# AIRR Community:
# Minimal Standards Working Group

**Update: December 4, 2017**

**NIAID**

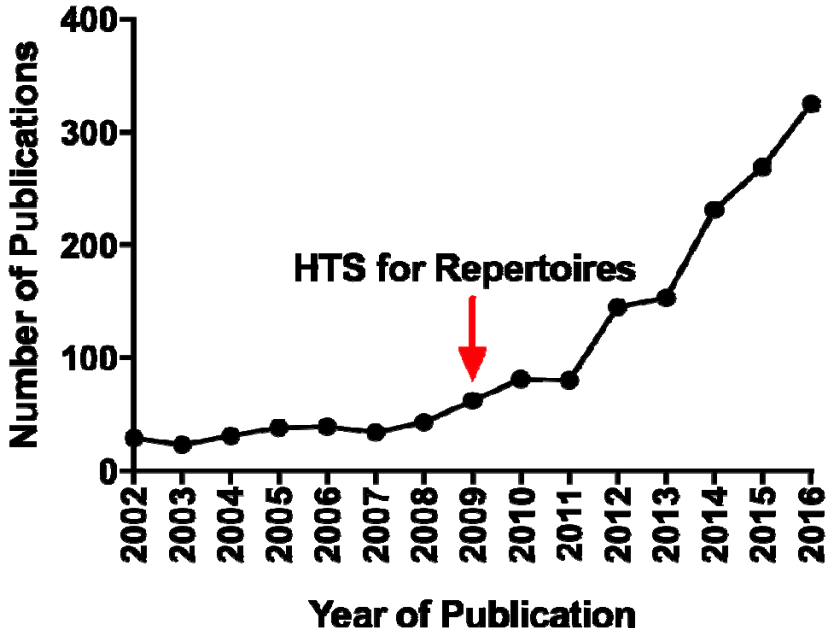
Nina Luning Prak
Steven H. Kleinstein
Florian Rubelt
Syed Ahmad Chan Bukhari
Christian Busse

# Overview

1. **WG's mission**
2. **Describe the progress**
3. **Future directions**

# Working Group Mission

Propose standard for data deposition with publication & sharing




Data should be described in sufficient detail such that a person skilled in the art of AIRR sequencing and data analysis will be able to reproduce the experiment and data analyses that were performed

# AIRR Community Minimal Standards WG

Set of data elements to describe data with publication/release

- Eline Luning Prak (co-chair), Steven Kleinstein (co-chair)
- Syed Ahmad Chan Bukhari, Brian Corrie, Bjoern Peters, Bojan Zimonja, Chaim Schramm, Christian Busse, Corey Watson, Encarnita Mariotti-Ferrandiz, Felix Breden, Florian Rubelt, Jean Bürckert, Jerome Jaglale, Lindsay Cowell, Marie-Paule Lefranc, Nishanth Marthandan, Richard Bruskiewich, Scott Boyd, Scott Christley, Uri Hershberg, Uri Laserson, William Faison, Brandon DeKosky

Monthly teleconferences, votes at AIRR Community annual meetings

COMMENT

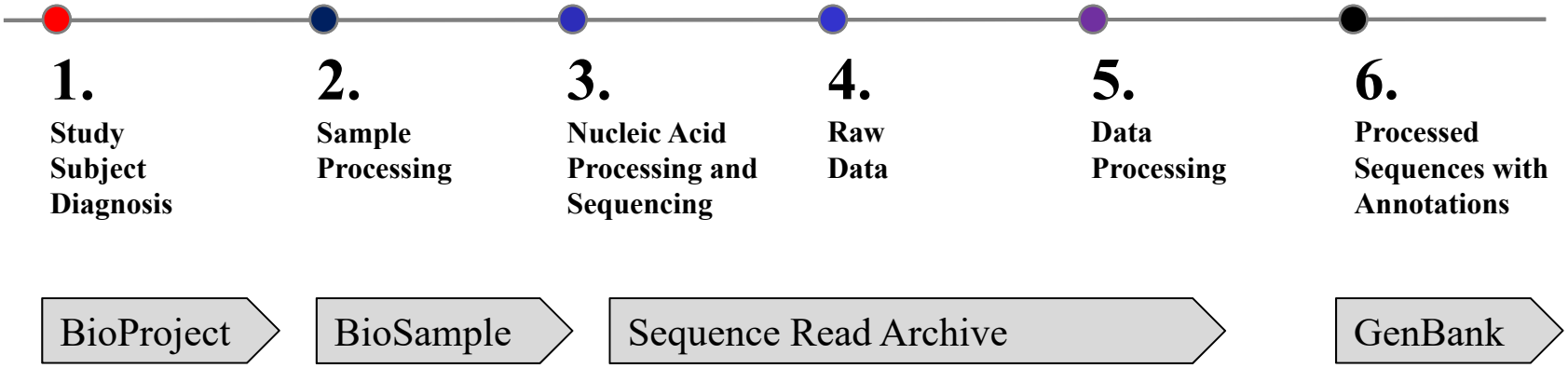
# Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data

