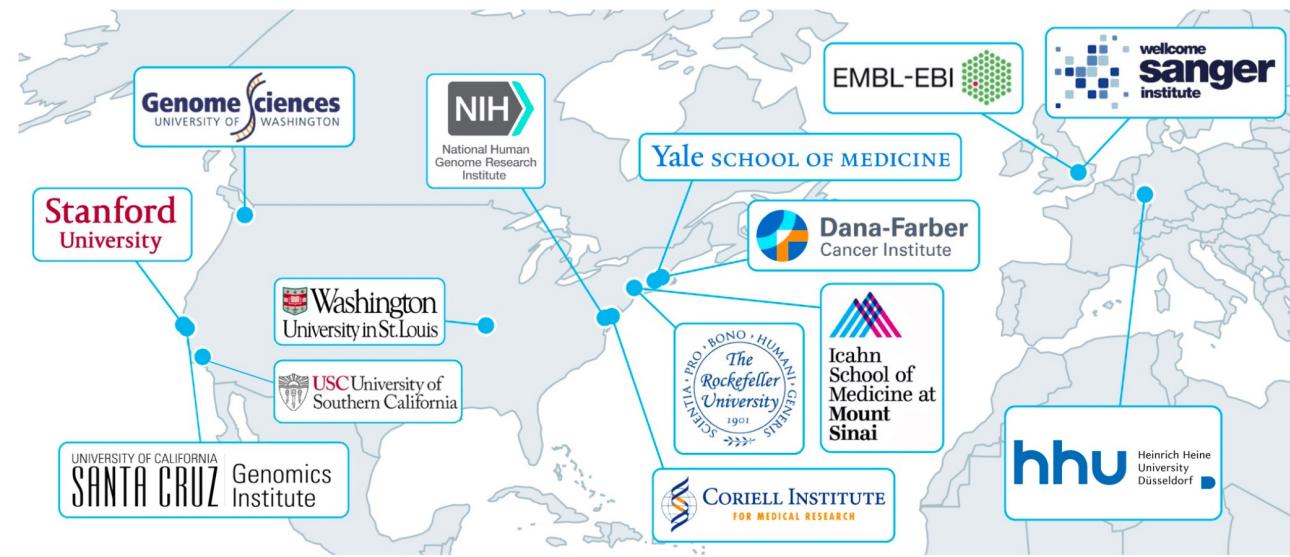


Pangenome Analysis of the KIR-Locus for 246 Haplotypes

Jason Chin¹, Qian Zeng², William Sibbett², Haipeng Liu², Brian Norvell², Lax Iyer², Leigh Smith², Cynthia Vierra-Green³, Angela Kenyon², Debbie Boles², Taylor J. Jensen², Stan Letovsky², Jonathan Williams², Martin Maiers³

1. Foundation of Biological Data Sciences & Omicsight, LLC
2. Laboratory Corporation of America Holdings
3. National Marrow Donor Program

Human Pangenome Reference Consortium



Milestones



With the similar haplotype read separation techniques, 47 haplotyped resolved genome assemblies generated and published.

Article

A draft human pangenome reference

<https://doi.org/10.1038/s41586-023-05896-x>

Received: 9 July 2022

Accepted: 28 February 2023

Published online: 10 May 2023

Open access

Check for updates

Wen-Wei Liao^{1,2,3,60}, Momin Asri^{4,60}, Jana Ebler^{5,6,60}, Daniel Doerr^{5,6}, Marina Haukness⁴, Glenn Hickey⁴, Shuangjia Lu^{1,2}, Julian K. Lucas⁴, Jean Monlong¹, Haley J. Abel⁷, Silvia Buonaiuto⁸, Xian H. Chang⁴, Haoyu Cheng^{9,10}, Justin Chu⁹, Vincenza Colonna^{8,11}, Jordan M. Eizenga⁴, Xiaowen Feng^{9,10}, Christian Fischer¹¹, Robert S. Fulton^{12,13}, Shilpa Garg¹⁴, Cristian Groza¹⁵, Andrea Guaracino^{11,16}, William T. Harvey¹⁷, Simon Heumos^{18,19}, Kerstin Howe²⁰, Miten Jain²¹, Tsung-Yu Lu²², Charles Markello⁴, Fergal J. Martin²³, Matthew W. Mitchell²⁴, Katherine M. Munson¹⁷, Moses Ngagi Mwaniki²⁵, Adam M. Novak⁴, Hugh E. Olsen¹, Trevor Pesout⁴, David Porubsky¹⁷, Pjotr Prins¹, Jonas A. Sibbesen²⁶, Jouni Sirén⁴, Chad Tomlinson¹², Flavia Villani¹¹, Mitchell R. Vollger^{17,27}, Lucinda L. Antonacci-Fulton¹², Gunjan Baid²⁸, Carl A. Baker¹⁷, Anastasiya Belyaeva²⁸, Konstantinos Billis²³, Andrew Carroll²⁹, Pi-Chuan Chang³⁰, Sarah Cody¹⁷, Daniel E. Cook²⁸, Robert M. Cook-Deegan²⁹, Omar E. Cornejo³⁰, Mark Diekhans⁴, Peter Ebert^{5,6,31}, Susan Fairley²³, Olivier Fedrigó³², Adam L. Felsenfeld³³, Giulio Formenti³², Adam Frankish²³, Yan Gao³⁴, Nanibaa' A. Garrison^{35,36,37}, Carlos Garcia Giron²³, Richard E. Green^{38,39}, Leanne Haggerty²³, Kendra Hoekzema¹⁷, Thibaut Hourlier²³, Hanlee P. Ji⁴⁰, Eimear E. Kenny⁴¹, Barbara A. Koenig⁴², Alexey Kolesnikov²⁹, Jan O. Korbel^{23,43}, Jennifer Kordosky¹⁷, Sergey Koren⁴⁴, HoJoon Lee⁴⁰, Alexandra P. Lewis¹⁷, Hugo Magalhães^{5,6}, Santiago Marco-Sola^{45,46}, Pierre Marjion^{5,6}, Ann McCartney⁴⁴, Jennifer McDaniel⁴⁷, Jacquelyn Mountcastle³², Maria Nattestad²⁹, Sergey Nurk⁴⁴, Nathan D. Olson⁴⁷, Alice B. Popejoy⁴⁸, Daniels Puju⁴⁹, Mikko Rautainen⁴⁴, Allison A. Regier¹², Arang Rhee⁴⁴, Samuel Sacco³⁰, Ashley D. Sanders⁵⁰, Valeria A. Schneider⁵¹, Baergen I. Schultz³³, Kishwar Shafin²⁸, Michael W. Smith³³, Heidi J. Sofia³³, Ahmad N. Abou Tayoun^{52,53}, Françoise Thibaud-Nissen⁵¹, Francesca Floriana Tricomi²³, Justin Wagner⁴⁷, Brian Walenz⁴⁴, Jonathan M. D. Wood²⁰, Aleksey V. Zimin^{49,54}, Guillaume Bourque^{55,56,57}, Mark J. P. Chaisson²², Paul Flicek²³, Adam M. Phillippy⁴⁴, Justin M. Zook⁴⁷, Evan E. Eichler^{17,58}, David Haussler^{4,58}, Ting Wang^{12,13}, Erich D. Jarvis^{32,58,59}, Karen H. Miga⁴, Erik Garrison^{11,53}, Tobias Marschall^{5,6,53}, Ira M. Hall^{1,2,53}, Heng Li^{19,10,53} & Benedict Paten^{4,53}

