Keeping your Data FAIR

Organizing, Managing, and Sharing your Data

Peter FitzGerald, PhD

Outline

- Principles of FAIR
- The practical aspects of adhering to a FAIR approach
- Data Management Strategies
  - Record Keeping, Storage, file formats
- Sharing
  - What, Where, When
- Repositories
- Journal Requirements
- Example RNA-Seq project
- Resources

FAIR Principles

https://www.go-fair.org/fair-principles/

F indable: The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

A ccessible: Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

I nteroperable: The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

R eusable: The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

FAIR Principles

In practice in your local environment (computer)

Findable: Clearly the first step in managing data it to be able to find it. This requires that the data be stored and annotated in a way that it is readily findable in days, weeks, or years in the future. Good data management practices are essential.

Accessible: Working in the NIH environment it is important that the data is accessible by you, your PI and future lab members who may need to expand on your work. This dictates that all *important* data reside somewhere other than you personal laptop computer.

Interoperable: This usually means staying away from proprietary data formats (that may not be supported in the future) and additionally, the data need to interoperate with applications or workflows for analysis, storage, and processing.

Reusable: In this case reusable could better be defined as reproducible. Adhering to good data management principles means that you (or others) will be readily able to reproduce your data workflow resulting in exactly the same results.

FAIR Principles

In practice for the Scientific Community

F indable: This again relates to good annotation using appropriate terminology and depositing the data in public database. Meaningful and accurate metadata is a must.

A ccessible: In accordance with the various rules/policies the data needs to deposited in the appropriate public repository.

I nteroperable: Again referring to the metadata, annotation should follow standards that make further analysis possible.

R eusable: The primary data and associated metadata, and custom software and methodology should be deposited in a way that allows others to successfully reproduce your experiment.

All the FAIR principles can only be achieve with good data management practices.

NIH has issued the Data Management and Sharing (DMS) policy (effective January 25, 2023) to promote the sharing of scientific data.

This means that all NIH PIs have submitted a data management and sharing plan in January of this year (2023). If you aren't aware of this you should discuss it with our PI to be aware what they have committed themselves (and by inference you) to.

NIH also has a Genomic Data Sharing Policy which has been in place since 2014

Learn more about NIH Sharing Policies at https://sharing.nih.gov/

Data Management Strategy

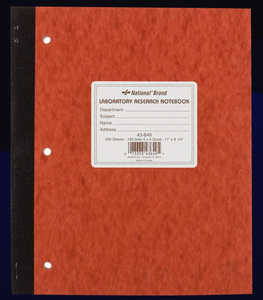
Hopefully by now it is clear that each one of us should have a data managrement strategy that will ensure that:

- All data in organized and stored in a way that it can be readily located by YOU your PI and future members of the Lab.
- Data is stored in a safe and secure location
- Metadata is capture in a timely manner
- Key aspects of software-use and reference material is capture early on

Sharing made easy

A good data management plan will ensure that complying with NIH data sharing policies is straight forward since all of the elements have already been captured and are readily locatable.

Record Keeping
Traditional vs Electronic



Traditionally the “Lab Notebook” was where everything was recorded and documented - each and every experiment was recorded. There are well established principles of what was and wasn’t necessary to go into the “Lab Notebook”. Generally well established procedures are in place for the storage (long or short) for lab reagents (cell-lines clones, etc) that might need to be share with the Scientific Community.

Record Keeping
Traditional **vs Electronic**

Unfortunately, in the current Electronic age there is no clear replacement for the “Lab Notebook”. There are numerous instances of “Electron Lab Notebooks”, but at the current time none of them have become the defect standard and NCI is only now moving in that direction, with there “beta progam. Additionally, modern research methods combine traditional “wet lab science” with many techniques that are purely computational. This has introduced many new challenges for the researcher who has to store, manage and ultimately share this purely electronic data and its associated metadata. (Metadata being defined as information about data).

Electronic Lab Notebook or Equivalent

If your Lab has a defined plan for managing electronic data you should follow it. If the Lab doesn't have a formal plan for file management find a system that works for you and stick with it!! The resource section has links to sites that suggest some strategies.

