


How I learned to quit worrying

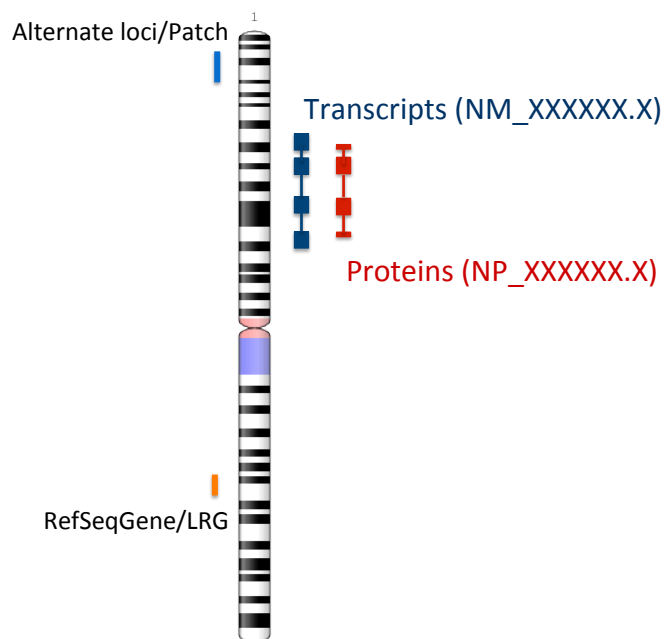
And love multiple coordinate systems

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



 @deannachurch

Short Course in Medical Genetics 2013



* Not drawn to scale

NCBI Genome Remapping Service

* Indicates required fields

Assembly-Assembly Clinical Remap Alt loci remap

Genome Information

Source Organism * Source Assembly Target Assembly *

Assembly-Assembly Clinical Remap Alt loci remap

Minimum ratio of bases that must be remapped: 0.5

Maximum ratio for difference between source length and target length: 2.0

Allow multiple locations to be returned: ☒

Merge Fragments: ☒

Data

Input format: Best Guess Output format: Same as input

Please upload a file or paste data here *

Upload a file: No file selected.

OR


Paste data here:

You can paste multiple lines into the text area.

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

NCBI Remap

[What is NCBI Remap?](#)
[About our alignments](#)
[FAQ](#)
[API Documentation](#)
[About Genome Workbench](#)



Assembly-Assembly

Genome Information

Source Organism * Source Assembly * Target Assembly *

Homo sapiens NCBI36 (hg18) GRCh37 (hg19)

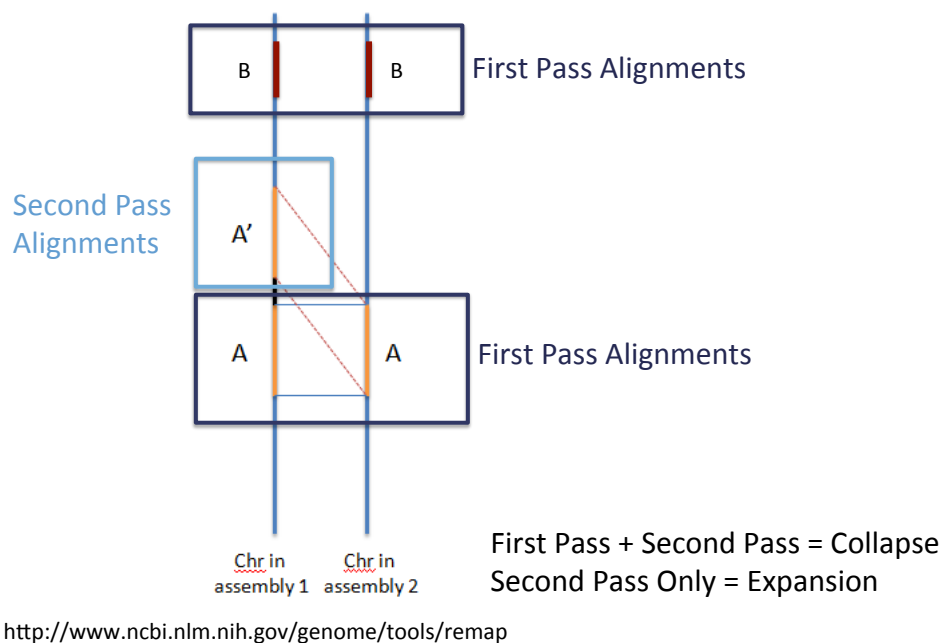
Start typing to get a list of available organisms


