

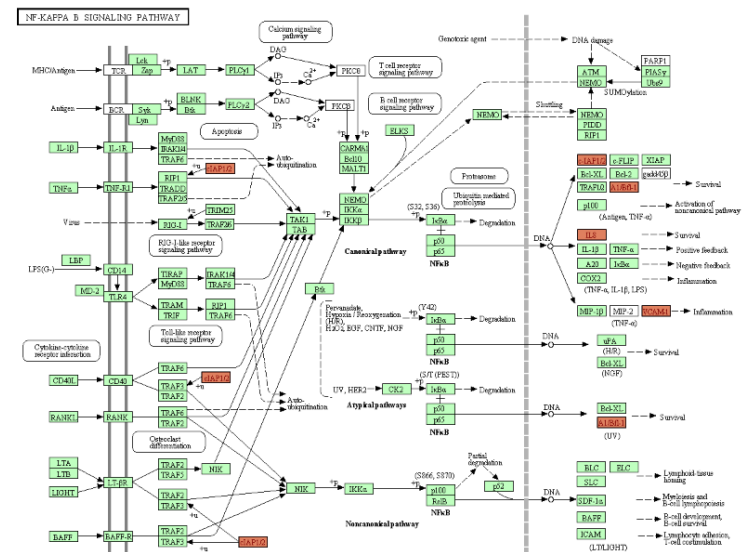
AS Life Science Library Training Course
2015/06/09

Chen, Shu-Hwa Ph.D.

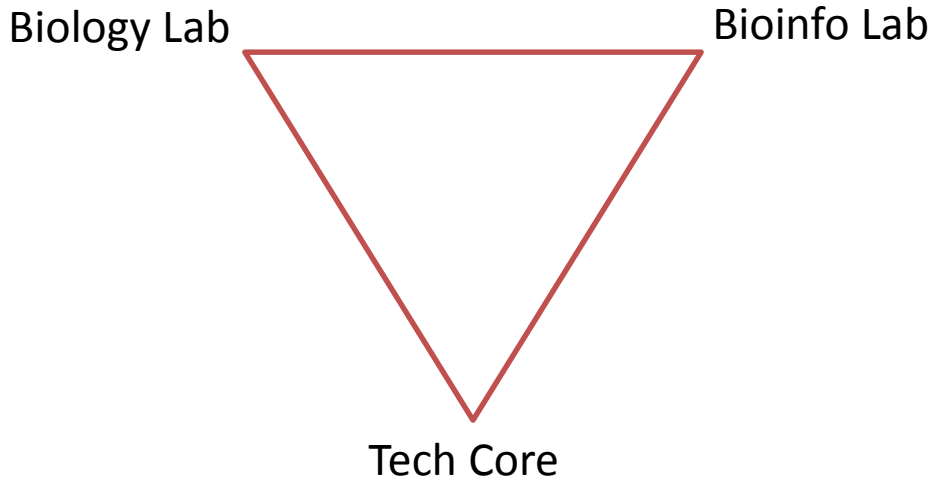
IIS, Academia Sinica

Outline

- Something about RNASeq
- Introduce MOLAS system
 - How to submit your data
 - How to view and analyze your data



High-throughput Methods



NGS TECHNOLOGY



Roche GS-FLX



Life Technologies SOLiD



Illumina HiSeq



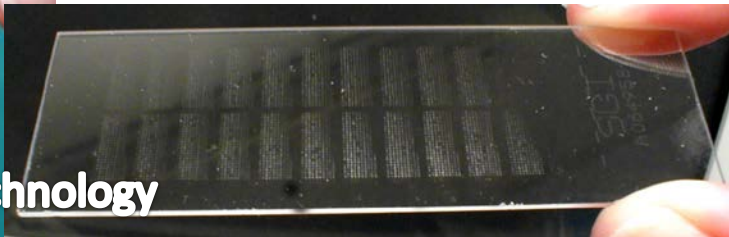
Life Technologies Ion Torrent



PacBio



Microarray Technology



RNASeq Data Analysis

Figure 1: RNA-Seq Data Analysis



- [2x50bp, 30million]'s View
- [2x150bp, 400 million]'s View
- [2x150bp, 6 billion]'s View

3 GB

1.3 MB

<1MB

Expression file (tab txt file):
~10Mb / human transcriptome

output as compressed FASTQ files of approximately 3 GB. These are put into secondary analysis, for aligning to reference genome, counting, and normalizing, generating about 1.3 MB. cSNPs, indels, and splice junctions are identified, and gene- and transcript-level expression and differential expression levels are calculated. Data output from these tertiary analyses vary, but are usually less than 1 MB.

Sequencing Platforms



ABI 3730xl
Sanger Sequencing



454 Life Sciences
pyrosequencing



SOLiD +
Illumina



Pacific Biosciences,
Oxford Nanopore etc
Single-molecule
sequencing

Length/read	800 bp	400 bp	100 bp	20 000+ bp
Reads/run	96	1 million	2 billion	5 million
Bases/run	60 kbp	400 Mbp	500 Gbp	100 Gbp
Speed	10 years/HG	1 month/HG	1 day/HG	10 min/HG

“old school”

“2nd gen”

“3rd gen”

FastQ format

- Start with “@”
- Four lines: “+” w/ or w/o seq head, quality scores

seq head	@ <u>EAS139</u> :136:FC706VJ:2:5:1000:12850 (1):N:18:ATCACG
seq letters	AA
+	+
quality score	BBBCCCC?<A?BC?7@@????????DBBA@@@@A@@

FASTQ files from CASAVA-1.8 Should have the following READ-ID format:

@<instrument>:<run number>:<flowcell ID>:<lane>:<tile>:<x-pos>:<y-pos>

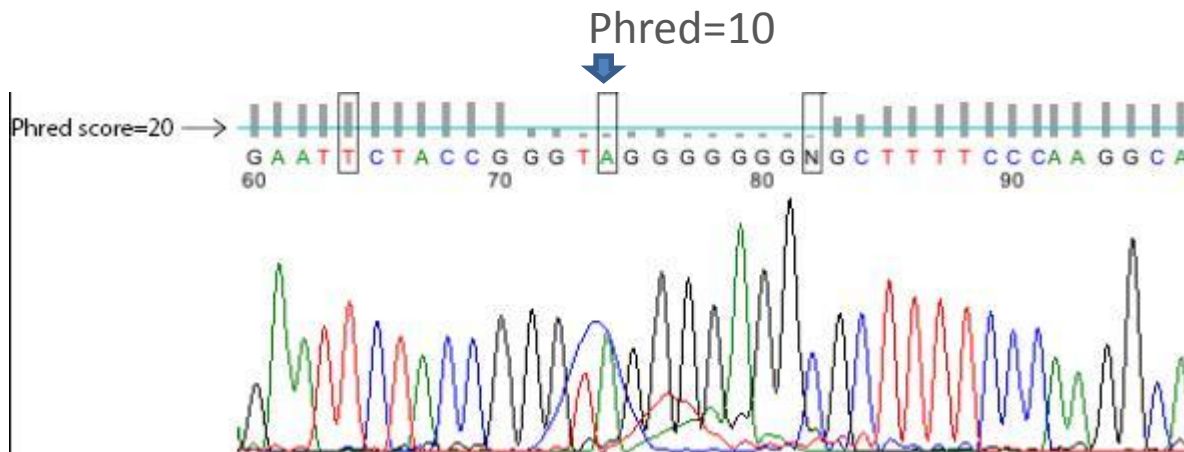
(<read>):<is filtered>:<control number>:<index sequence>