The most important parts of a plan are to capture all the relevant information and store them in a way that is secure and findable

Some Options

Formal Electronic Lab Notebook
One Note - Rigid Hierarchy of named directories and Files
Word Documents
PDF Documents
Any strategy will likely dictate that all data, associated metadata and documents should be stored in organized defined file/folder hierarchies

Elements of a good file management plan

- Early on document (and revise as needsed) your organization strategy so you (and others) know what goes where.
- Some level of Index
- Good structured file naming conventions (e.g. Illumina sequencing run name 230623_VH00687_218_AACV5NMM5)
- Organized file/folder hierarchy - e.g. by time YYYY-MM-DD or YYYYMMDD, by file type, by experiment or project.
- File Hierarchies should not got too long.
- Use open formats where possible - text files, css files, etc.
- For tabular data make sure all columns have meaningful headers
- Separate primary (raw) data from derived data and protect if accordingly

Plan - Demands discipline

Where and how to store all files

- Decide what can be stored “locally” and what must be put in “permanent storage” - working copies vs raw data or final product
- Separate raw data from derived data
- Devise a system to distinguish data exploration and side tracks from the final results
- Use structured naming conventions (many to chose from) - adopt one and ***stick to it***
- Plan how will you share the data etc. with others in the lab, and future members

Proprietary Format	Preferred Format
Excel (.xls, .xlsx)	Comma Separated Values (.csv or .tsv)
Word (.doc, .docx)	plain text (.txt), PDF (.pdf) formatted
PowerPoint (.ppt, .pptx)	PDF (.pdf)
Photoshop (.psd)	TIFF (.tif, tiff), JPEG 9.JPG,.jpeg), PNG (.png), PDF (.pdf)
Quicktime (.mov)	MPEG-4 (.mp4)

Computational Biologists

Computational experiments occur on a different time scale than traditional wet lab experiments and tend to generated more output.
While some tasks are computationally time intensive and take long periods to complete. Others can be run in a matter of minutes, where different approaches, programs or parameters are explore. This avalanche of data (some good some bad) greatly complicates the data management issues.

An added problem for those involved in the field of bioinformatic is the distributed nature of computing. Some compute intensive tasks may be done on the HPC system (Biowulf/Helix or FRCE) or a Cloud resource (Galaxy, DNAnexus, PartekFlow) while others may be done on a local desktop or laptop machine

Some software/workflows impose their own structure on the data, so the challenge is how to fit into your structure and or move the files

File Structures

- Organize your data hierarchically, and identify ways to divide your data into categories (or attributes):
  - Project
  - Time
  - Location
  - File type
- Within folders, files can be maintained chronologically, by classification or code, or alphabetically (depending on the types of files)
- Folder and subfolder names should reflect the content of the folder, not the names of researchers or staff
- Document your file directory structure and describe the kinds of records that should be maintained in those folders to ensure compliance
- Include basic information, such as project titles, dates, and some type of unique identifier (such as a grant number)
- If appropriate develop a strategy for keep track of files on remote systems
- Remember to keep PII data separately and secure

