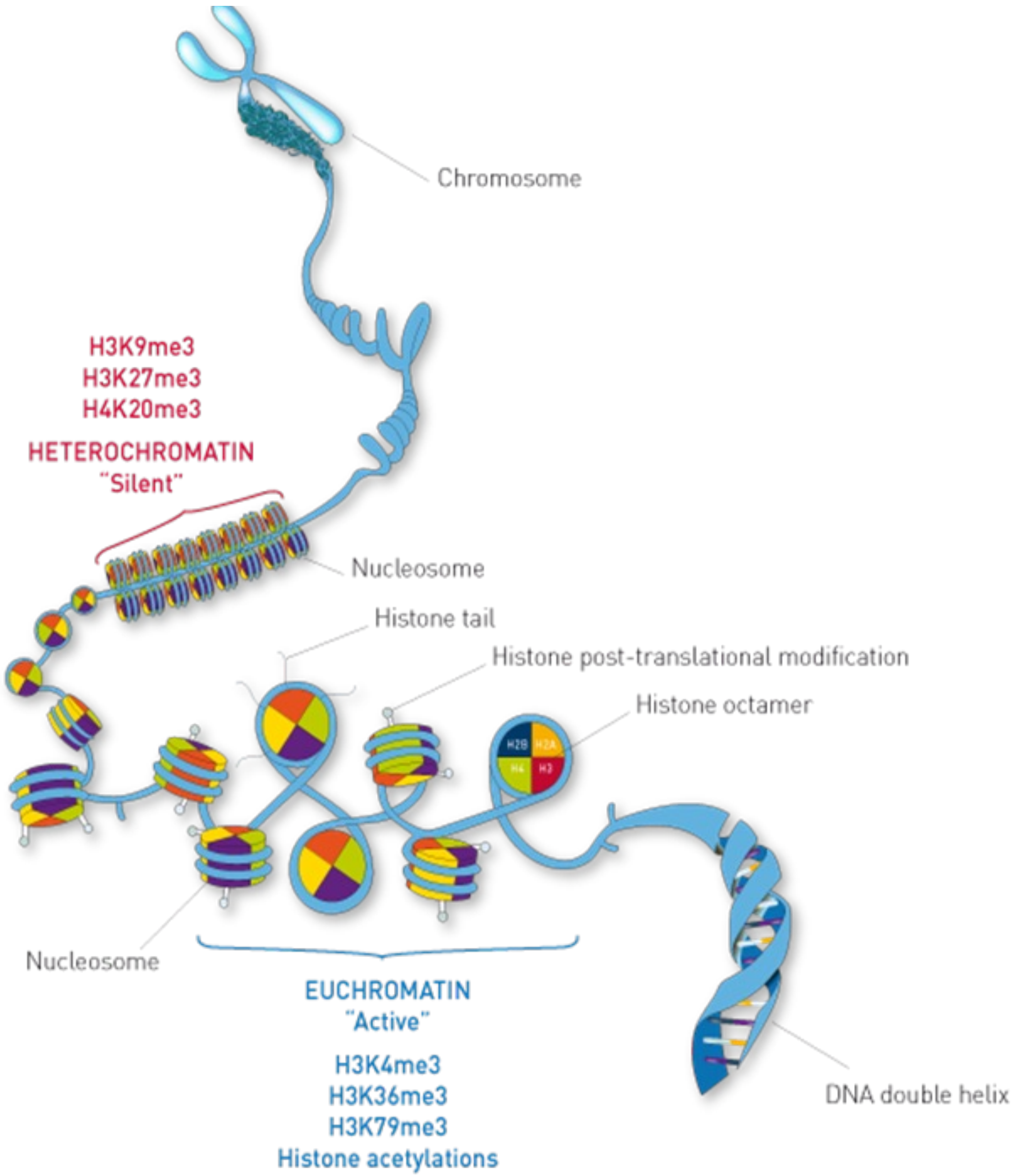


DIFFERENCIAL PEACK EXPRESIONS USING CHIP SEQUENCING (CHIP-SEQ) ANALYSIS

RIMA ZINJUWADIA

MSC SEM-4

BIOINFORMATICS

Introduction of ChIP-Seq

- ChIP-seq is a wonderful technique that allows us to interrogate the physical binding interactions between protein and DNA using next-generation sequencing.

What is chromatin immunoprecipitation?

- Chromatin immunoprecipitation (ChIP) allows us to determine protein-binding sites on DNA. Chromatin is the complex of DNA packaged with histone proteins into nucleosomes. ChIP makes use of reversible cross-links made between DNA and associated proteins by formaldehyde fixation of cells or tissue.

- The fixed chromatin is physically sheared and DNA fragments associated with a particular protein are selectively immunoprecipitated and analysed. Analysis can be on a locus-by-locus basis using PCR, but more commonly ChIP is interrogated with microarrays (ChIP-chip) or next-generation sequencing (ChIP-seq)

How does ChIP-seq work?

- Chromatin immunoprecipitation sequencing, or ChIP-seq, combines ChIP with next-generation sequencing .

- ChIP-seq protocols have been adapted from ChIP-chip methods: proteins are cross-linked to their bound DNA by formaldehyde treatment, cells are homogenized, and chromatin is sheared and immunoprecipitated with antibody-bound magnetic beads.

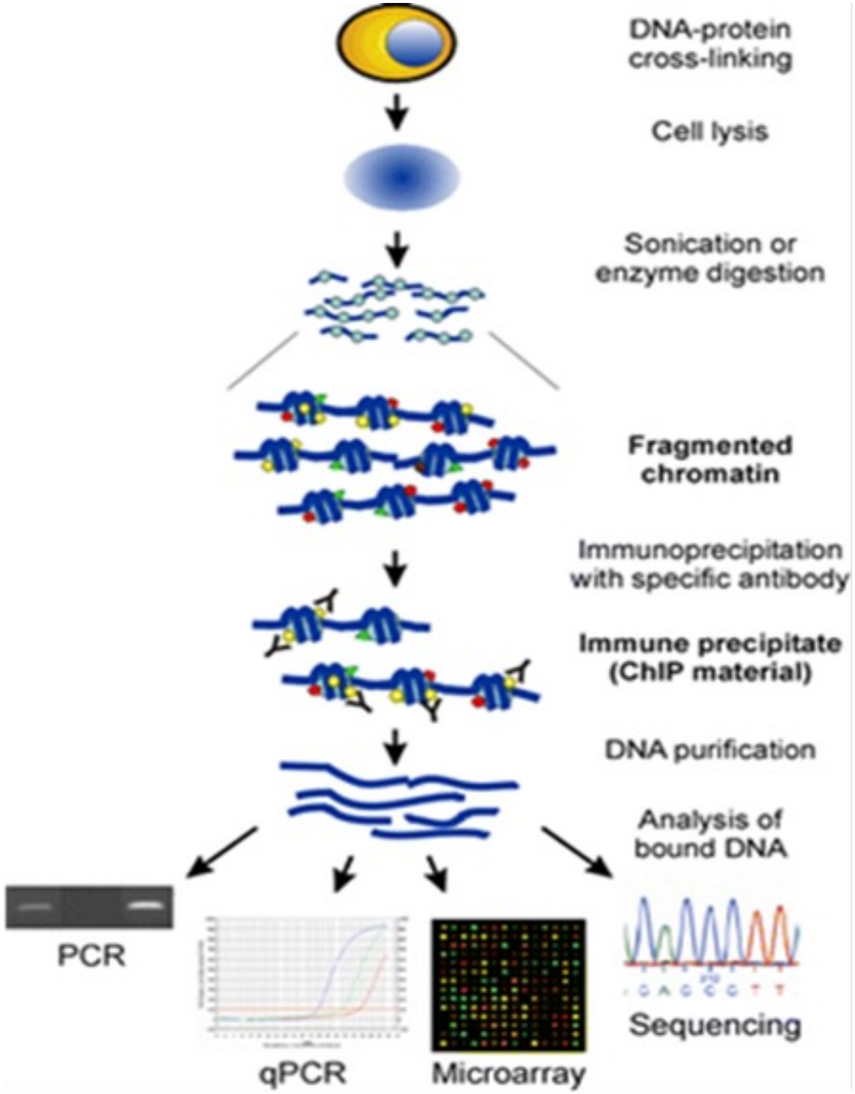
- The immunoprecipitated DNA is then used as the input for a next-generation sequencing library prep protocol, where it is sequenced and analysed for DNA binding sites.

How does ChIP-seq work?

- Chromatin immunoprecipitation sequencing, or ChIP-seq, combines ChIP with next-generation sequencing .

- ChIP-seq protocols have been adapted from ChIP-chip methods: proteins are cross-linked to their bound DNA by formaldehyde treatment, cells are homogenized, and chromatin is sheared and immunoprecipitated with antibody-bound magnetic beads.

- The immunoprecipitated DNA is then used as the input for a next-generation sequencing library prep protocol, where it is sequenced and analysed for DNA binding sites.


DNA-protein cross-linking
Cell lysis
Sonication or enzyme digestion
Fragmented chromatin
Immunoprecipitation with specific antibody
Immune precipitate (ChIP material)
DNA purification
Analysis of bound DNA
PCR
qPCR
Microarray
Sequencing

Advantages of ChIP-Seq

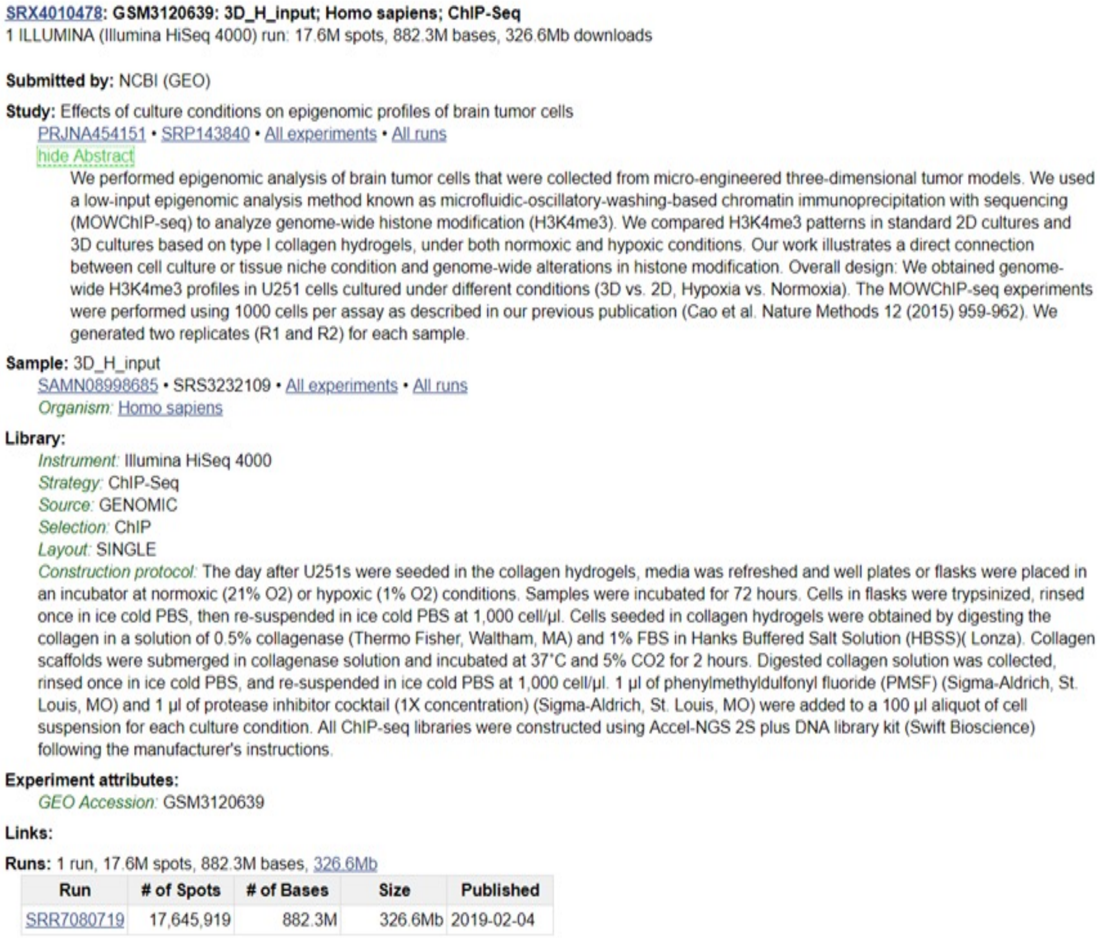
- Captures DNA targets for transcription factors or histone modifications across the entire genome of any organism
- Defines transcription factor binding sites
- Reveals gene regulatory networks in combination with RNA sequencing and methylation analysis
- Offers compatibility with various input DNA samples

Computational analysis of ChIP-Seq

- As with many high-throughput sequencing approaches, ChIP-seq generates extremely large data sets, for which appropriate computational analysis methods are required. To predict DNA-binding sites from ChIP-seq read count data, peak calling methods have been developed. The most popular method is MACS which empirically models the shift size of ChIP-Seq tags, and uses it to improve the spatial resolution of predicted binding sites.

