



The Genomic Remap Service from NCBI

Mapping genomic coordinates from one genome assembly to another for selective organisms

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

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope

Updates of genome assemblies or the availability of multiple assemblies for a single organism present an annotation comparison challenge. Performing a *de novo* annotation, especially for some datasets, can be computationally and time intensive. One way to quickly address this issue is to align the assemblies to each other and then ‘map’ the features from one coordinate system to the other using these alignments as a guide. While not a replacement for *de novo* annotation, this ‘remap’ process can be useful when looking at individual regions. When performed on annotations across the genome, this process can provide useful information on the degree of difference between the two assemblies. The Remap Service is a tool from NCBI that makes remapping features/annotations simple and straightforward.

Access to the service

The Remap Service is accessible online through the genome tools collection at <https://www.ncbi.nlm.nih.gov/genome/tools/remap/>. The Assembly-Assembly Remap Service allows for remapping of features from one assembly to another. For human, “Clinical Remap” allows for the remapping of features (including HGVS-defined sequence variations) to RefSeqGene sequences, including the underlying transcript and protein sequences. It also maps from the RefSeqGene sequences to an assembly.

The input form

The Remap Service takes input from a web form (as shown). The type of remap to be performed can be selected using tabs at the top (A). For Assembly-Assembly remap, organism and source/target genomes can be selected using the lists in the “Genome Information” section (B). Mapping criteria can be adjusted using parameters in the “Remapping Options” (C). Increasing the “minimum ratio of bases” and decreasing the “maximum ratio of difference” improve the stringency of the mapping process. Popups linked from question mark icons (D) provide context-specific help.

The actual list of coordinates to be remapped can be uploaded as a text file using the “Choose File” button (E) or pasted directly in the textbox (F). In the example, a set of features from GRCh37 in GFF3 format is pasted in the input box, see page 2 for details. Multiple formats for input/output coordinates are supported with specific formats selectable using the pull-down list (G). A message box at the bottom (not shown) identifies fields needing input data. It disappears when all requirements have been satisfied. The “Submit” button (H) will then become available.

NCBI Genome Remapping Service
* indicates required fields.

Assembly-Assembly Clinical Remap Alt loci remap **A**

Genome Information

Source Organism * **B**
Start typing to get a list of available organisms

Source Assembly *

Assm Name	Assm SeqID	Date
GRCh38.p2	GCF_000001405.28	2014/12/05
GRCh38.p1	GCF_000001405.27	2014/10/03
GRCh38 (hg38)	GCF_000001405.26	2013/12/17
GRCh37.p13	GCF_000001405.25	2013/06/28
GRCh37.p12	GCF_000001405.24	2013/03/26

Target Assembly *

Assm Name	Assm SeqID	Date
GRCh38.p8	GCF_000001405.34	2016/06/30
GRCh38.p7	GCF_000001405.33	2016/03/21
GRCh38.p6	GCF_000001405.32	2015/12/21
GRCh38.p2	GCF_000001405.28	2014/12/05
GRCh38.p1	GCF_000001405.27	2014/10/03

Alignments performed: September 27, 2016, software version: 1.7

Alignments	Source Assembly Coverage (GRCh37.p13, GCF_000001405.25)	Target Assembly Coverage (GRCh38.p8, GCF_000001405.34)	Percent Identity
First Pass(Reciprocal Best Hits)	97.273	94.705	99.997
Total	99.817	97.587	99.980

Remapping Options **C**

Minimum ratio of bases that must be remapped: **D**
Maximum ratio for difference between source length and target length:
Allow multiple locations to be returned:
Merge Fragments:

During the remapping, what is the minimum amount of the feature that must be able to be remapped? The default is 50%.

Data

Input format: Best Guess Output format: GFF3 **G** results in GFF3

Upload a file: No file chosen **E**
OR
Paste data here: **F**

You can paste multiple lines into the text area.

```
#GFF3 for remap from GRCh37 to GRCh38
#COLS: sid,src(SO:id),var,start,stop,score,ori,phase,attr(ID=val,etc)
chr1 . snp 154437613 154437613 . + . ID=rs148996840
chr1 . snp 169519894 169519894 . + . ID=rs6015
chr1 . snp 10942437 10942437 . + . ID=rs1551278630
chr6 . snp 10940942 10940942 . + . ID=rs150746645
chr6 . snp 10941008 10941008 . + . ID=rs149939366
chr9 . snp 130698043 130698043 . + . ID=rs6781
```