Data Storage

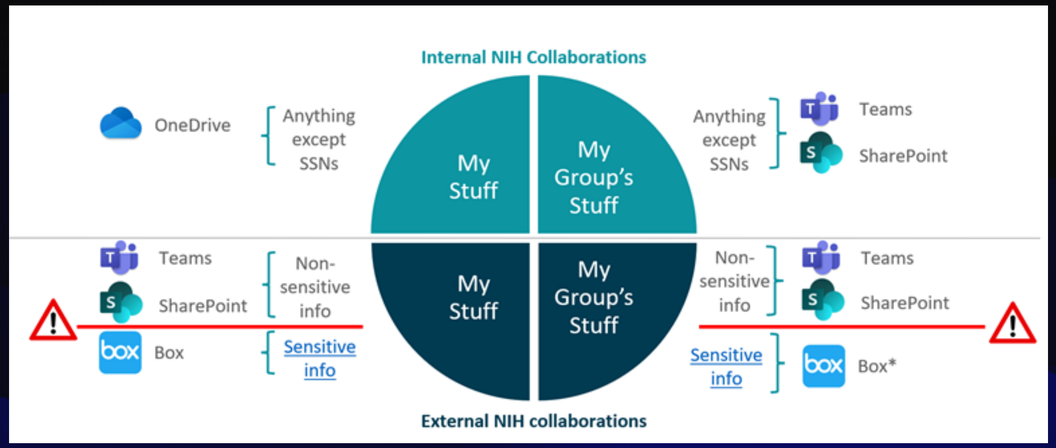
- Desktops/laptops - only home directory routinely backed up, not shareable - poor choice
- Networked drives - personal or shared Lab space - should be secure
- Lab server - is it backup up
- ~~External hard drives~~ -
- Optical storage - typically too slow and too small
- ~~Cloud storage~~ -
- ~~Flash drives~~ -
- HPC computer (Biowulf/Helix) - stable but NOT backed up - does have snapshots
- Many “good solutions” not appropriate for NIH security policies - (IT frowns on iCloud, Google Drive, Dropbox etc)
- Be aware of your “***permission***” choices and where possible limit others to “***read only***” and even yourself for critical data.

The 3-2-1 methodology - hard to achieve in NIH environment

A simple, commonly used storage system is the 3-2-1 methodology. This methodology suggests the following strategic recommendations: 3: Store three copies of your data, 2: using two types of storage media, 1: with one of them stored offsite.

The dangers that need to be mitigated include: Accidental erasure, disk failure, hacking and data being ***lost***

Data Storage



	Internal (NIH) Collaborators	External (non-NIH) Collaborators
Documents that contain sensitive information	• Teams • SharePoint Online	• Box*
Documents that do not contain sensitive information	• Teams • SharePoint Online	• Teams • SharePoint Online

https://service.cancer.gov/ncisp?id=nci_kb_article&sysparm_article=NCI-KB0014654

Data Storage

Features	OneDrive	Teams	SharePoint	Box	Network Storage / Shared Drive
Available without NIH network connection or VPN	yes	yes	yes	yes	no
Share sensitive information with external collaborators	no	no (but probably in the future)	no (but probably in the future)	yes	no
Share non-sensitive information with external collaborators	no	yes	yes	no	no
Individual File Size Limit	250GB	250GB	250GB	150GB	100GB+
Total File Storage	5TB	25TB per team	25TB per site	unlimited	25TB+
Sharing and Access Control	yes	yes	yes	yes	limited
Real-Time Co-Authoring	yes	yes	yes	yes	no
Syncing	yes	yes	yes	yes	no
Versioning	yes	yes	yes	yes	no
Restore	yes	yes	yes	yes	no
Desktop App	yes	yes	yes	yes	yes
Web & Mobile App	yes	yes	yes	yes	no
Conversations	no	yes	no	no	no
Chat	no	yes	no	no	no
Voice/Video	no	yes	no	no	no
Other Info	Additional Microsoft apps, including Planner, Lists, Meeting Notes, OneNote, Wiki, Recordings			Box comments, Box tasks, Box notes	

https://service.cancer.gov/ncisp?id=nci_kb_article&sysparm_article=NCI-KB0014654

Data Management Environment (DME)

The NCI Data Management Environment (DME) platform for the storage and management of scientific research datasets. It eliminates the need to maintain redundant copies of large heterogenous data and provides the ability to annotate, retrieve, and share datasets for further research, analysis, and collaboration. ***Most of the primary data generated by CCR cores is currently deposited into DME.***

If you have an NIH account, the NCI Data Vault team can give you access to DME. For access requests or any other questions, contact NCIDataVault@mail.nih.gov. The user guide found at **https://wiki.nci.nih.gov/display/DMEdoc** describes the capabilities of DME.

The system can be accessed via a Command Line Interface (CLI), a web-based graphical user interface (GUI), as well as programmatically via a set of API calls and utilities.

Reasons for Sharing Data

There are a number of policies which mandate sharing of different data.

