How I learned to quit worrying

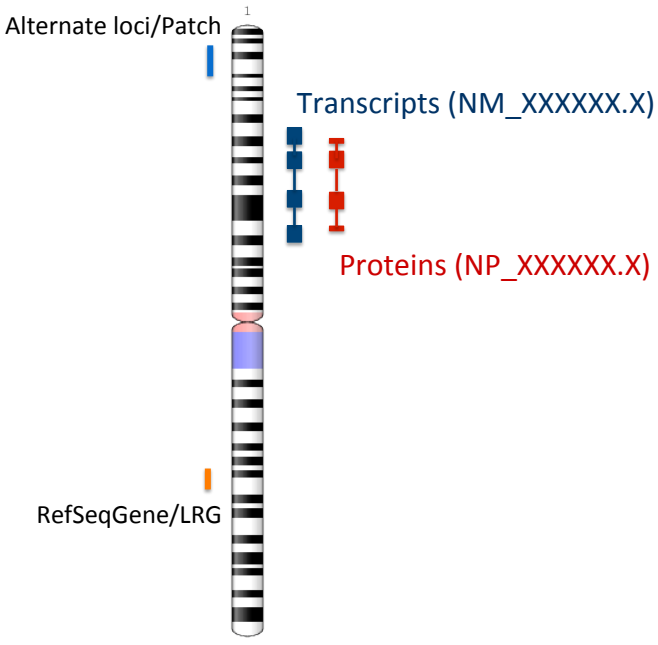
And love multiple coordinate systems

Deanna M. Church
Senior Director of Genomics and Content
Personalis, Inc

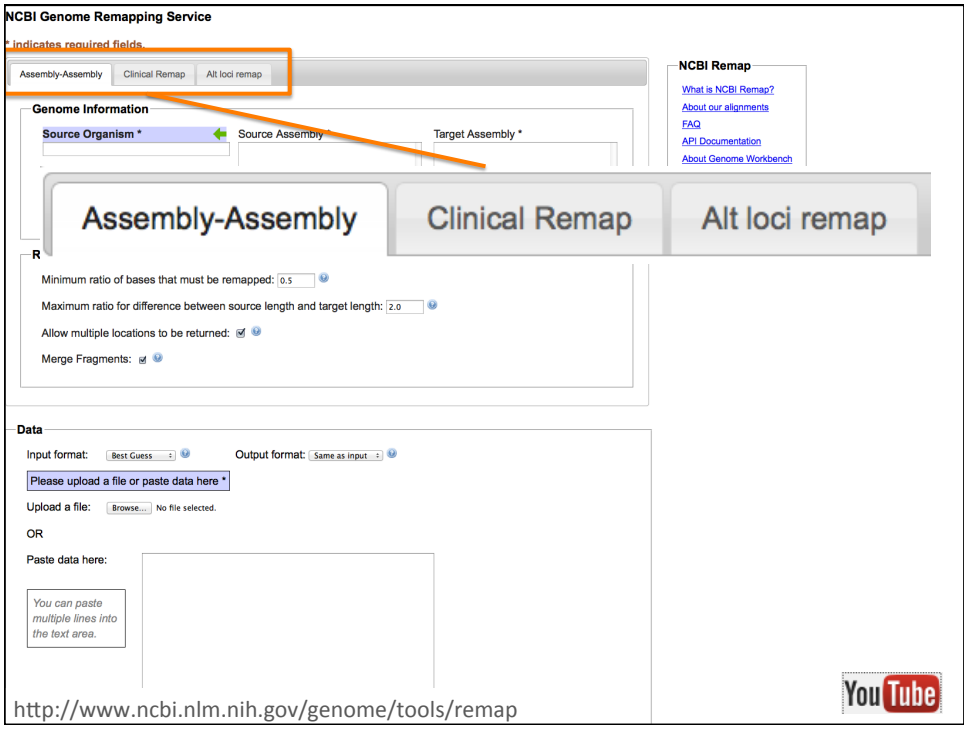
Personalis

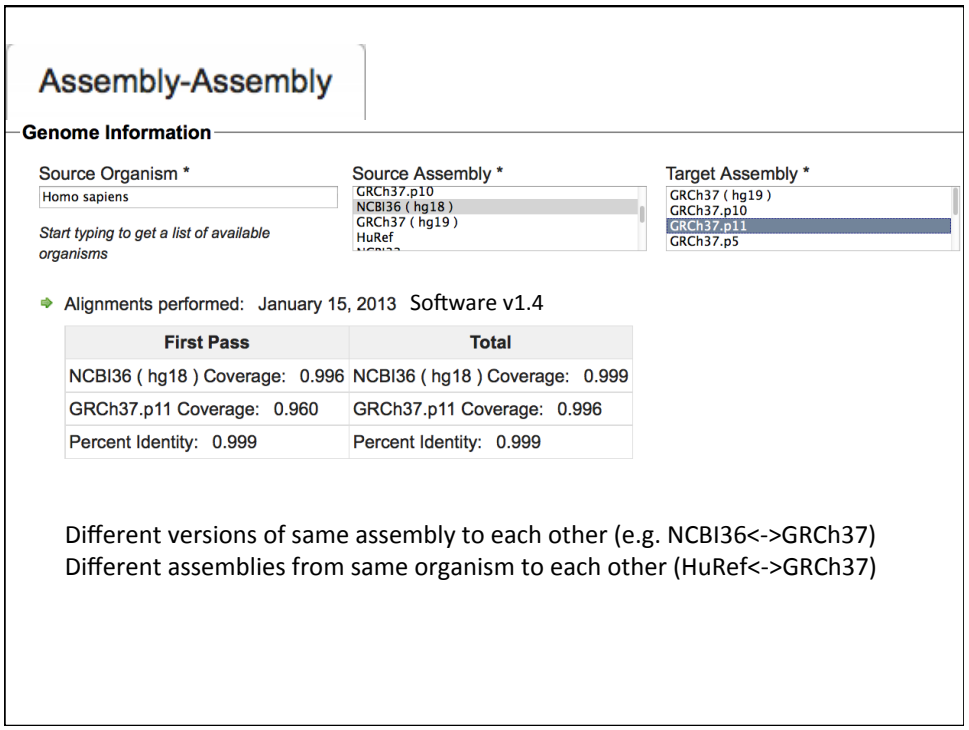
@deannachurch