Alignments performed: January 15, 2013 Software v1.4

First Pass	Total
NCBI36 (hg18) Coverage: 0.996	NCBI36 (hg18) Coverage: 0.999
GRCh37.p11 Coverage: 0.960	GRCh37.p11 Coverage: 0.996
Percent Identity: 0.999	Percent Identity: 0.999


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





NCBI36 


↑


GRCh37  Remap failure: low coverage (<50%)


Remapping Options

Minimum ratio of bases that must be remapped: 


Maximum ratio for difference between source length and target length: 

Allow multiple locations to be returned: ☒ 

Merge Fragments: ☒ 

NCBI36  100 bp

↓

GRCh37  400 bp

Remap failure: expansion (target length/source length >2)

Merge Fragments: ☒

GRCh37

Chromosome 22: GRCh37.p2 primary reference assembly

Tiling Path (Components): other segments: 2

Graphs, Subtracks: 0 on, 616 off

Alignments, Subtracks: 1 on, 2 off

+ First Pass Alignments, total 2 objects shown

Genes, Subtracks: 1 on, 3 off

+ NCBI genes: 1 gene model shown

Other features, Subtracks: 1 on, 4 off

+ nesc_feature, Subtracks: 1 on, 2 off

+ Reremapped Annots, total 1 linked feature group shown

rev11133

rev10227

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

Merge Fragments: ☒

Beware: Second Pass alignments and Merge

+ Sequence NC_000023.8: Homo sapiens chromosome X, complete sequence

+ SNP, density (seq-feats)

+ Tiling Path (Components): finished: 2

+ Genes, Subtracks: 1 on, 2 off

+ NCBI genes, 3 gene models shown

NM_001468.1 GAGE1

NM_001472.1 GAGE2

GAGE3

NP_001459.1

+ Structural Variants, Subtracks: 1 on, 3 off

+ Source Annots, total 1 feature shown

nsrv6329

+ Alignments, Subtracks: 3 on, 4 off

+ First Pass Alignments, total 1 object shown

+ Second Pass Alignments, total 42 objects shown

+ FeatureAlignments, No data

feat_name	source_id	mapped_id	source_id	mapped_id	source_length	mapped_length	source_start	source_stop	source_mapped_start	source_mapped_stop	mapped_str	coverage	recip	asm_unit
dbVar: nsrv6329	1	1	NC_000023.8	NC_000023.9	23483	23454	49049434	49072916	+	49233167	49256620	+	0.99877	First Pass
dbVar: nsrv6329	1	2	NC_000023.8	NC_000023.9	23483	32747	49049434	49072916	+	49095963	49128709	+	1.3945	Second Pass
dbVar: nsrv6329	1	3	NC_000023.8	NC_000023.9	23483	50554	49049434	49072916	+	49078155	49128708	+	2.15279	Second Pass
dbVar: nsrv6329	1	4	NC_000023.8	NC_000023.9	23483	53160	49049434	49072916	+	49076782	49129941	+	2.26377	Second Pass
dbVar: nsrv6329	1	5	NC_000023.8	NC_000023.9	23483	51931	49049434	49072916	+	49076782	49128712	+	2.21143	Second Pass
dbVar: nsrv6329	1	6	NC_000023.8	NC_000023.9	23483	53160	49049434	49072916	+	49076782	49129941	+	2.26377	Second Pass

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 

Clinical Remap

Genome Information

Available only for human

I have data on *

GRCh37 (hg19)
GRCh37.p12
NCBI36 (hg18)
RefSeqGene

I want to map data to *

RefSeqGene

* LRG soon

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

Alt loci remap

Genome Information

Organism *

Homo sapiens

Select Assembly *

GRCh37.p13

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.