Florian Rubelt[1,21], Christian E Busse[2,21], Syed Ahmad Chan Bukhari[3,21], Jean-Philippe Bürckert[4], Encarnita Mariotti-Ferrandiz[5], Lindsay G Cowell[6], Corey T Watson[7], Nishanth Marthandan[8], William J Faison[9], Uri Hershberg[10], Uri Laserson[11], Brian D Corrie[12,13], Mark M Davis[1,14], Bjoern Peters[15], Marie-Paule Lefranc[16], Jamie K Scott[8,12,17], Felix Breden[12,13], The AIRR Community[18], Eline T Luning Prak[19,22] & Steven H Kleinstein[3,20,22]

**High-throughput sequencing of B and T cell receptors is routinely being applied in studies of adaptive immunity. The Adaptive Immune Receptor Repertoire (AIRR) Community was formed in 2015 to address issues in AIRR sequencing studies, including the development of reporting standards and the sharing of data sets.**

# MiAIRR Standard + implementation @ NCBI

AIRR-seq data can be deposited over 4 linked NCBI data repositories



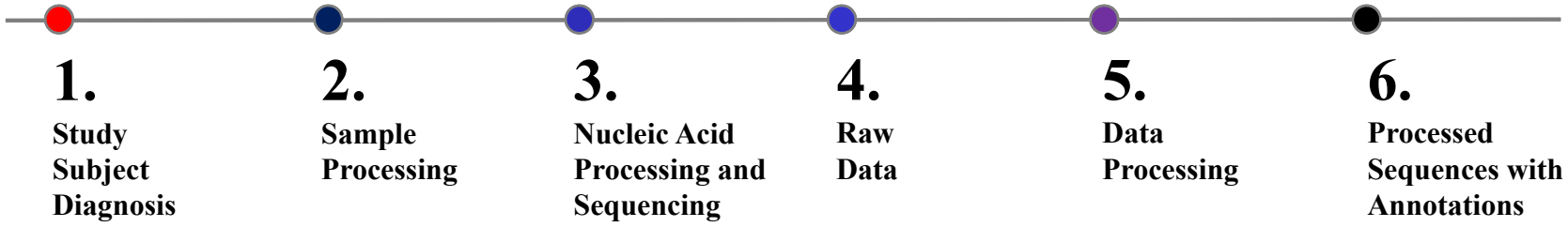

**All of these repositories are linked by identifiers at NCBI.**

**To deposit data at SRA, first need to create BioProject and BioSample**

Details are available at: http://airr-community.org

# MiAIRR includes processed sequences

Deposition of raw and processed sequences facilitates secondary analysis




MiAIRR

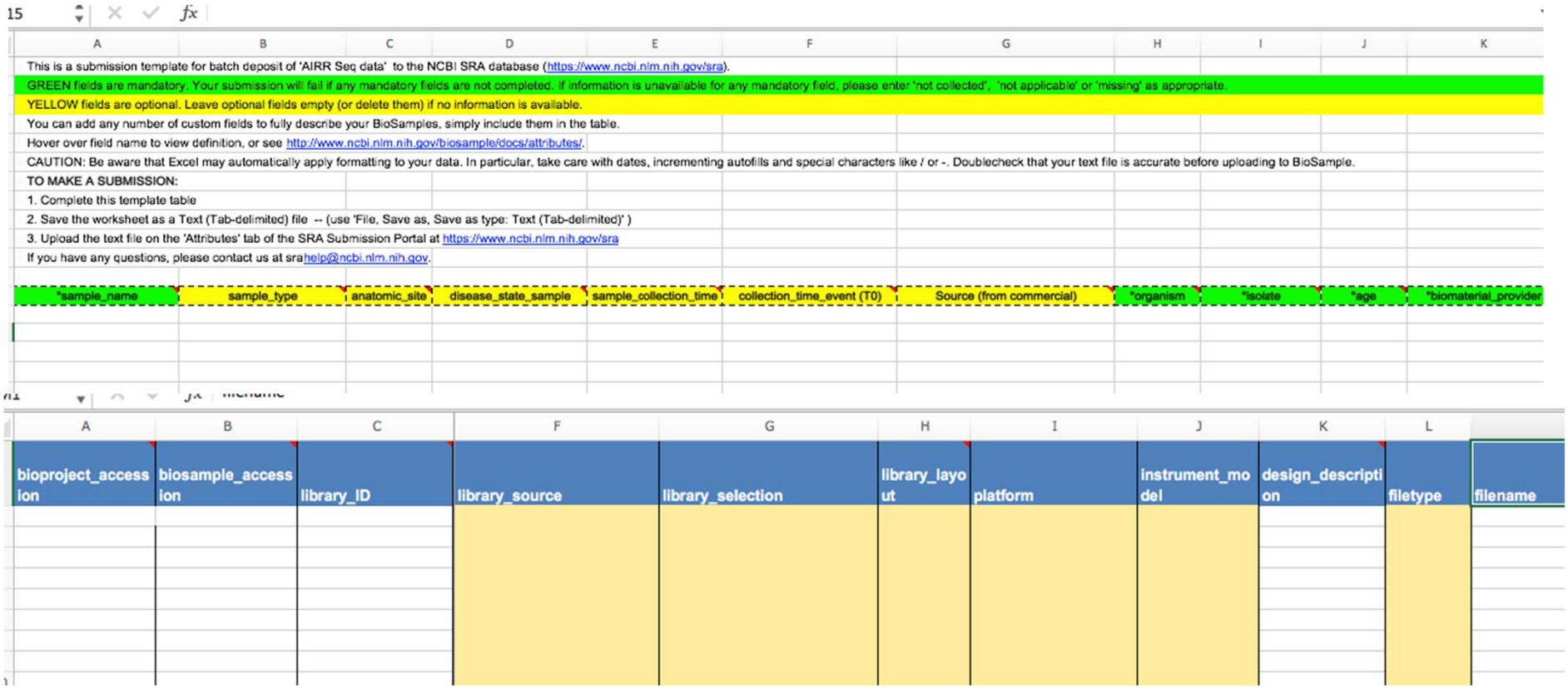
- VDJ germline reference database
- Cell index
- V gene
- D gene
- J gene
- C region
- IMGT-JUNCTION nucleotide sequence
- IMGT-JUNCTION amino acid sequence
- Read count

GenBank

MiAIRR can be implemented by multiple repositories beyond NCBI

# Implementation of AIRR Standard @ NCBI

MiAIRR-compliant templates have been developed in collaboration with NCBI



