SRA: Sequence Read Archive

Collection of sequence data from next-generation sequencing technology for different organisms
http://www.ncbi.nlm.nih.gov/sra/, http://www.ncbi.nlm.nih.gov/Traces/sra/

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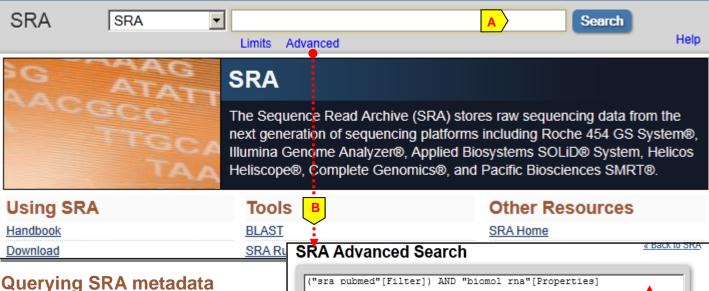
Scope

Sequence Read Archive (SRA) is the NCBI database which stores sequence data obtained from next generation sequence (NGS) technology. Through this database, the metadata for those sequences can be queried to locate the sequence dataset for subsequent download and further analysis. Specifically, SRA:

- Archives of raw oversampling NGS data for various genomes from several platforms;
- Shares NGS data with EMBL and DDBJ;
- Serves as a starting point for "secondary analysis";
- Provides access to data from human clinical samples to authorized users who agree to the dataset's privacy and usage mandates.
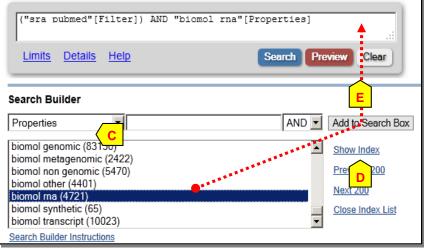
Data access

Metadata from SRA can be queried from Entrez SRA page and the SRA project list plus sequence data can be browsed, searched and downloaded from its homepage at http://www.ncbi.nlm.nih.gov/sra/ and http://www.ncbi.nlm.nih.gov/Traces/sra/, respectively. For sequence-based search against certain subsets of SRA reads (long reads from 454 platform) using BLAST, a link is listed under the "Specialized BLAST" section in the BLAST homepage: http://blast.ncbi.nlm.nih.gov/ .

Editors:

P. Cooper, R. Morris, L. Szilagyi

Layout Designer:

W. Ratzat

Layout Editor:

T. Tao

Contributors:

W. Matten, S. McGinnis, M. Romiti, E. Sayers, M. Shumway, M. Valjavec-Gratian

Contact:

info@ncbi.nlm.nih.gov

(301) 496-2475

Querying SRA metadata

Querying SRA metadata can be performed through the Entrez SRA page by entering desired terms and clicking the "Search" button (**A**). Complex query can be constructed using functions provided by Advanced (**B**) page, where indexed field and available values can be examined using the pull-down list (**C**) and the "Show Index" link (**D**). A selected term can be added to the search box using "Add to Search Box" button (**E**). An example set of query terms thus constructed in in the search box.

Metadata result page

Initial search result is displayed in summary format (**A**). Clicking a title opens the record (**B**) to provide more details on the experiment. In the detailed display, the summary of the experiment is given at the top (**C**), followed by data downloading links (**D**) and links to details of individual runs in SRA browsers through the SRR accessions (**E**). Entries in other databases related to this experiment, such as the publication (PubMed) describing the experiment are given in the "Related Information" portlet (**F**).

Display Settings: ⊙ Summary **A** Send to: ⊙

Results: 3

☐ Epigenetics of house mouse **B**
1. 2 ILLUMINA (Illumina Genome Analyzer) runs: 11.6M spots, 426.1M bases, 494.9MB downloads
Accession: SRX002662

☐ Epigenetics of house m
2. 3 ILLUMINA (Illumina Geno
Accession: SRX002661

☐ Epigenetics of house m
3. 2 ILLUMINA (Illumina Geno
Accession: SRX002660

Display Settings: ⊙ Full Send to: ⊙

Accession: SRX002662
Experiment design: n/a
Submission: SRA008170 by GEO
Study summary: ChIP-seq Accurately Predicts Tissue-Specific Activity of Enhancers (SRP000600) •
Study • All experiments (more...) **C**
Sample: source: Mouse embryonic day 11.5 (e11.5) midbrain; description: Tissue pooled from ~150
embryos Chromatin IP against p300 (SRS002167) (more...)
Library: Midbrain_p300_ChIP-seq (more...)
Platform: Illumina (more...)
Processing:
  Base calls: Base Space,
  Quality score: , 0x0.0E0
Spot descriptor:
  [forward]
  1