http://en.wikipedia.org/wiki/FASTQ_format

http://cancan.cshl.edu/labmembers/gordon/fastq_illumina_filter/

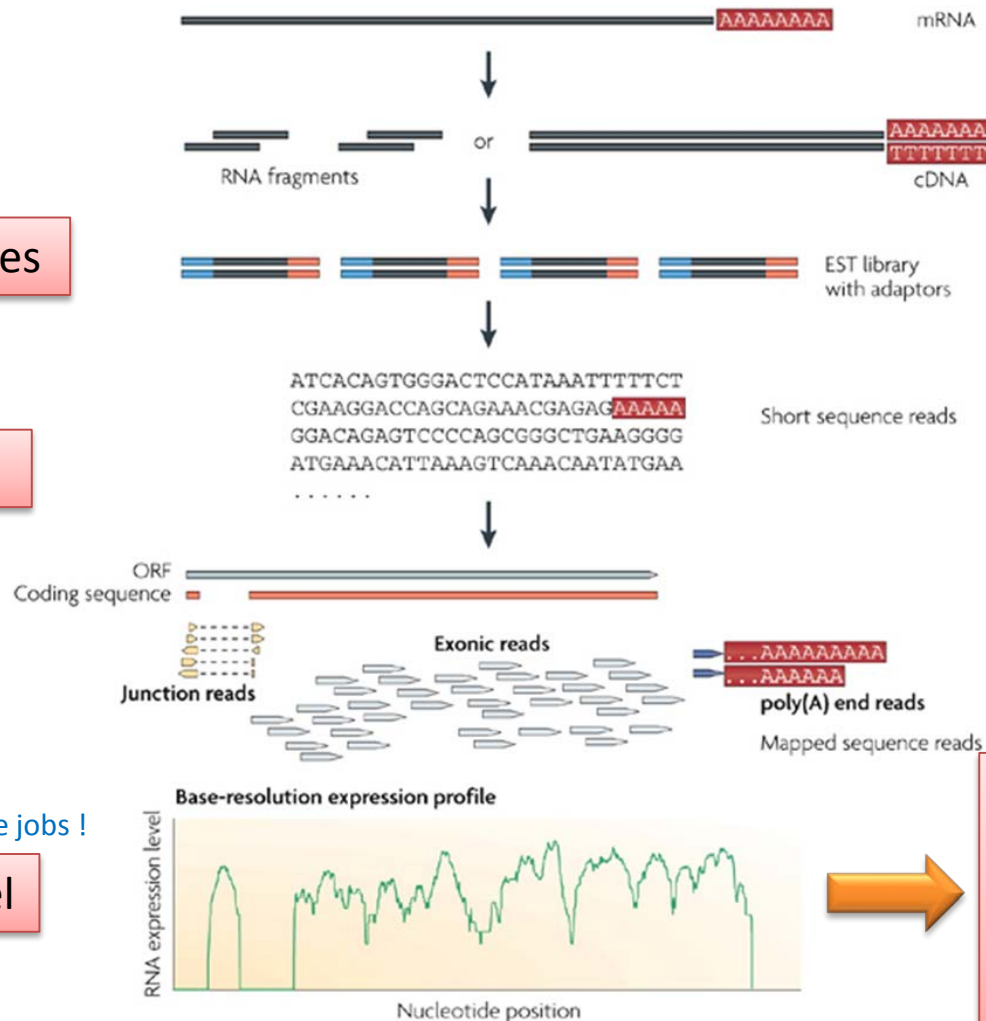
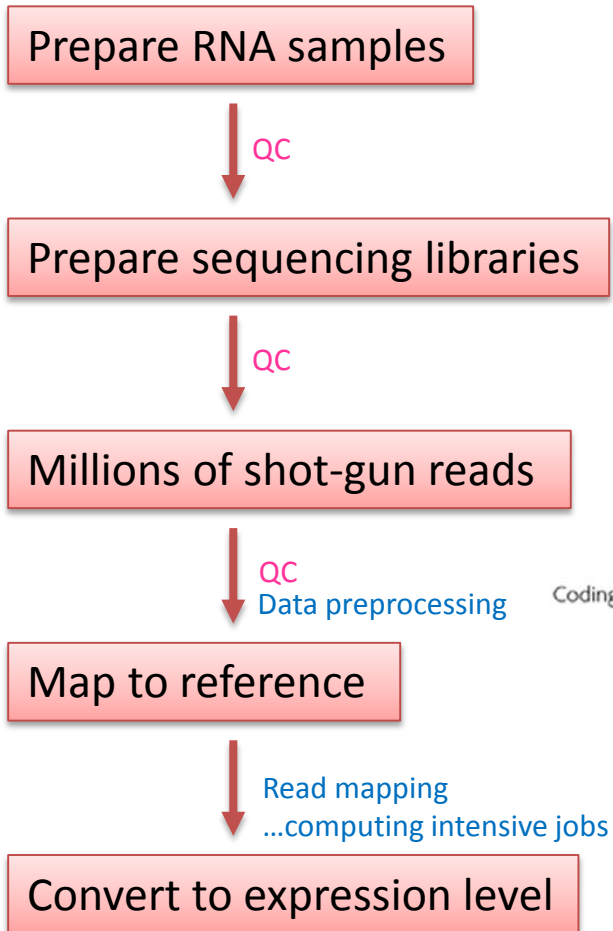
Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%