Selection of ChIP-Seq Data (NCBI-SRA)

Treated SRR708018



SRX4010478: GSM3120639: 3D_H_input; Homo sapiens; ChIP-Seq
1 ILLUMINA (Illumina HiSeq 4000) run: 17.6M spots, 882.3M bases, 326.6Mb downloads

Submitted by: NCBI (GEO)

Study: Effects of culture conditions on epigenomic profiles of brain tumor cells
PRJNA454151 • SRP143840 • All experiments • All runs
hide Abstract

We performed epigenomic analysis of brain tumor cells that were collected from micro-engineered three-dimensional tumor models. We used a low-input epigenomic analysis method known as microfluidic-oscillatory-washing-based chromatin immunoprecipitation with sequencing (MOWChIP-seq) to analyze genome-wide histone modification (H3K4me3). We compared H3K4me3 patterns in standard 2D cultures and 3D cultures based on type I collagen hydrogels, under both normoxic and hypoxic conditions. Our work illustrates a direct connection between cell culture or tissue niche condition and genome-wide alterations in histone modification. Overall design: We obtained genome-wide H3K4me3 profiles in U251 cells cultured under different conditions (3D vs. 2D, Hypoxia vs. Normoxia). The MOWChIP-seq experiments were performed using 1000 cells per assay as described in our previous publication (Cao et al. Nature Methods 12 (2015) 959-962). We generated two replicates (R1 and R2) for each sample.

Sample: 3D_H_input
SAMN08998685 • SRS3232109 • All experiments • All runs
Organism: Homo sapiens

Library:
Instrument: Illumina HiSeq 4000
Strategy: ChIP-Seq
Source: GENOMIC
Selection: ChIP
Layout: SINGLE
Construction protocol: The day after U251s were seeded in the collagen hydrogels, media was refreshed and well plates or flasks were placed in an incubator at normoxic (21% O2) or hypoxic (1% O2) conditions. Samples were incubated for 72 hours. Cells in flasks were trypsinized, rinsed once in ice cold PBS, then re-suspended in ice cold PBS at 1,000 cell/μl. Cells seeded in collagen hydrogels were obtained by digesting the collagen in a solution of 0.5% collagenase (Thermo Fisher, Waltham, MA) and 1% FBS in Hanks Buffered Salt Solution (HBSS)(Lonza). Collagen scaffolds were submerged in collagenase solution and incubated at 37°C and 5% CO2 for 2 hours. Digested collagen solution was collected, rinsed once in ice cold PBS, and re-suspended in ice cold PBS at 1,000 cell/μl. 1 μl of phenylmethyldulfonyl fluoride (PMSF) (Sigma-Aldrich, St. Louis, MO) and 1 μl of protease inhibitor cocktail (1X concentration) (Sigma-Aldrich, St. Louis, MO) were added to a 100 μl aliquot of cell suspension for each culture condition. All ChIP-seq libraries were constructed using Accel-NGS 2S plus DNA library kit (Swift Bioscience) following the manufacturer's instructions.

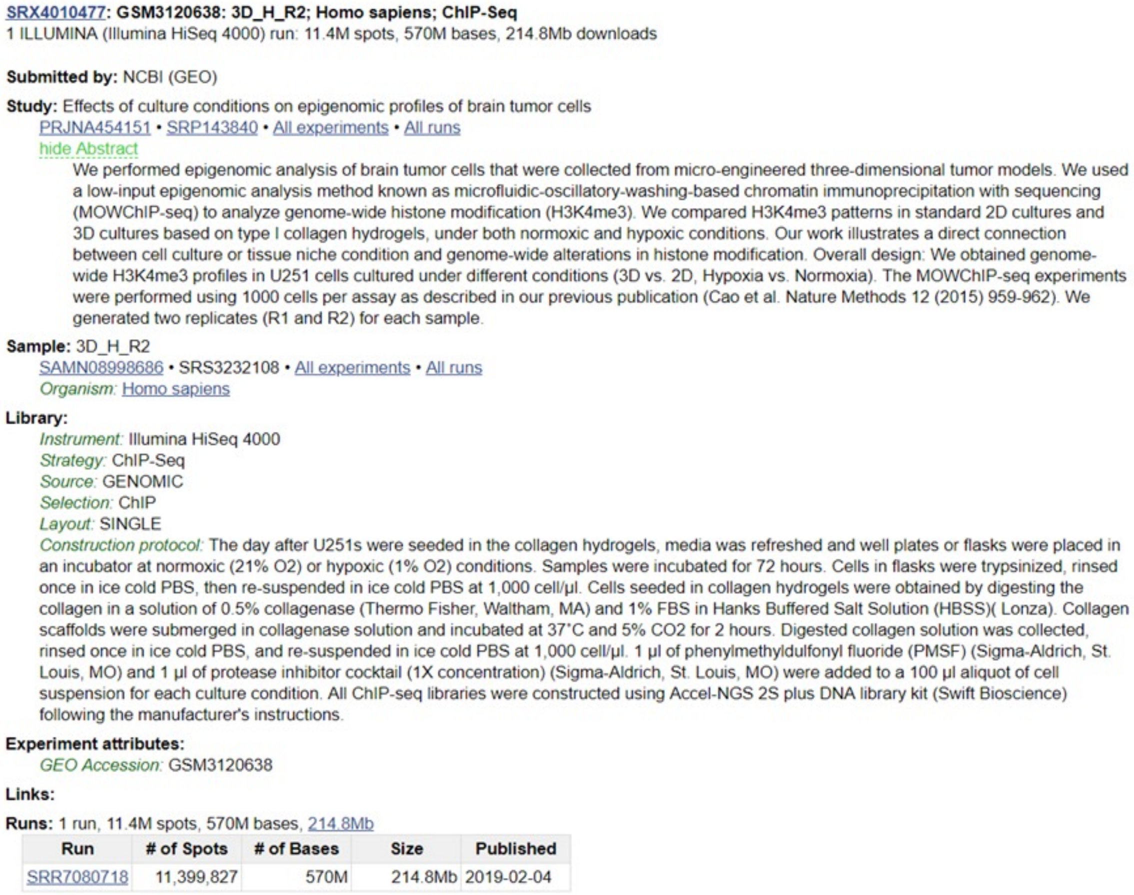
Experiment attributes:
GEO Accession: GSM3120639

Links:

Runs: 1 run, 17.6M spots, 882.3M bases, 326.6Mb

Run	# of Spots	# of Bases	Size	Published
SRR7080719	17,645,919	882.3M	326.6Mb	2019-02-04

Control SRR7080719



SRX4010477: GSM3120638: 3D_H_R2; Homo sapiens; ChIP-Seq
1 ILLUMINA (Illumina HiSeq 4000) run: 11.4M spots, 570M bases, 214.8Mb downloads

Submitted by: NCBI (GEO)

Study: Effects of culture conditions on epigenomic profiles of brain tumor cells
PRJNA454151 • SRP143840 • All experiments • All runs
hide Abstract

We performed epigenomic analysis of brain tumor cells that were collected from micro-engineered three-dimensional tumor models. We used a low-input epigenomic analysis method known as microfluidic-oscillatory-washing-based chromatin immunoprecipitation with sequencing (MOWChIP-seq) to analyze genome-wide histone modification (H3K4me3). We compared H3K4me3 patterns in standard 2D cultures and 3D cultures based on type I collagen hydrogels, under both normoxic and hypoxic conditions. Our work illustrates a direct connection between cell culture or tissue niche condition and genome-wide alterations in histone modification. Overall design: We obtained genome-wide H3K4me3 profiles in U251 cells cultured under different conditions (3D vs. 2D, Hypoxia vs. Normoxia). The MOWChIP-seq experiments were performed using 1000 cells per assay as described in our previous publication (Cao et al. Nature Methods 12 (2015) 959-962). We generated two replicates (R1 and R2) for each sample.

Sample: 3D_H_R2
SAMN08998686 • SRS3232108 • All experiments • All runs
Organism: Homo sapiens

Library:
Instrument: Illumina HiSeq 4000
Strategy: ChIP-Seq
Source: GENOMIC
Selection: ChIP
Layout: SINGLE
Construction protocol: The day after U251s were seeded in the collagen hydrogels, media was refreshed and well plates or flasks were placed in an incubator at normoxic (21% O2) or hypoxic (1% O2) conditions. Samples were incubated for 72 hours. Cells in flasks were trypsinized, rinsed once in ice cold PBS, then re-suspended in ice cold PBS at 1,000 cell/μl. Cells seeded in collagen hydrogels were obtained by digesting the collagen in a solution of 0.5% collagenase (Thermo Fisher, Waltham, MA) and 1% FBS in Hanks Buffered Salt Solution (HBSS)(Lonza). Collagen scaffolds were submerged in collagenase solution and incubated at 37°C and 5% CO2 for 2 hours. Digested collagen solution was collected, rinsed once in ice cold PBS, and re-suspended in ice cold PBS at 1,000 cell/μl. 1 μl of phenylmethyldulfonyl fluoride (PMSF) (Sigma-Aldrich, St. Louis, MO) and 1 μl of protease inhibitor cocktail (1X concentration) (Sigma-Aldrich, St. Louis, MO) were added to a 100 μl aliquot of cell suspension for each culture condition. All ChIP-seq libraries were constructed using Accel-NGS 2S plus DNA library kit (Swift Bioscience) following the manufacturer's instructions.

Experiment attributes:
GEO Accession: GSM3120638

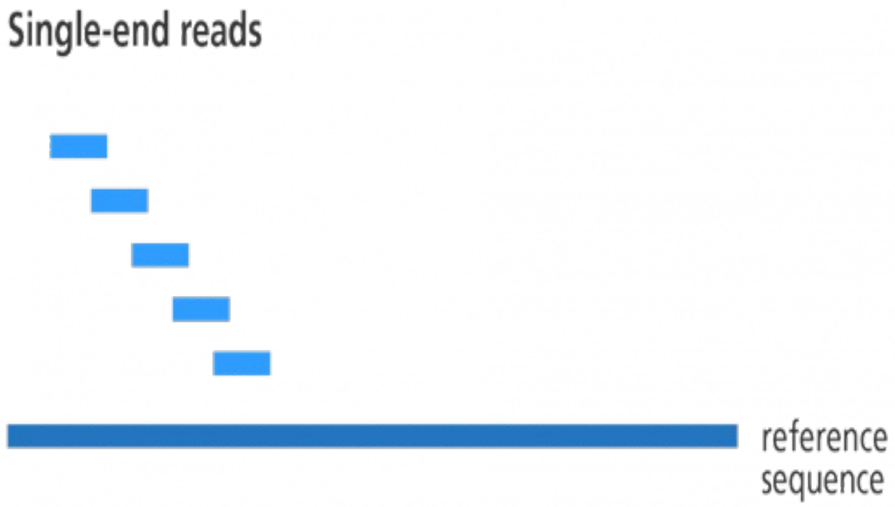
Links:

Runs: 1 run, 11.4M spots, 570M bases, 214.8Mb

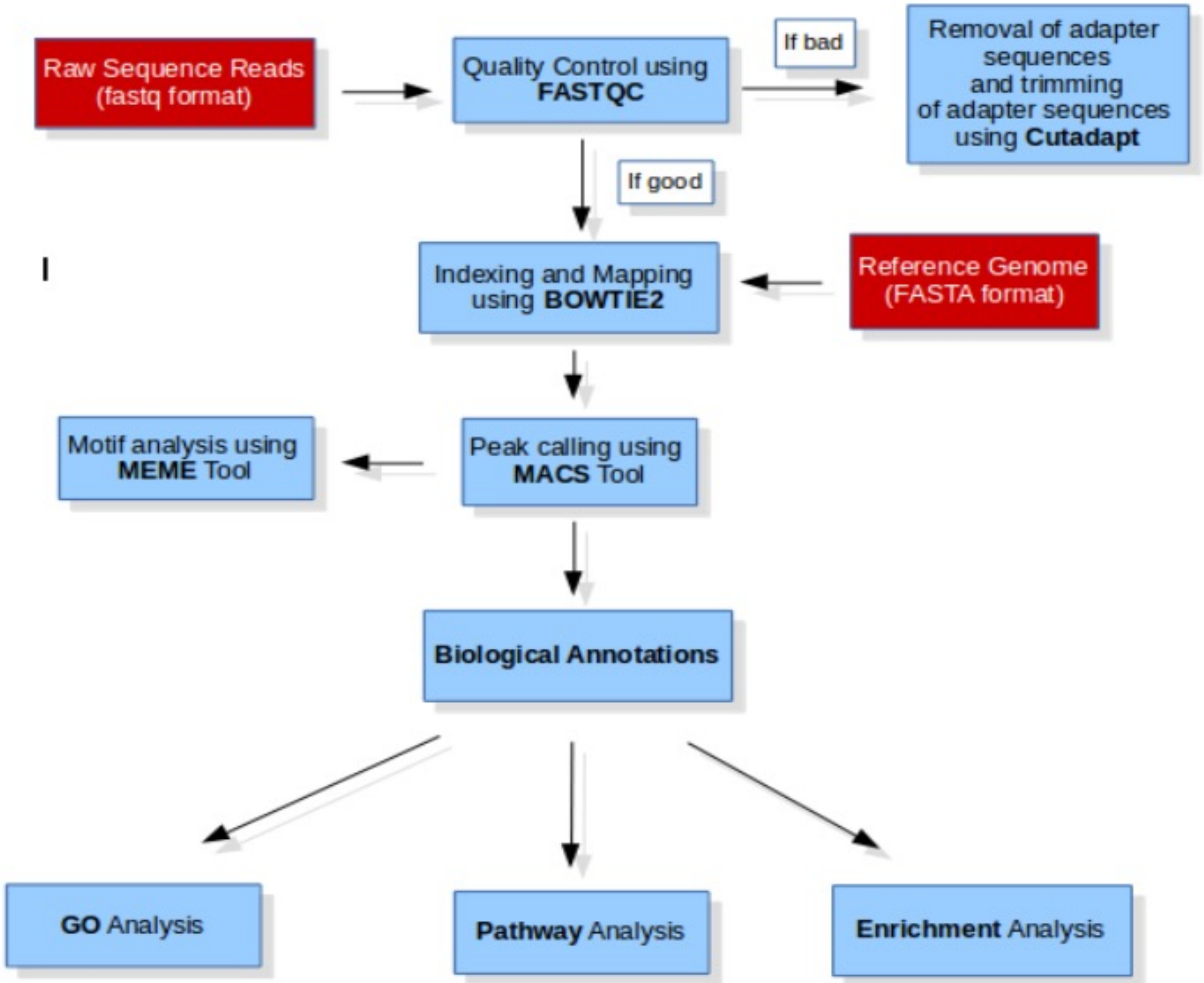
Run	# of Spots	# of Bases	Size	Published
SRR7080718	11,399,827	570M	214.8Mb	2019-02-04

Single End

- In single-end reading, the sequencer reads a fragment from only one end to the other, generating the sequence of base pairs. In this less sequencing is required and it is used for general purposes like differential expression analysis. Single-read sequencing can be a good choice for certain methods such as small RNA-Seq or chromatin immunoprecipitation sequencing (ChIP-Seq).