Under the Data Management & Sharing (DMS) Policy, NIH requires researchers to prospectively plan for how scientific data will be preserved and shared through submission of a Data Management and Sharing Plan. The DMS policy applies to all NIH-supported research that results in the generation of scientific data, regardless of funding mechanism. Investigators are required to: Submit a Data Management and Sharing plan outlining how scientific data and any accompanying metadata will be managed and shared, taking into account any potential restrictions or limitations.
Comply with the Data Management and Sharing plan approved by the funding Institute or Center (IC). January 2023

NIH expects the broad and responsible sharing of human as well as non-human genomic data resulting from NIH-funded research because the timely sharing of research results can accelerate discoveries that improve our ability to diagnose, treat, and prevent disease.

It is good science… just like publishing is important to share your research sharing data is also critical in so much as it permits other to validate and/or build on your results

Your ability to comply with these policies will be directly effected by your data management strategies

Sharing

- What to Share
- When to Share
- Where to Share

The simple answer is everything that is needed to reproduce your finding, in a public repository, as soon as possible.

Controlled Access Data

Data is in public repositories but access is restricted to “authorized” investigators. Primary data from human subjects typically falls into this category (e.g. dbGap) - remember to anonymize patient data

Public Access Data

Data is deposited in open repositories which allow access with out restrictions. Most basic research data falls into this category (e.g. Genbank)

Journal Policies

- All data deposited at time of Submission
- All data available to reviewer (upon request)
- All Data available upon approval - i.e. before it hits the public

Types of data that must be shared

- Raw data
- Processed data
- Workflows
- Software (license type)

Genomic Repositories

Listing of Repositories

Mandatory deposition	Suitable repositories
Protein sequences	Uniprot
DNA and RNA sequences	Genbank DNA DataBank of Japan (DDBJ) EMBL Nucleotide Sequence Database (ENA)
DNA and RNA sequencing data	NCBI Trace Archive NCBI Sequence Read Archive (SRA)
Genetic polymorphisms	dbSNP dbVar European Variation Archive (EVA)
Linked genotype and phenotype data	dbGAP The European Genome-phenome Archive (EGA)
Macromolecular structure	Worldwide Protein Data Bank (wwPDB) Biological Magnetic Resonance Data Bank (BMRB) Electron Microscopy Data Bank (EMDB)
Microarray data (must be MIAME compliant)	Gene Expression Omnibus (GEO) ArrayExpress
Crystallographic data for small molecules	Cambridge Structural Database

General Repositories

Requirement	Dataverse	Dryad	figshare	GigaScience	Mendeley Data	OSF	Vivli	Zenodo
Data Size and Format								
Hosting of common file formats (e.g. csv, tsv, xls, xlsx, doc, pdf)	✓	✓	✓	✓	✓	✓	✓	✓
Hosting of proprietary file formats (e.g. raw image files)	✓	✓	✓	✗	✓	✓	✓	✓
Unlimited size per file	✗	✓	✗	✓	✗	✗	✓	✗
Unlimited total dataset size	✓	✓	✓	✓	✓	✓	✓	✓
Data Licensing								
CC0 waiver	recommended	required	recommended	required	available	available	available	available
Data Attribution and Citation Tools								
Assignment of dataset DOIs	✓	✓	✓	✓	✓	✓	✓	✓
User Access Controls								
Tiered access (e.g. administrator-level, collaborator-level, curator-level)	✓	✗	✓	✗	✓	✓	✓	✗
Journal-integrated, anonymous access (for peer review pre-publication)	✓	✓	✓	✓	✓	✓	✗	✗
Optional embargo to data release following publication	✗	✓	✓	✓	✓	✓	✓	✓
Data Access Tools								
Comprehensive data and metadata search tools	✓	✗	✗	✗	✓	✓	✓	✗
Data access via direct download	✓	✓	✓	✓	✓	✓	✓	✓
Data downloading via API	✓	✓	✓	✗	✓	✓	✗	✓
Built-in tools for reading proprietary file formats	✗	✗	✓	✗	✗	✗	✓	✗
Integrated data analysis tools	✓	✗	✗	✓	✗	✗	✓	✗
Cost								
Data deposition fees	none	tiered	none	none	none	none	membership	none
Data maintenance fees	none	none	none	none	none	none	membership	none