http://en.wikipedia.org/wiki/Phred_quality_score

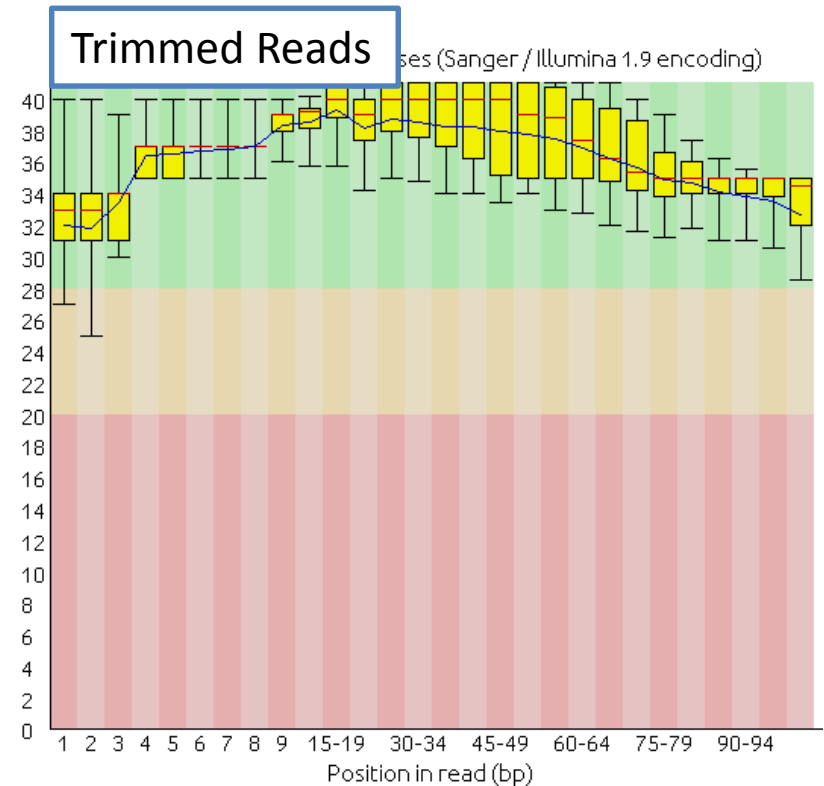
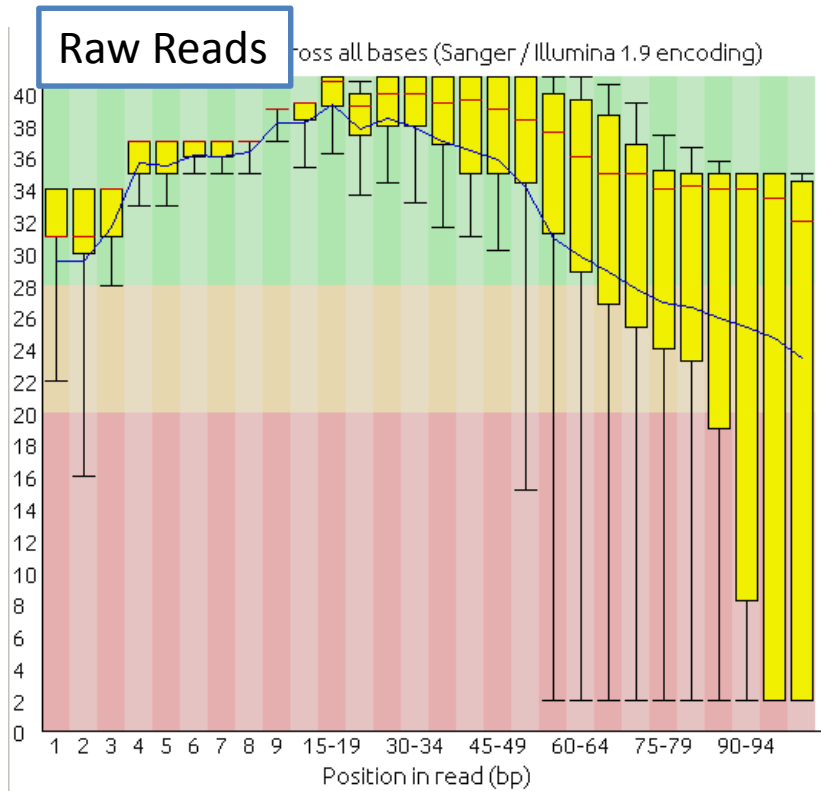
A typical RNA-Seq experiment