Workflow of the ChIP-Seq Analysis



Raw Sequence Reads (fastq format)

Quality Control using **FASTQC**

If bad

Removal of adapter sequences and trimming of adapter sequences using **Cutadapt**

If good

Indexing and Mapping using **BOWTIE2**

Reference Genome (FASTA format)

Peak calling using **MACS** Tool

Motif analysis using **MEME** Tool

Biological Annotations

GO Analysis

Pathway Analysis

Enrichment Analysis

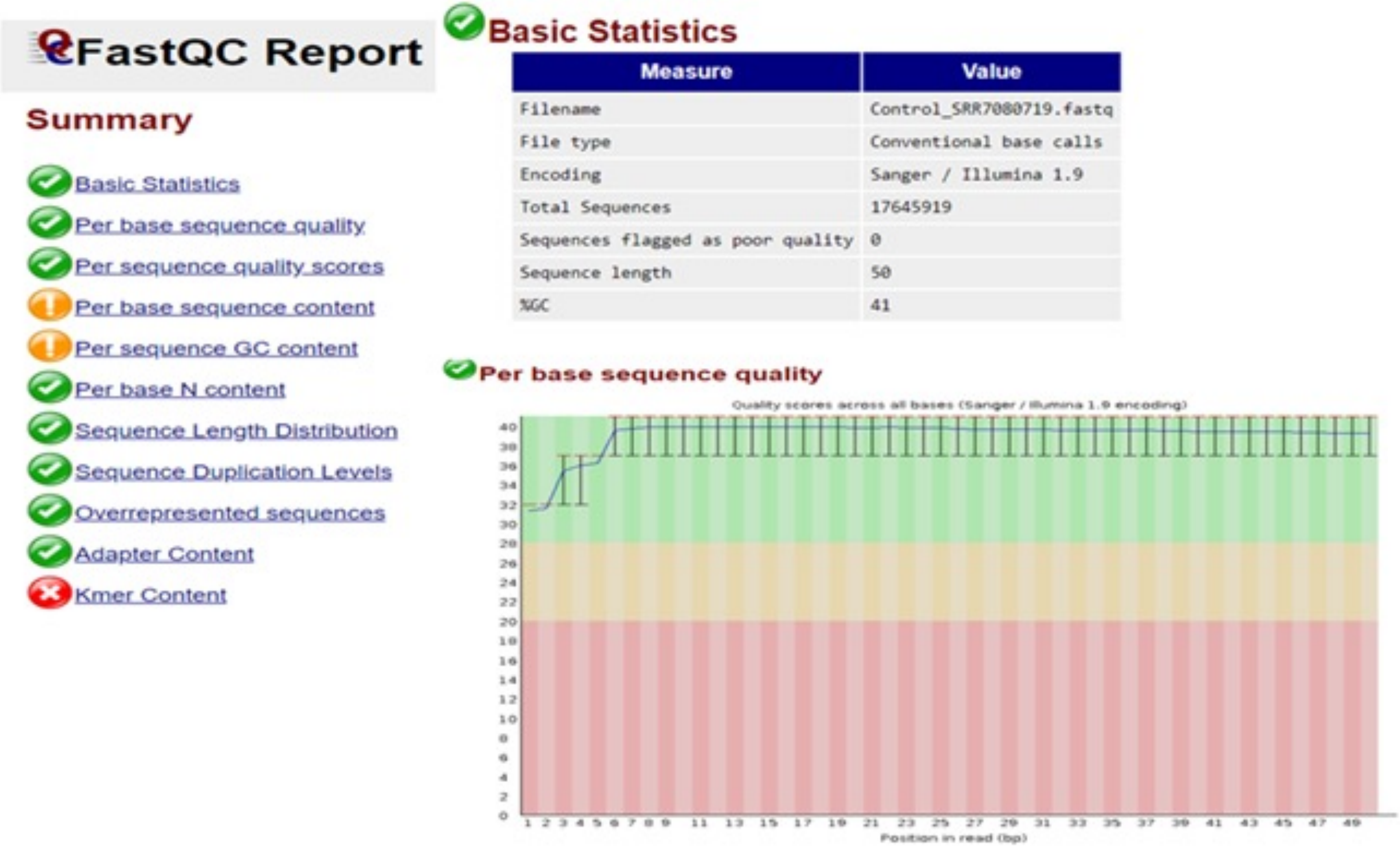
FastQC

- Modern high throughput sequencers can generate hundreds of millions of sequences in a single run. Before analysing this sequence to draw biological conclusions you should always perform some simple quality control checks to ensure that the raw data looks good and there are no problems or biases in your data which may affect how you can usefully use it.

- Most sequencers will generate a QC report as part of their analysis pipeline, but this is usually only focused on identifying problems which were generated by the sequencer itself. FastQC aims to provide a QC report which can spot problems which originate either in the sequencer or in the starting library material.

- FastQC can be run in one of two modes. It can either run as a stand alone interactive application for the immediate analysis of small numbers of FastQ files, or it can be run in a non-interactive mode where it would be suitable for integrating into a larger analysis pipeline for the systematic processing of large numbers of files.

FastQC result of Control - SRR7080719



FastQC Report

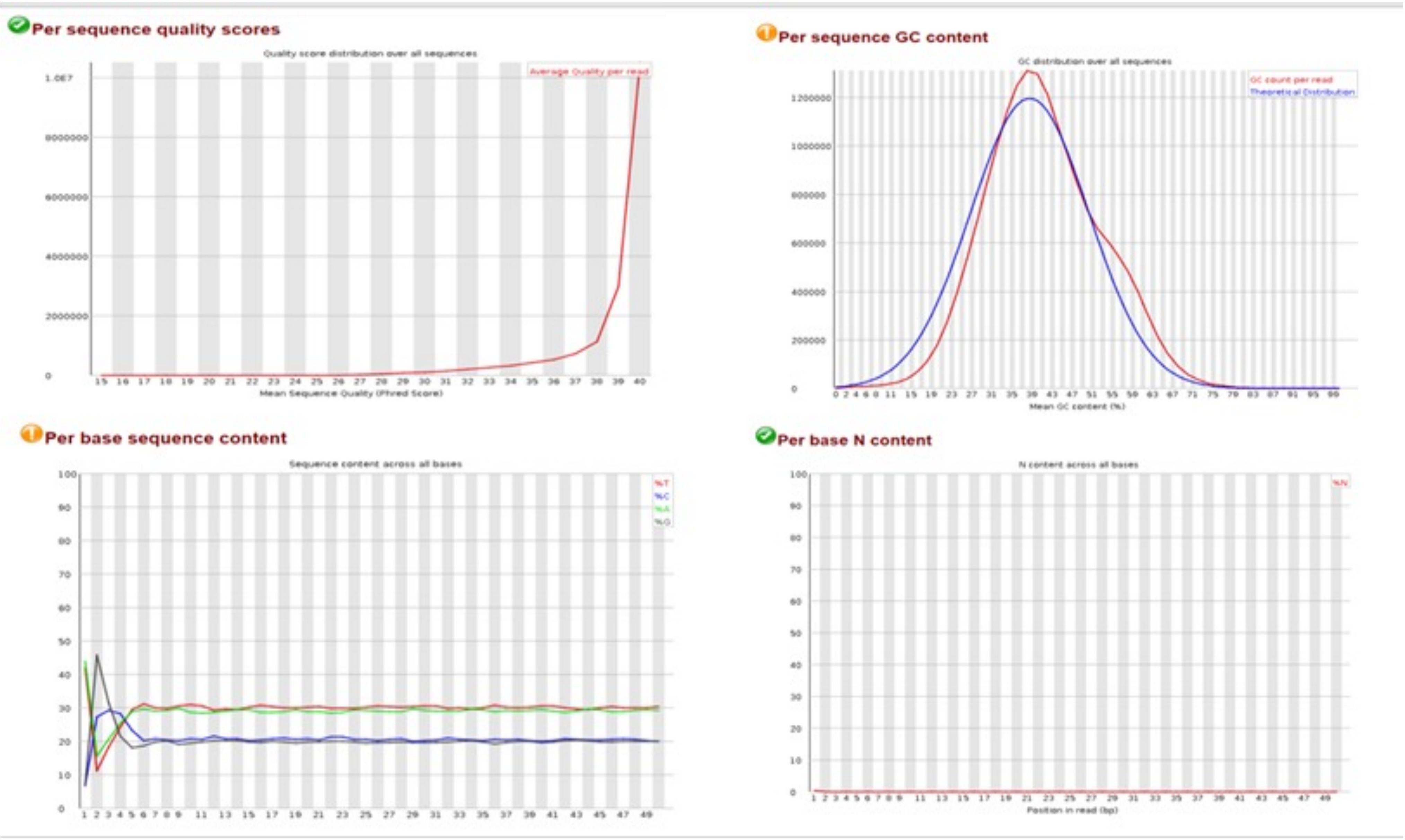
Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per sequence quality scores
- ! Per base sequence content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✓ Sequence Duplication Levels
- ✓ Overrepresented sequences
- ✓ Adapter Content
- ✗ Kmer Content

✓ Basic Statistics

Measure	Value
Filename	Control_SRR7080719.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	17645919
Sequences flagged as poor quality	0
Sequence length	50
%GC	41