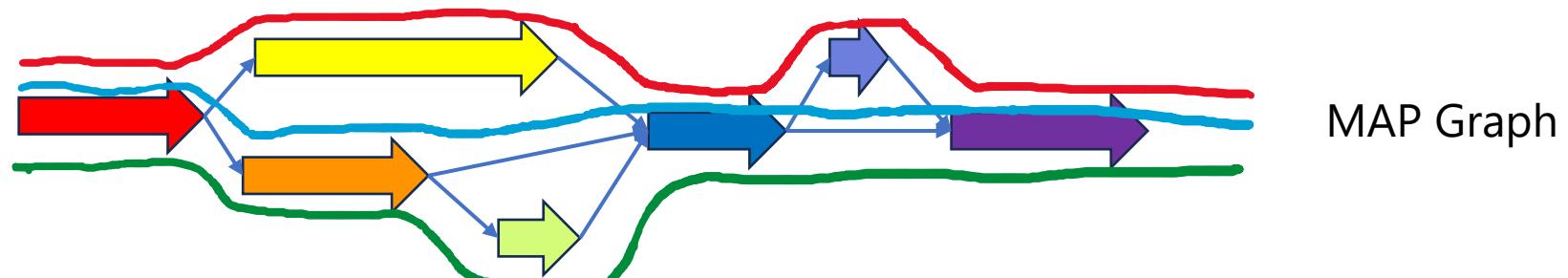
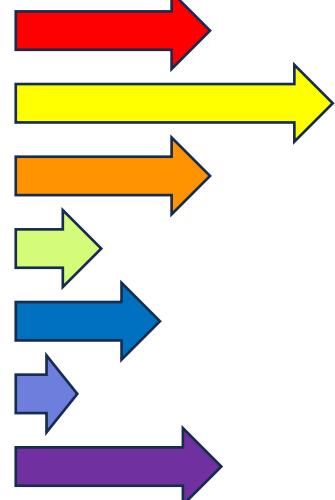
Human Pangenome Collection

- 47 Diploid Genome Sequenced, assembled with whole genome haplotype resolution (more to come in the coming years)
- Many MHC and KIR regions are well assembled revealing additional new full DNA sequence level haplotypes
- We develop new bioinformatics tools to help analyzing many complete haplotypes at once
 - Extract all interested sequences from the pangenome collection
 - New algorithm and visualization to comparing/organizing sequences

A Primer to Pangenome Graph

Building a graph collapsed sequence region similar to each other, and we can show linear path of each haplotype which parts are homologous to others and which part is more unique.