Intensive analysis to Interpret Biological Meanings

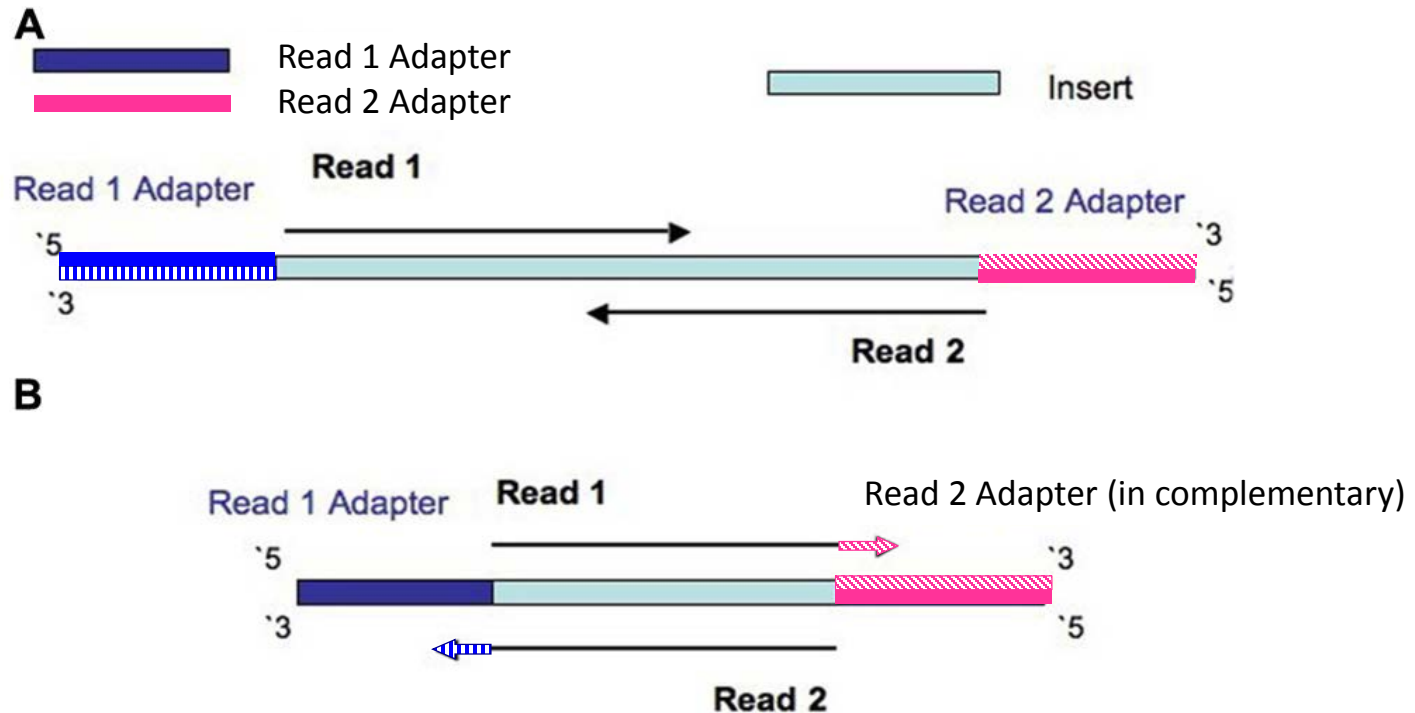
Read preprocessing

- Trimming: by base quality score

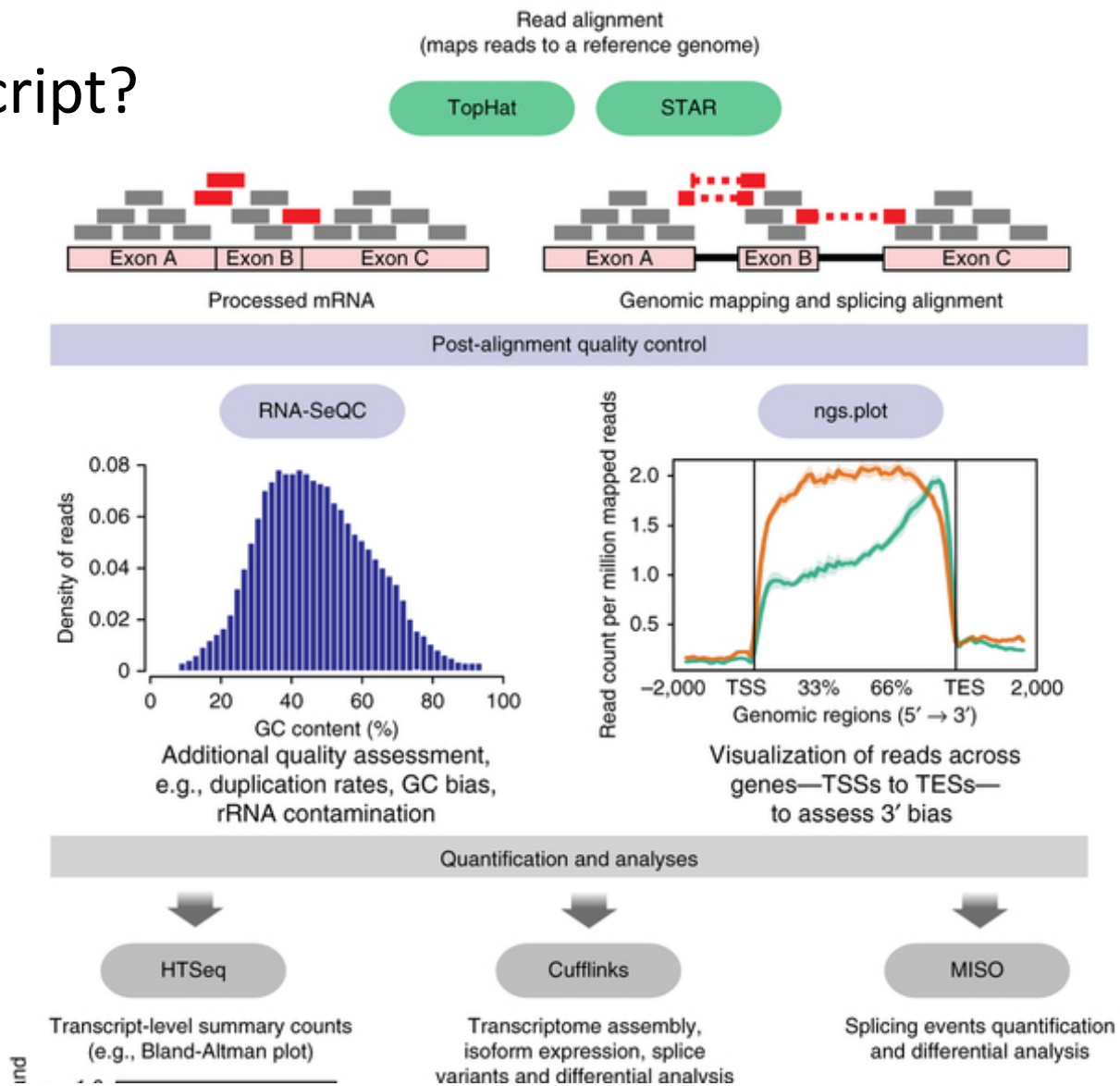
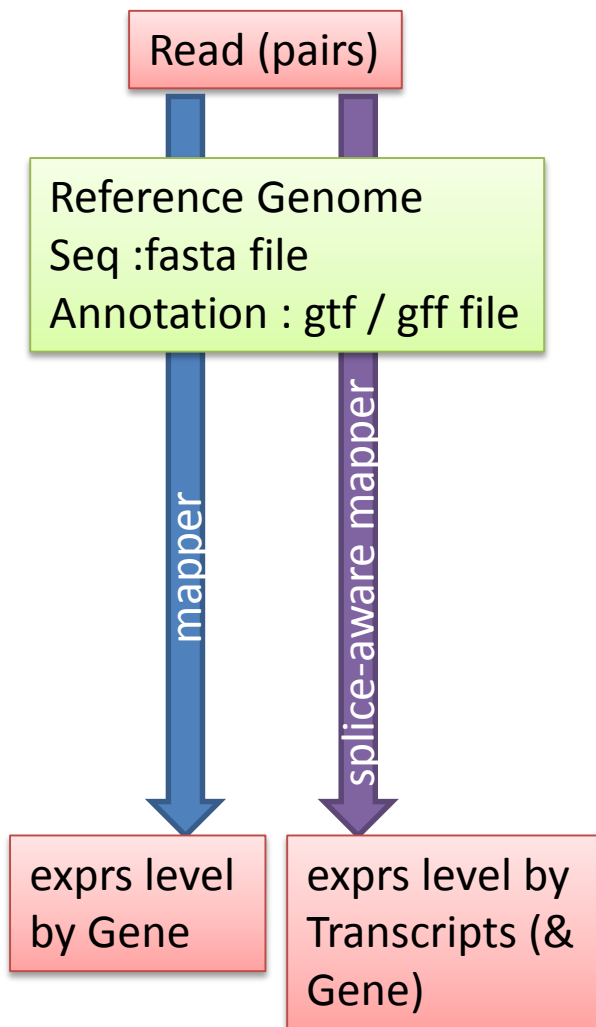


Read preprocessing

- Trimming: adapter contamination



Expression Level by Gene or by Transcript?



