

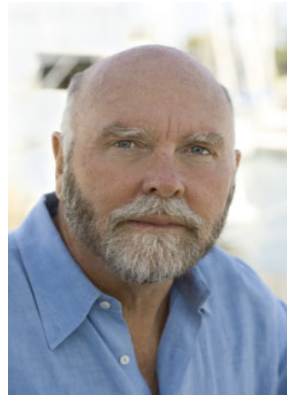


Introduction to Galaxy

Sequencing of the human genome, 2001

- Many years of hard work
- BAC based approaches towards WGS
- Amplification in bacterial culture
- More than 20,000 BAC clones
- Each containing about 100kb fragment
- Together provided a tiling path through each human chromosome
- Isolation, select pieces about 2-3 kb
- Subcloned into plasmid vectors, amplification, isolation
- recreate contigs
- Refinement, gap closure, sequence quality improvement
- (less 1 error/ 40,000 bases)
- 2.7 billion dollars

Sequencing of individual genomes, 2007-



\$100 million



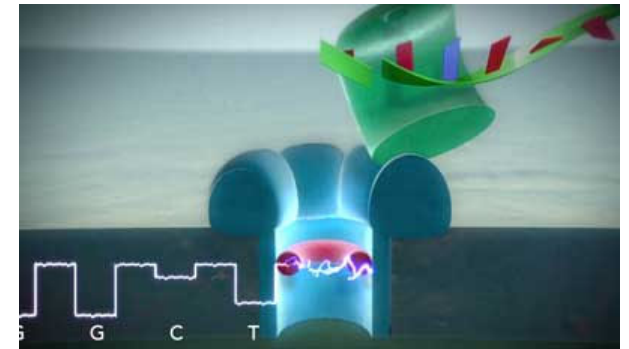
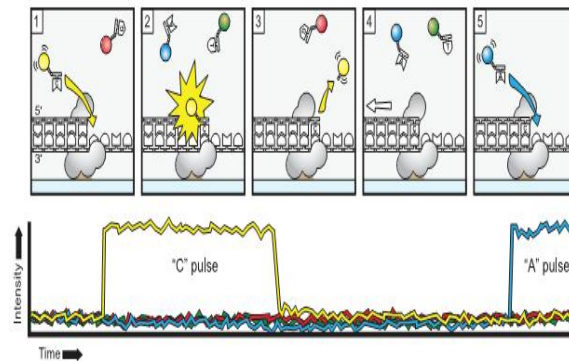
\$1 million



\$1000

2nd and 3rd generation sequencing

- Illumina/Solexa Genome Analyzer: 2006
- Pacific Biosciences SMRT: launching 2010
- Oxford Nanopore MinION 2014