Software

- **Existing Software**
  - Proprietary
  - Open Source
    - Version
    - Parameters
    - Reference
    - Command line example
- **Home Grown**
  - Related libraries and versions
  - Documentation
  - Source or Binary
  - License
  - Test data set
  - Operating system
  - Language
  - Share code on https://github.com/
  - Use Version Control

Open Source Workflow
A simple reference to a workflow may substitute for a laundry list of individual components

- **Command line interface (CLI)**
  - Harder to use
  - Easier to document
  - Check out the “shell’s history options”
  - Use scripts rather then typed commands
  - On biowulf use explicit module versions (e.g. module load samtools**/1.17**
  - e.g. Star, samtools

- **Graphic User Interface (GUI)**
  - Easier to use
  - Harder to Document
  - e.g. Partek Flow, Geneius

- **Interactive computing platform. Integrated development environment (IDE)**
  - Varies by use
  - Easier to document (automatic history)
  - e.g. Jupyter Notebook, R Studio

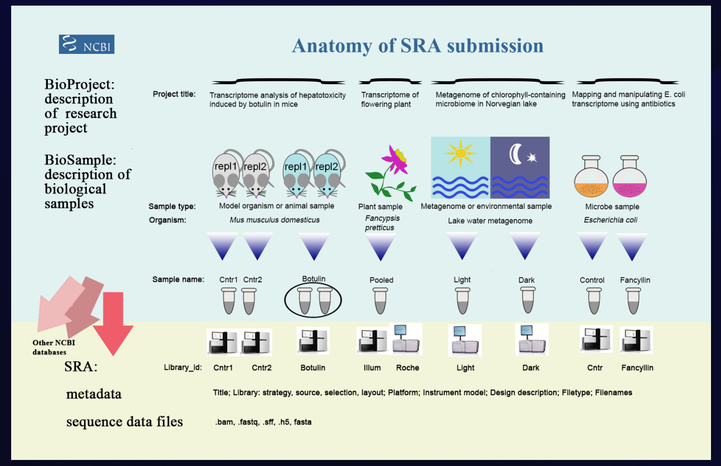
- **Web Interface**
  - Easier to use
  - Variable degree of automatic documentation
  - e.g. NIDAP, DNAnexus

Software Licences

CCO 1.0 CC0 1.0 Universal (CC0 1.0) Public Domain Dedication
https://creativecommons.org/publicdomain/zero/1.0/
CC BY 4.0 Attribution 4.0 International (CC BY 4.0)
https://creativecommons.org/licenses/by/4.0/
Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
CC BY NC 3.0 Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0)
https://creativecommons.org/licenses/by-nc/3.0/
NonCommercial — You may not use the material for commercial purposes.
MIT
Apache 2.0
https://www.apache.org/licenses/LICENSE-2.0
BSD 3-Clause
BSD 2-Clause
GPL v3 - GNU General Public License
https://www.gnu.org/licenses/gpl-3.0.en.html
GPL v2
LGPL
MPL-2.0,
CeCILL
CeCILL-B
CERN OHL

Metadata
(a set of data that describes and gives information about other data.)

- **Primary Data (SRA/dbGAP - Input Request) - FASTQ**
  - Experiment type (RNASeq, WGS, ChIPSeq)
  - Organism (Human, Mouse)
  - Cell Line
  - Instrumentation (Illumina, nanopore)
  - Machine Version (NextSeq 550, 2000, NovaSeq)
  - Library Prep - PCR amplified
  - Sample Type - whole RNA or mRNA etc
  - Other factors - rRNA depletion method



https://www.ncbi.nlm.nih.gov/sra/docs/submitmeta/

- **Software**
  - Version
  - Parameters
  - Reference
  - Command line example
- **Workflow**
  - Genome Version - sequence
  - Genome Version - Annotation
  - Pathway Versions
  - Naming conventions (Proper gene name, IDs etc)

RNASEQ Workflow - Software

~~BCLs~~ - **FASTQs** - SAMs - **BAMs** - Count Matrices (Gene/Transcript) - DEG - Function/Pathway Enrichment