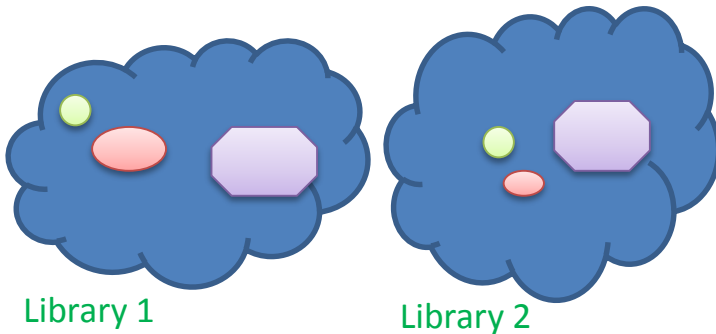
Other issues

- Stranded or not?
- PolyA tailed or rRNA depletion?
- Special protocols that need extra bioinformatical works?
- Trimmed read length? Low complexity repeats? Other sources of contamination?
- Have reference genome?
- Novel transcripts?

The Normalized Expression Level

- Between sample:

Library Size Effect



- Within sample:

length effect



- Count the mapped read number, normalized to **library size**

cpm: count per million reads

- Count the mapped read number, normalized to BOTH **library size** and **(target seq) length**

- ✓ **TPM**: transcripts per million reads
- ✓ **RSEM**: RNA-Seq by Expectation-Maximization
- ✓ **RPKM**: reads Per kilobase of exon per million mapped reads
- ✓ **FPKM**: fragments per kilobase of exon per million fragments mapped

The art of Normalization

- Borrowing information across genes to get a better estimate.
- **Count-based** model
 - edgeR, DESeq etc.
 - Use “read count” (or estimated count from RSEM) and enforced a normalization model to fit data to the statistic assumption
 - Want to retain statistical power
- Programs like SAMSeq (rank-based model, only applicable for large replicates) and limma are fine with continuous values (like FPKM). Limma takes more cares about weak mean-variance relationship (stabilizing variation).




The Usage


Demo: <http://molas.iis.sinica.edu.tw/grch38/>


多重體學線上分析平台

Multi-Omics onLine Annotation System (MOLAS)



MOLAS About MOLAS Browse Projects New Submission Check Submitted jobs

 demo
Human, grch38

 demo
Mouse, grcm38

Upload expressed profiling in FPKM in tab file. Example dataset for download:
For grch38, grcm38

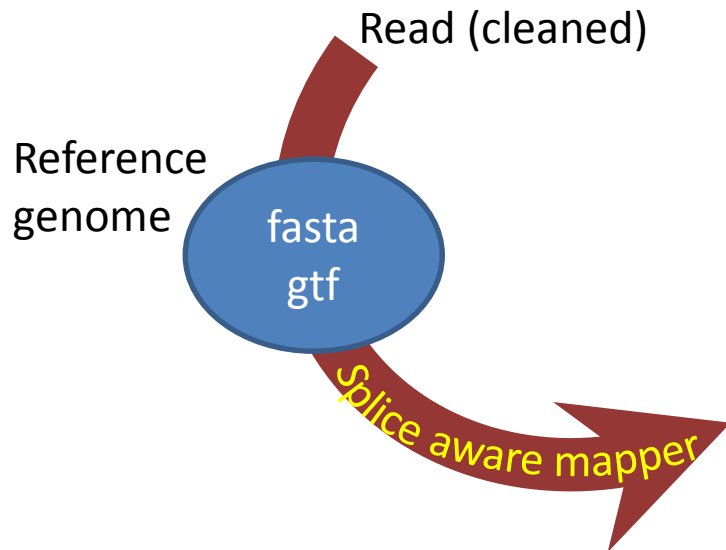
Important: Please read this before submission

To view and analyse your RNASeq experiment

All you need is an expression file

Input file

- A tab-delimited text file generated by other software (e.g. cufflink, EdgeR, RSEM) in ensembl transcript id (grch38 and grcm38)



#tracking_id	GA120-2_0	GA120-3_0
ENST00000591062	0	0.159246
ENST00000376259	0	3.96794
ENST00000235878	0.287651	0
ENST00000299596	0.0300576	0.0146675
ENST00000625158	6.08204	7.03465
ENST00000321949	4.24507	4.28616
ENST00000258484	0	6.00768
ENST00000625157	0.0134854	0.00783917
ENST00000321944	6.44635	5.25123
ENST00000321945	0.907242	1.13444

GTF: the Gene Transfer Format

```
1 ensembl_havana transcript 4344146 4360314 . - . gene_id "ENSMUSG00000025900"; gene_version "6"; transcript_id "ENSMUST00000027032"; transcript_version "5"; gene_name "Rp1"; gene_source "ensembl_havana"; gene_biotype "protein_coding"; transcript_name "Rp1-001"; transcript_source "ensembl_havana"; transcript_biotype "protein_coding"; tag "CCDS"; ccds_id "CCDS14804";
```