Shared Blocks



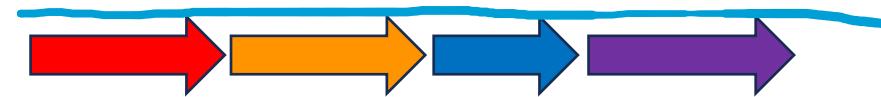
MAP Graph

The shared blocks along with each haplotype

Each block represents
a set of closely related
similar sequences



Haplotype I

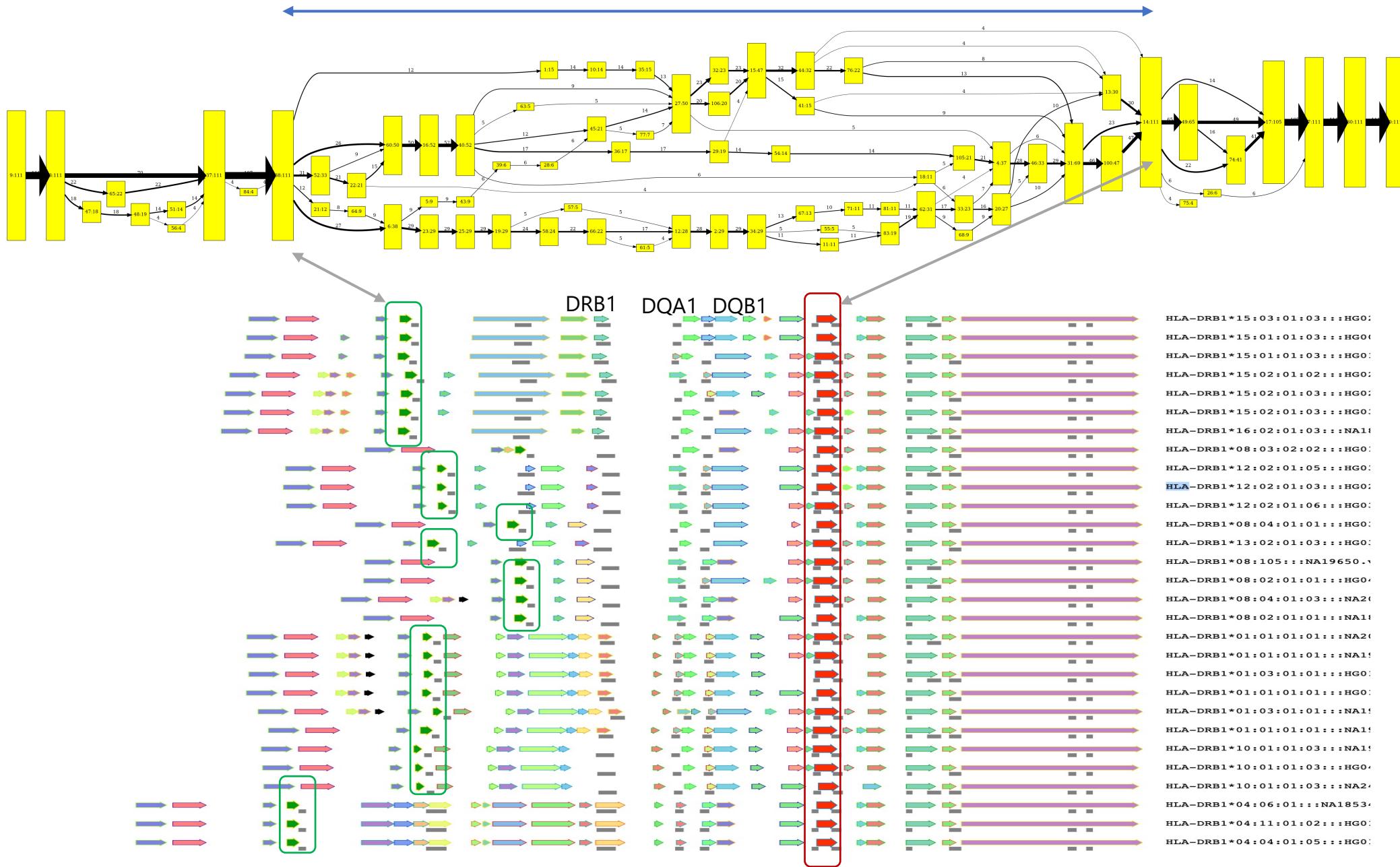


Haplotype II

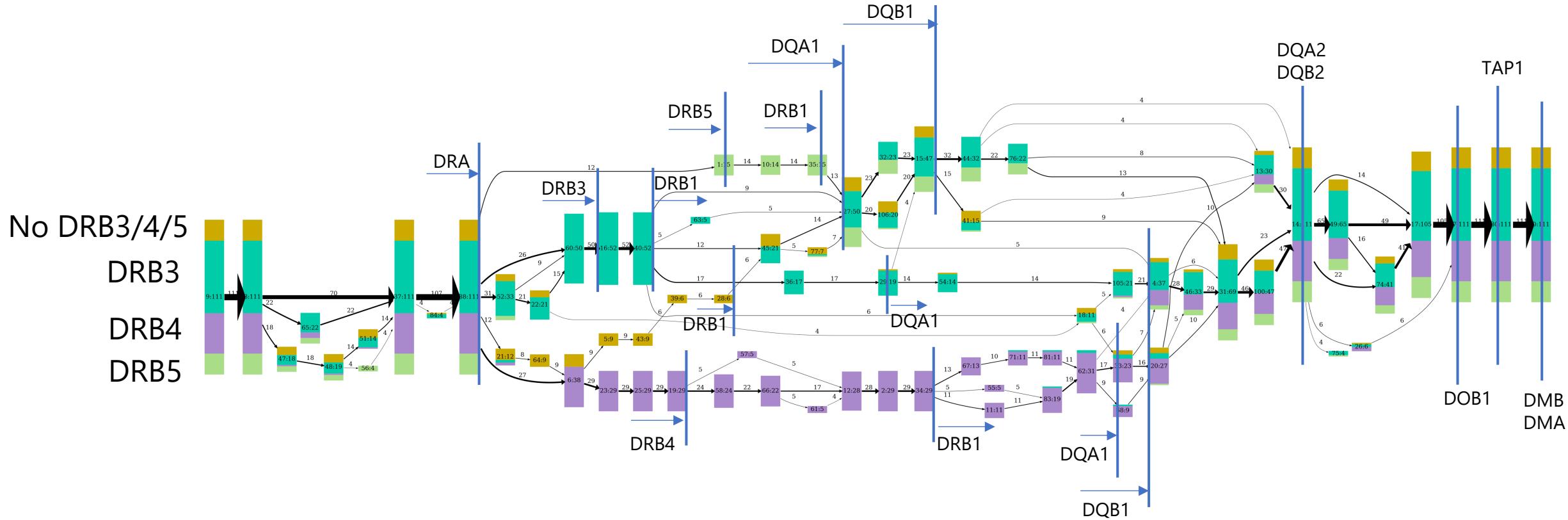


Haplotype III

DRB3/4/5 → DRB1 → DQA1→ DQB1 region



MHC Class II Region As Example



No DRB3/4/5: DRB1*01, DRB1*10, DRB1*08

DRB3: DRB1*03, DRB1*11, DRB1*13, DRB1*12, DRB1*14

DRB4: DRB1*04, DRB1*07, DRB1*09

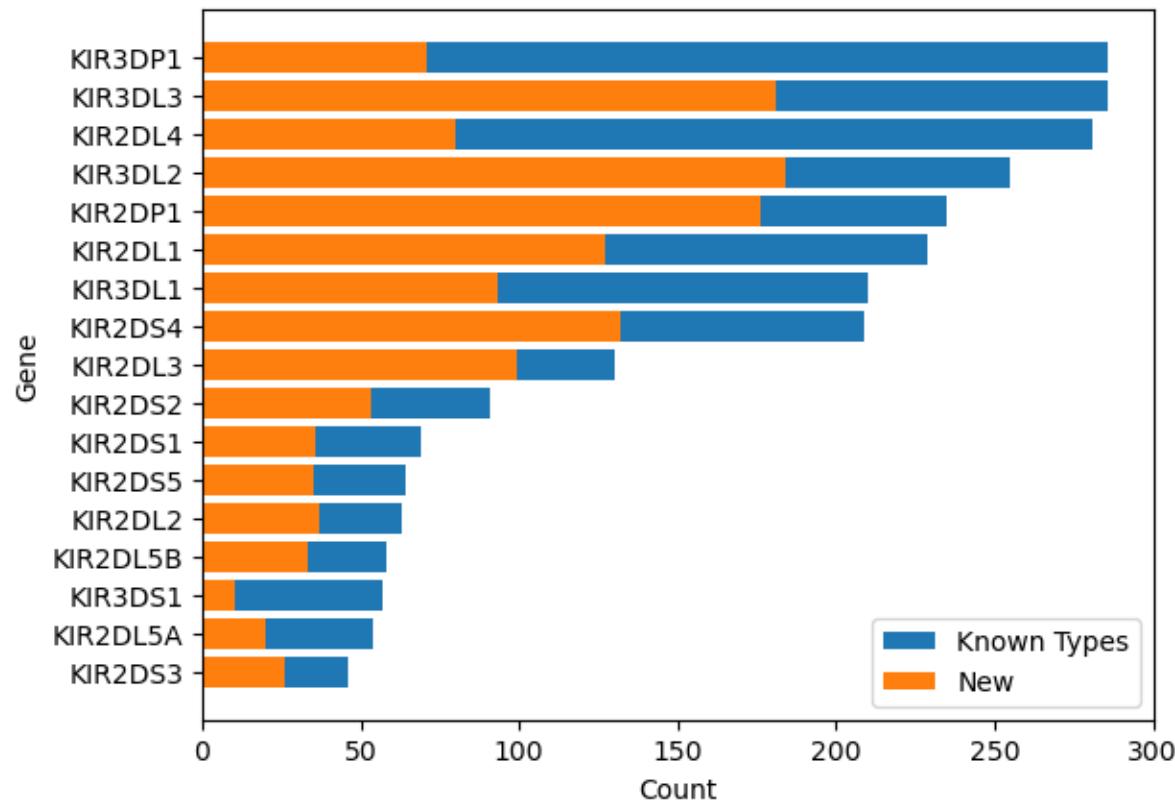
DRB5: DRB1*15, DRB1*16

Full KIR Haplotype Sequence Collection

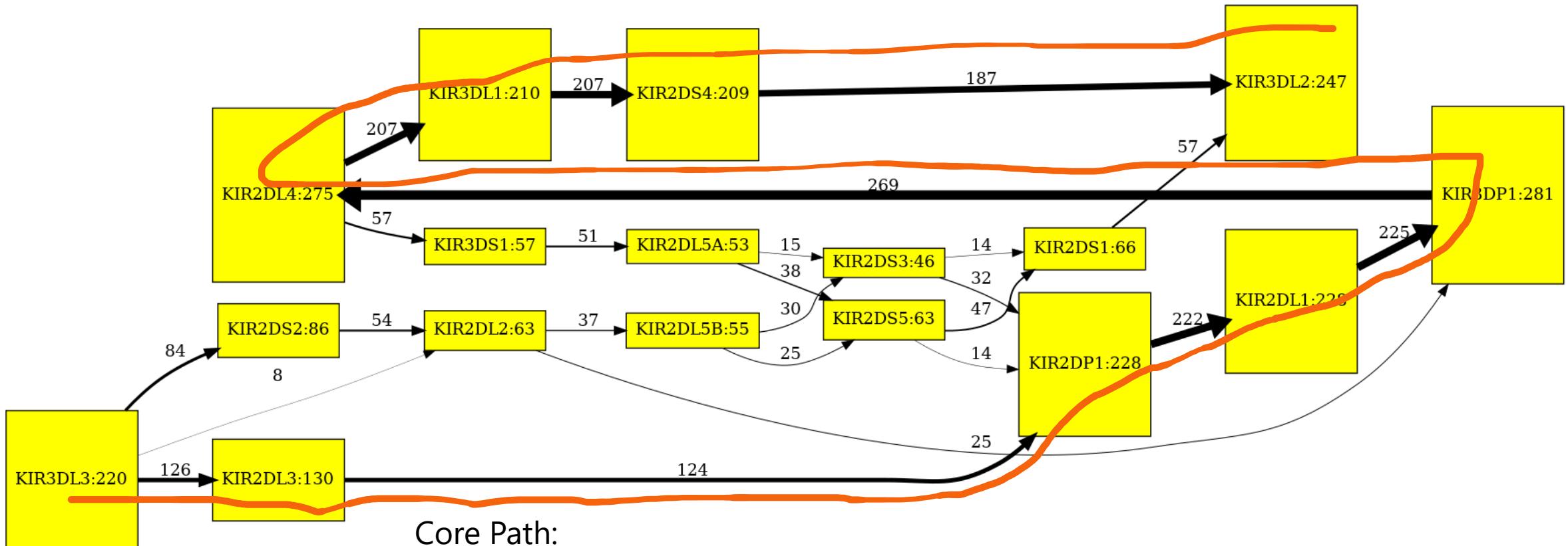
- Data Sets