- **FASTQ**
  - Experiment type - RNASeq
  - Organism - Human
  - Cell Line - MCF7
  - Instrumentation - Illumina,
  - Machine Version - NextSeq 2000
  - Library Prep - PCR amplified
  - Sample Type - mRNA
  - Other factors -rRNA depletion
  - RNAQuality - RIN 8.9
  - Sample Concentration 300 ngul
  - Sample Volume 10ul
  - Read Length - 2x100
  - Sample ID 220303_A00430_0573_BHWTW7DRXY
  - Run Date - 03/03/2022

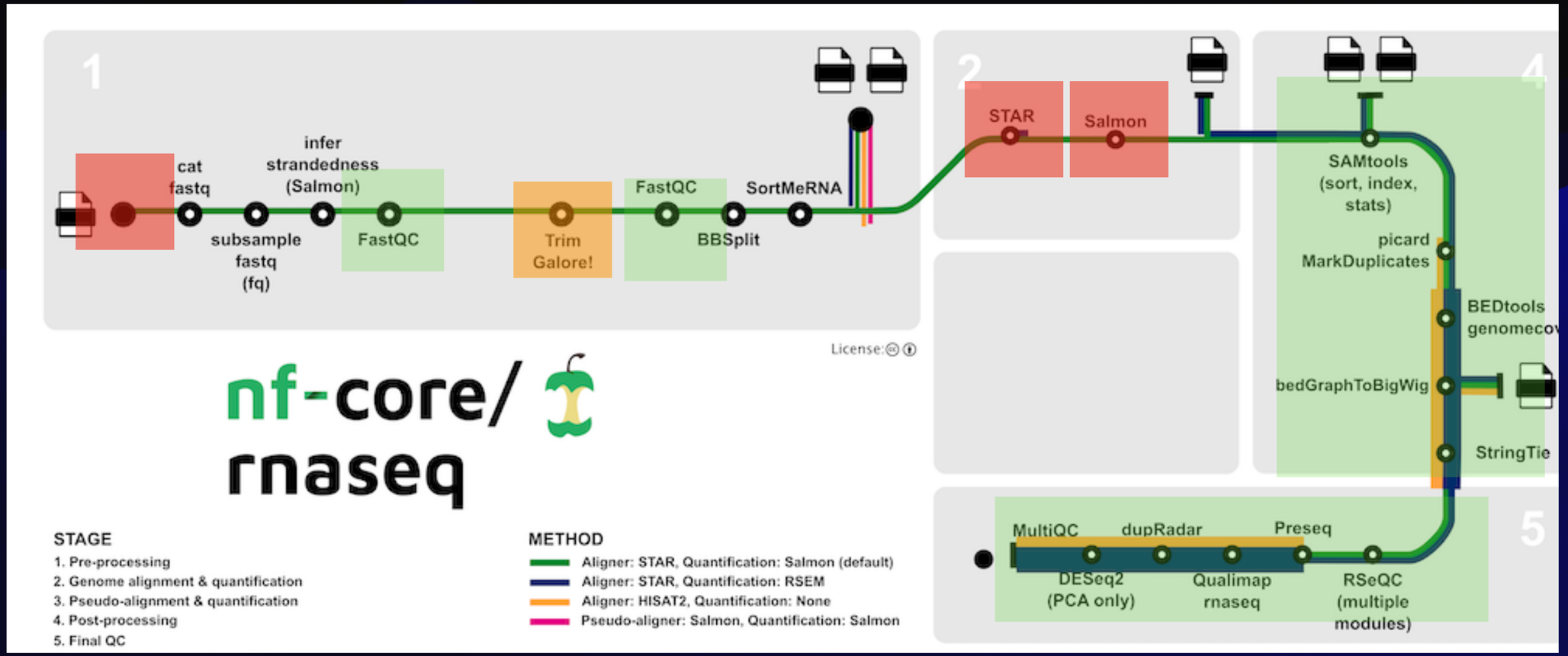
- **BAM**
  - Experiment type - RNASeq
  - Organism - Human
  - Genome Version - GRCh38 (NCBI)
  - Genome Annotation - Gencode 30
  - Run Date - 03/06/2022
  - Aligner - STAR (Parameters - should be in bam file)
    (Version -)
  - Extra - Picard Markduplicates
    (Version -)
    (Parameters -)