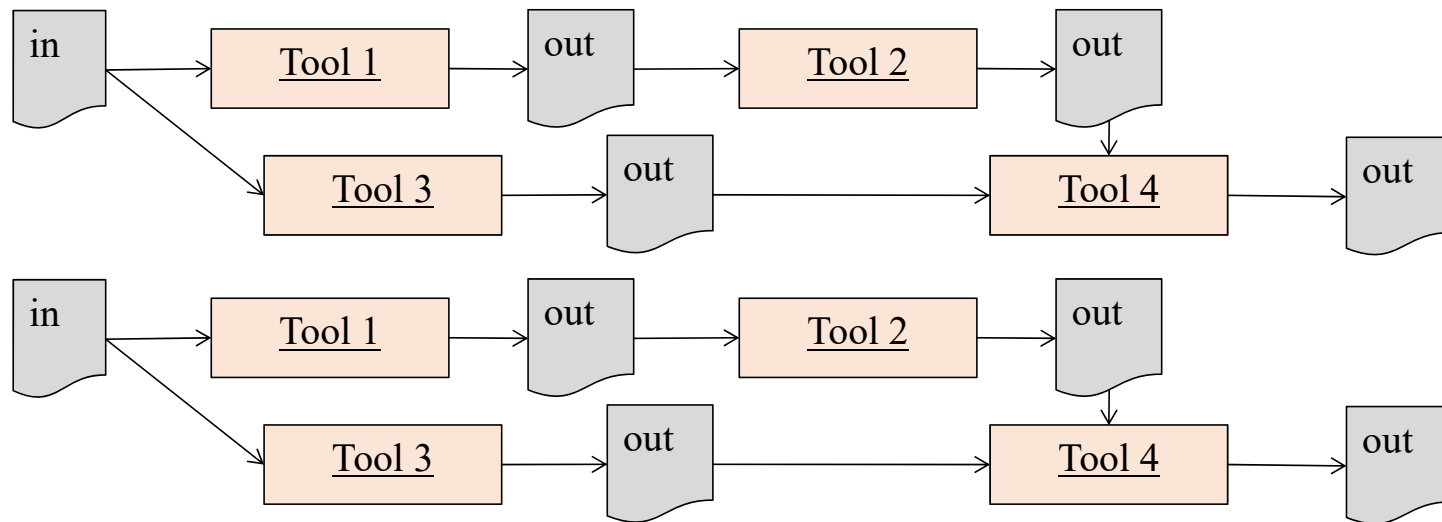


What is Galaxy?

- Galaxy is a free web-based scientific analysis platform that can be used by scientists across the world
- Galaxy's tools assist in the analysis of large biological datasets that are common in fields such as genomics or bioinformatics



Motivation of Galaxy



Sequencers throughput require parallel processing of multiple samples
⇒ how do you efficiently monitor all these workflow executions ?

What makes Galaxy special?



Free



Available world-wide



Massive toolbox for data analysis

More than 5500 tools available



Free training materials

Useful for learning the user interface and several tools



Helpful community and forums



Reproducibility

“Workflows” can be saved to show step-by-step uses of tools on data

How to Start

- Go to usegalaxy.org and click the "Login or Register" tab at the top
- Click on the "Register here" link below the login to begin creating an account
 - Once done filling in the required information, click "Create"
 - You will be required to verify your account's email address before being able to use any of the tools or data available
- After creating and verifying your account, you are ready to begin your introduction into Galaxy



