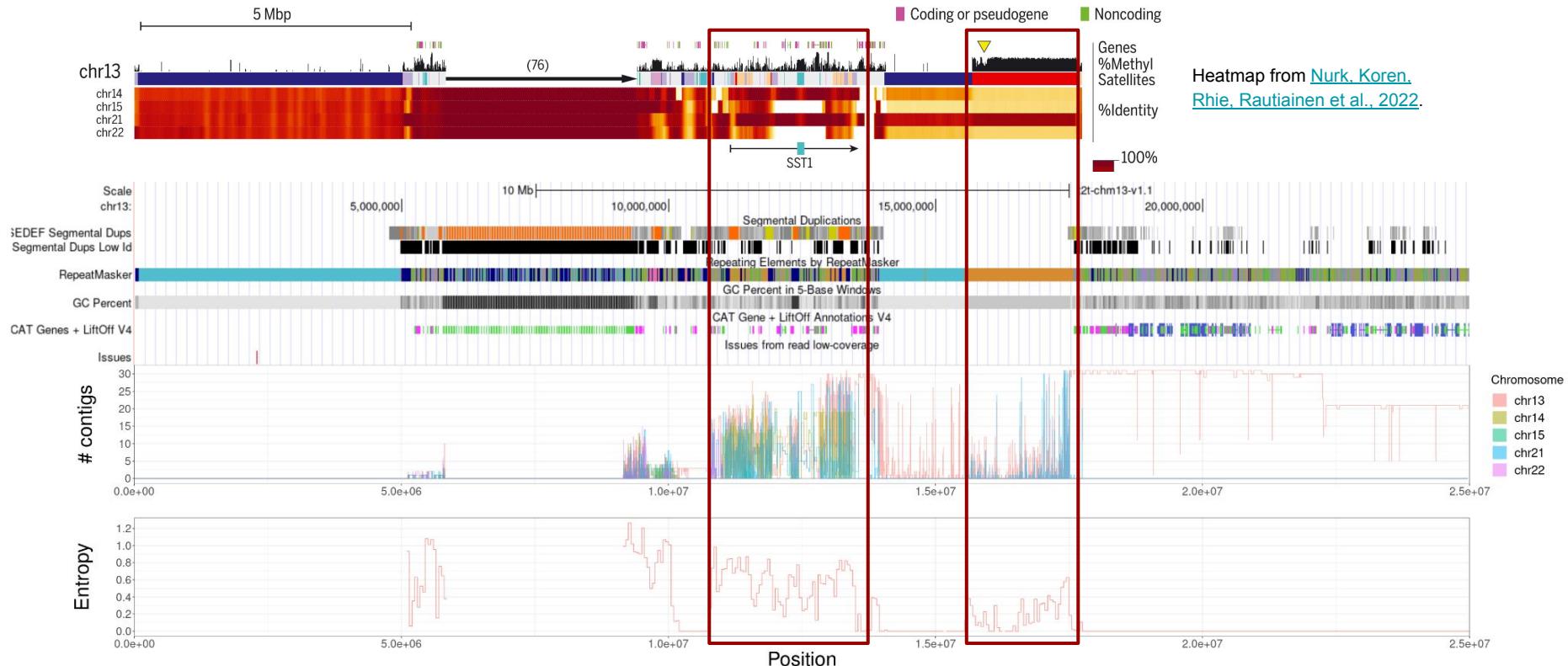
Untangling results - Aggregated - chr13



Untangling results - Aggregated - chr13



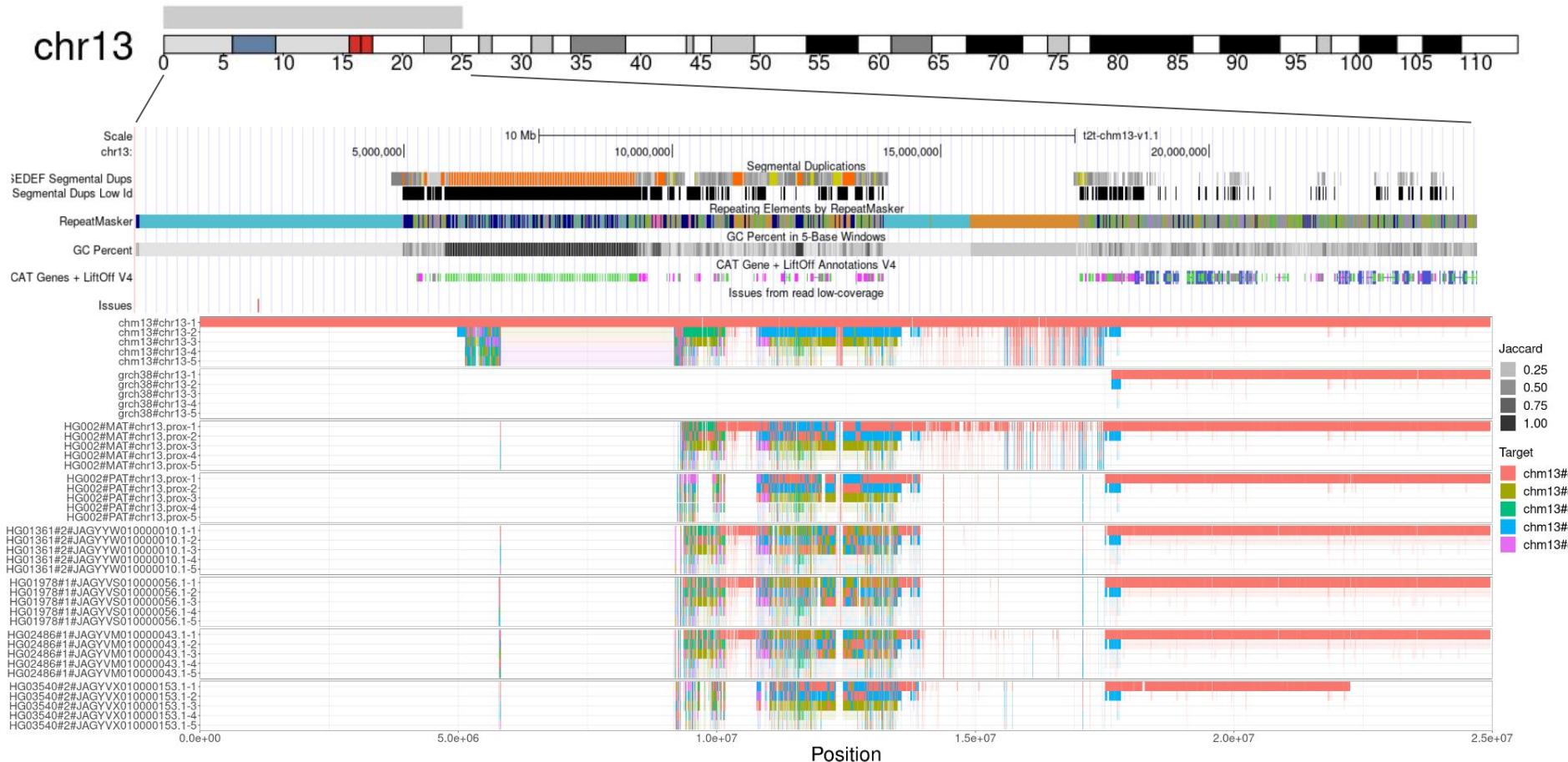
Untangling results - Aggregated - chr13