This is a submission template for batch deposit of 'AIRR Seq data' to the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra).
GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed. If information is unavailable for any mandatory field, please enter 'not collected', 'not applicable' or 'missing' as appropriate.
YELLOW fields are optional. Leave optional fields empty (or delete them) if no information is available.
You can add any number of custom fields to fully describe your BioSamples, simply include them in the table.
Hover over field name to view definition, or see http://www.ncbi.nlm.nih.gov/biosample/docs/attributes/.
CAUTION: Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate before uploading to BioSample.
TO MAKE A SUBMISSION:
1. Complete this template table
2. Save the worksheet as a Text (Tab-delimited) file -- (use 'File, Save as, Save as type: Text (Tab-delimited)' )
3. Upload the text file on the 'Attributes' tab of the SRA Submission Portal at https://www.ncbi.nlm.nih.gov/sra
If you have any questions, please contact us at srahelp@ncbi.nlm.nih.gov.

| *sample_name | sample_type | anatomic_site | disease_state_sample | sample_collection_time | collection_time_event (T0) | Source (from commercial) | *organism | *isolate | *age | *biomaterial_provider |
|---|---|---|---|---|---|---|---|---|---|---|

| bioproject_accession | biosample_accession | library_ID | library_source | library_selection | library_layout | platform | instrument_model | design_description | filetype | filename |
|---|---|---|---|---|---|---|---|---|---|---|

