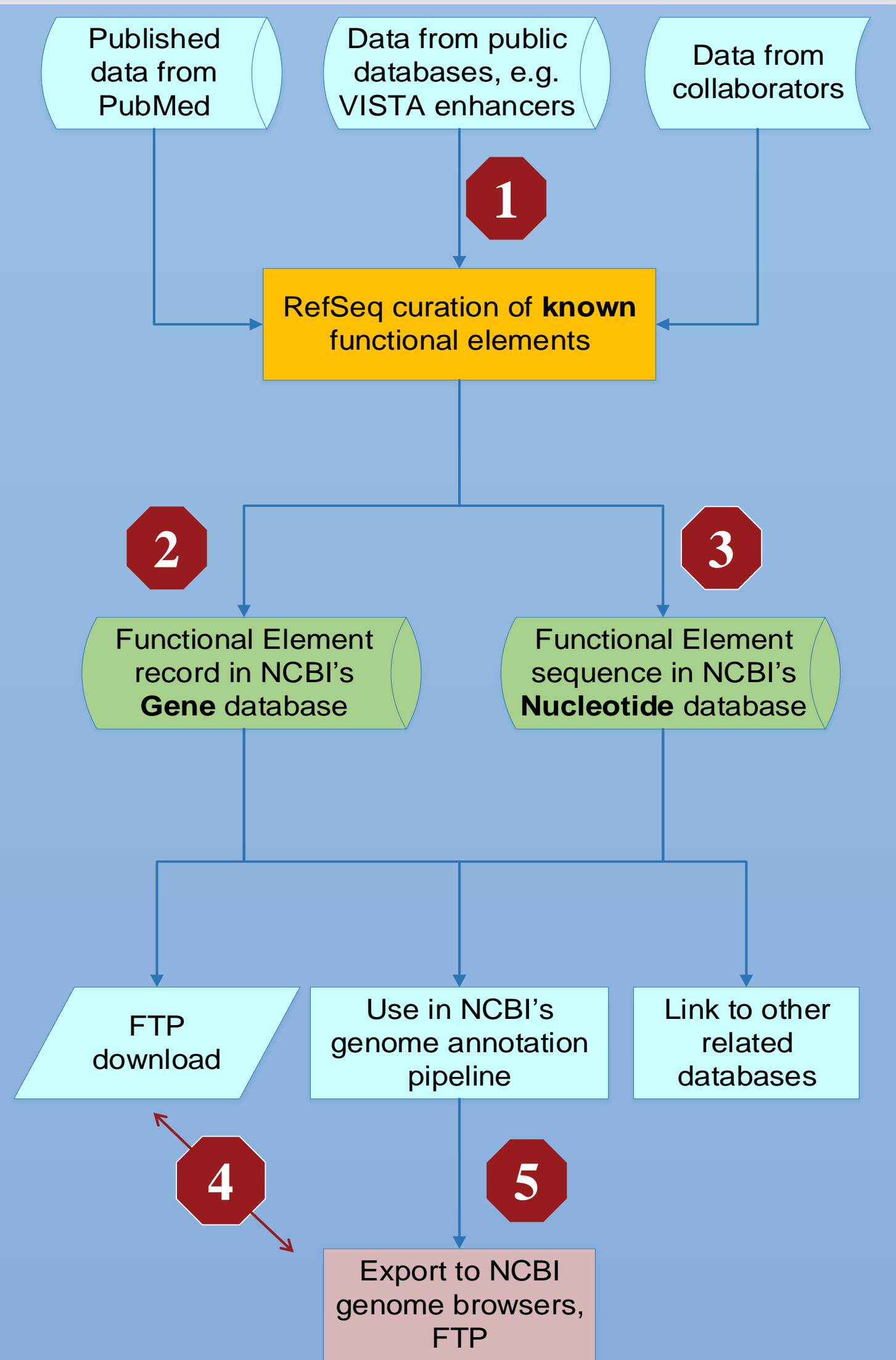


**Abstract:** The mouse genome contains many non-genic elements that function in diverse biological processes, such as gene regulation, chromosome organization, recombination or replication. Large-scale epigenomic mapping projects can predict the locations of gene regulatory elements, but those data are difficult to interpret in the absence of specialized research knowledge, are not generally visible in traditional genome annotation, have not been reconciled with experimental data in the literature, and do not always show function when tested experimentally. NCBI has therefore introduced a higher-confidence and more accessible dataset, RefSeq Functional Elements ([ncbi.nlm.nih.gov/refseq/functionalelements/](https://ncbi.nlm.nih.gov/refseq/functionalelements/)), which are annotated on the mouse genome alongside conventional genes. The dataset, which is comprised of richly annotated RefSeq records and descriptive records in the Gene database ([ncbi.nlm.nih.gov/gene/](https://ncbi.nlm.nih.gov/gene/)), includes known enhancers, silencers, and other non-genic regions with known function. Data are derived from experimental evidence in the literature, either based on individual locus studies or on experimentally validated subsets from larger-scale studies, such as the positive subset of VISTA enhancers. The dataset is publicly available for FTP download ([ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/vertebrate\\_mammalian/Mus\\_musculus/latest\\_assembly\\_versions/](ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/vertebrate_mammalian/Mus_musculus/latest_assembly_versions/)) and can be visualized in the 'Biological regions' track available in NCBI graphical displays, including the Genome Data Viewer. These known elements are valuable for basic discovery of gene regulatory regions, or as known positive controls for genome-wide studies aimed at new discovery. This presentation will encompass analyses of mouse RefSeq Functional Elements, including their overlap with genes and a variety of predicted gene regulatory, chromatin state segmentation and conserved element datasets. The results reveal how these known elements are organized relative to genes, and indicate that this dataset incorporates gene regulatory elements that have not been found in epigenomic predicted datasets, including several developmental-specific elements that have been elucidated in the literature.