 - 67 previously published and annotated haplotypes from GenBank
 - 95 haplotypes from the genome assemblies collected by the HPRC (and other collections in GRCh38 / hg19)
 - 84 additional assembled haplotypes, exceeding 75 kb in length, from 48 samples collected from NMDP donors sequenced through a KIR locus DNA enrichment protocol
- ~ 246 Haplotype sequences

Summary of The Annotated Genes In the Dataset

Result from Immunannot



Gene Flow Graph From The Dataset



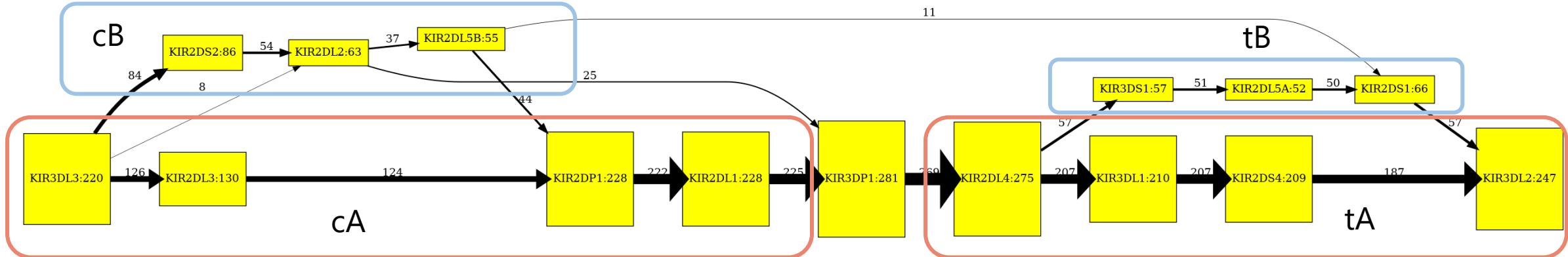
Core Path:

KIR3DL3 → KIR2DL3 → KIR2DP1 → KIR2DL1 → KIR3DP1 → KIR2DL4 →
KIR3DL1 → KIR2DS4 → KIR3DL2

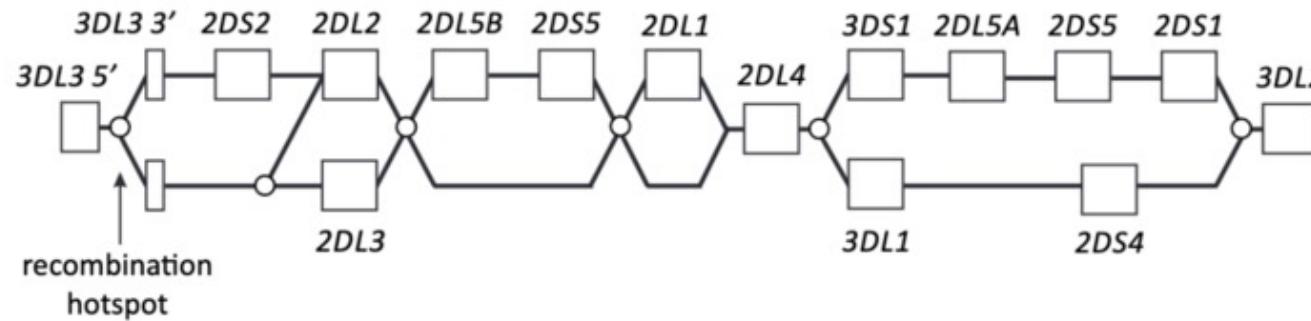
91 assembly contigs have this haplotype (~ cA01~tA01 haplotype)