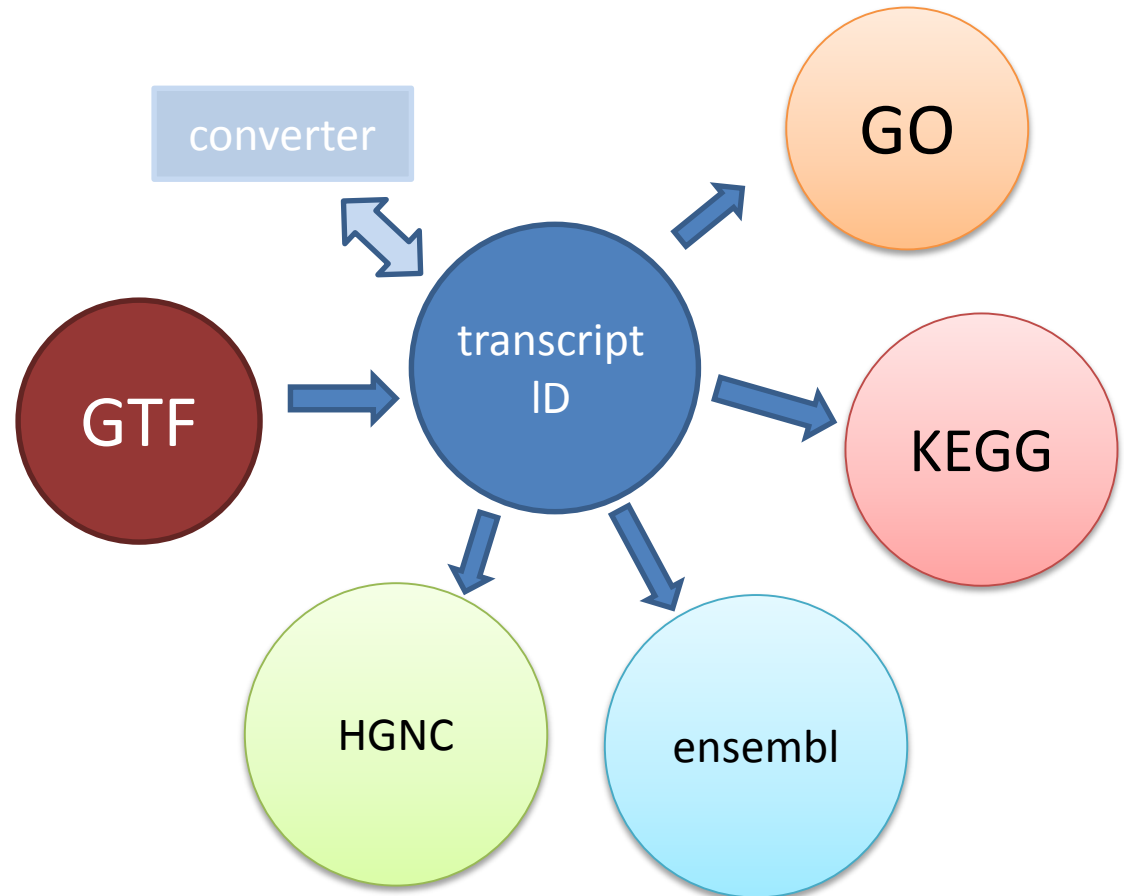
MOLAS compatible GTF



grch38
grch37



gcm38



New Submission



Upload expressed profiling in FPKM in tab file: Example dataset for download:
For grch38, grcm38

選擇檔案 未選擇任何檔案

***Important: Please read this before submission**

Submit Clear All

New Submission

There are 208244 transcripts annotated in human genome,ensembl grch38.78. In MOLAS, 197912 transcripts are in the database (transcripts of "small non-coding genes" are excluded. [Link to Details](#))
197523 data entries are found in the uploaded file,in which 14 ensembl transcriptid (0.01%, 14/197523) can not mapped to MOLAS database.
197509 MOLAS database transcript id are mapped (99.8%, calculated by mapped id / molas id: 197509/197912)

FPKM file top 5 lines :

#tracking_id	Sample_1	Sample_2	Sample_3	operation
				Sample_4
ENST00000380075	0	0	0.909464	1.0386
ENST00000380071	320.788	208.653	269.647	421.71
ENST00000380079	160.909	71.0702	63.7214	0
ENST00000563164	11.2517	15.5313	7.45358	14.1989
ENST00000563166	0	0	0	1.99288

Select library:

Present Selected:

Dataset	operation
Sample_1, Sample_2, Sample_3, Sample_4	<input type="radio"/> modify <input type="radio"/> delete

Selecting Dataset:

Sample_1 Sample_2 Sample_3 Sample_4

Update
Reset

Create New Project (Provide a static link for submission revised and shared for 12 months)

Just a try without Project creation (Just a dynamic link available for a week)

Clear All

Project Profile



This project is a transcriptome study on grch38 reference genome (transcripts #:197523,library#:2)

Project Info

Project Name (limit to 50 words)


Brief on this Project :

Upload an website logo (image file in jpg,gif,or png format)

未選擇任何檔案




Name of Sub-directory: 

Contact E-mail as Account: 

Password: 

Open to Public:

Yes

No share this project data to my friends with this secret word: 

Deployment Success

About MOLAS

Browse Projects

New Submission

Check Submitted jobs

Dear User:

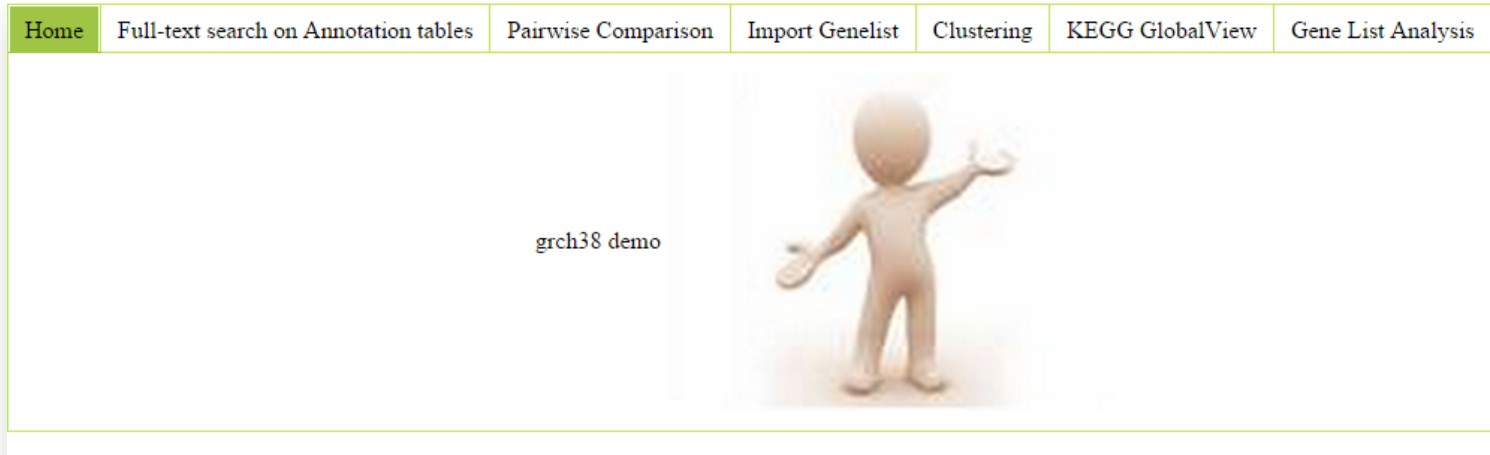
You have completed the submission. There are 8 libraries in your submission.
The whole system will be ready few minutes later after data deployment.
Please check the website below to start your journey on data analysis.

<http://molas.iis.sinica.edu.tw/grch38> - **Data Deployment Success!**

Thanks for your using our platform to deep your research.

MOLAS administrator

Browse project and



http://molas.iis.sinica.edu.tw/human_grch38_demo/

Enrichment Analysis

1 Enter contig(s) (one id per line) example
 Q2184373.1.1040
 L04143.1.1702
 S06143.0323

Or upload from file:
 選擇文件 本地文件 [download example](#)

Save to file: L370402556143 done!