RNASEQ Workflow - File types

~~BCLs~~ - **FASTQs** - SAMs - **BAMs** - Count Matrices (Gene/Transcript) - DEG -

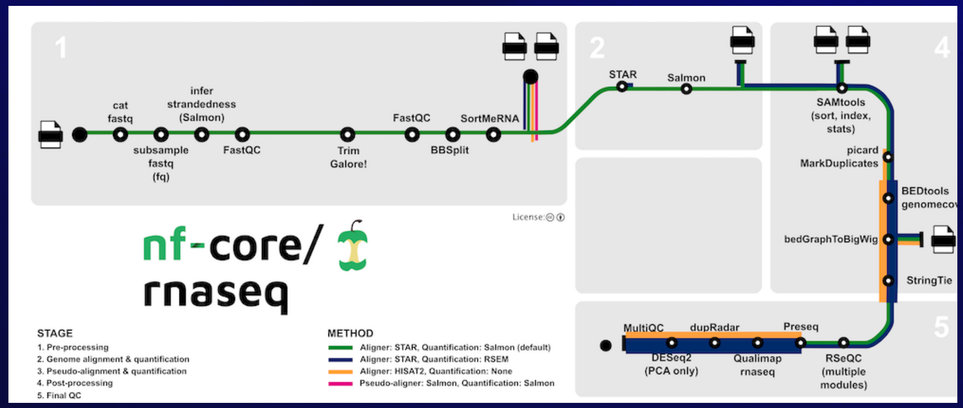
- **Count Matrics**
  Input source - from aligned bam file
  Software - salmon -version, parameters
  Gene/Transcripts
  Genome Annotations - Gencode 30

- **DEG (Differentially Expressed Gene List)**
  Input source - Raw count matrices
  Software - R Studios (2022.02.0 Build 443) - EdgeR version 3.36.0 within R version 4.3.1
  Normalization method - TMM
  Cutoff Parameters - > 2 fold, P <0.05


1
cat fastq
infer strandedness (Salmon)
subsample fastq (fq)
FastQC
Trim Galore!
FastQC
BBSplit
SortMeRNA
2
STAR
Salmon
4
SAMtools (sort, index, stats)
picard MarkDuplicates
BEDtools genomecov
bedGraphToBigWig
StringTie
5
MultiQC
dupRadar
Preseq
DESeq2 (PCA only)
Qualimap rnaseq
RSeQC (multiple modules)
License:
nf-core/rnaseq
STAGE
1. Pre-processing
2. Genome alignment & quantification
3. Pseudo-alignment & quantification
4. Post-processing
5. Final QC
METHOD
Aligner: STAR, Quantification: Salmon (default)
Aligner: STAR, Quantification: RSEM
Aligner: HISAT2, Quantification: None
Pseudo-aligner: Salmon, Quantification: Salmon