H

Example input format

The Remap Service supports several input formats. The example shown below is a list of SNPs from GRCh37p13 (hg19) in GFF3 format, which can be used as test input to the Remap Service to map them forward to GRCh38p2. For testing purposes, source and target assemblies can be switched.

```
#GFF3 for remap from GRCh37 to GRCh38
#COLS:sid,src(SO:id),var,start,stop,score,ori,phase,attr(ID=;etc)
chr1 . snp 154437613 154437613 . + . ID=rs148996840
chr1 . snp 169519894 169519894 . + . ID=rs6015
chr6 . snp 10942437 10942437 . + . ID=rs1551278630
chr6 . snp 10940942 10940942 . + . ID=rs150746645
chr6 . snp 10941008 10941008 . + . ID=rs149939366
chr9 . snp 130698043 130698043 . + . ID=rs6781
```

Displaying and downloading remap results

A completed remap request is displayed in a summary format with four main sections: Summary Data, Mapping Report, Annotation Data, and Genome Workbench Files.

The Summary section (A) provides a synopsis of the remapped features grouped by chromosome. Complete summary is downloadable using the disk icon. Up to ten remapped entries are shown in the Mapping Report section (B), with the complete set in the downloadable file marked by the disk icon (C). Clicking the icon activates a download window (D) from where the result can be saved or opened in Excel (under Firefox, other browsers may behave differently).

All features are in this report whether they successfully remapped or not. Additionally, this report provides interval-by-interval comparison of the features on the source and the target assemblies. Lastly, the coverage score is calculated by dividing the length of the feature in the target assembly by the length of the feature in the source assembly. A value of "1.00000" is perfect, while a value of less than 1 indicates a deletion in the target assembly and a value of greater than 1 indicates an insertion in the target assembly. This sort of metric does not assess single base changes within the region.

The link in the Annotation Data section (E) provides the remap data for the target assembly only. The link in the Genome Workbench Files section (F) provides the data in a format that can be readily imported into the standalone Genome Workbench tool (<https://www.ncbi.nlm.nih.gov/projects/gbench/>) from NCBI. Online links to help documents are also available from the result page (G) to provide more detailed explanation on the file format. Additional requests for technical assistance and feature additions to this tool can be emailed to NCBI using the "Write to the Help Desk" link (H).

API access

The Remap Service also provides API access. An example Perl script along with example command line are also provided. Detailed explanation is available in the online document:

<https://www.ncbi.nlm.nih.gov/genome/tools/remap/docs/api?page=result>

NCBI Remap Results
Feature remapping information from GRCh37.p13 (GCF_000001405.25) to GRCh38.p8 (GCF_000001405.34)

Summary Data (A)

ID	Source Features	Remapped Features	Source Intervals	Remapped Intervals
chr1	2	2	2	2
chr6	3	3	3	3
chr9	1	1	1	1

Mapping Report (sample) (B)

Feature	Src Intervals	Remap Intervals	Src Location	Src Length	Map Location	Map Length	Coverage
rs6781	1	1	chr9:130698043	1	chr9:127985764	1	1.00000
rs15512...	1	1	chr6:10942437	1	chr6:10942204	1	1.00000
rs15074...	1	1	chr6:10940942	1	chr6:10940709	1	1.00000
rs14993...	1	1	chr6:10941008	1	chr6:10940775	1	1.00000
rs14899...	1	1	chr1:154437613	1	chr1:154437613	1	1.00000
rs6015	1	1	chr1:169519894	1	chr1:169519894	1	1.00000

Annotation Data (E)

Genome Workbench Files (F)

Opening report_remapped_textarea.xls

You have chosen to open:
report_remapped_textarea.xls
which is: XLS file
from: http://www.ncbi.nlm.nih.gov

What should Firefox do with this file?

Open with: Microsoft Excel (default)

Save File

Do this automatically for files like this from now on.

OK Cancel



The Genome Decoration Page from NCBI

A tool to create publication quality graphics for chromosomes with custom annotations

<https://www.ncbi.nlm.nih.gov/genome/tools/gdp/>

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope and Access

The Genome Decoration Page (GDP) is a service that allows users to upload their own annotation data and create publication quality graphic showing on the ideogram of the chromosomes. Annotation can be shown as a track, or set of tracks, along the ideogram. Alternatively, an annotation track can be shown as an ideogram 'Overlay' on top of the chromosome representation. This service currently supports the decoration of certain builds of human and mouse genomes.