## RefSeq Functional Elements scope

**Definition:**

- Any **non-genic** genomic element that has functional significance **based on experimental support**

**Types:**

- Gene regulatory elements**, e.g. enhancers, silencers, promoters, protein binding sites
- Known structural elements**, e.g. boundary elements, chromatin conformation-associated regions
- Other elements of functional importance**, e.g. well-defined recombination hotspots or replication origins

**Annotation scope:**

- Elements that have been **experimentally validated**
- Human and mouse** elements

## Gene database records

[ncbi.nlm.nih.gov/gene/](https://ncbi.nlm.nih.gov/gene/)

The screenshot shows the NCBI Gene database record for Hbb-ar (hemoglobin, activating region) in Mus musculus. The record includes a summary, genomic context, and various annotations. A red box highlights the 'Other curated content' section, which includes 'Genomic regions, transcripts, and products', 'Variation', 'Interactions', 'General gene information', 'Homology', 'NCBI Reference Sequences (RefSeq)', and 'Related sequences'. The 'Genomic context' section shows the location of the gene on chromosome 7, with a scale bar indicating the distance from the gene to the centromere and telomere.

## RefSeq records

- Genomic accessions with NG\_ prefixes
- Data standard ontologies from:
  - International Nucleotide Sequence Database Collaboration (INSDC)
  - Sequence Ontology (SO)
  - Evidence & Conclusion Ontology (ECO)
- Links to publications in PubMed
- INSDC qualifiers with functional and descriptive metadata

**Example:** Segment of NG\_055003.1, *Hbb-ar*, GeneID:15128

```
regulatory      388..596
                /regulatory_class="promoter"
                /experiment="EXISTENCE:reporter gene assay evidence
                [ECO:0000049][PMID:12920040]"
                /note="-158 to +52 epsilon-y promoter fragment"
                /function="Hbb-y gene minimal promoter in transfected MEL
                cells"
                /db_xref="GeneID:15128"

regulatory      540..596
                /regulatory_class="transcriptional_cis_regulatory_region"
                /experiment="EXISTENCE:reporter gene assay evidence
                [ECO:0000049][PMID:12920040]"
                /note="sequence between -100 and -158 that includes a
                CACCC motif; based on difference between the activities of
                the HS2 -100 and HS2 -158 promoter constructs"
                /function="region critical for epsilon-y promoter function
                in uninduced MEL cells"
                /db_xref="GeneID:15128"

protein_bind     540..562
                /experiment="EXISTENCE:protein binding evidence
                [ECO:0000024][PMID:12920040]"
                /note="22-bp epsilon-y-CACCC BKLF- and EKLF-binding
                oligomer"
                /bound_moiety="Krueppel like factors 1 and 3"
                /function="necessary for Hbb-y promoter activity"
                /db_xref="GeneID:15128"
```

## Data access

Find multiple ways to access the data on our website:

[ncbi.nlm.nih.gov/refseq/functionalelements/](https://ncbi.nlm.nih.gov/refseq/functionalelements/)

### NCBI RefSeq Functional Elements

- [Overview](#)
- [RefSeq Functional Element Records](#)
  - [RefSeq Functional Element Feature Annotation](#)
  - [Feature Annotation Glossary](#)
- [Data Access](#)
  - [Access via Gene](#)
  - [Access via Nucleotide](#)
  - [Access via BLAST](#)
  - [Access via BioProject](#)
  - [Access via NCBI Graphical Displays](#)
  - [Access via FTP](#)
    - [Feature table](#)
    - [Feature extraction examples](#)
- [Feedback](#)

## NCBI genome browser view

View feature annotation for RefSeq Functional Elements in the 'Biological regions' track in an NCBI genome browser, e.g. Genome Data Viewer ([ncbi.nlm.nih.gov/genome/gdv/](https://ncbi.nlm.nih.gov/genome/gdv/))

The screenshot shows the NCBI Genome Data Viewer for the Hbb-ar gene. The 'Biological regions' track is highlighted, showing various RefSeq Functional Elements such as enhancers, promoters, and protein binding sites. A red box highlights the 'Biological regions, aggregate, NCBI Mus musculus Annotation Release 108, 2019-08-10' track. A tooltip for the 'enhancer' feature is shown, providing details about its location, length, and function. Another tooltip for the 'protein\_bind' feature is shown, providing details about its location, length, and function.

**Example:** Genome Data Viewer image showing *Zrs1* (GeneID: 105804842) feature annotation within an intron of *Lmbr1* (GeneID: 56873). Tooltips are displayed for select features.