✓ Per base sequence quality



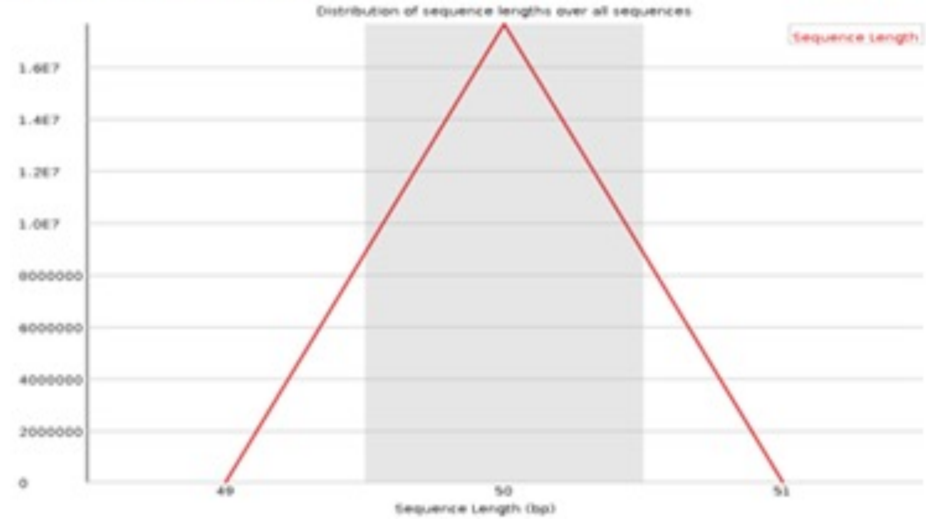
Per sequence quality scores

Per sequence GC content

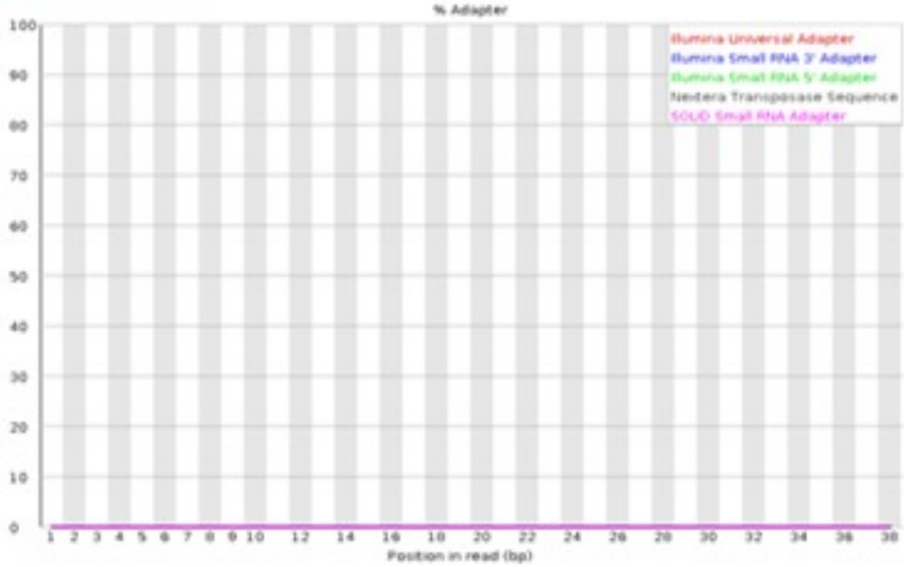
Per base sequence content

Per base N content

✓ Sequence Length Distribution



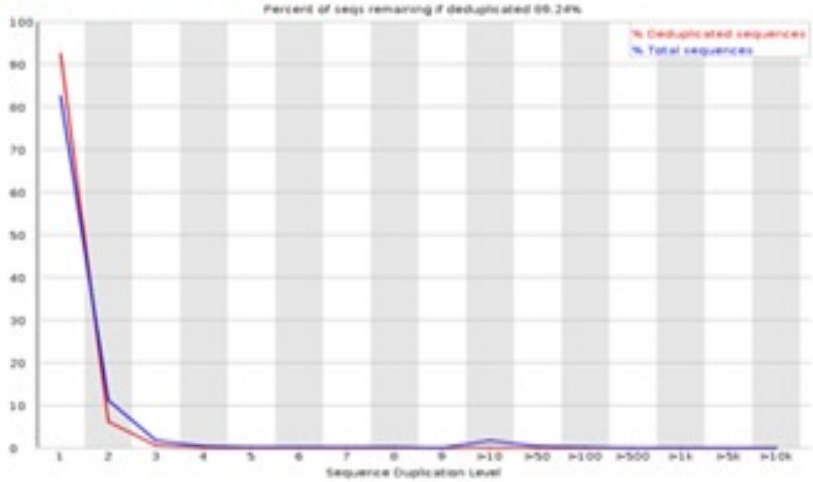
✓ Adapter Content



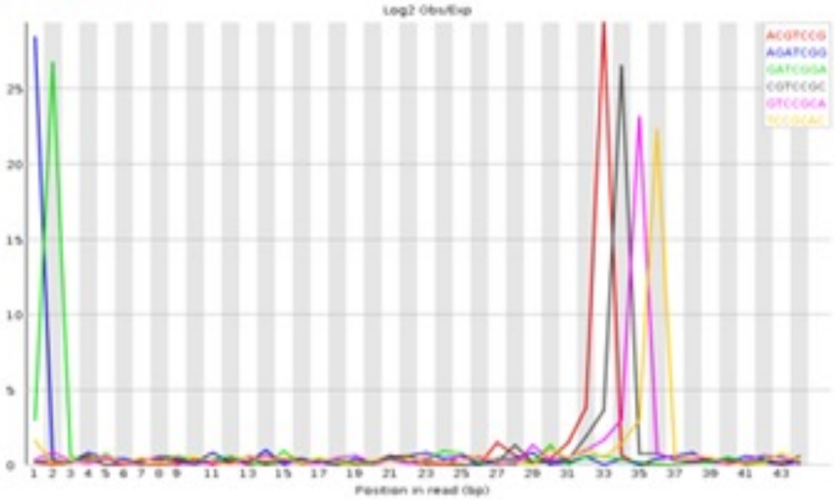
✓ Overrepresented sequences

No overrepresented sequences

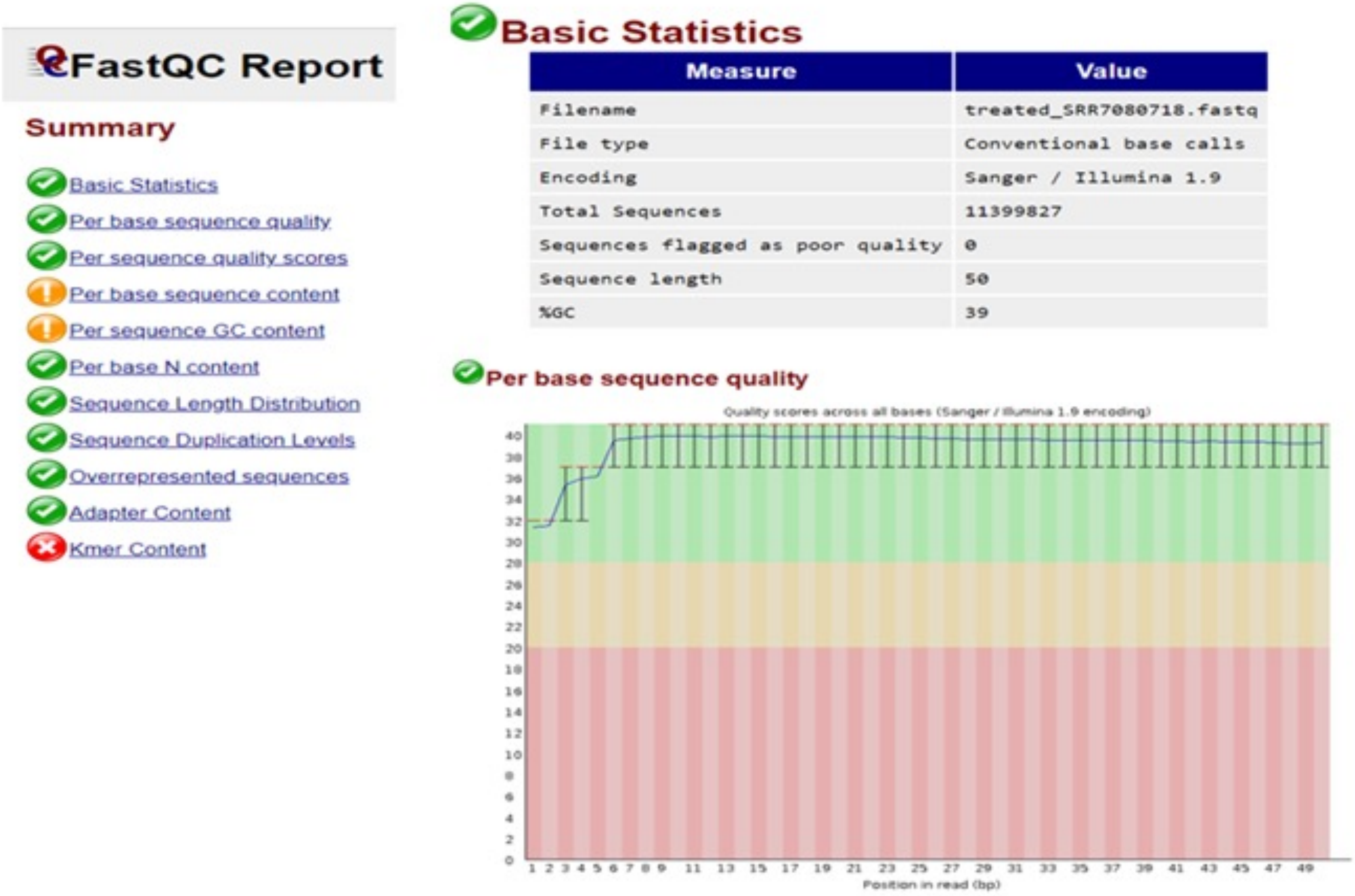
✓ Sequence Duplication Levels



✗ Kmer Content



FastQC result of Treated - SRR7080718



FastQC Report

Summary

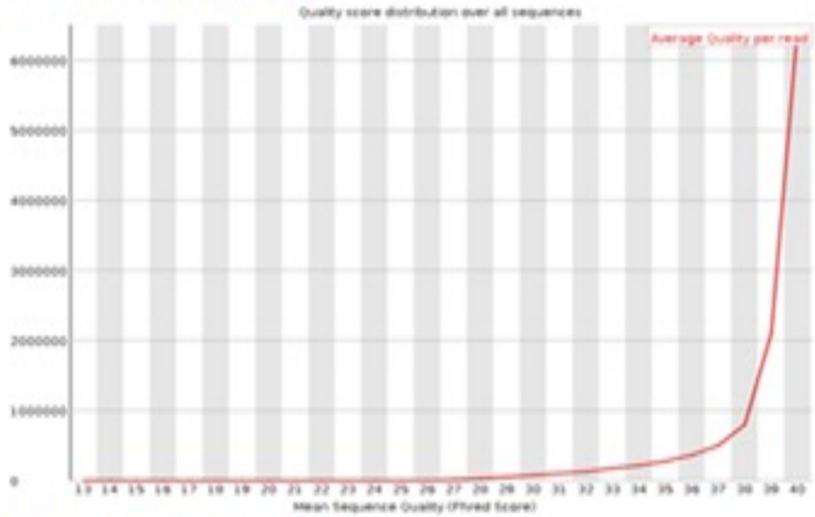
- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per sequence quality scores
- ! Per base sequence content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✓ Sequence Duplication Levels
- ✓ Overrepresented sequences
- ✓ Adapter Content
- ✗ Kmer Content

✓ Basic Statistics

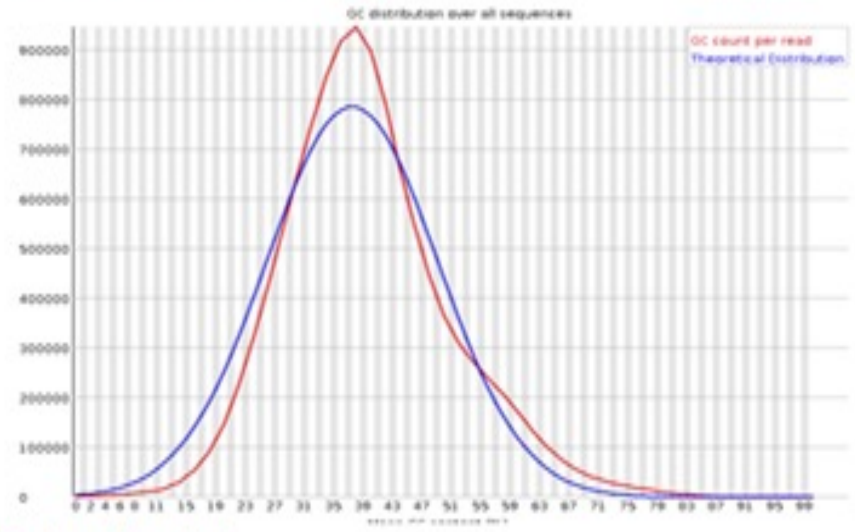
Measure	Value
Filename	treated_SRR7080718.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	11399827
Sequences flagged as poor quality	0
Sequence length	50
%GC	39