2 Select program:
 * KEGG
 GO

send reset

Total 3 input contigs: 6k2 used: 6k1 excluded

Show 10 entries Search: CSV PDF

Pathway name	Numbers frequency	Background frequency	P-value	Contig associated to the term
Lipid acid metabolism	1 out of 2 numbers	3 out of 3263 numbers	0.00763	comp100076_c0_seq1
Parkinson's disease	1 out of 2 numbers	52 out of 3263 numbers	0.05527	comp100076_c0_seq1
Oxidative phosphorylation	1 out of 2 numbers	96 out of 3263 numbers	0.05764	comp100076_c0_seq1
Metabolic pathways	2 out of 2 numbers	819 out of 3263 numbers	0.00219	comp100076_c0_seq1

Showing 1 to 4 of 4 entries

Functional Enrichment

Text Search

Home Full-text search on Annotation tables Sequence Search / BLAST Library Compare Logout

Enrichment Analysis Clustering KEGG GlobalView

Fuzzy search

Enter your keywords:

Search DB: blastNR KEGG GO pfAM
 Show data: SignalP tmHMM

send reset

Show 10 entries Search: FASTA

contig	protein_id	SignalP	tmHMM	description	species
comp100076_c0_seq1	m180609	N	Y	"A Chain A, Crystal Structure of Anopheles Gambiae Adenosine Kinase in Complex With P1-P4-Diadenosine(3') Tetraphosphate"	Anopheles gambiae

Dynamic comparison like DDD

1. Select library (any tree)
 PL HPT HEMO
 HC 0 hr 3 hr 6 hr 24 hr 48 hr
 Hpt 0 hr 3 hr 6 hr 24 hr 48 hr

2. Select group:
 PoolA * PoolB (# of overlaps)
 PoolA FPKM: PoolB FPKM:
 10 select fold change:
 PoolA Expressed only (FPKM value in PoolB=0 # contigs)
 PoolA FPKM: PoolB FPKM:
 The number of selected contigs = 508

3. Select Analytic Approach:
 Show Contig List
 Functional enrich
 GO
 Total 508 input contigs: 6k276 used: 6k1230 excluded

send reset Show 10 entries Search: CSV PDF

Pathway name	Numbers frequency	Background frequency	P-value	Contig associated to the term
Ribosome	77 out of 247 numbers	120 out of 2866 numbers	1.53e-56	comp101446_c0_seq1 comp116750_c0_seq1
Proteasome	19 out of 247 numbers	33 out of 2866 numbers	7.36e-13	comp105011_c0_seq1 comp107926_c0_seq1
Pathogenic Escherichia coli infection	7 out of 247 numbers	19 out of 2866 numbers	0.00063	comp08844_c0_seq1 comp03366_c0_seq1

Pairwise Comparison

Pathway View

Home Full-text search on Annotation tables Sequence Search / BLAST Enrichment Analysis KEGG GlobalView

KEGG GlobalView

Network Hierarchy (26179)
 KEGG Database (320214238)
 KEGG Module (1381)
 Profile Hierarchy: metabolism (948)
 Biosynthesis (229)
 Citrate Metabolism (225)
 Lipid Metabolism (134)
 Lipid Biosynthesis (134)
 Phenylalanine (10)
 Protein Metabolism (10)
 Citric Acid Cycle (10)

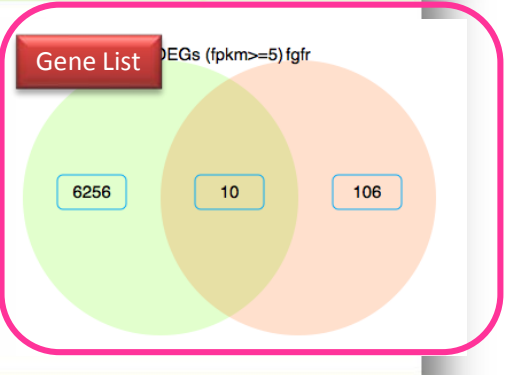
Show 10 entries Search: CSV PDF

Pathway id	Pathway name	Reactions	Frequency
00010	Glycolysis / Gluconeogenesis	26 / 80	24 / 84
00019	Citrate cycle (TCA cycle)	21 / 57	21 / 57
00043	Phenylalanine pathway	1 / 7	1 / 7
00044	Phenylalanine degradation	1 / 7	1 / 7
00051	Fatty acid biosynthesis: microorganisms	22 / 76	22 / 76
00052	Gluconeogenesis	18 / 64	18 / 64
00053	Aspartate and asparagine metabolism	1 / 7	1 / 7
00054	Fatty acid biosynthesis	1 / 30	1 / 30
00055	Fatty acid elongation	1 / 21	1 / 21
00071	Fatty acid metabolism	31 / 49	31 / 49

Showing 1 to 10 of 337 entries

Full-text search on Annotation tables Pairwise Comparison Import Genelist Clustering KEGG GlobalView Gene List Analysis

grch38 demo



Clustering

Clustering Result

1 2 3 4

5 6

Clustering

transcripts info 00326324: FGFR1-008

Transcript: FGFR1-008

Transcript Central View

One of FGFR1 transcript: ENS00000326324

FGFR1-008	fibroblast growth factor receptor 1	Location: chr8: 36,412,732-36,468,616 reverse strand.
Ensembl ID	Transcript_Biotype	
ENS00000326324	protein_coding	

This transcript is derived from gene ENSG00000077782 which encodes 41 transcript(s)

Gene info

Fuzzy Search

Home **Full-text search on Annotation tables** Pairwise Comparison Import Genelist Clustering KEGG GlobalView Gene List Analysis

Fuzzy search

Enter your keywords:

Search : Genesymbol description KEGG

Showing 1 to 10 of 116 entries (filtered from 197,443 total entries) Show entries Search:

Transcriptid	Genesymbol	Description	KEGG
ENST00000229395	FGFR1OP2	FGFR1 oncogene partner 2	
ENST00000260795	FGFR3	fibroblast growth factor receptor 3	/ fibroblast growth factor receptor 3 [EC:]
ENST00000264748	FGFRL1	fibroblast growth factor receptor-like 1	
ENST00000292414	FGFR4	fibroblast growth factor receptor 4	/ fibroblast growth factor receptor 4 [EC:]
ENST00000326324	FGFR1	fibroblast growth factor receptor 1	/ fibroblast growth factor receptor 1 [EC:]
ENST00000327214	FGFR1OP2	FGFR1 oncogene partner 2	
ENST00000335922	FGFR1	fibroblast growth factor receptor 1	/ fibroblast growth factor receptor 1 [EC:]
ENST00000336553	FGFR2	fibroblast growth factor receptor 2	/ fibroblast growth factor receptor 2 [EC:]
ENST00000340107	FGFR3	fibroblast growth factor receptor 3	/ fibroblast growth factor receptor 3 [EC:]
ENST00000341462	FGFR1	fibroblast growth factor receptor 1	/ fibroblast growth factor receptor 1 [EC:]

Showing 1 to 10 of 116 entries (filtered from 197,443 total entries) ◀ Previous Next ▶