Welcome to Galaxy

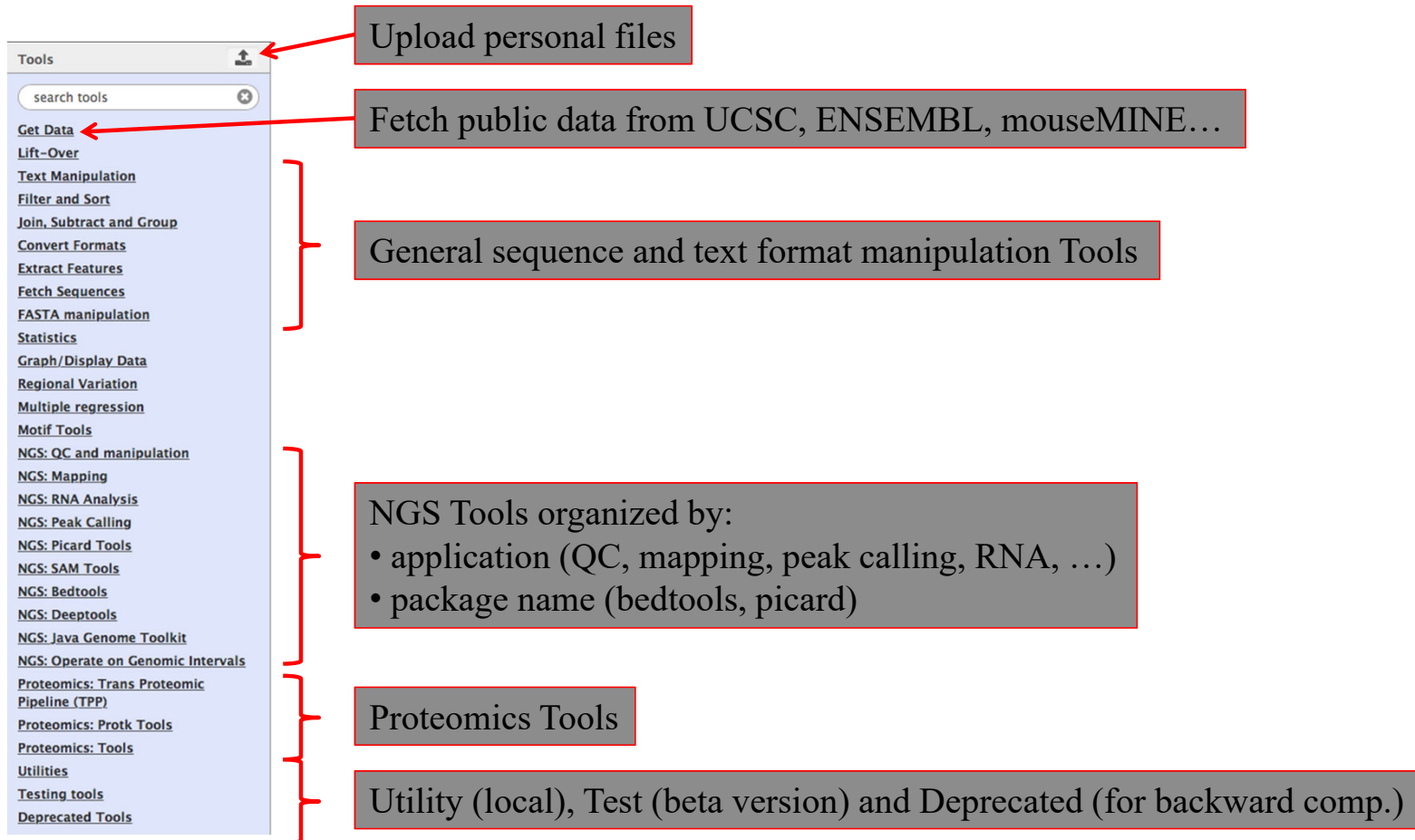
The screenshot shows the Galaxy web interface with several annotations:

- Search tools:** A red circle highlights the "search tools" input field in the top left.
- Tools are organized by Categories:** A red arrow points from this text to the "Contents" section on the left sidebar.
- Click a Category to see all tools:** A red arrow points from this text to the "Finding your data" category in the left sidebar.
- Personal workflows == pipelines:** A red arrow points from this text to the "Workflow" menu item in the top navigation bar.
- Main Panel : Launch Jobs / View Results:** A red arrow points from this text to the main content area of the interface.
- Tools:** A red arrow points from this text to the "Tools" section in the left sidebar.
- History:** A red arrow points from this text to the "History" section on the right sidebar.

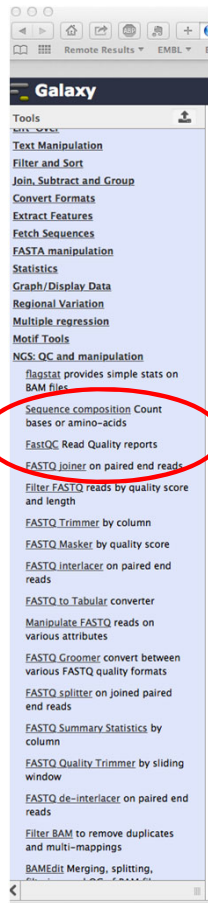
The interface includes a top navigation bar with links like "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". The left sidebar contains a "Tools" section with a search bar and a list of categories. The main panel displays a welcome message and a "Data Libraries" table.