✓ Per base sequence quality

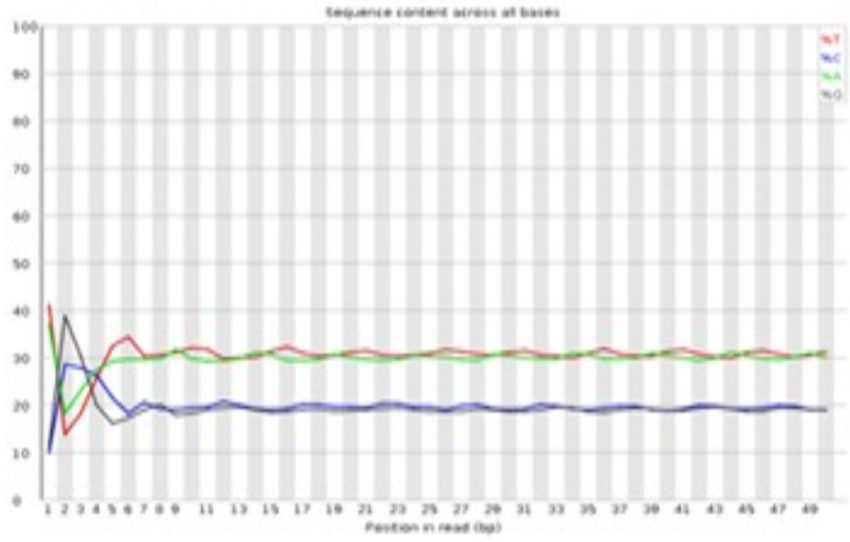
✓ Per sequence quality scores



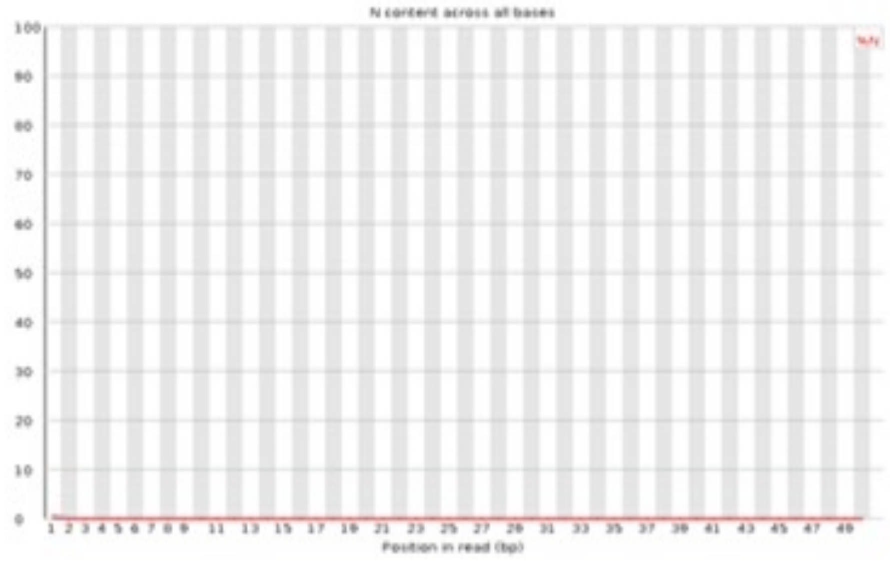
! Per sequence GC content



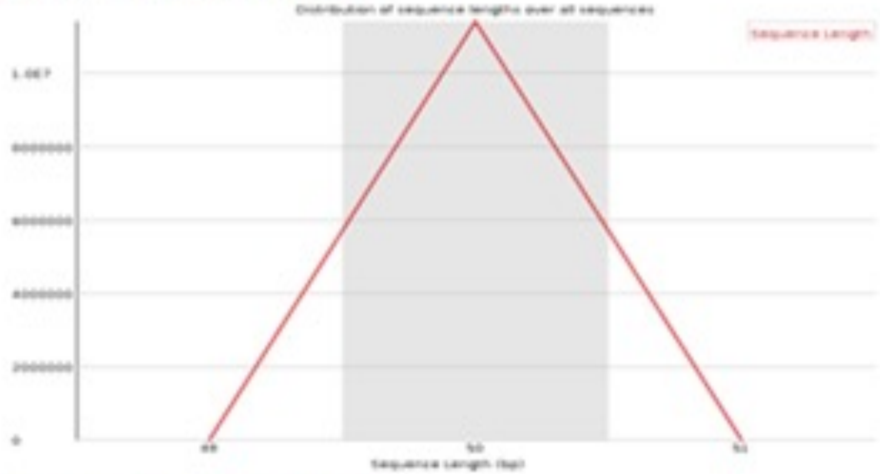
! Per base sequence content



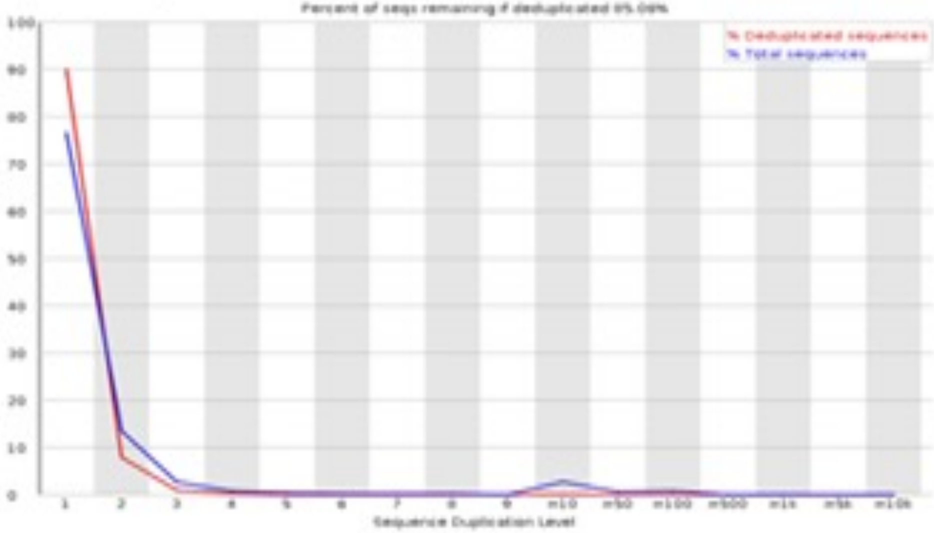
✓ Per base N content



✓ Sequence Length Distribution



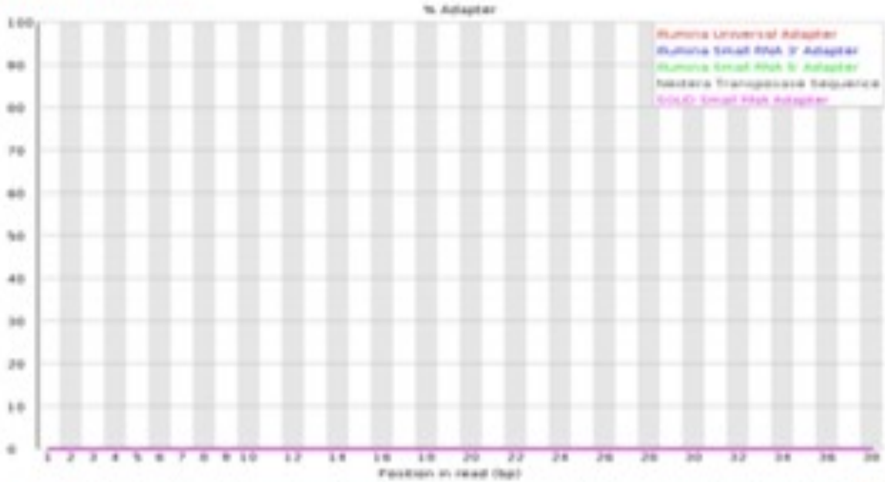
✓ Sequence Duplication Levels



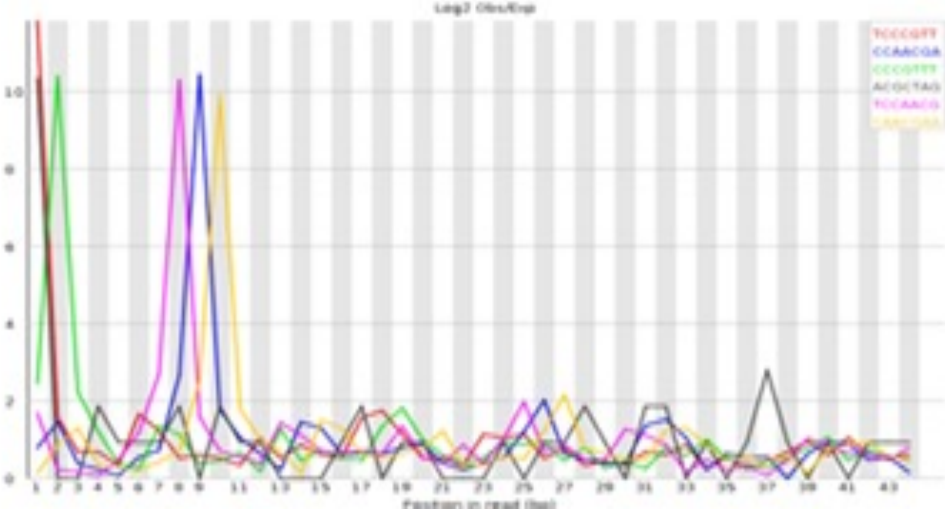
✓ Overrepresented sequences

No overrepresented sequences

✓ Adapter Content



✗ Kmer Content



MAPPING ON REFERENCE GENOME

Use Bowtie2 to align reads.

http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml#options

Bowtie 2

- Bowtie 2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences. It is particularly good at aligning reads of about 50 up to 100s of characters to relatively long (e.g. mammalian) genomes.

- Bowtie 2 outputs alignments in SAM format, enabling interoperation with a large number of other tools (e.g. SAMtools, GATK) that use SAM. Bowtie 2 is distributed under the GPLv3 license, and it runs on the command line under Windows, Mac OS X and Linux

Bowtie 2

- Bowtie 2 is often the first step in pipelines for comparative genomics, including for variation calling, ChIP-seq, RNA-seq, BS-seq.

- Bowtie uses indexed genome for the alignment in order to keep its memory footprint small. Because of time constraints we will build the index only for one chromosome of the human genome. For this we need the chromosome sequence in fasta format.

Bowtie 2

Steps:

- Indexing using Bowtie2
- Mapping using Bowtie2
- Align the Control_SRR7080719 reads using Bowtie2
- Align the treated_SRR7080718 reads using Bowtie2
- Converting SAM to BAM using samtools

BAM File Formate

- A BAM file (*.bam) is the compressed binary version of a SAM file that is used to represent aligned sequences up to 128 Mb. SAM and BAM formats are described in detail at https://samtools.github.io/hts-specs/SAMv1.pdf.

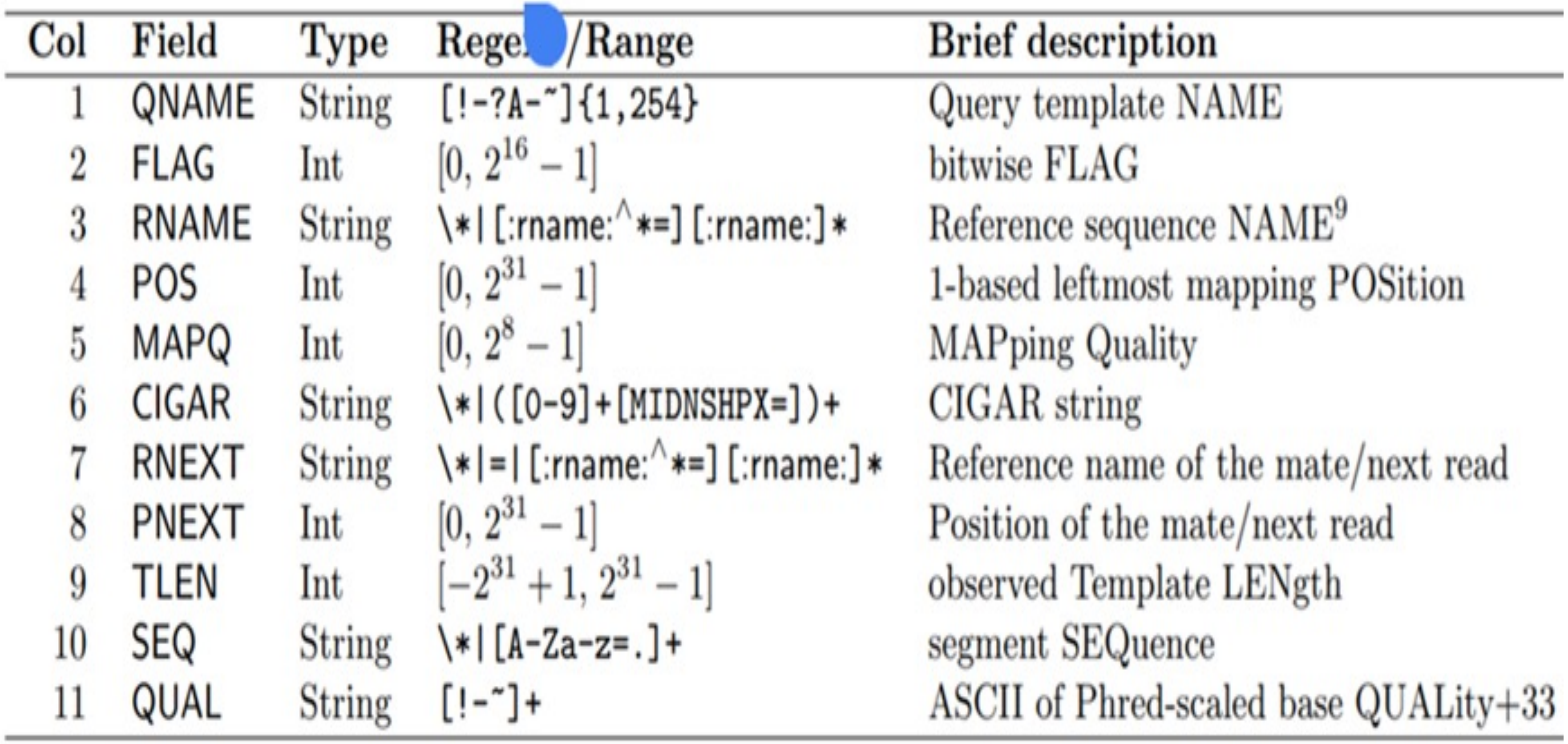
- BAM files use the file naming format of SampleName_S#.bam, where # is the sample number determined by the order that samples are listed for the run. In multi-node mode, the S# is set to S1, regardless the order of the sample.