Variation Set: KIR2DS2 / KIR2DL2 / KIR2DL5B / **KIR2DS5** / KIR3DS1 /
KIR2DL5A / **KIR2DS3**

Remove KIR2DS3 and KIR2DS5



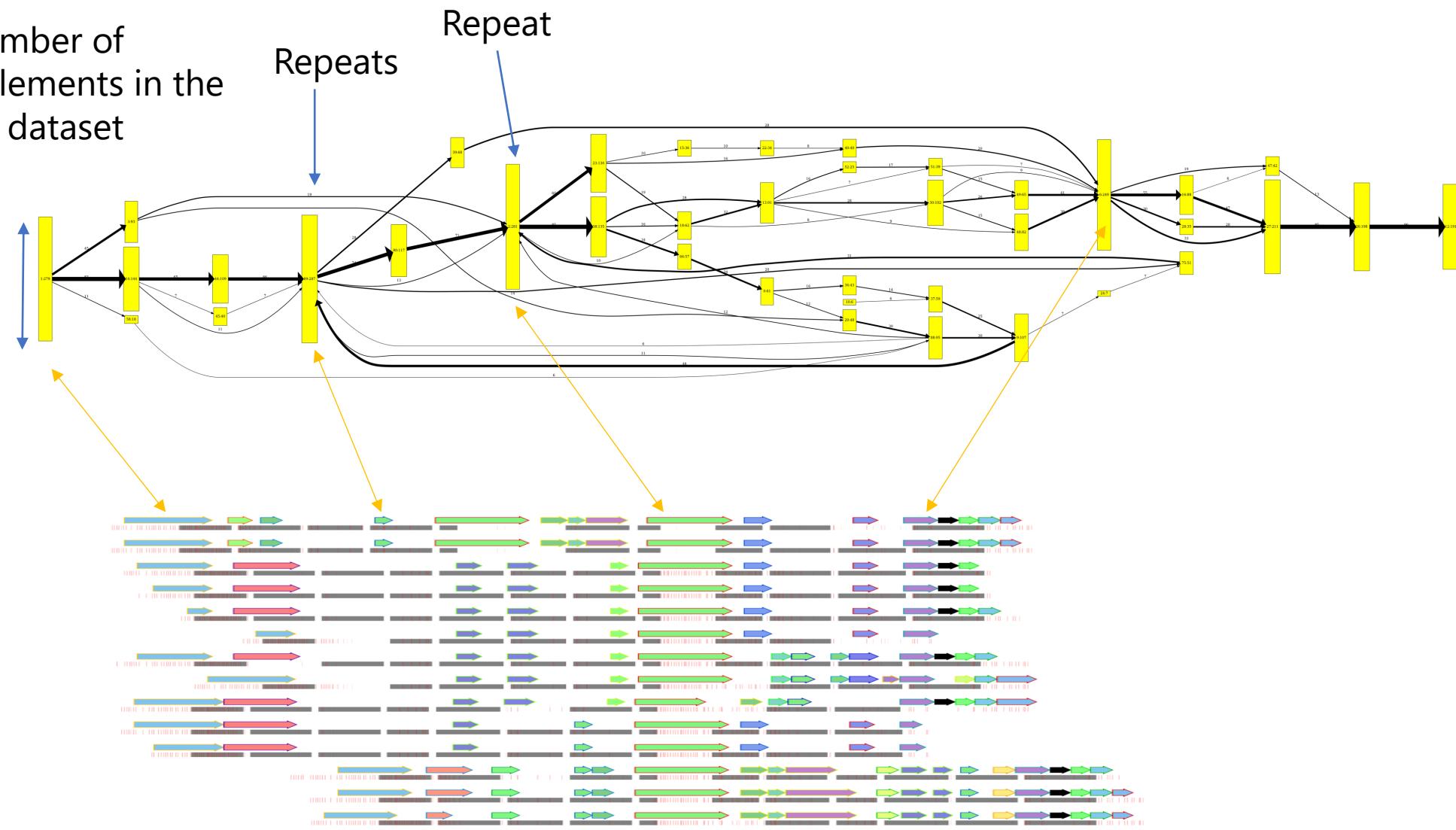
A



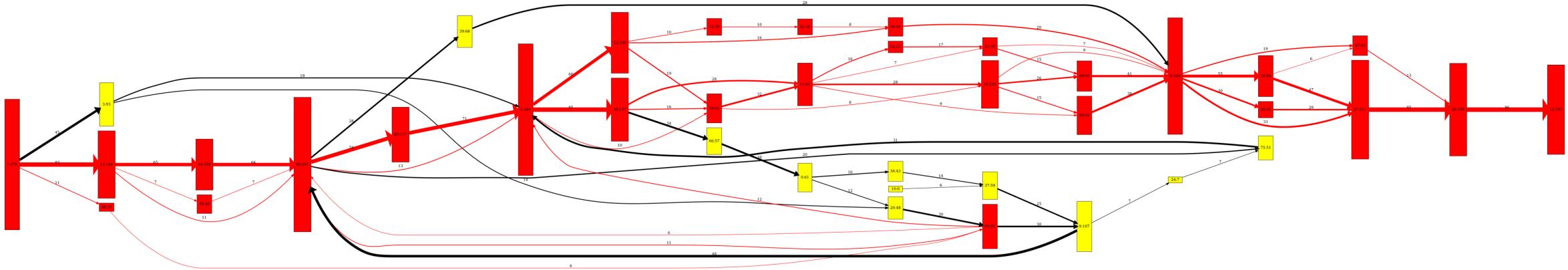
Leaton 2020

KIR Haplotype Diversity Beyond Gene Level

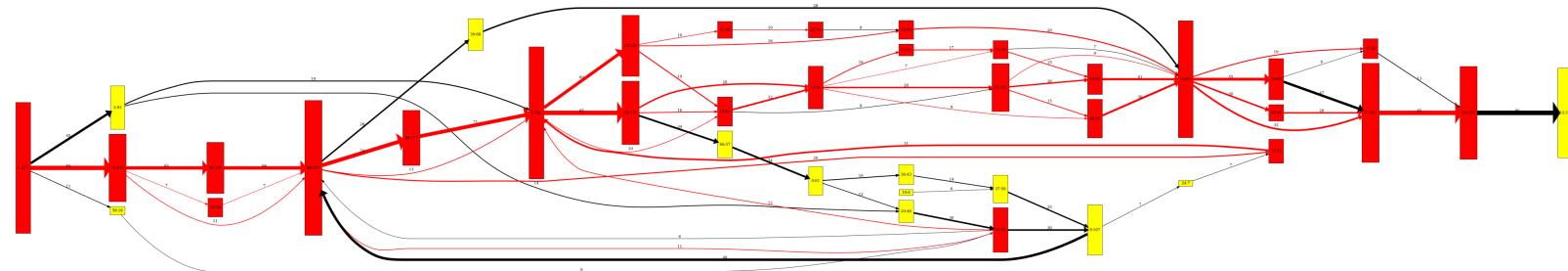
Height ~ number of
conserved elements in the
pangenome dataset