Data library name	Data library description
Furlong Lab	Furlong Lab Data
Heisler's Lab	Data from the Heisler's lab
SRA	Short Read Archive
Steinmetz Lab	Steinmetz Lab Data

Noticeable Tool Categories



Running a Tool is easy



Click a tool to bring it up in the middle panel eg FastQC

Running a Tool is easy

The screenshot shows the Galaxy web interface with the FastQC tool selected. A red circle highlights the 'Short read data from your current history' input field, which contains a file named '14: allreads_marked-dups_H7H2FBGXX_8_15s009121-1-2_Cerese_lane115s009121_1_sequence_bowtie2_mm9'. A red arrow points from this input field to a list of steps: (1) Select input files, (2) Position parameters, and (3) Click Execute. Another red arrow points from the 'Execute' button to a status bar on the right side of the interface, which shows the job is currently running. A large grey box with a red border contains text explaining the job submission process and the meaning of the status colors: Green for 'Successfully completed', Yellow for 'Running', Grey for 'Waiting', and Red for 'Failed job'.

Run the tool on many files is easy too !

- (1) Select input files
- (2) Position parameters
- (3) Click Execute

Job is submitted to compute cluster
=> a new dataset block is added in the active history

Green : Successfully completed
Yellow : Running
Grey : Waiting
Red : Failed job

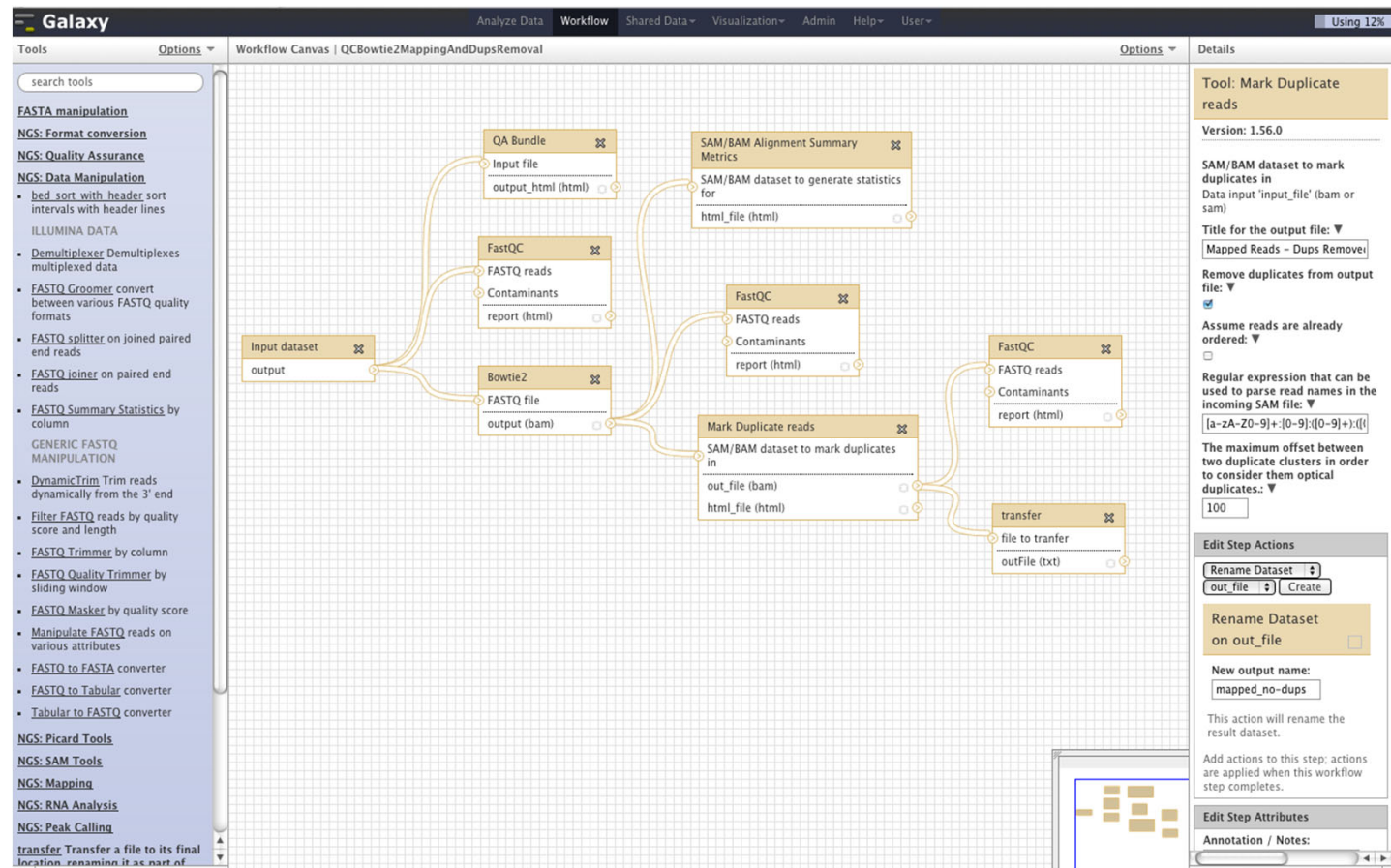
Tool summary



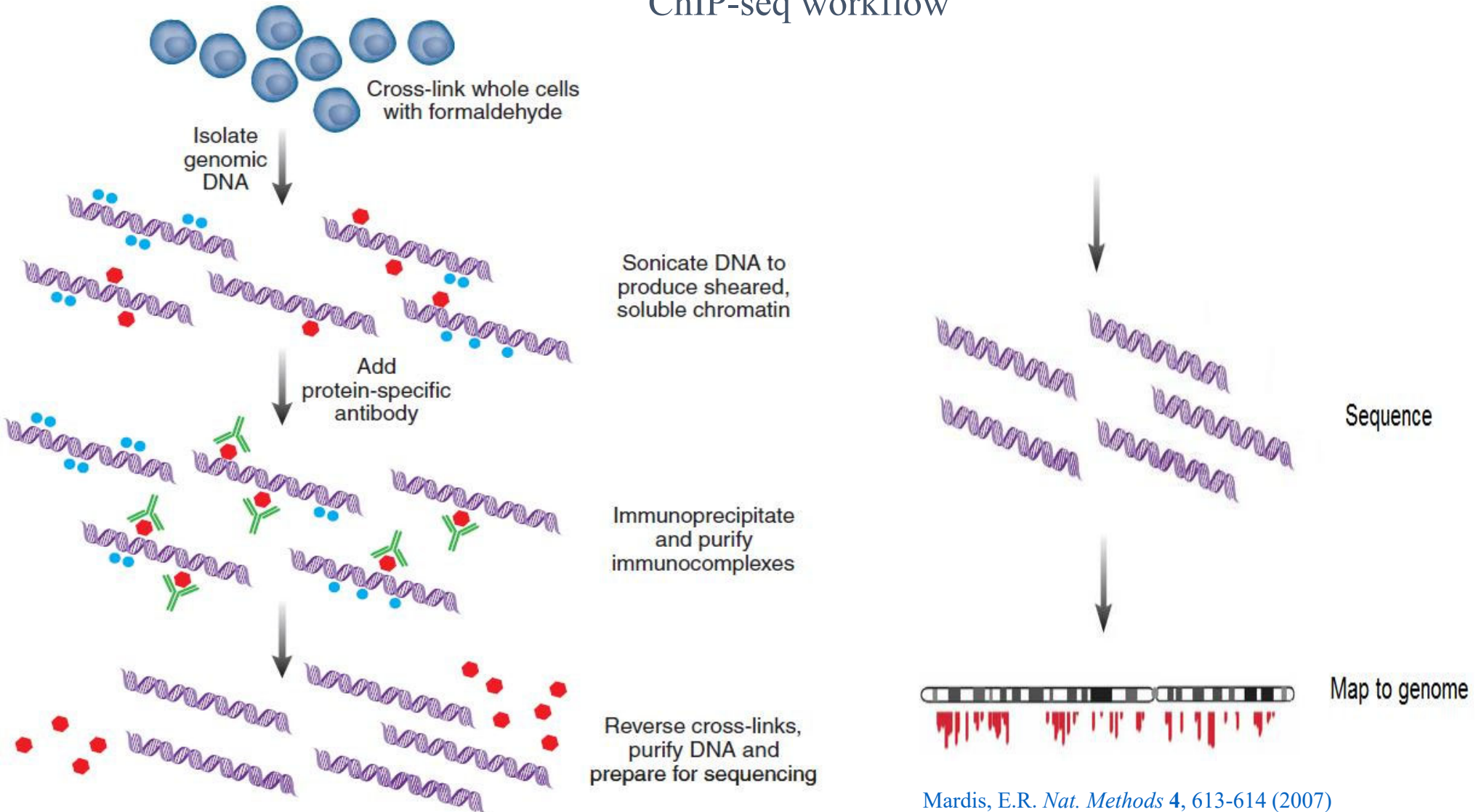
More than >1000 tools available !