Process Name	Software	Version
BEDTOOLS GENOMECOV	bedtools	2.30.0
CUSTOM DUMPSOFTWAREVERSIONS	python	3.11.0
	yaml	6.0
CUSTOM GETCHROMSIZES	getchromsizes	1.16.1
DESEQ2 QC STAR SALMON	bioconductor-deseq2	1.28.0
	r-base	4.0.3
DUPRADAR	bioconductor-dupradar	1.28.0
	r-base	4.2.1
FASTQC	fastqc	0.11.9
GTF2BED	perl	5.26.2
GTF GENE FILTER	python	3.9.5
GUNZIP FASTA	gunzip	1.10
GUNZIP GTF	gunzip	1.10
MAKE TRANSCRIPTS FASTA	rsem	1.3.1
	star	2.7.10a
PICARD MARKDUPLICATES	picard	3.0.0
QUALIMAP RNASEQ	qualimap	2.2.2-dev
RSEQC BAMSTAT	rseqc	3.0.1
RSEQC INFEREXPERIMENT	rseqc	3.0.1
RSEQC INNERDISTANCE	rseqc	3.0.1
RSEQC JUNCTIONANNOTATION	rseqc	3.0.1
RSEQC JUNCTIONSATURATION	rseqc	3.0.1
RSEQC READDISTRIBUTION	rseqc	3.0.1
RSEQC READDUPLICATION	rseqc	3.0.1
SALMON QUANT	salmon	1.10.1
SALMON SE GENE	bioconductor-summarizedexperiment	1.24.0
	r-base	4.1.1
SALMON TX2GENE	python	3.9.5
SALMON TXIMPORT	bioconductor-tximeta	1.12.0
	r-base	4.1.1
SAMPLESHEET CHECK	python	3.9.5
SAMTOOLS FLAGSTAT	samtools	1.16.1
SAMTOOLS IDXSTATS	samtools	1.16.1
SAMTOOLS INDEX	samtools	1.16.1
SAMTOOLS SORT	samtools	1.16.1
SAMTOOLS STATS	samtools	1.16.1
STAR ALIGN	gawk	5.1.0
	samtools	1.16.1
	star	2.7.9a
STRINGTIE STRINGTIE	stringtie	2.2.1
TRIMGALORE	cutadapt	3.4
	trimgalore	0.6.7
UCSC BEDCLIP	ucsc	377
UCSC BEDGRAPHTOBIGWIG	ucsc	377
UNTAR SALMON INDEX	untar	1.30
UNTAR STAR INDEX	untar	1.30
Workflow	Nextflow	22.10.7
	nf-core/rnaseq	3.11.2

nf-core/rnaseq is a bioinformatics pipeline that can be used to analyse RNA sequencing data obtained from organisms with a reference genome and annotation. It takes a samplesheet and FASTQ files as input, performs quality control (QC), trimming and (pseudo-)alignment, and produces a gene expression matrix and extensive QC report. https://nf-co.re/rnaseq



Resources

- NIH Data Sharing Site - https://sharing.nih.gov/
- Harvard Biomedical Data Management - https://datamanagement.hms.harvard.edu/
- Harvard File and Directory Structure Tutorial - https://datamanagement.hms.harvard.edu/plan-design/directory-structure
- Open Science Framework - https://osf.io/
- Keeping a Lab Notebook - https://www.training.nih.gov/assets/Lab_Notebook_508_(new).pdf
- Electronic lab notebooks CBIIT - https://service.cancer.gov/ncisp?id=nci_kb_article&sysparm_article=NCI-KB0014880
- CBIIT Storage - https://service.cancer.gov/ncisp?id=nci_kb_article&sysparm_article=NCI-KB0014654
- DME Users Guide - https://wiki.nci.nih.gov/display/DMEdoc
- GitHub Software Repository - https://github.com/
- DRYAD Best Practices - https://datadryad.org/stash/best_practices#organize
- GLOBUS Data Transfer - https://hpc.nih.gov/docs/globus/
- Protocols Repository - http://protocols.io
- Plasmid Repository https://www.addgene.org

- **Journal Instructions:**
- Nature Communications Guide to authors - https://www.nature.com/ncomms/submit/guide-to-authors
- Brief guide for submission to Nature Communications - https://www.nature.com/documents/ncomms-submission-guide.pdf

Upcoming Talks

Jun
28
2023

Realizing FAIR principles and Reproducible Computational Workflows with the Arvados Platform

When: **Wed, Jun 28, 2023 - 11:00 am - 12:00 pm**

Delivery: **Online**

Presented By: **Brett Smith Senior Software Engineer Curii**

Jul
27
2023

Data Management and Sharing: Part 1

When: **Thu, Jul 27, 2023 - 1:00 pm - 2:00 pm**

Delivery: **Online**

Presented By: **Raisa Ionin (NIH Library)**

Jul
28
2023

Data Management and Sharing: Part 2

When: **Fri, Jul 28, 2023 - 1:00 pm - 2:00 pm**

Delivery: **Online**

Presented By: **Raisa Ionin (NIH Library)**

Jul
25
2023

Managing Bioinformatics Projects with Jupyter Notebook

Part Of: Introduction to Bioinformatics Summer Series Course

When: **Tue, Jul 25, 2023 - 1:00 pm - 2:00 pm**

Delivery: **Online**

Presented By: **Amy Stonelake (BTEP)**