- BAM files contain a header section and an alignment section.
- Header —Contains information about the entire file, such as sample name, sample length, and alignment method. Alignments in the alignments section are associated with specific information in the header section.
- Alignments—Contains read name, read sequence, read quality, alignment information, and custom tags. The read name includes the chromosome, start coordinate, alignment quality, and the match descriptor string

SAM file format

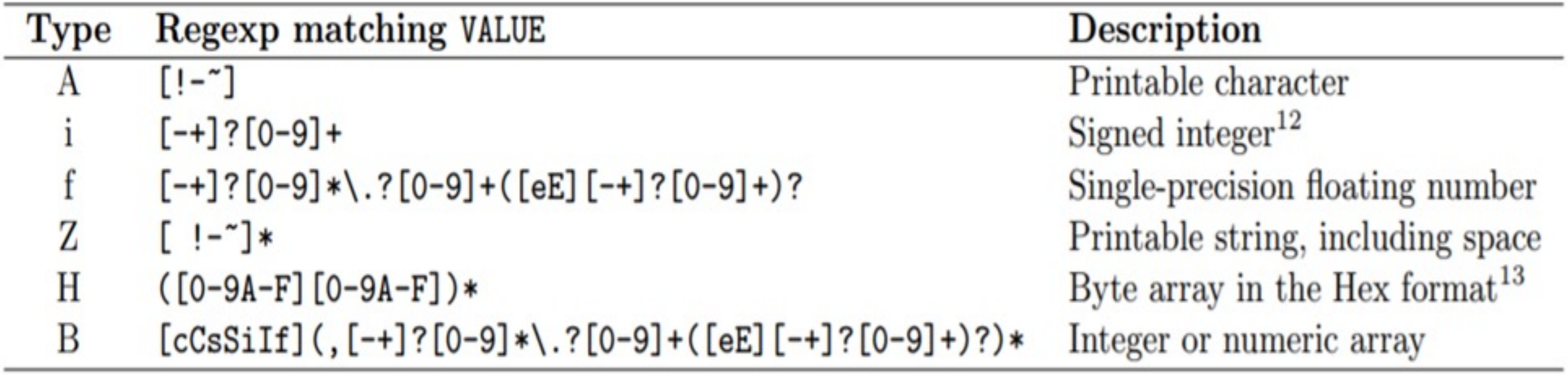
- SAM stands for Sequence Alignment/Map format. It is a TAB-delimited text format consisting of a header section, which is optional, and an alignment section. If present, the header must be prior to the alignments. Header lines start with ‘@’, while alignment lines do not. Each alignment line has 11 mandatory fields for essential alignment information such as mapping position, and variable number of optional fields for flexible or aligner specific information.

The alignment section: mandatory fields



Col	Field	Type	Regex/Range	Brief description
1	QNAME	String	[!-?A-~]{1,254}	Query template NAME
2	FLAG	Int	$[0, 2^{16} - 1]$	bitwise FLAG
3	RNAME	String	*|[:rname:^*=][:rname:]*	Reference sequence NAME[9]
4	POS	Int	$[0, 2^{31} - 1]$	1-based leftmost mapping POSition
5	MAPQ	Int	$[0, 2^{8} - 1]$	MAPping Quality
6	CIGAR	String	*|([0-9]+[MIDNSHPX=])+	CIGAR string
7	RNEXT	String	*|=|[:rname:^*=][:rname:]*	Reference name of the mate/next read
8	PNEXT	Int	$[0, 2^{31} - 1]$	Position of the mate/next read
9	TLEN	Int	$[-2^{31} + 1, 2^{31} - 1]$	observed Template LENgth
10	SEQ	String	*|[A-Za-z=.]+	segment SEQuence
11	QUAL	String	[!-~]+	ASCII of Phred-scaled base QUALity+33

The alignment section: optional fields



Type	Regexp matching `VALUE`	Description
A	`[!-~]`	Printable character
i	`[-+]?[0-9]+`	Signed integer[12]
f	`[-+]?[0-9]*\.?[0-9]+([eE][-+]?[0-9]+)?`	Single-precision floating number
Z	`[ !-~]*`	Printable string, including space
H	`([0-9A-F][0-9A-F])*`	Byte array in the Hex format[13]
B	`[cCsSiIf](,[-+]?[0-9]*\.?[0-9]+([eE][-+]?[0-9]+)?)*`	Integer or numeric array

SAMtools

- Samtools is a set of utilities that manipulate alignments in the BAM format. It imports from and exports to the SAM (Sequence Alignment/Map) format, does sorting, merging and indexing, and allows to retrieve reads in any regions swiftly.

- Samtools is designed to work on a stream. It regards an input file \`-' as the standard input (stdin) and an output file \`-' as the standard output (stdout). Several commands can thus be combined with Unix pipes. Samtools always output warning and error messages to the standard error output (stderr).

- Samtools is also able to open a BAM (not SAM) file on a remote FTP or HTTP server if the BAM file name starts with \`ftp://' or \`http://'. Samtools checks the current working directory for the index file and will download the index upon absence. Samtools does not retrieve the entire alignment file unless it is asked to do so.

SAMtools

Steps:

- sort sam file into bam sorted
- Remove duplicate reads
- Only pick reads which is mapping uniquely
- Indexing of .bam

PEAK CALLING

MACS2

MACS2

- With the improvement of sequencing techniques, chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq) is getting popular to study genome-wide protein-DNA interactions.

- To address the lack of powerful ChIP-Seq analysis method, we present a novel algorithm, named Model-based Analysis of ChIP-Seq (MACS), for identifying transcript factor binding sites.

- MACS captures the influence of genome complexity to evaluate the significance of enriched ChIP regions, and MACS improves the spatial resolution of binding sites through combining the information of both sequencing tag position and orientation. MACS can be easily used for ChIP-Seq data alone, or with control sample with the increase of specificity.

MOTIF-ANALYSIS

MEME-ChIPtool

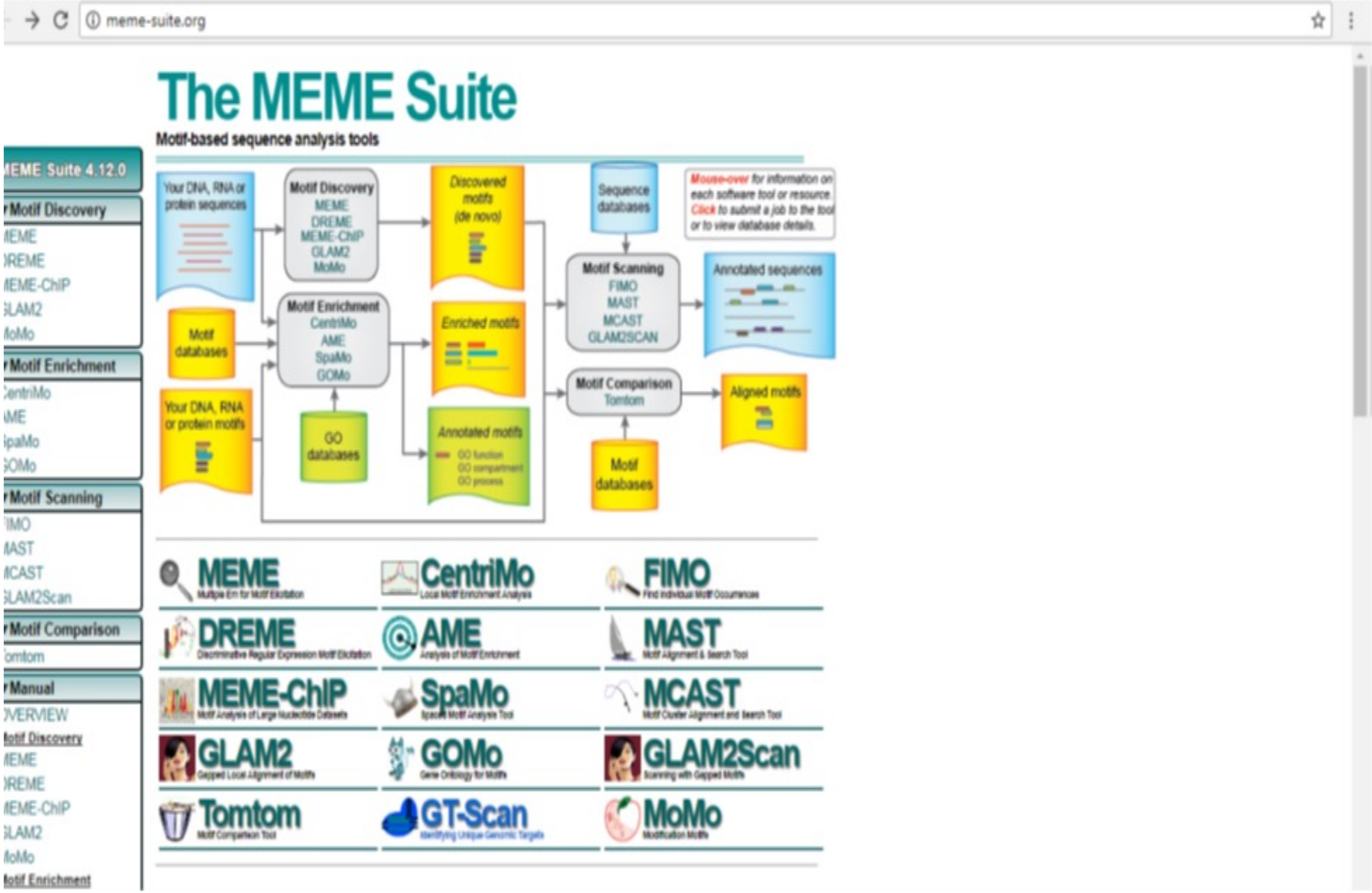
MEME-ChIP

- In genetics, a sequence motif is a nucleotide or amino-acid sequence pattern that is widespread and has, or is conjectured to have, a biological significance. For proteins, a sequence motif is distinguished from a structural motif, a motif formed by the three-dimensional arrangement of amino acids which may not be adjacent.

- ‘Motif discovery’ (or ‘motif finding’) in biological sequences can be defined as the problem of finding short similar sequence elements (building the ‘motif’) shared by a set of nucleotide or protein sequences with a common biological function.

MEME-ChIP

- All the peaks fasta sequences retrieved for the motif analysis can be used online for MEME-Chip tool for further motif analysis.
- MEME-Chip performs motif discovery, motif enrichment analysis and clustering on large nucleotide datasets.
- First go to MEME suite from your browser



meme-suite.org

The MEME Suite

Motif-based sequence analysis tools

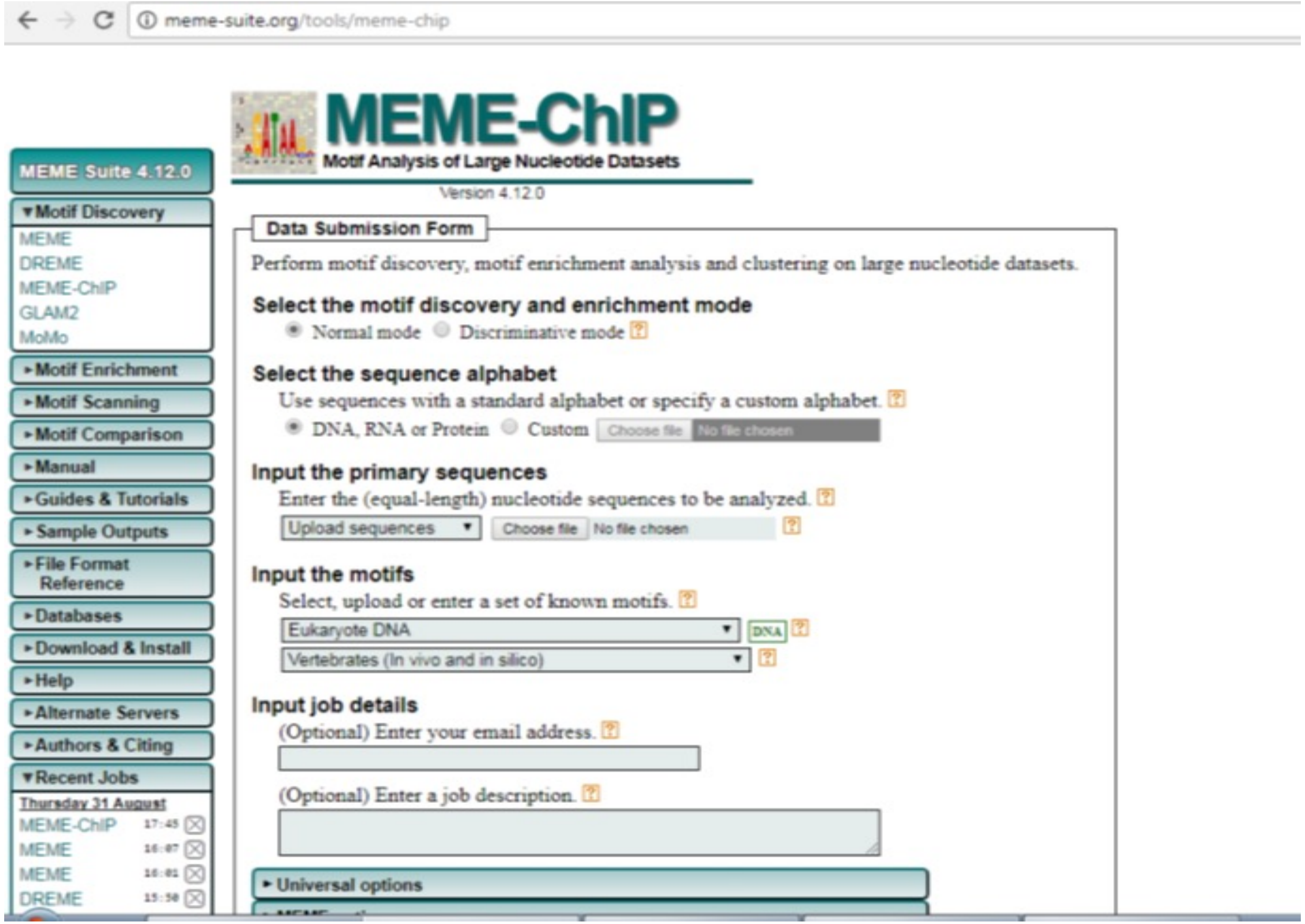
EME Suite 4.12.0

- Motif Discovery
  - IEME
  - REME
  - IEME-ChIP
  - LAM2
  - loMo
- Motif Enrichment
  - entriMo
  - ME
  - paMo
  - OMo
- Motif Scanning
  - IMO
  - IAST
  - ICAST
  - LAM2Scan
- Motif Comparison
  - omtom
- Manual
  - VERVIEW
  - lotif Discovery
  - IEME
  - REME
  - IEME-ChIP
  - LAM2
  - loMo
  - lotif Enrichment

Mouse-over for information on each software tool or resource. *Click* to submit a job to the tool or to view database details.