- ✓ All results can be downloaded or directly transferred to your project folder
- ✓ Missing Tools can be easily integrated
- ✓ Easy way to add a GUI to your own script
- ✓ Parameters for cluster submission can be adjusted for each tool (and even be dynamically computed)

Tools can be assembled into workflows



ChIP-seq workflow



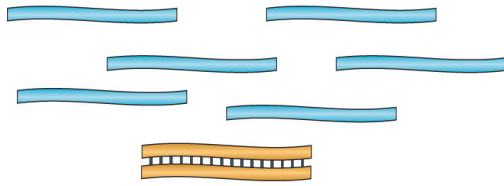
Major steps in ChIP-seq data analysis

- Trimmomatic or Cutadapt to remove adaptor sequences and filter low quality reads
- Fastqc to evaluate the quality
- Bowtie 2 to map reads to references
- MACS2 to define peaks

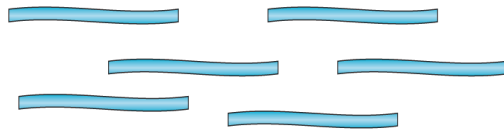
RNA-seq workflow

a Data generation

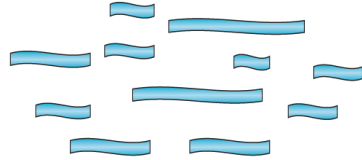
① mRNA or total RNA



② Remove contaminant DNA

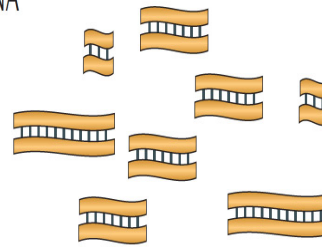


③ Fragment RNA

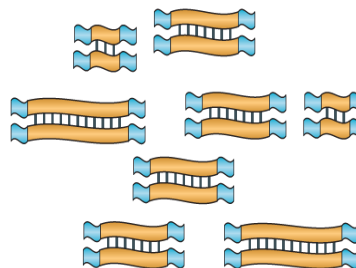


Remove rRNA?
Select mRNA?