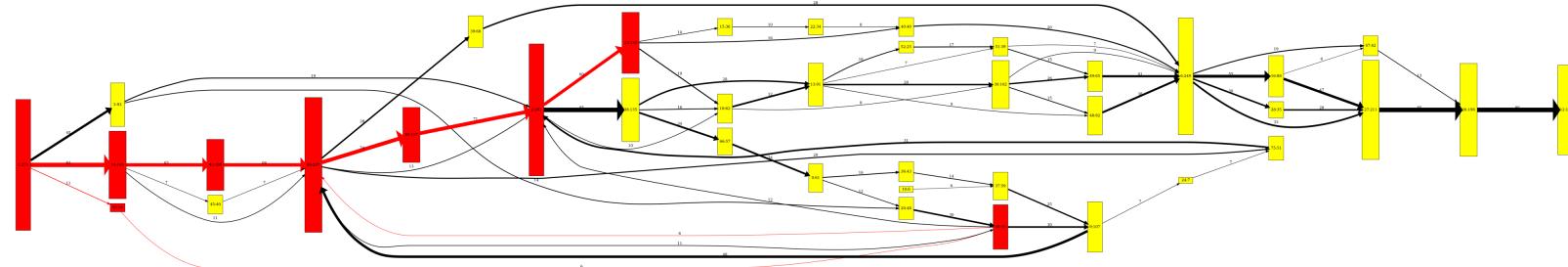
Sequence Level Variation Along the Main Gene Haplotypes



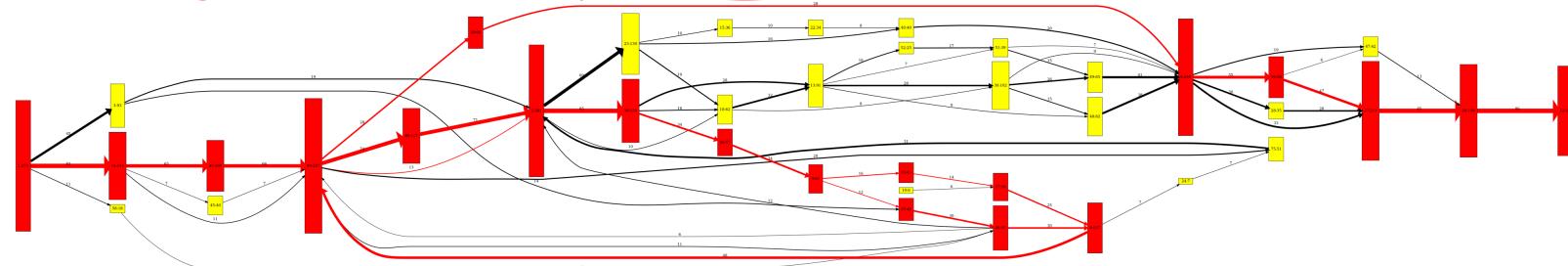
The red nodes are the 91 contigs with the main gene flow haplotypes