Total: 2 runs. 11.6M spots. 426.1M bases **D**
Download reads for this experiment in sra (494.8M) *or sra-lite* (494.8M) *formats* ⓘ

#	Run	# of Spots	# of Bases	
1.	SRR013594	3,709,655	141M	**E**
2.	SRR013595	7,921,562	285.2M	

Related information ▲ **F**
BioSample
GEO DataSets
PubMed
Taxonomy

Recent activity ▲
Turn Off Clear

🔍 SRA Links for PubMed (Select 19212405) (3)
  SRA

See more...

Getting the sequence data via links

Often interests of a specific set of SRA data is prompted by a published paper. Entrez indexing makes this a relatively straight forward process. A filter term "pubmed_sra[filter]" (**G**) retrieves PubMed record with link to SRA, such as this ChIP-seq paper displayed in Abstract format (**H**). Clicking the SRA link (**I**) under the "Related Information" retrieves all the relevant SRA experiments under this reported project and displays them in summary format as shown above (**A**).

PubMed.gov | PubMed ▼ | pubmed_sra[filter] AND nature[jour] AND "Chip-seq"[title] ⊗ **G** | Search
US National Library of Medicine
National Institutes of Health   Help

🔊 RSS Save search Limits Advanced

Display Settings: ⊙ Abstract **H** Send to: ⊙

Nature. 2009 Feb 12;457(7231):854-8.

ChIP-seq accurately predicts tissue-specific activity of enhancers.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA.

Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

Abstract
A major yet unresolved quest in decoding the human genome is the identification of the regulatory sequences that control the spatial and temporal expression of genes. Distant-acting transcriptional enhancers are particularly challenging to uncover because they are scattered among the vast non-coding portion of the genome. Evolutionary sequence constraint can facilitate the discovery of enhancers, but fails to predict when and where they are active in vivo. Here we present the results of chromatin immunoprecipitation with the enhancer-associated protein p300 followed by massively parallel sequencing, and map several thousand in vivo binding sites of p300 in mouse embryonic forebrain, midbrain and limb tissue. We tested 86 of these sequences in a transgenic mouse assay, which in nearly all cases demonstrated reproducible enhancer activity in the tissues that were predicted by p300 binding. Our results indicate that in vivo mapping of p300 binding is a highly accurate means for identifying enhancers and their associated activities, and suggest that such data sets will be useful to study the role of tissue-specific enhancers in human biology and disease on a genome-wide scale.

PMID: 19212405 [PubMed - indexed for MEDLINE] PMCID: PMC2745234 **Free PMC Article**

Images from this publication. See all images (5) Free text

⊕ Publication Types, MeSH Terms, Substances, Secondary Source ID, Grant Support
⊕ LinkOut - more resources

nature | FREE Author Manuscript in PubMed Central

Section removed for clarity.

Related information ▲
Related Citations
GEO DataSets
Gene
Gene (OMIM)
HomoloGene
Nucleotide (RefSeq)
Nucleotide (Weighted)
OMIM (calculated)
OMIM (cited)
Protein (RefSeq)
Protein (Weighted)
References for this PMC Article
SRA ● **I**
Substance (MeSH Keyword)
Taxonomy via GenBank
UniGene
GEO Profiles
Free in PMC

Additional display

SRA homepage (**J**) provides additional display and browsing functionality to allow the examination of SRA data objects at different levels using easy to recognized tabs. SRA-specific documents and software toolkit are also available under the "Documentation" and "Software" tabs, respectively. Detailed SRA help document is in the NCBI Bookshelf at http://www.ncbi.nlm.nih.gov/books/NBK47528/. Comments and feedback should be sent to: sra@ncbi.nlm.nih.gov

J **Sequence Read Archive** http://www.ncbi.nlm.nih.gov/Traces/sra/

| Main | Browse | Search | Download | Submit | Documentation | Software |

| Studies | Samples | Analyses | Run Browser | Provisional SRA |