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/.

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).

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 straightforward 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).

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