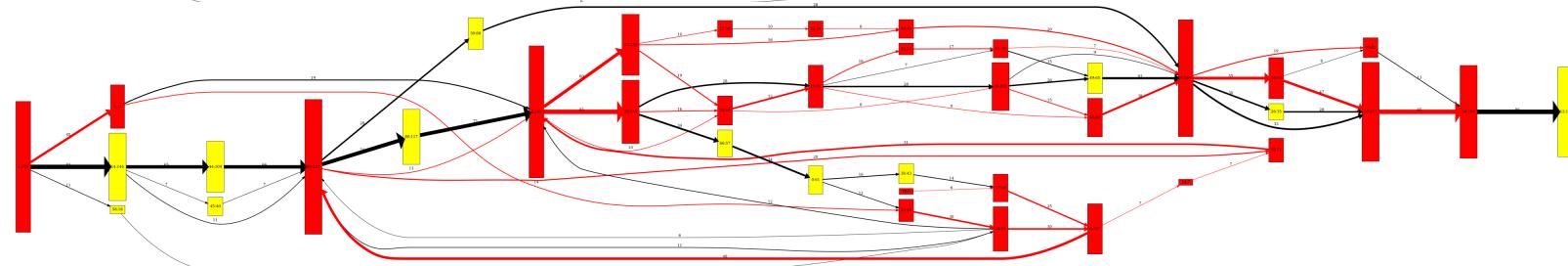
cA01~tA01



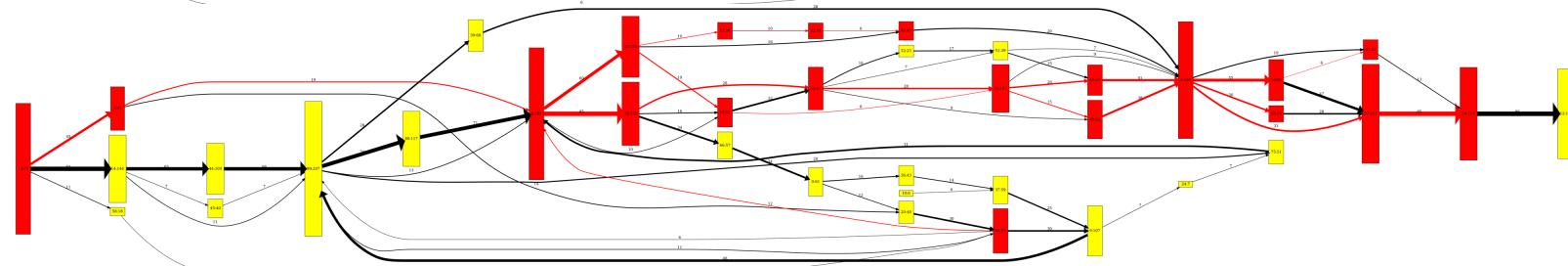
cA01~tA02



cA01~tB01

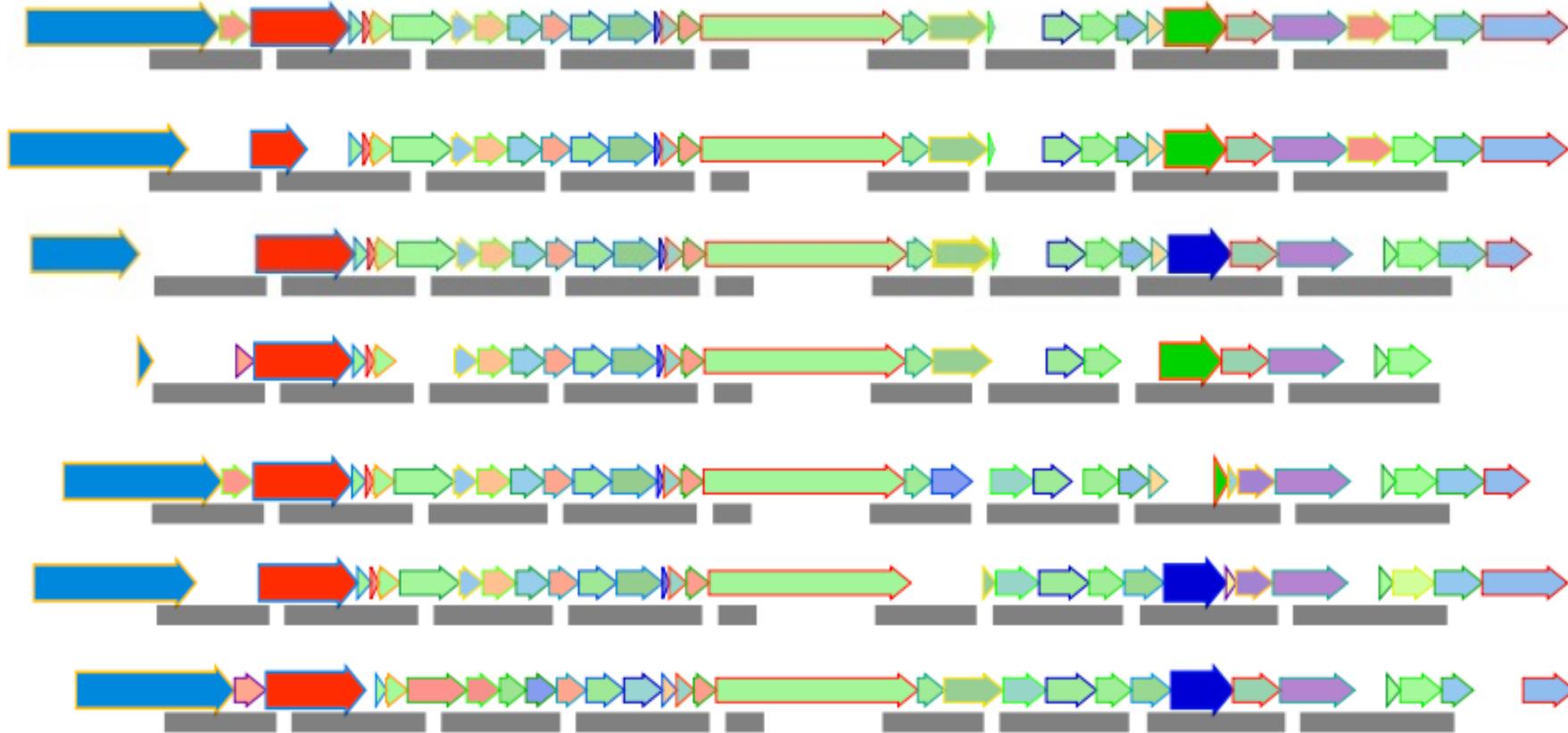


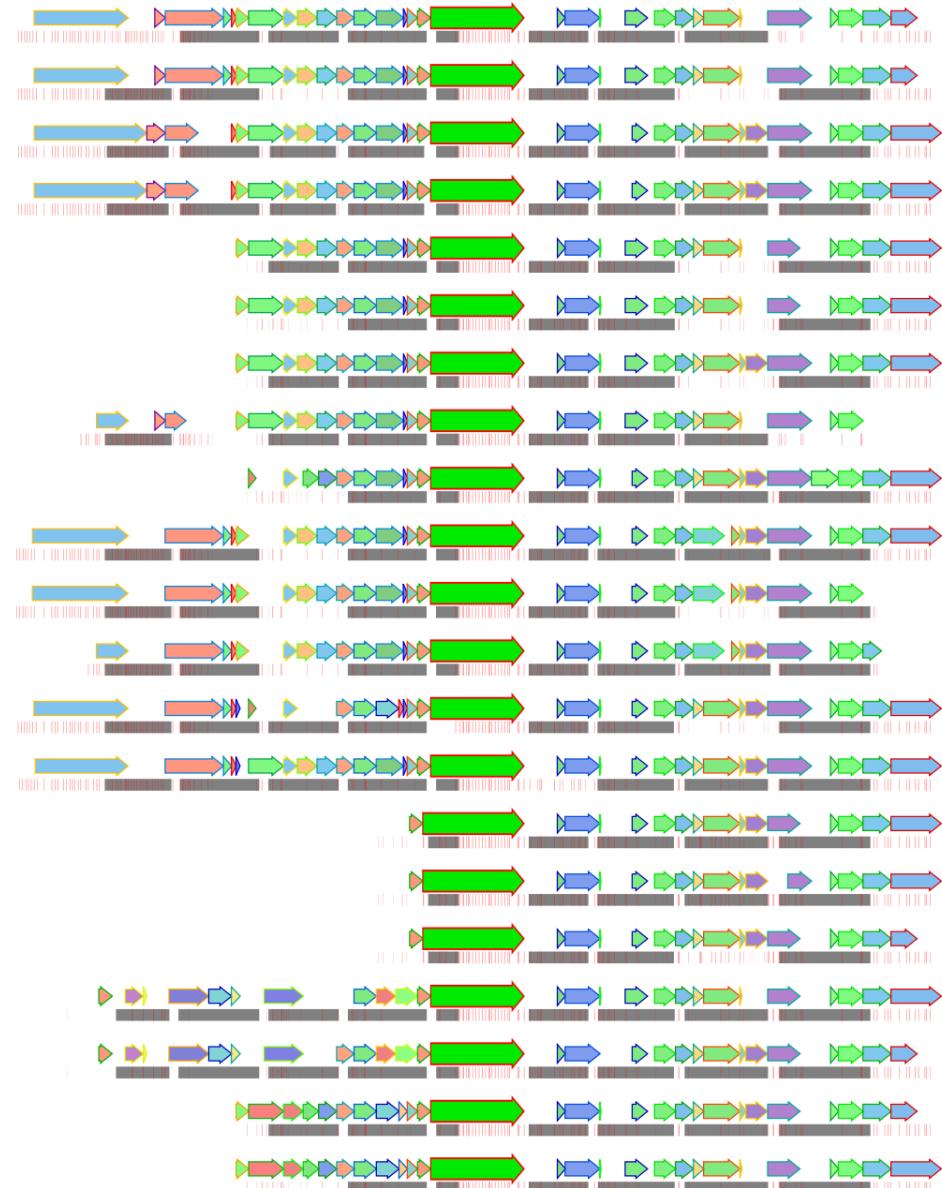
cB01~tA01



cB02~tA01

cA01~tA01 Sequence Level Varieties





HG03516.paternal.f1_assembly_v2_genbank::HG03516#
 HG02109.paternal.f1_assembly_v2_genbank::HG02109#
 hg38_tagged::chr19_KI270938v1_alt_hg38_690037_863
 hg19_tagged::chr19_hg19_55218923_55392486_0
 KIR48JW_47_asm_ctg_m::ctg000000_P_217_136066_0
 KIR48JW_43_asm_ctg_m::ctg000000_P_195_136151_0
 KIR48JW_29_asm_ctg_m::ctg000000_P_189_136154_0
 cA01-tA01_MN167513/1147345
 KIR48JW_14_asm_ctg_m::ctg000000_P_214_136182_0
 hg38_tagged::chr19_hg38_54706823_54881030_0
 hg38_tagged::chr19_KI270919v1_alt_hg38_604_162310.
 cA01-tA01_NC000019.10/1148971
 HG02572.paternal.f1_assembly_v2_genbank::HG02572#
 HG01361.paternal.f1_assembly_v2_genbank::HG01361#
 KIR48JW_48_asm_ctg_m::ctg000000_P_2184_108242_1
 KIR48JW_2_asm_ctg_m::ctg000001_P_290_106356_0
 KIR48JW_44_asm_ctg_m::ctg000001_P_211_106272_0
 KIR48JW_7_asm_ctg_m::ctg000000_P_1684_165811_1
 KIR48JW_35_asm_ctg_m::ctg000000_P_233_164361_0
 KIR48JW_41_asm_ctg_m::ctg000000_P_3270_139127_1