MEME Multiple Em for Motif Elicitation	**CentriMo** Local Motif Enrichment Analysis	**FIMO** Find Individual Motif Occurrences
DREME Discriminative Regular Expression Motif Elicitation	**AME** Analysis of Motif Enrichment	**MAST** Motif Alignment & Search Tool
MEME-ChIP Motif Analysis of Large Nucleotide Datasets	**SpaMo** Spaced Motif Analysis Tool	**MCAST** Motif Cluster Alignment and Search Tool
GLAM2 Gapped Local Alignment of Motifs	**GOMo** Gene Ontology for Motifs	**GLAM2Scan** Scanning with Gapped Motifs
Tomtom Motif Comparison Tool	**GT-Scan** Identifying Unique Genomic Targets	**MoMo** Modification Motifs

Go to Motif discovery and select for MEME-CHIP and upload the peak fasta sequences file at the section "Input the primary sequences"and click on "start search



Perform motif discovery, motif enrichment analysis and clustering on large nucleotide datasets.

Select the motif discovery and enrichment mode

● Normal mode ○ Discriminative mode

Select the sequence alphabet

Use sequences with a standard alphabet or specify a custom alphabet.

● DNA, RNA or Protein ○ Custom Choose file No file chosen

Input the primary sequences

Enter the (equal-length) nucleotide sequences to be analyzed.

Upload sequences ▾ Choose file No file chosen

Input the motifs

Select, upload or enter a set of known motifs.

Eukaryote DNA ▾ DNA

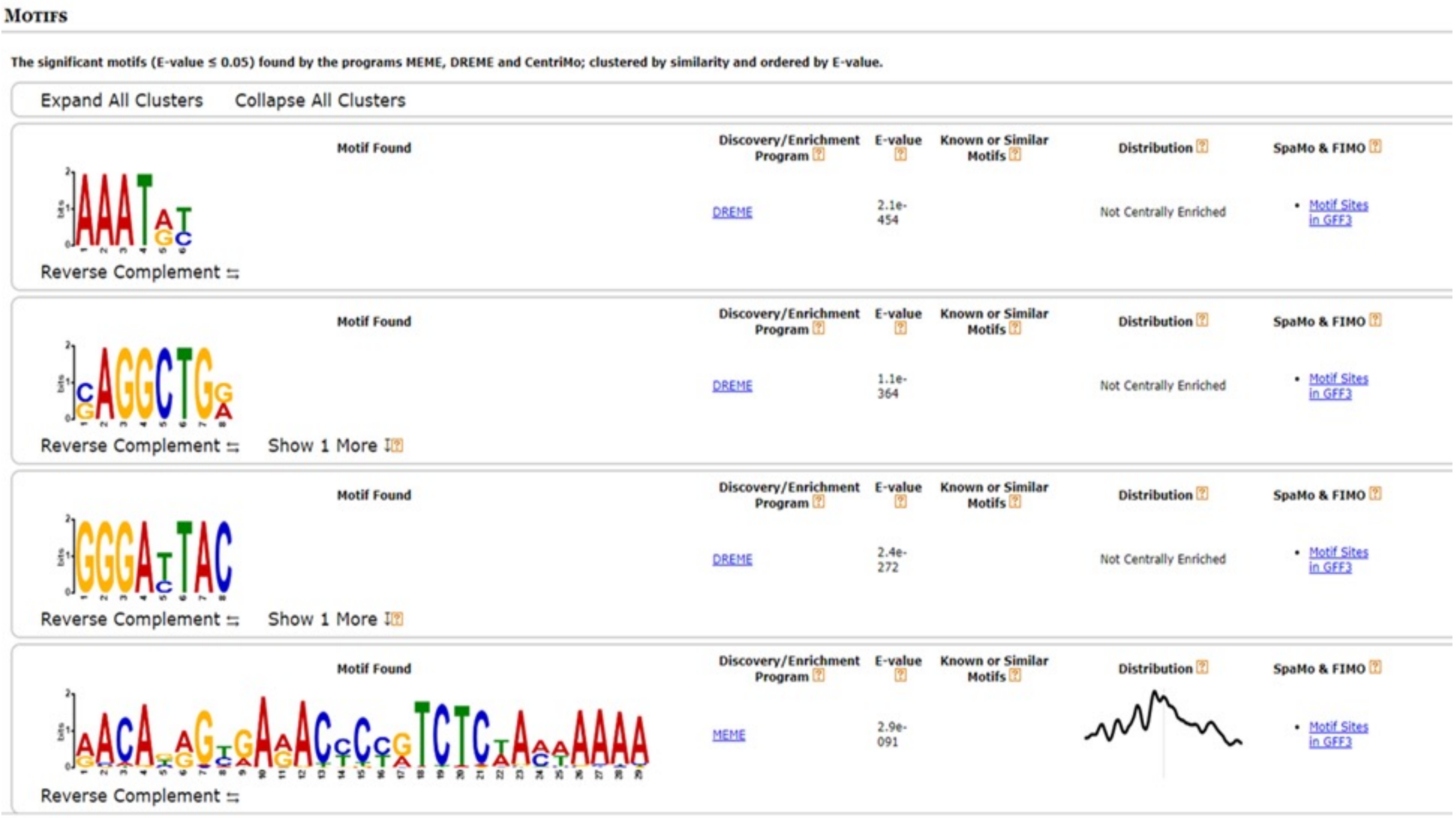
Vertebrates (In vivo and in silico) ▾

Input job details

(Optional) Enter your email address.

(Optional) Enter a job description.

Results from MEME-CHIP :

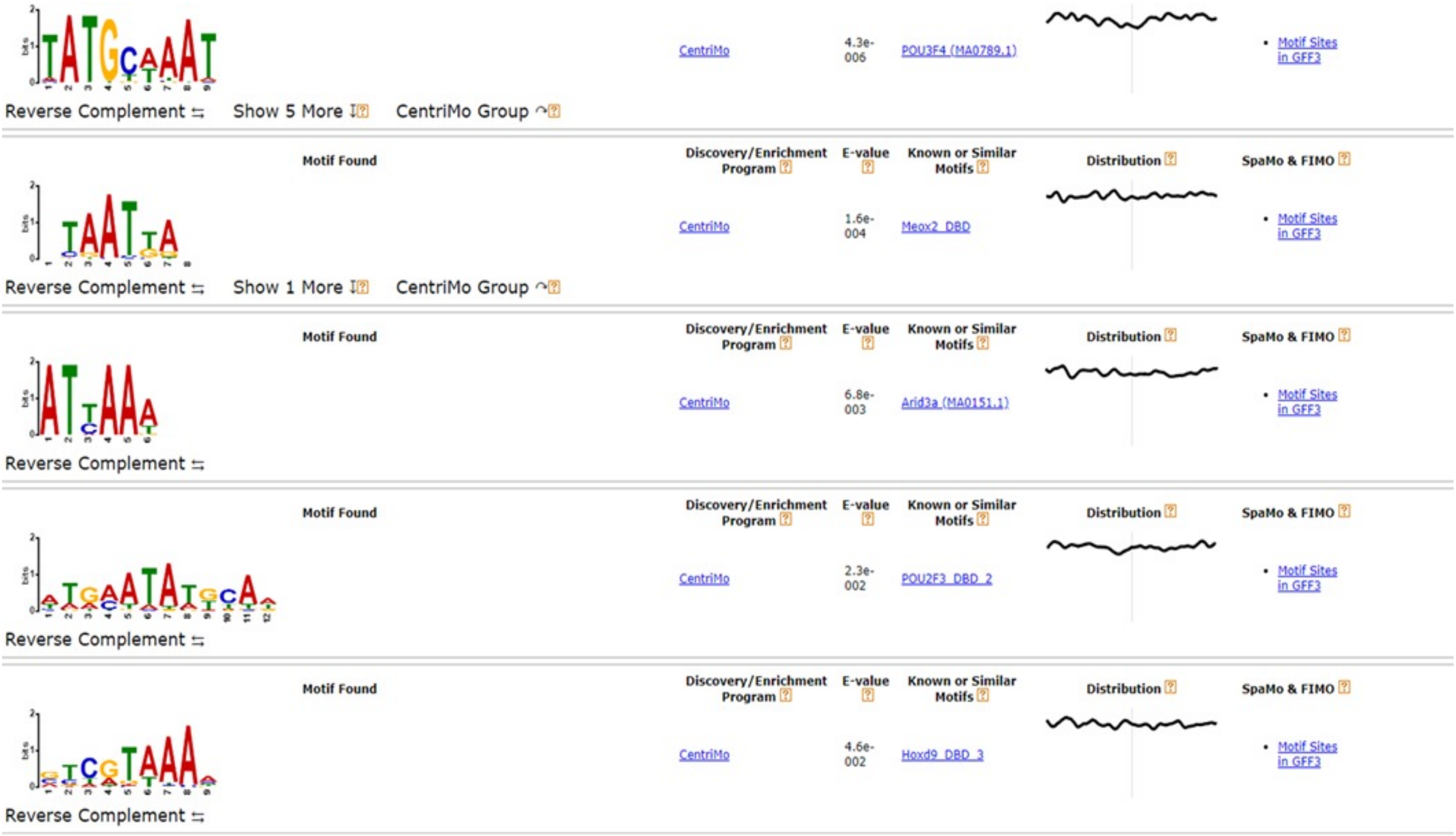


Motifs

The significant motifs (E-value ≤ 0.05) found by the programs MEME, DREME and CentriMo; clustered by similarity and ordered by E-value.

Expand All Clusters Collapse All Clusters

Motif Found	Discovery/Enrichment Program	E-value	Known or Similar Motifs	Distribution	SpaMo & FIMO
Reverse Complement ⇆	DREME	2.1e-454		Not Centrally Enriched	Motif Sites in GFF3
Reverse Complement ⇆ Show 1 More	DREME	1.1e-364		Not Centrally Enriched	Motif Sites in GFF3
Reverse Complement ⇆ Show 1 More	DREME	2.4e-272		Not Centrally Enriched	Motif Sites in GFF3
Reverse Complement ⇆	MEME	2.9e-091			Motif Sites in GFF3



Motif Found	Discovery/Enrichment Program	E-value	Known or Similar Motifs	Distribution	SpaMo & FIMO
Reverse Complement ⇆ Show 5 More ⇣ CentriMo Group ↷	CentriMo	4.3e-006	POU3F4 (MA0789.1)		- Motif Sites in GFF3
Reverse Complement ⇆ Show 1 More ⇣ CentriMo Group ↷	CentriMo	1.6e-004	Meox2_DBD		- Motif Sites in GFF3
Reverse Complement ⇆	CentriMo	6.8e-003	Arid3a (MA0151.1)		- Motif Sites in GFF3
Reverse Complement ⇆	CentriMo	2.3e-002	POU2F3_DBD_2		- Motif Sites in GFF3
Reverse Complement ⇆	CentriMo	4.6e-002	Hoxd9_DBD_3		- Motif Sites in GFF3

CHIP PEAK ANNOTATION, COMPARISON, AND VISUALIZATION

ChIPseeker

Chip Peak Annotation, Comparison, And Visualization

- Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) has become standard technologies for genome wide identification of DNA-binding protein target sites. After read mappings and peak callings, the peak should be annotated to answer the biological questions.

- Annotation also create the possibility of integrating expression profile data to predict gene expression regulation. ChIPseeker was developed for annotating nearest genes and genomic features to peaks.

- ChIP peak data set comparison is also very important. ChIPseeker (Yu, Wang, and He 2015) support statistical testing of significant overlap among ChIP seq data sets, and incorporate open access database GEO for users to compare their own dataset to those deposited in database.Converting genome coordinations from one genome version to another is also supported, making this comparison available for different genome version and different species.