GDP is accessible on the web at: [gdp/](https://www.ncbi.nlm.nih.gov/genome/tools/gdp/)

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

The help documentation is online at:

<https://www.ncbi.nlm.nih.gov/genome/tools/gdp/help>

Answers to Frequently Asked Questions are at:

<https://www.ncbi.nlm.nih.gov/genome/tools/gdp/faq>

The GDP Start Page

The start page of GDP (shown below) consists of an input section to the left (A) and a set of simplified instructions (B) on how to use this service to the right. The input section consists of a "Genome Information" (C) subsection for selecting a target genome, a "Layout" (D) subsection for ideogram size and orientation, and an "Annotations" (E) subsection for uploading data for different tracks. The "Overlay" (F) option is for uploading data to create a track that is displayed directly on top of the ideogram. GDP supports close over two dozen model organisms. Requests to add additional genomes can be submitted through the "Support Center" link at the lower right (not shown).

The screenshot shows the NCBI Genome Decoration Page interface. At the top, there are navigation links for 'Genome Decoration Page', 'Bulk Conversion Service', and 'Ideogram Data'. The main heading is 'Genome Decoration Page' with the subtext 'Place your annotations on the genome'. The interface is divided into two main columns. The left column contains three main sections: 'Genome Information' (C), 'Layout' (D), and 'Annotations' (E). 'Genome Information' includes dropdowns for 'Organism' and 'Representation', and radio buttons for 'Chromosomes' (all or selected). 'Layout' includes a 'Font' dropdown and options for 'Ideogram' (vertical/horizontal), 'Band labels', and 'Legends'. 'Annotations' has a 'Track (no file)' section with 'Upload file', 'Color', 'Shape', and 'Width' options, and an 'Add track' button. Below it is an 'Overlay (no file)' section with similar options and an 'Add track' button. The right column contains instructions and a preview window. It starts with 'Step 1: Select Genome Information' and shows two ideogram icons labeled 'Cytogenetic' and 'Sequence'. Below this is 'Step 2: Set-up your layout' with a list of options. Then 'Step 3: Upload your Annotations' with instructions on file formats (GTF, GFF, GFF3, GVF, BED, HGVS). Finally, 'Step 4: Download your Figure' with instructions on downloading the figure in various formats. A 'Sign in to NCBI' link is in the top right corner.

Basic Controls

Controls in the “Genome Information” and “Layout” (A) determine the genome and basic layout. With a target genome selected, clicking the “Refresh” (B) button in the right panel generates a basic layout of the genome (C) when custom data is not provided. You can download this image in various formats (D).

The screenshot shows the 'Genome Decoration Page' interface. On the left, the 'Genome Information' section (A) includes controls for Organism (Homo sapiens), Representation (Cytogenetic | GRCh38 | 850), and Chromosomes (all selected). The 'Layout' section includes controls for Font (Courier), Ideogram (Vertical • Align: centromere), Band labels (Orientation: Lengthwise, Settings: Default), and Legends (Position: Below, Text: Horizontal, Columns: 1, 2, 3, Shape size: 5 px, Font size: 12 pt, Text color: black). On the right, the 'Ideogram Preview' panel (C) shows a basic ideogram for Cytogenetic: GRCh38.p7 (GCF_000001405.33) Resolution: 850. The 'Refresh' button (B) is highlighted. A 'Download format' dropdown (D) is set to PNG, with other options like JPEG, TIFF, EPS, SVG, and PDF visible. A 'Download' button is also present.

An Example Result

The “Annotations” control (E) allows upload of custom data files for different tracks. Clicking the “Refresh” button generates the decorated ideogram with the uploaded tracks (F). The service will display warning messages (G) if it detects issues in the uploaded file. In this case, it is an issue with unrecognized comment line that can be ignored. Clicking “Download” (H) brings out a dialog box to allow file size and name customization before download.

The screenshot shows the 'Annotations' section (E) on the left, with three tracks: 'Track (no file)' (gene_list.gff), 'Track (no file)' (snp_list.bed), and 'Overlay (no file)' (cyto_band.txt). The 'Ideogram Preview' panel (F) shows the decorated ideogram with the uploaded tracks. The 'Refresh' button is highlighted. A warning message (G) is displayed at the bottom: 'Warning InvalidRegion invalid band '#translated regions' - region ignored'. A 'Download' button (H) is highlighted, which opens a dialog box for file name and image size customization. The dialog box shows 'File name: ideogram.png' and 'Image size: 659 x 446'. The 'Download' button in the dialog is highlighted.

Example files used to generate the above image are available online:

https://ftp.ncbi.nlm.nih.gov/pub/factsheets/GDP_SNPList.bed

https://ftp.ncbi.nlm.nih.gov/pub/factsheets/GDP_GeneList.gff

https://ftp.ncbi.nlm.nih.gov/pub/factsheets/GDP_CytoBand.txt

Feedback