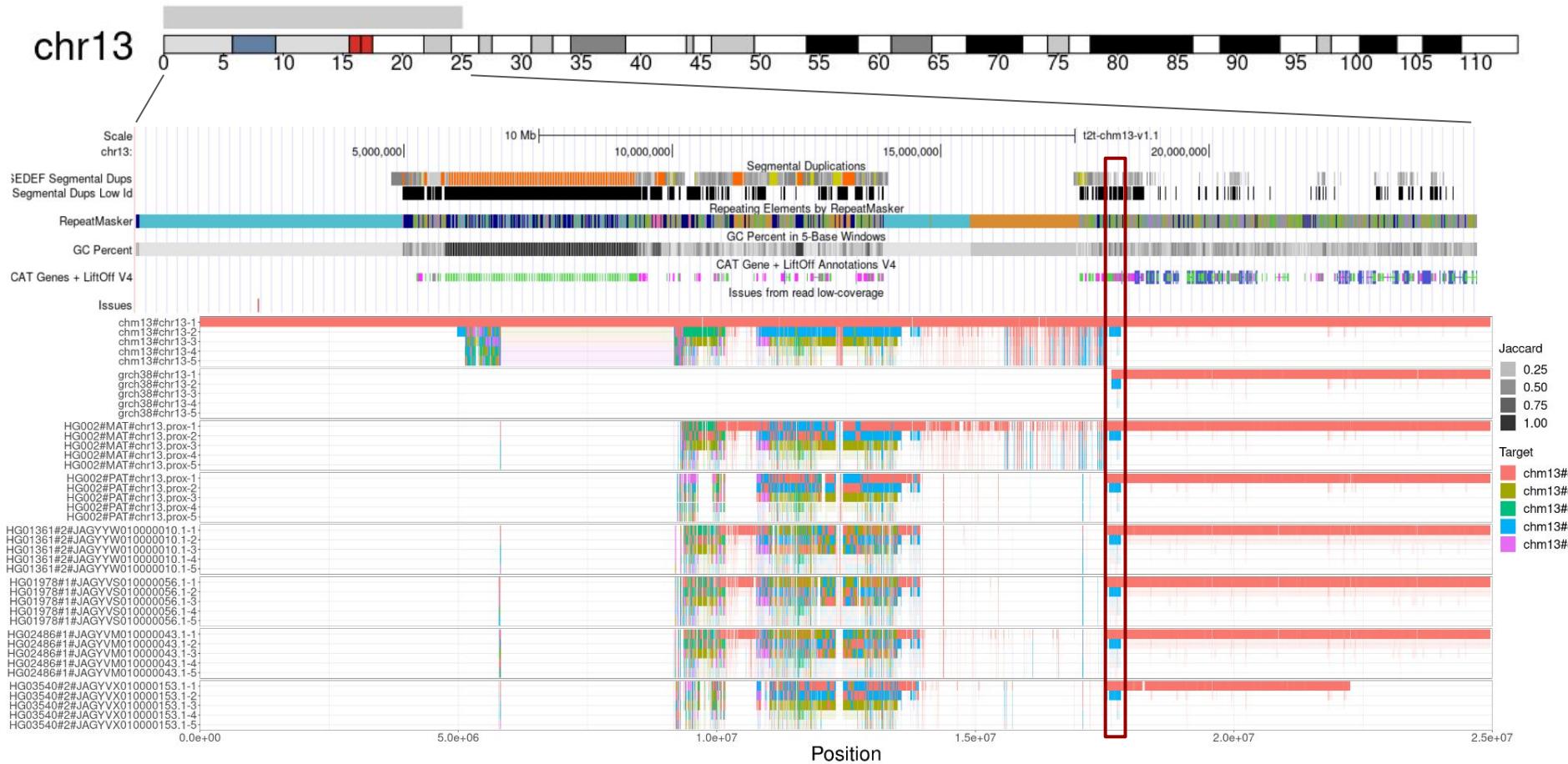
Heatmap from [Nurk, Koren, Rie, Rautiainen et al., 2022](#).

Increasing untangle entropy as we go towards the rDNA indicates sequence exchange between “non-homologous” chromosomes.