 KIR48JW_13_asm_ctg_m::ctg000000_P_246_136098_0

Enrichment probes (red) are mapped to all contigs. we can analyze the density to have better tiling for the probe density on new sequences

Identify All Genes In The Haplotype Set

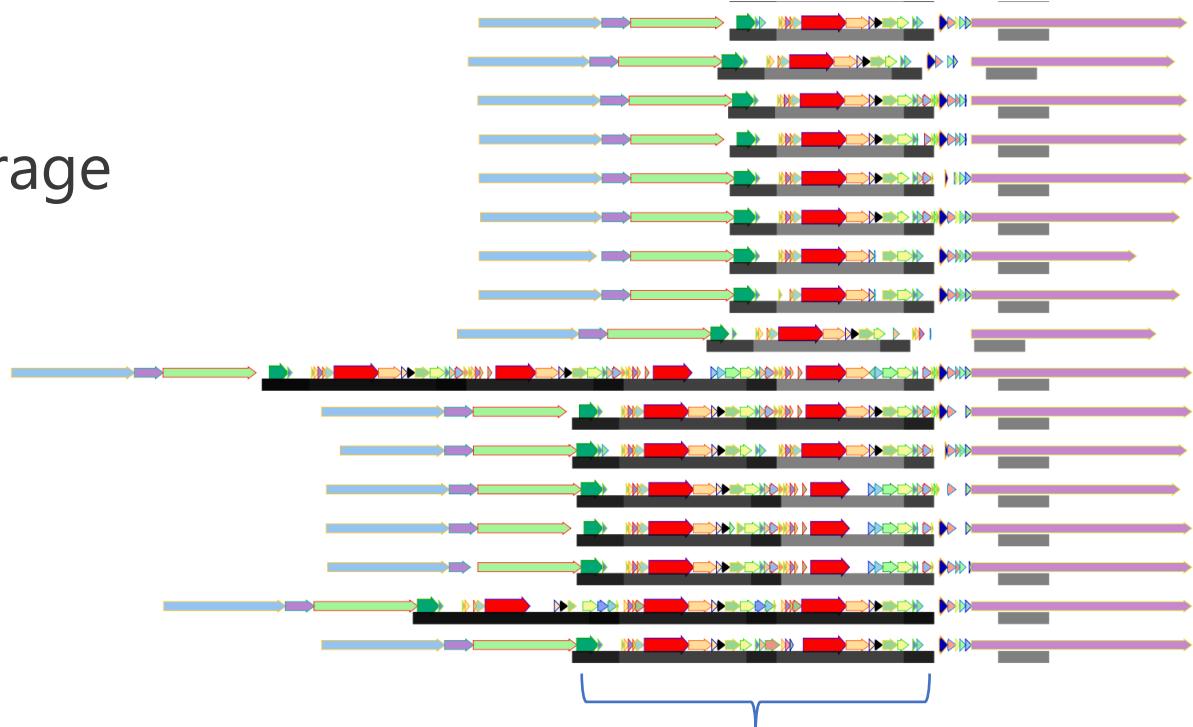
A screenshot of a GitHub repository page. The top navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Wiki, and Settings. Below the navigation, the repository path is shown as `KIR-NMDP-Analysis / KIR / assembly_alignmap / results /`. A dropdown menu indicates the branch is `main`. A message from user `cschin` with the status `move` is displayed. The main content is a table listing files and their last commit messages:

Name	Last commit message
..	
KIR-annotation.bed	move
KIR2DL1_gene.fa.gz	move
KIR2DL2_gene.fa.gz	move
KIR2DL3_gene.fa.gz	move
KIR2DL4_gene.fa.gz	move
KIR2DL5A_gene.fa.gz	move
KIR2DL5B_gene.fa.gz	move
KIR2DP1_gene.fa.gz	move
KIR2DS1_gene.fa.gz	move
KIR2DS2_gene.fa.gz	move

<https://github.com/cschin/KIR-NMDP-Analysis>

Next

- KIR Haplotype Nomenclature
- Sequence Classification (May leverage LLM approach)
- LILR / LAIR analysis
- Acknowledgment
 - Sponsored by NMDP, Martin Maiers



LILRB3 has different repeat domains