Short Course in Medical Genetics 2013



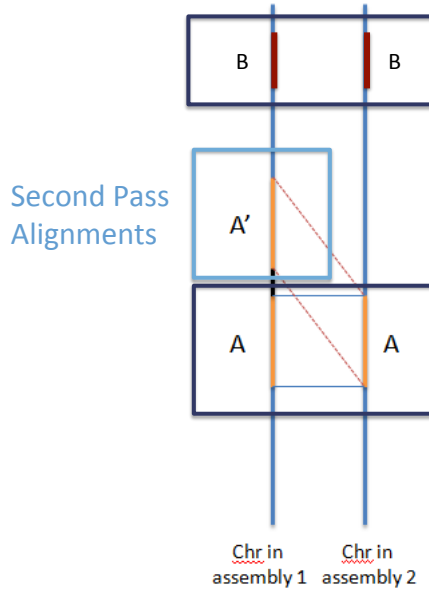
* Not drawn to scale





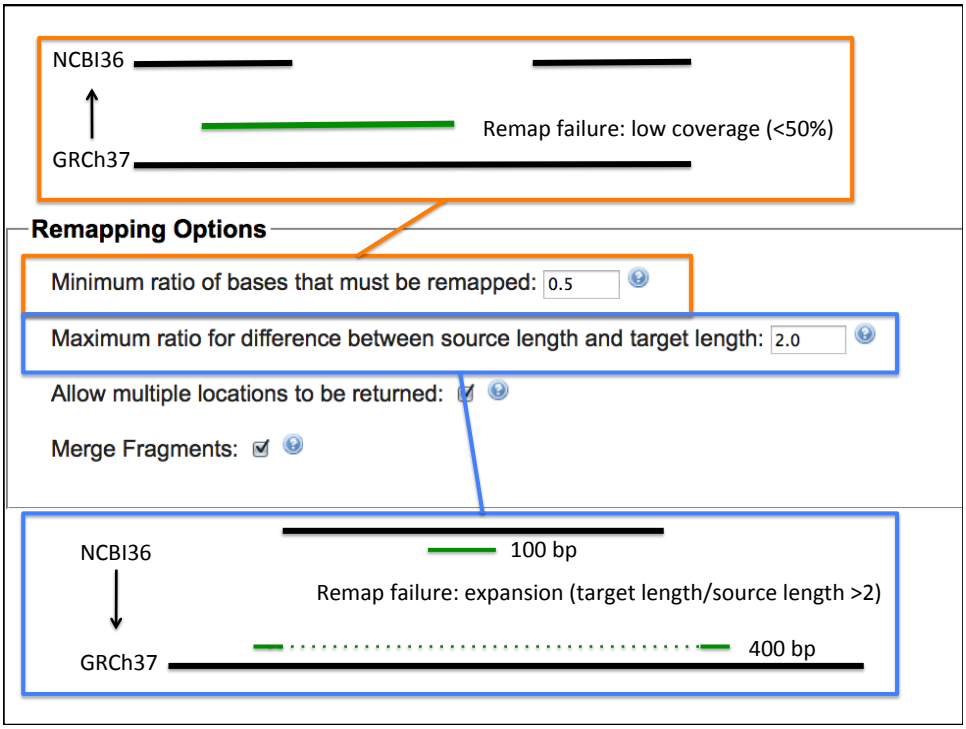
Different versions of same assembly to each other (e.g. NCBI36<->GRCh37)
Different assemblies from same organism to each other (HuRef<->GRCh37)

Producing Assembly-Assembly Alignments

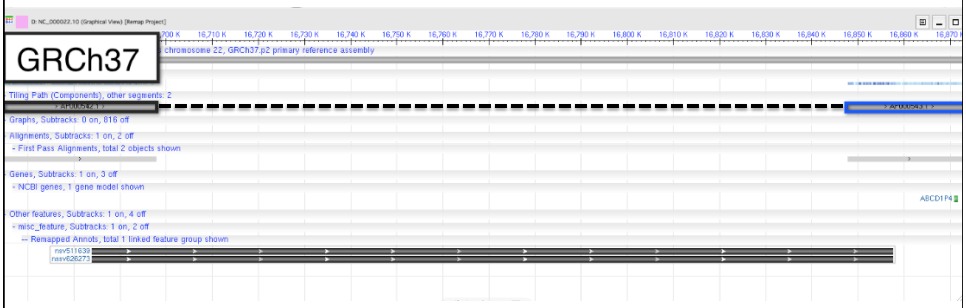


First Pass + Second Pass = Collapse
Second Pass Only = Expansion

http://www.ncbi.nlm.nih.gov/genome/tools/remap



Merge Fragments:



Helps rescue features that cross a gap
(common for CNVs/Structural Variants)

Merge Fragments:

Beware: Second Pass alignments and Merge

#feat_name	source_int	mapped_int	source_id	mapped_id	source_length	mapped_length	source_start	source_stop	sourc	mapped_start	mapped_stop	mapped_str	coverage	recip	asm_unit
dbVar: nssv6329	1	1	NC_000023.8	NC_000023.9	23483	23454	49049434	49072916	+	49233167	49256620	+	0.99877	First Pass	Primary Assembly
dbVar: nssv6329	1	2	NC_000023.8	NC_000023.9	23483	32747	49049434	49072916	+	49095963	49128709	+	1.3945	Second Pass	Primary Assembly
dbVar: nssv6329	1	3	NC_000023.8	NC_000023.9	23483	50554	49049434	49072916	+	49078155	49128708	+	2.15279	Second Pass	Primary Assembly
dbVar: nssv6329	1	4	NC_000023.8	NC_000023.9	23483	53160	49049434	49072916	+	49076782	49129941	+	2.26377	Second Pass	Primary Assembly
dbVar: nssv6329	1	5	NC_000023.8	NC_000023.9	23483	51931	49049434	49072916	+	49076782	49128712	+	2.21143	Second Pass	Primary Assembly
dbVar: nssv6329	1	6	NC_000023.8	NC_000023.9	23483	53160	49049434	49072916	+	49076782	49129941	+	2.26377	Second Pass	Primary Assembly

Remap Output

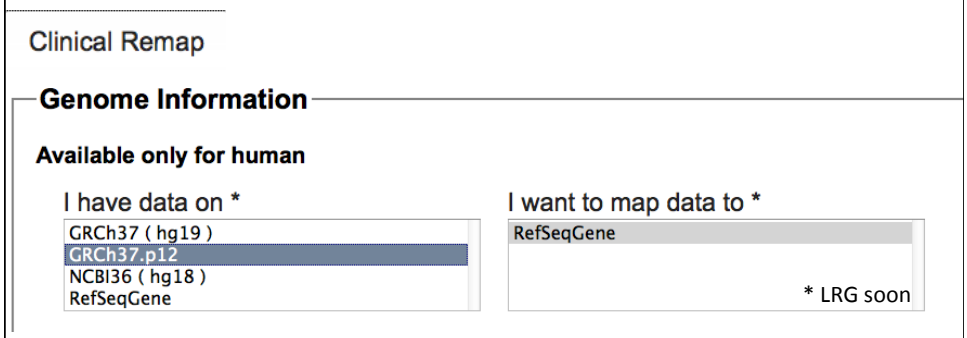
Summary data: Quick overview of how well your features mapped

Mapping report: Detailed report containing all of your input features and their source location, target location (or reason for failure) and coverage score.

Annotation File: An annotation file of only the features that successfully remapped. Suitable for loading to most browsers.

Genome Workbench file: A file formatted for loading to Genome Workbench (a client side browser). Includes assembly-assembly alignments for review.

Genome Workbench videos YouTube

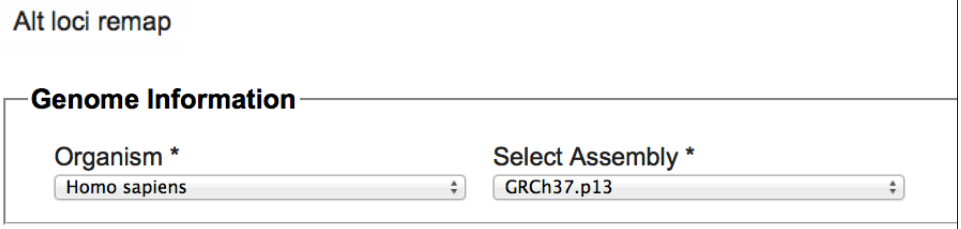


When mapping to RefSeqGene

Genomic location (NG_XXXXXX.X)
Transcript location(s) (NM_XXXXXX.X)
Protein location(s) (NP_XXXXXX.X)

] Optional, but checked by default

No second pass alignments, only one 'best' alignment



Maps features:

From Primary Assembly -> Alternate Loci/Patches (common)
From Alternate Loci/Patches->Primary Assembly

No second pass alignments, only one 'best' alignment

Take home messages

- Tools are available for mapping features from one coordinate system to another.
  - Assembly <-> Assembly
  - Assembly <-> RefSeqGene
  - Primary Assembly <-> Alternate Loci/Patches
- Feature remapping is NOT a substitute for *de novo* annotation.