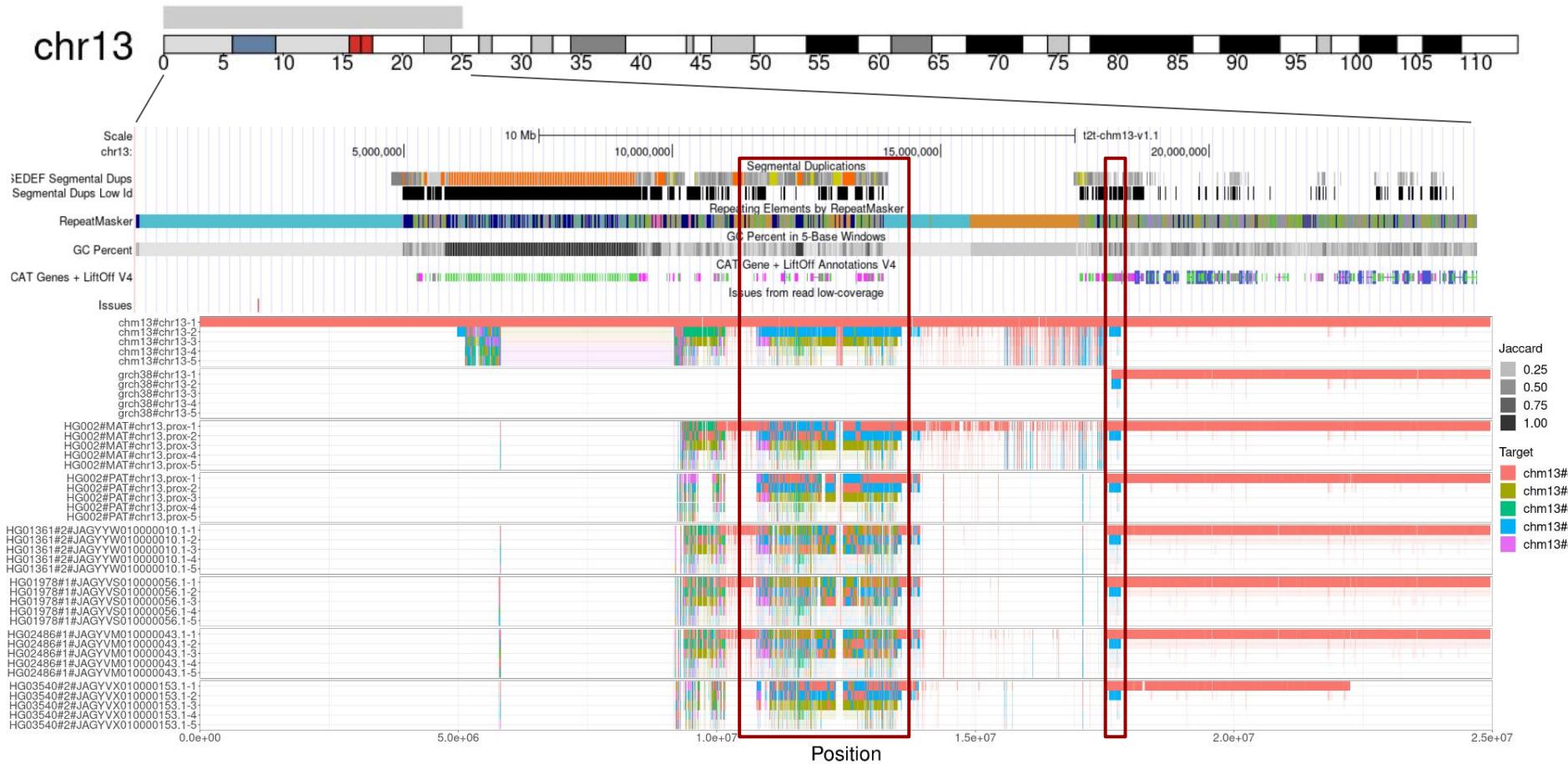
Untangling results - By contig - chr13 - 5 best hits



Untangling results - By contig - chr13 - 5 best hits



Untangling results - By contig - chr13 - 5 best hits



Towards traces of recombination

The high level of homology of the acrocentric chromosomes could be due to **non-homologous recombination**.

High-quality *de novo* assemblies and pangenomic approaches will shed light on the most difficult regions of the human genomes.

Volume 16 Number 4 1988

Nucleic Acids Research

Homologous alpha satellite sequences on human acrocentric chromosomes with selectivity for chromosomes 13, 14 and 21: implications for recombination between nonhomologues and Robertsonian translocations

K.H.Choo*, B.Vissel, R.Brown, R.G.Filby and E.Earle

ABSTRACT

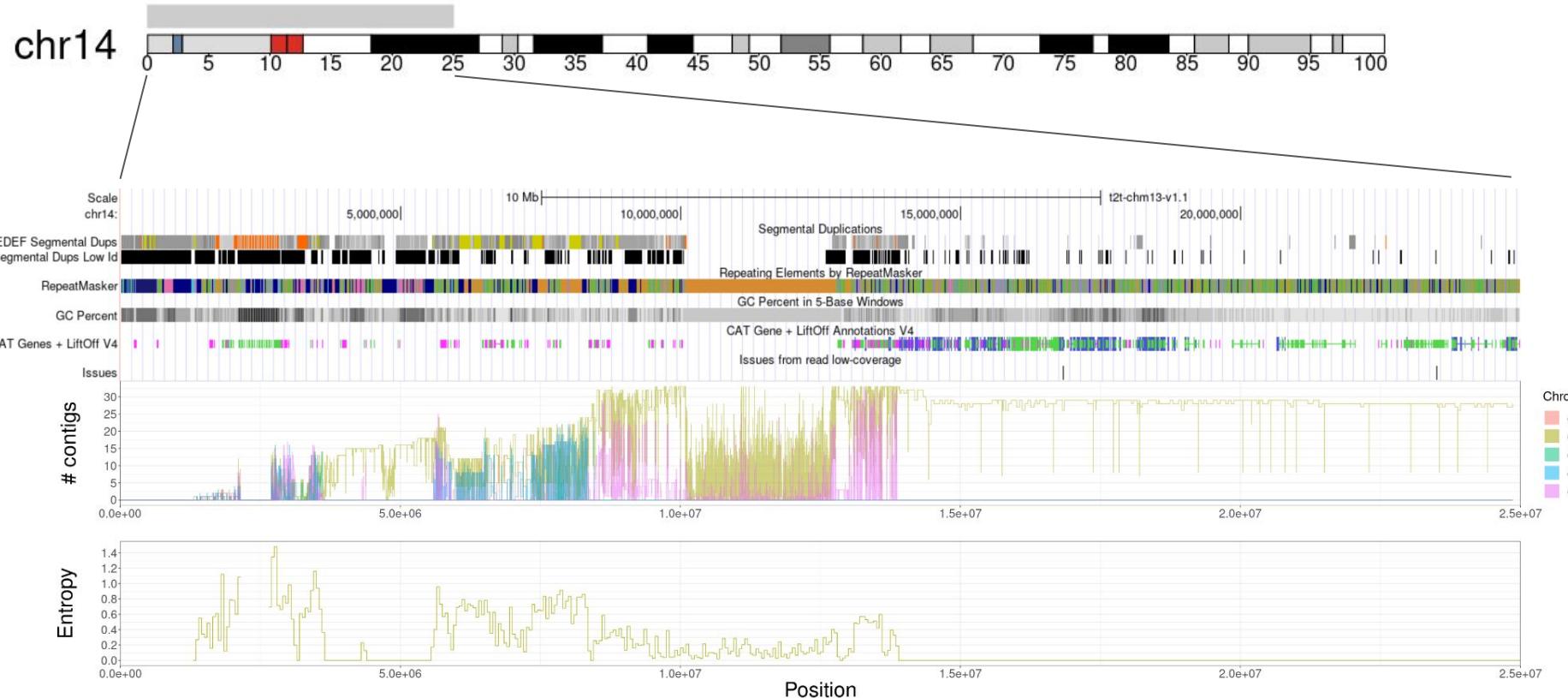
We report a new subfamily of alpha satellite DNA (pTRA-2) which is found on all the human acrocentric chromosomes. The alploid nature of the cloned DNA was established by partial sequencing. Southern analysis of restriction enzyme-digested DNA fragments from mouse/human hybrid cells containing only human chromosome 21 showed that the predominant higher-order repeating unit for pTRA-2 is a 3.9 kb structure. Analysis of a "consensus" *in situ* hybridisation profile derived from 13 normal individuals revealed the localisation of 73% of all centromeric autoradiographic grains over the five acrocentric chromosomes, with the following distribution: 20.4%, 21.5%, 17.1%, 7.3% and 6.5% on chromosomes 13, 14, 21, 15 and 22 respectively. An average of 1.4% of grains was found on the centromere of each of the remaining 19 nonacrocentric chromosomes. These results indicate the presence of a common subfamily of alpha satellite DNA on the five acrocentric chromosomes and suggest an evolutionary process consistent with recombination exchange of sequences between the nonhomologues. The results further suggests that such exchanges are more selective for chromosomes 13, 14 and 21 than for chromosomes 15 and 22. The possible role of centromeric alpha satellite DNA in the aetiology of 13q14q and 14q21q Robertsonian translocations involving the common and nonrandom association of chromosomes 13 and 14, and 14 and 21 is discussed.