Details are available at: http://airr-community.org

# Example GenBank Record

Processed sequence data, including V(D)J assignments and CDR3 sequence

9 Example record (GenBank format)

```
LOCUS       AB123456                420 bp    mRNA    linear   EST 01-JAN-2015
DEFINITION  <free text description>
ACCESSION   AB123456
VERSION     AB123456.7
KEYWORDS    <other keywords>; AIRR.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
            Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   Stibbons,P.
  TITLE     Section 5 information for experiment FOO1
  JOURNAL   published (01-JAN-2000) on Figshare
  REMARK    DOI:10.1000/0000-12345678
REFERENCE   2  (bases 1 to 420)
  AUTHORS   Stibbons,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2000) Center for Transcendental Immunology, Unseen
            University, Ankh-Morpork, 12345, DISCWORLD
DBLINK      BioProject: PRJNA000001
            BioSample: SAMN000001
            Sequence Read Archive: SRR0000001
FEATURES             Location/Qualifiers
     source          1..420
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /citation=[1]
                     /rearranged
                     /note="AIRR_READ_COUNT:123"
     CDS             1..420
                     /codon_start=3
                     /translation="PGASVKMSCKASGYTFTDYNIHWVKQSHGKSLEWIAYINPNNGGYG
                     YNDKFRDKATLTVDRSSNTAYMGLRSLTSEDSAVYYCARAGVYDGYTMDYWGQGTSVTVS
                     SAKTTAPSVYPLAPVCGGTTGSSVTLGCLVKGN"
     V_region        1..324
     V_segment       1..257
                     /gene="IGHV1-34"
                     /allele="01"
                     /db_xref="IMGT/LIGM:AC073565"
     D_segment       266..272
```

13

```
     J_segment       291..324
                     /gene="IGHJ4"
                     /allele="01"
                     /db_xref="IMGT/LIGM:V00770"
     misc_feature    258..290
                     /function="CDR3"
                     /inference="COORDINATES:nucleotide motif:IgBLAST:1.6"
     C_region        325..420
                     /gene="Ighg2c"
ORIGIN
        1 agcctggggc ttcagtgaag atgtcctgca aggcttctgg ctacacattc actgactata
       61 acatacactg ggtgaagcag agccatggaa agagccttga gtggattgca tatattaatc
      121 ctaacaatgg tggttatggc tataacgaca agttcaggga caaggccaca ttgactgtcg
      181 acaggtcatc caacacagcc tacatggggc tccgcagcct gacctctgag gactctgcag
      241 tctattactg tgcaagagcg ggagtttacg acggatatac tatggactac tggggtcaag
      301 gaacctcagt caccgtctcc tcagccaaaa caacagcccc atcggtctat ccactggccc
      361 ctgtgtgtgg aggtacaact ggctcctcgg tgactctagg atgcctggtc aagggcaact
```

Detailed “how-to” document available on AIRR Community website

# 5-steps to submit MiAIRR-compliant data

## MiAIRR:Minimum information about an Adaptive Immune Receptor Repertoire Sequencing Experiment

MiAIRR Paper

Templates How to Guide?

**Submission of AIRR sequencing data and metadata to NCBI's public data repositories consists of five sequential steps:**

1. Submit study information to NCBI BioProject using the NCBI web interface.
2. Submit sample-level information to the NCBI BioSample repository using the AIRR-BioSample templates.
3. Submit raw sequencing data to NCBI SRA using the AIRR-SRA data templates.
4. Generate a DOI for the protocol describing how raw sequencing data were processed using Zenodo or an equivalent DOI-granting service.
5. Submit processed sequencing data with sequence-level annotations to GenBank using AIRR feature tags.

Contact **steven.kleinstein@yale.edu** or **ahmad.chan@yale.edu**, if you need help in preparing or submitting your data according to MiAIRR standards to the NCBI.

# Do you have AIRR-seq data to deposit?

We need users for MiAIRR and the NCBI data submission system…


(and are willing to help deposit your data)

MiAIRRhelp@googlegroups.com –or- @miairrhelp

# AIRR standards 2018 - MiAIRR 1.1

## Make it known, make it easy, demonstrate its utility

**KNOWN**

- Reach out to and assist other labs to submit their data
- Bug-fix MiAIRR-NCBI implementation
- Refine MiAIRR 1.0 data fields as absolutely necessary
- **Must** remain compatible with current NCBI submissions

**EASY**

- Further develop submission + retrieval toolkit/pipeline
- Make available as NCBI package
- Identify ontologies for a limited number of key data elements
- Evaluate submission to other INDSC repos (EBI/ENA)

**UTILITY**

- Showcase with a meta-analysis using multiple data sets

**Main Goal:** Increase # of public data sets

# AIRR Standards 2018: Proposed Work Products

**Paper : Analysis & comparison of AIRR data**

- Requesting collaborators with data sets for this project: Autoimmunity, Infectious Disease, Cancer.
- Data can be previously published, but would need to be submitted via MiAIRR standard.