④ Reverse transcribe
into cDNA

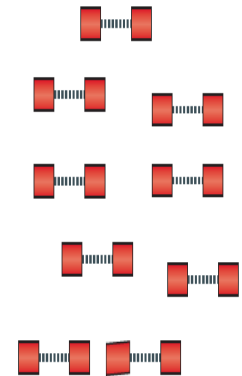


⑤ Ligate sequence adaptors



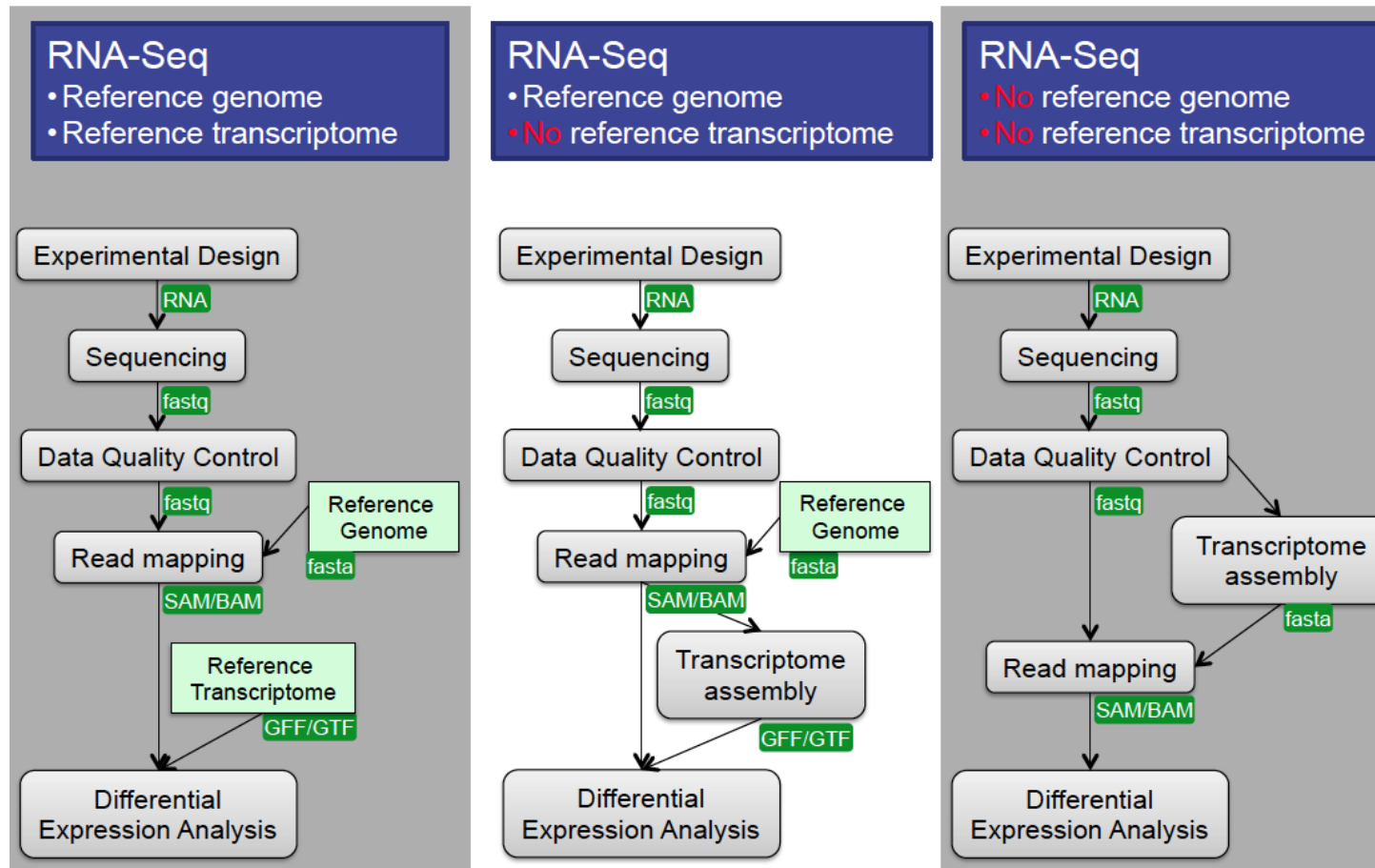
Strand-specific RNA-seq

⑦ Sequence cDNA ends



Differential Gene Expression

Options for DGE analysis



How does one pick the right tools?

1. Quality Check - **FASTQC**
2. Trimming – **Trimmomatic, cutadator**
3. Splice-aware alignment – **STAR, hisat2**
Bacterial alignment - **BWA** or **Novoalign**
4. Counting reads per gene – **featureCounts, stringtie**
 - Counting reads per isoform - **Salmon**
5. DGE Analysis – **edgeR, DESeq limma**

De novo transcriptome assembly - **Trinity**

References cont.

- <https://galaxyproject.github.io/>
- <https://galaxyproject.org/usegalaxy/>
- usegalaxy.org