[Chroo et al.. 1988](#)

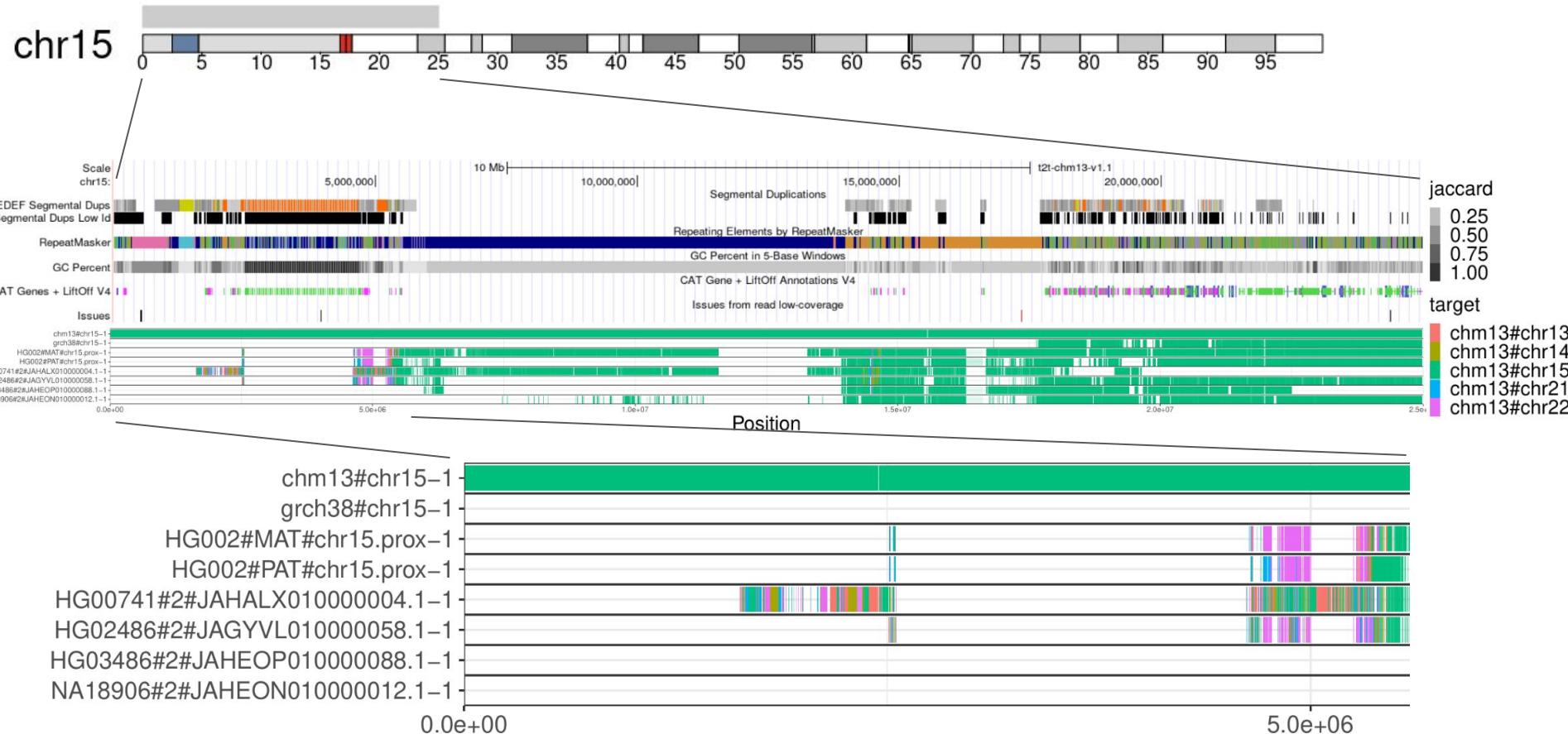
Untangling results - By contig - chr14



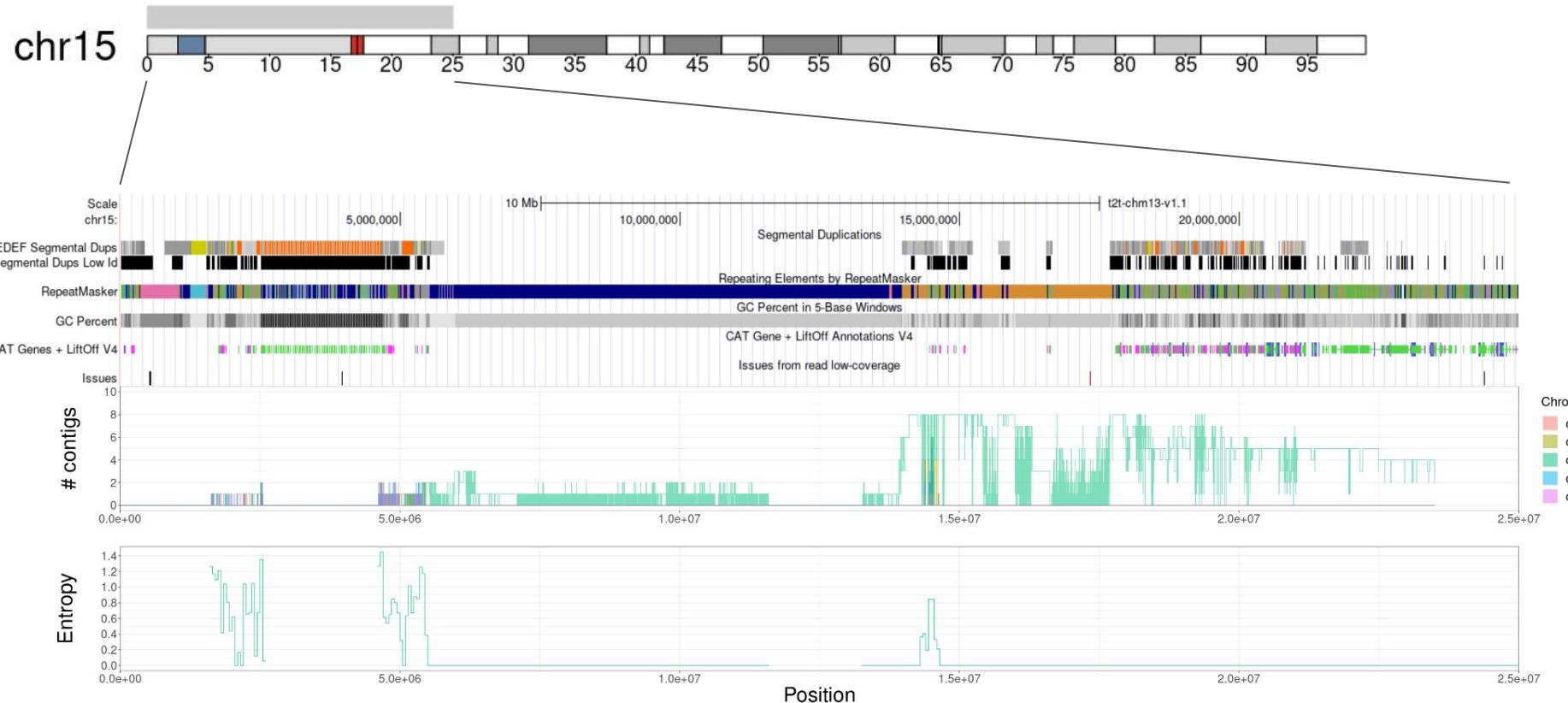
Untangling results - Aggregated - chr14