*Please contact us!*

# AIRR standards >2018 - MiAIRR 2.0

- develop mechanisms how AIRR studies can report data related to single cells:
  - cell phenotypes (e.g. flow cytometry)
  - Ig/TCR reactivities and functional properties
  - structural information
- Identify repositories able to host such data (together with ComRepo WG)

# Thank you…

To join the AIRR Standards Working Group, please contact us
join@airr-community.org


**Contribute MiAIRR data for meta-analysis paper...**
luning@pennmedicine.upenn.edu –or- steven.Kleinstein@yale.edu

The End

# Published MiAIRR data standard

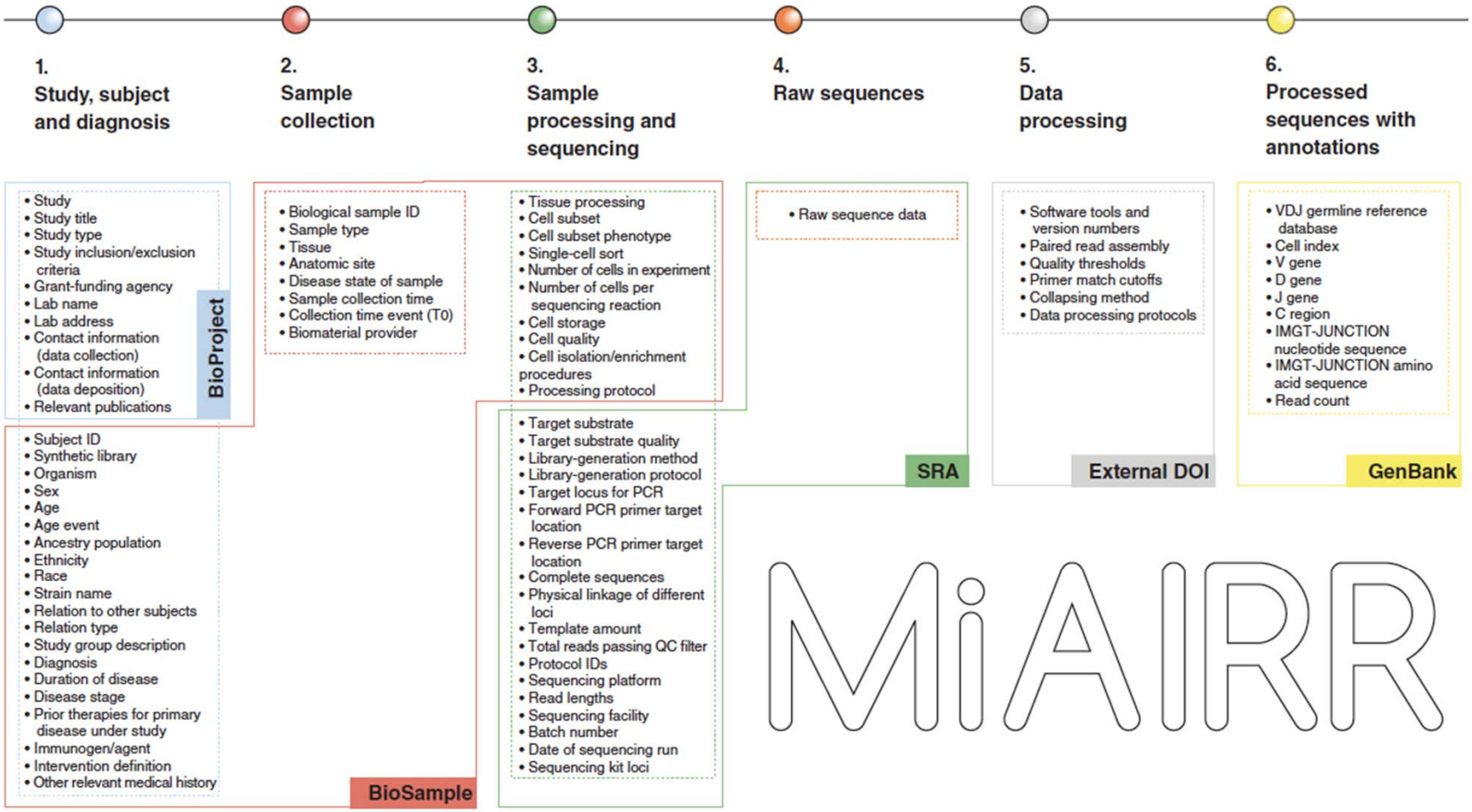



**Figure 2** An overview of the six MiAIRR sets and associated data fields, along with their target submission repositories at NCBI.