Chip Peak Annotation, Comparison, And Visualization

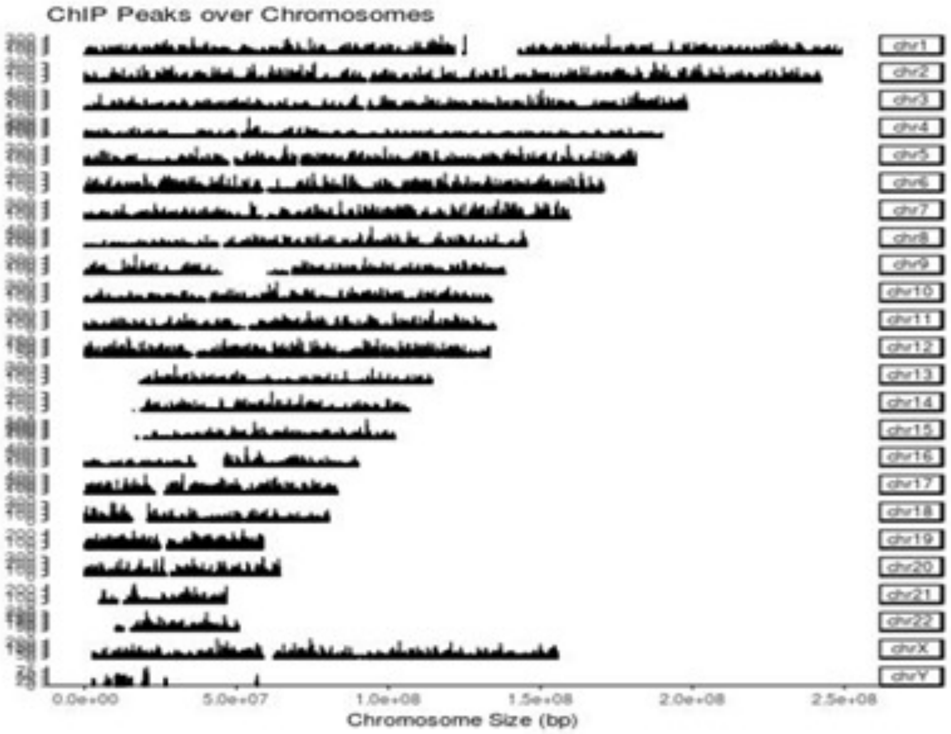
- Several visualization functions are implemented to visualize the coverage of the ChIP seq data, peak annotation, average profile and heatmap of peaks binding to TSS region.

- Functional enrichment analysis of the peaks can be performed by my Bioconductor packages DOSE(Yu et al. 2015), ReactomePA(Yu and He 2016), clusterProfiler(Yu et al. 2012).

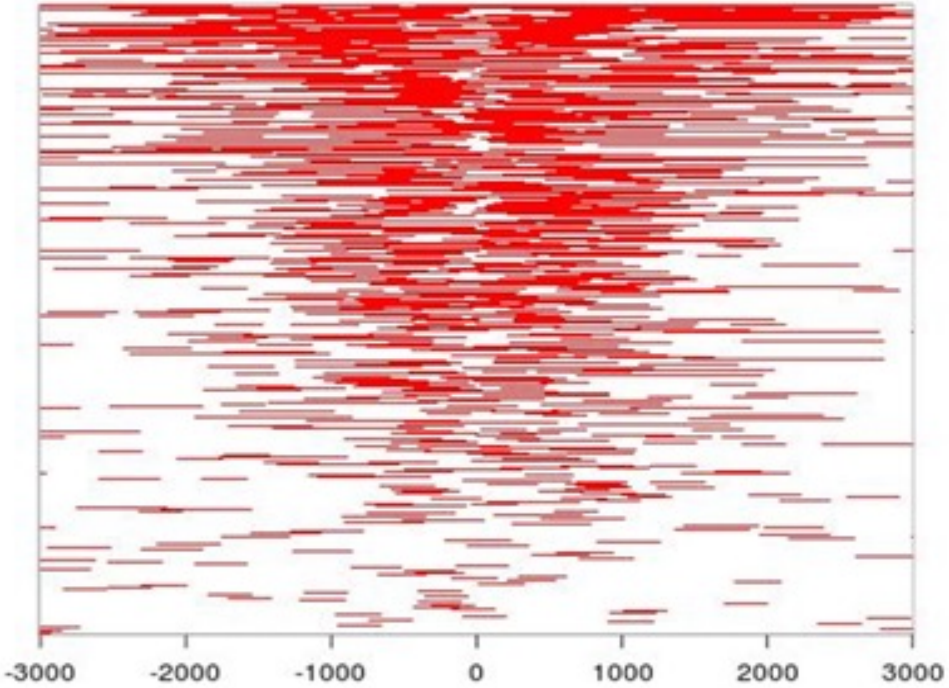
ChIPseeker

- ChIPseeker is an R package for annotating ChIP-seq data analysis. It supports annotating ChIP peaks and provides functions to visualize ChIP peaks coverage over chromosomes and profiles of peaks binding to TSS regions. Comparison of ChIP peak profiles and annotation are also supported. Moreover, it supports evaluating significant overlap among ChIP-seq datasets.

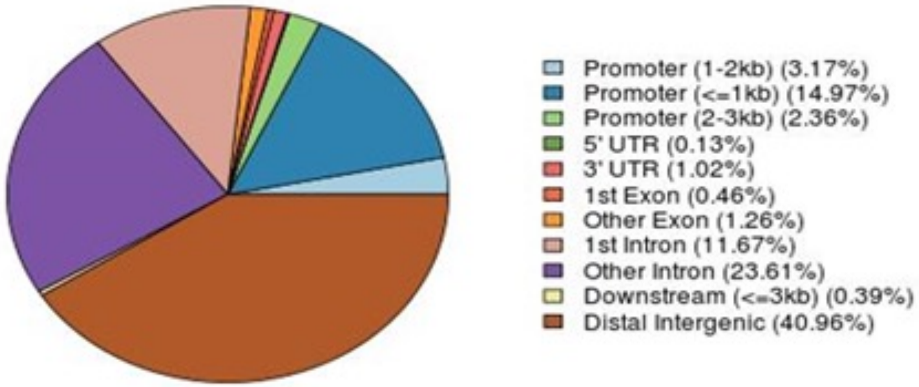
Results generated after running ChIPseeker tool:



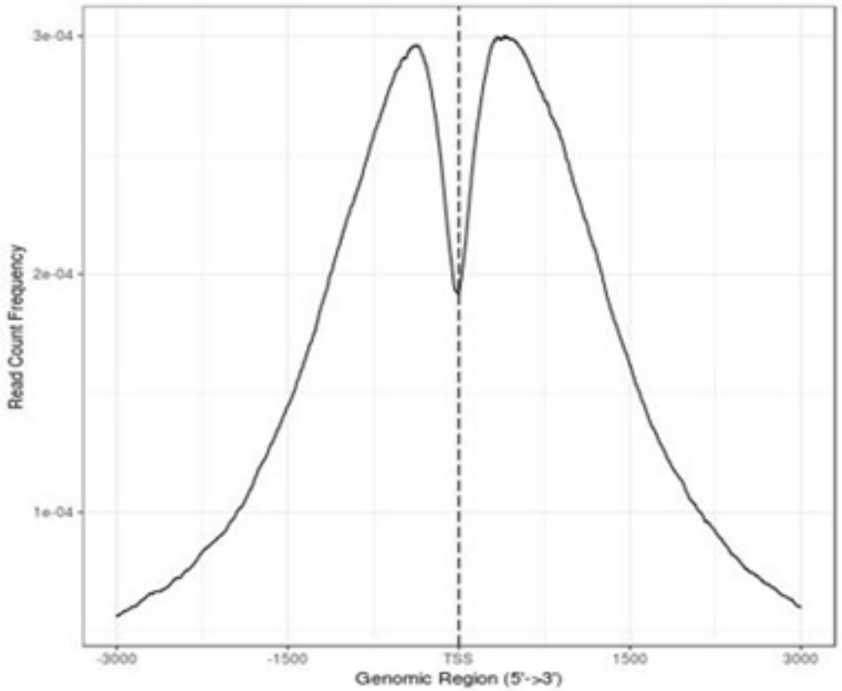
covplot



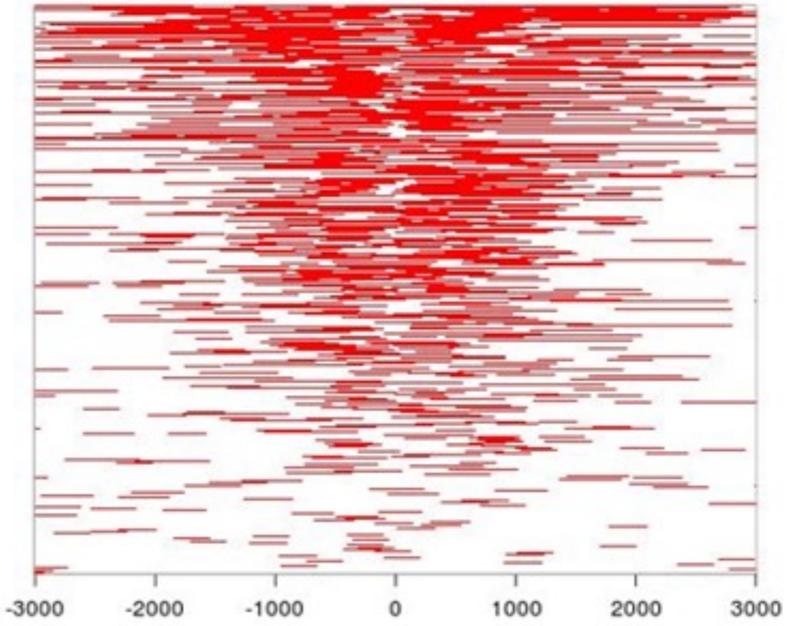
peakHeatmap



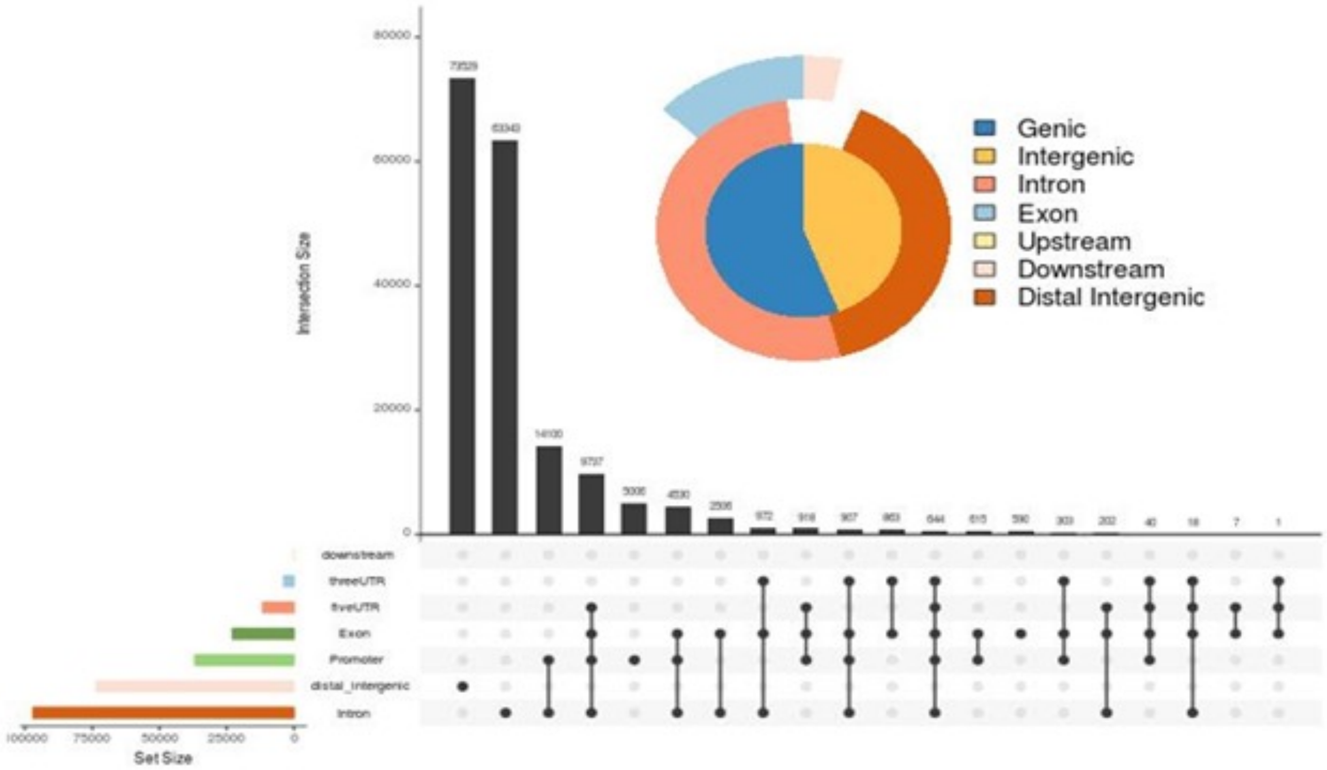
plotAnnoPie



plotAvgProf2



tagHeatmap



upsetplot