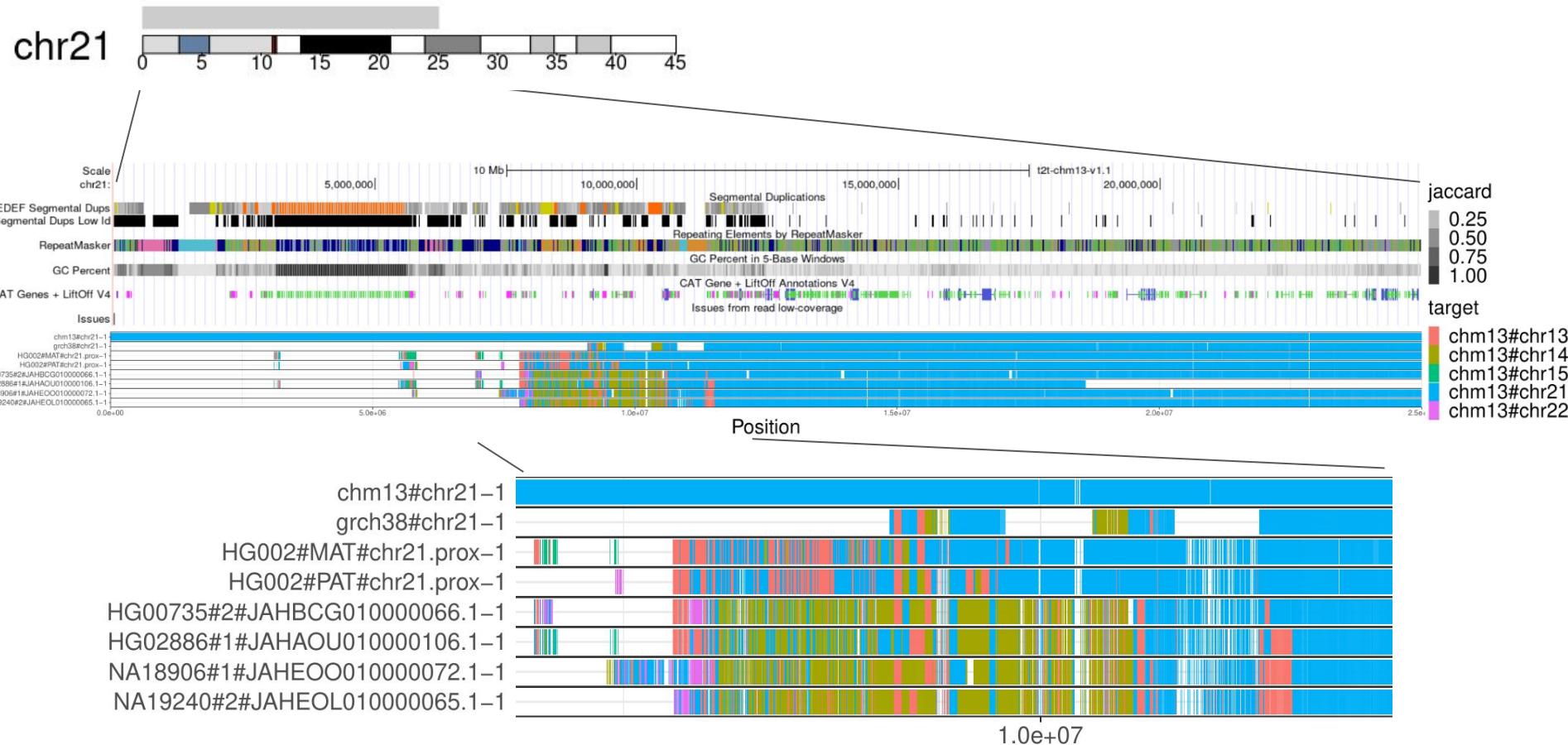
Untangling results - By contig - chr15



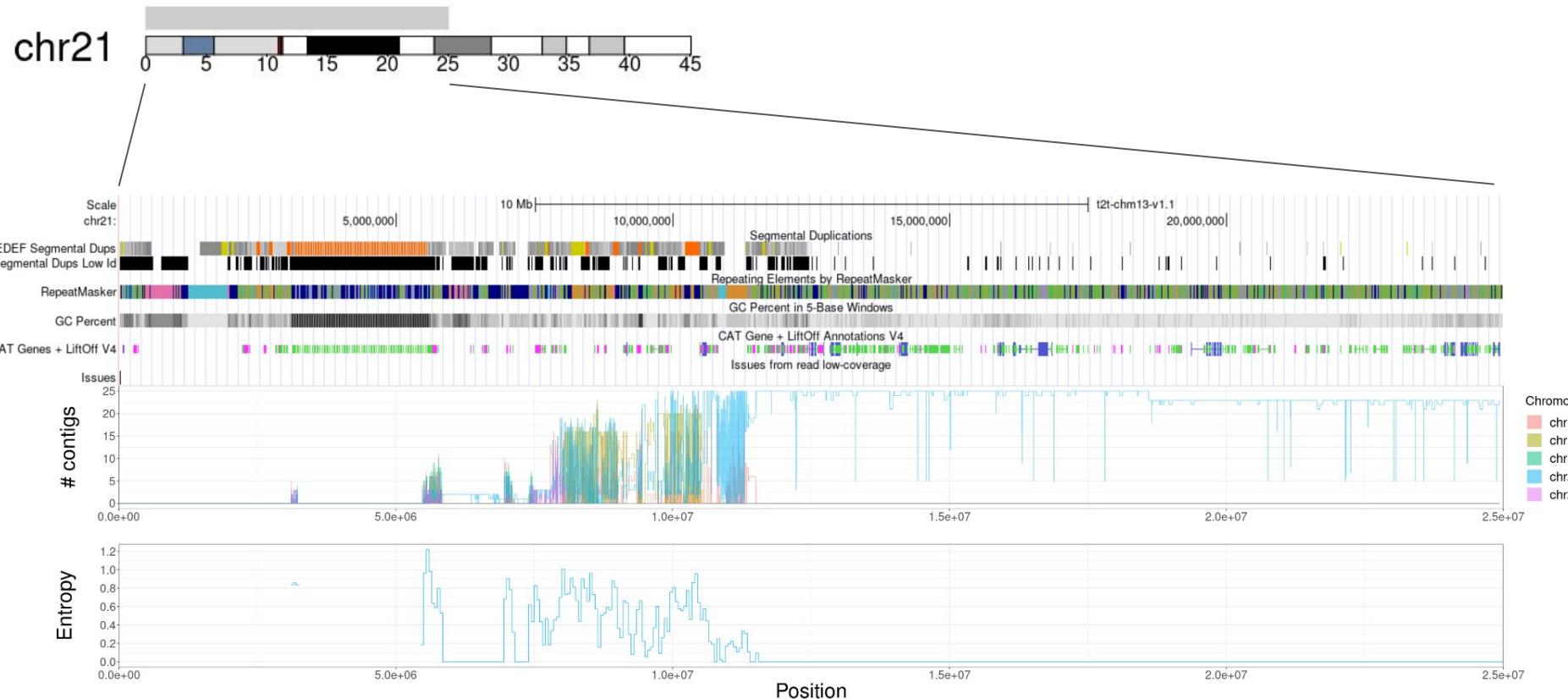
Untangling results - Aggregated - chr15



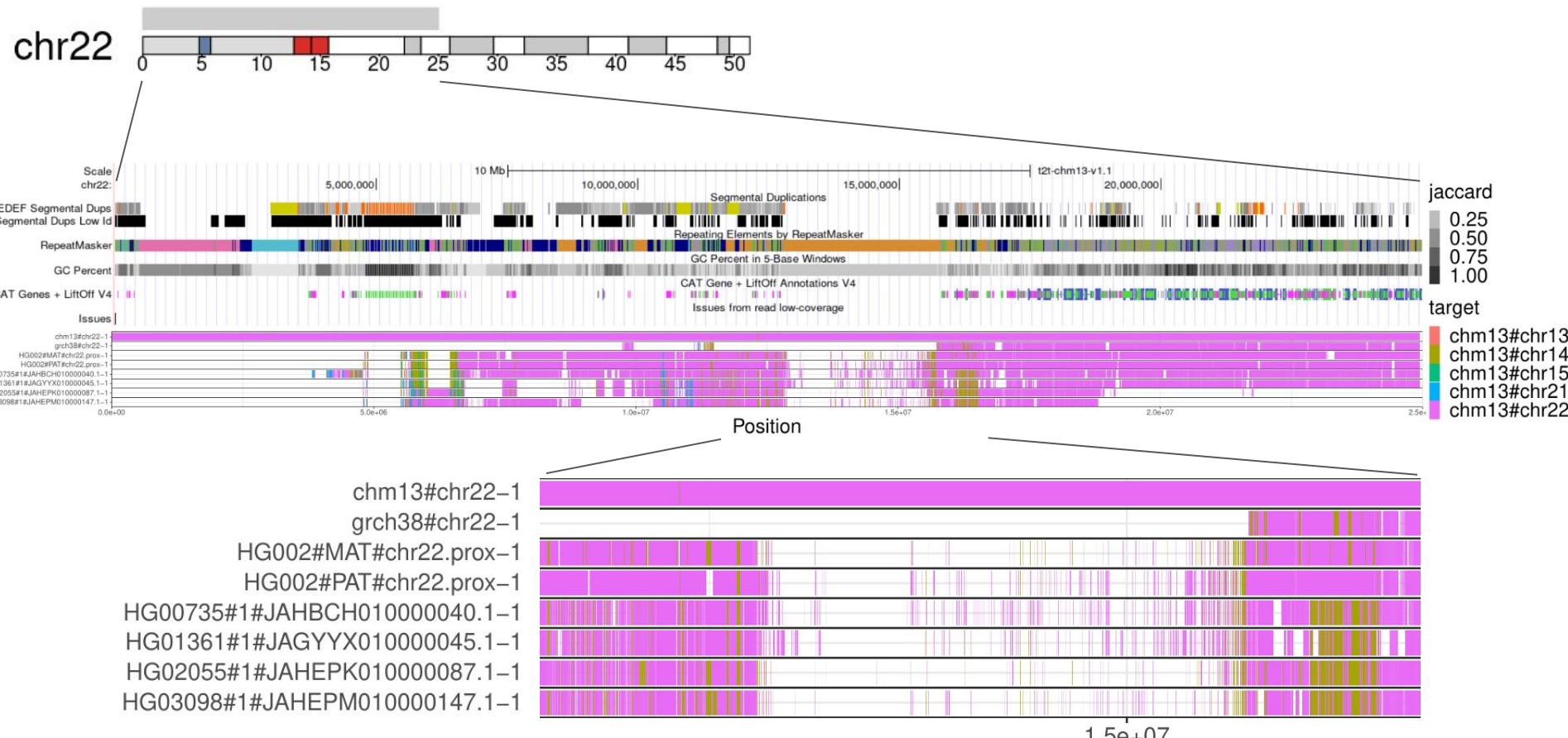
Untangling results - By contig - chr21



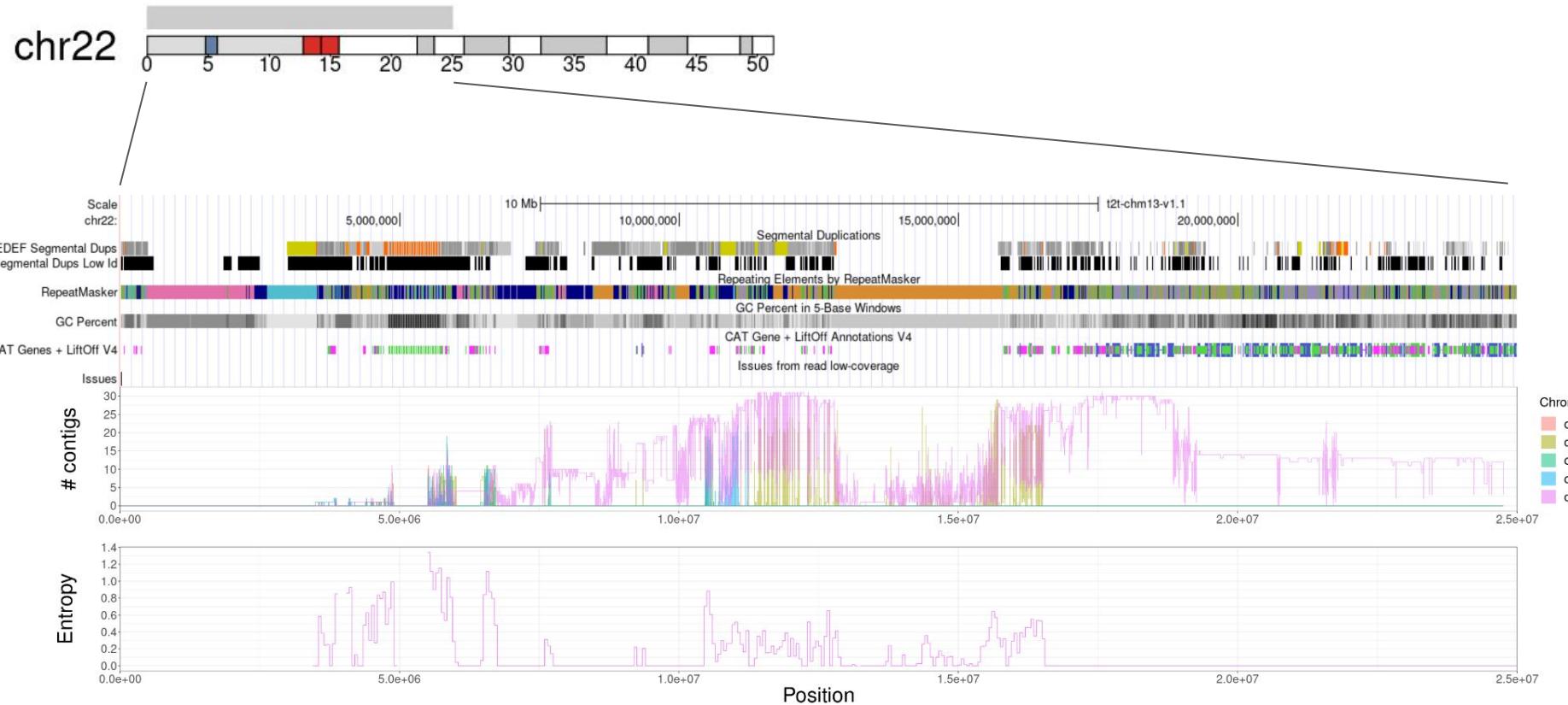
Untangling results - Aggregated - chr21



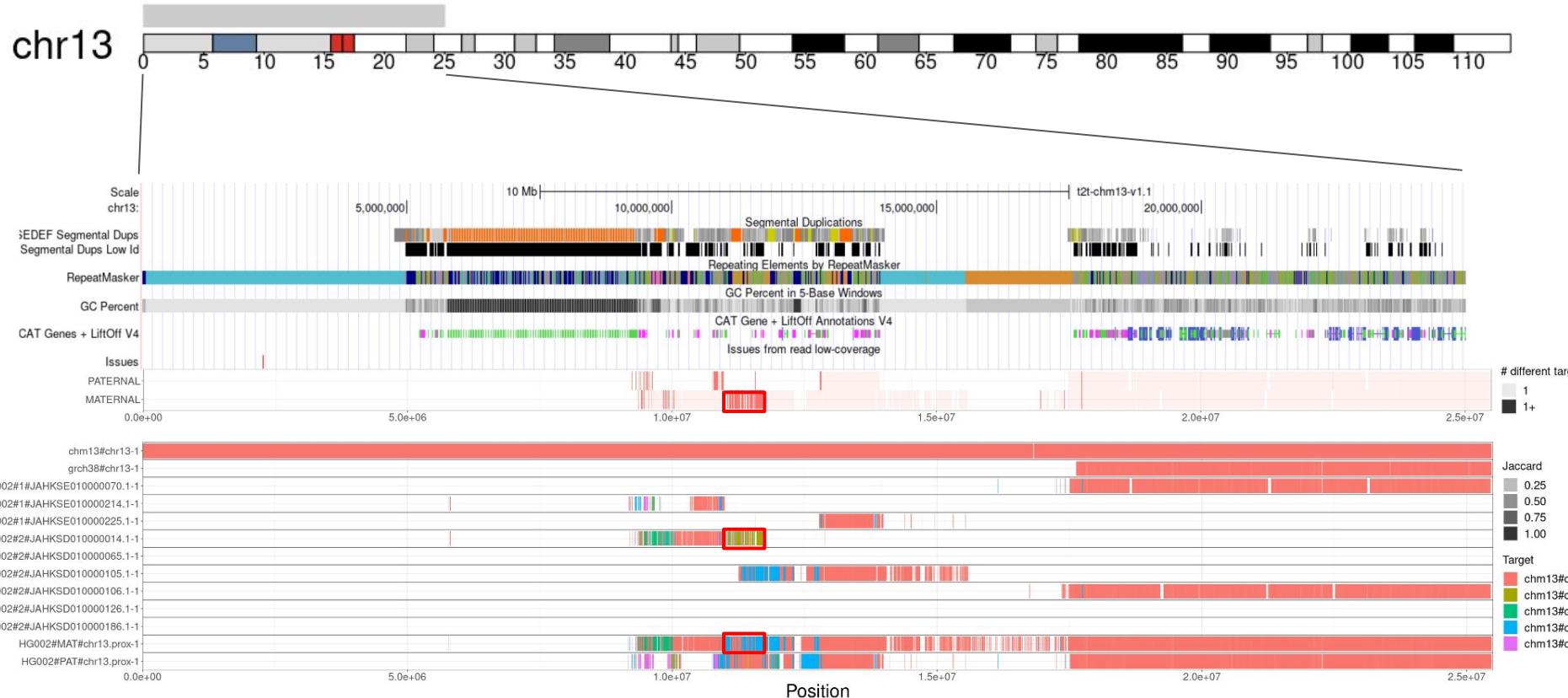
Untangling results - By contig - chr22



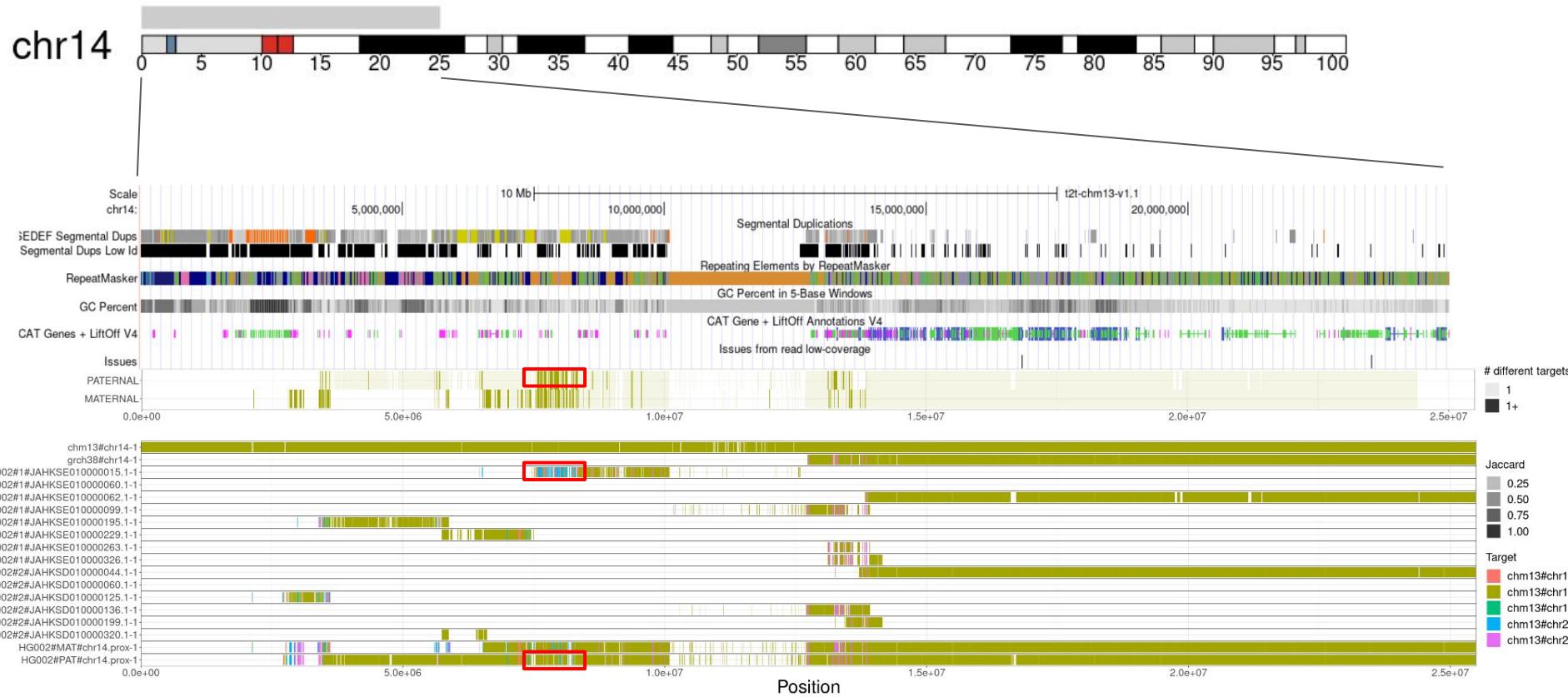
Untangling results - Aggregated - chr22



Hifiiasm (HiFi) / Verkko (HiFi + ONT) concordance - chr13



Hifiiasm (HiFi) / Verkko (HiFi + ONT) concordance - chr14



Hifiiasm (HiFi) / Verkko (HiFi + ONT) concordance - chr15



Hifiiasm (HiFi) / Verkko (HiFi + ONT) concordance - chr21



Hifiiasm (HiFi) / Verkko (HiFi + ONT) concordance - chr22