ENST00000326324: FGFR1-008

Transcript Central View

Transcript: FGFR1-008

Transcript Central View

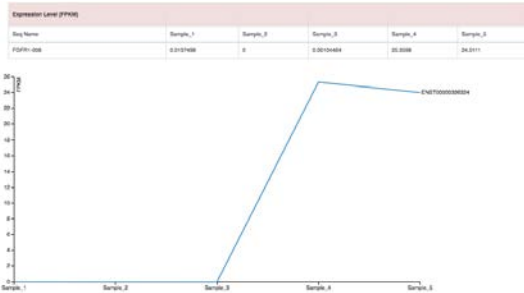
One of FGFR1 transcript: ENST00000326324

FGFR1-008 fibroblast growth factor receptor 1	Location: chr8 : 38,412,732-38,468,616 reverse strand.
Ensembl ID ENST00000326324	Transcript_Biotype protein_coding

Short cut to the Gene Central View

descriptions about this transcript

This transcript is derived from gene [ENSG00000077782](#) which encodes **41** transcript(s)



The expression level (the user uploaded data)

External Links to get more info

External Links for further info

Link to ensembl for

- the translated product ENSP00000327229 of transcript FGFR1-008/ENST00000326324
- transcript summary of FGFR1-008/ENST00000326324
- gene summary of FGFR1/ENSG00000077782 or Viewing the gene / transcript structure

Link to NCBI for

- gene centered view of FGFR1 in Entrez Gene (Gene ID:)
- the transcript cluster of human FGFR1 in UniGene

GO Information

GO information		
Biological Process		
	GO:0010518	positive regulation of phospholipase activity
	GO:0010863	positive regulation of phospholipase C activity
	GO:0014088	positive regulation of phosphatidylinositol 3-kinase signaling
	GO:0001501	skeletal system development
	GO:0018477	cell migration
	GO:0000185	MAPK cascade
	GO:0001784	neuron migration
	GO:0018108	peptidyl-tyrosine phosphorylation
	GO:0002376	immune system process
	GO:0034641	cellular nitrogen compound metabolic process
	GO:0038095	Fc-epsilon receptor signaling pathway
	GO:0040011	locomotion

GO terms tagged in this gene

Gene Central View

Gene: FGFR1

Gene Central View

FGFR1 fibroblast growth factor receptor 1	
Ensembl ID	Gene Biotype
ENSG0000077782	protein_coding
Synonym/ prev Symbol	chromosome location
	chr8: 38,411,138-38,468,834 reverse strand.

descriptions about this gene

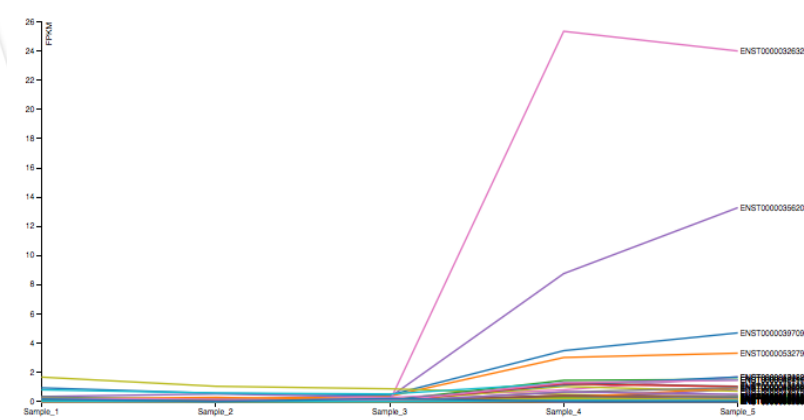
This gene is known with 41 transcript(s).

The expression level of all the possible transcripts:

Expression Level (FPKM)						
Seq Name Transcript ID	Transcript Biotype	Sample_1	Sample_2	Sample_3	Sample_4	Sample_5
ENST00000425967 [Ensembl]	protein_coding	0	0	0	0.0404075	0.378136
ENST00000334333 [Ensembl]	antisense	0.0255699				

The expression level of all transcripts of this gene (the user uploaded data)

ENST00000434187 [Ensembl]	protein_coding	0				
ENST00000531196 [Ensembl]	nonsense_mediated_decay	0				
ENST00000526688 [Ensembl]	processed_transcript	0				
ENST00000504508 [Ensembl]	retained_intron	0.152759				
ENST00000326324 [Ensembl]	protein_coding	0.0137498	0	0.00104404	25.3598	2
ENST00000530658 [Ensembl]	protein_coding	0	0	0.0290441	0	0
			0	0	0.251281	0



Human GRCh38/hg38 8:38,411,466..38,469,162

External Links for further info

- Link to Ensembl for gene summary of human FGFR1 ENSG0000077782 or Viewing the gene / transcript structure
- Link to NCBI for gene centered view of human FGFR1 in Entrez Gene (Gene ID), or the transcript cluster in UniGene
- Link to HGNC for retrieving gene FGFR1 (HGNC ID: 3688) related info.
- Link to The Human Protein Atlas (version 13) for the
 - normal tissue <http://www.proteinatlas.org/ENTP011138-FGFR1>
 - cancerous tissue <http://www.proteinatlas.org/ENTP011138-FGFR1>

Genome browser & Gene modeling

Neighboring Genes

Drop down the list to select a range. Given this gene, its downstream (e.g. - 1Kb) and upstream (e.g. + 1Kb) flanking genes are extracted as a list.

10 Kb

Neighboring Genes (-10000 bps ~ +10000 bps)					Save list
1	ENSG00000165046 [Ensembl]	LETM2	protein_coding	8: 38386207 - 38409527, forward strand	
2	ENSG00000254961 [Ensembl]	RP11-350N15.3	antisense	8: 38400536 - 38401683, reverse strand	
3	ENSG00000272159 [Ensembl]	RP11-350N15.6	antisense	8: 38408048 - 38408742, reverse strand	
4	ENSG0000077782 [Ensembl]	FGFR1	protein_coding	8: 38411138 - 38468834, reverse strand	
5	ENSG00000255201 [Ensembl]	RP11-350N15.4	antisense	8: 38421889 - 38428096, forward strand	
6	ENSG00000239218 [Ensembl]	RPS20P22	transcript_processed_pseudogene	8: 38434347 - 38439664, reverse strand	

Pairwise Comparison

Pairwise Comparison

1. Select the data for comparison
Present grouping:

Pool	Dataset
pool a:	sample_1,sample_2
pool b:	sample_3,sample_4,sample_5

2. Set the comparing scheme

PoolA expression level (fpkm)

PoolB expression level (fpkm)

fold change cutoff:

expression pattern:

The number of selected transcriptid = 5383

3. Select Analytic Approach:

- Show Gene List
- Calculate GO term enrichment
- Draw heatmap with 2D clustering

Clustering

If some samples have similar properties, clustering can help group them together and perform gene expression profile analysis.

Home Full-text search on Annotation tables Library Compare Enrichment Analysis **Clustering**

KEGG GlobalView

Clustering [View Clustering Result](#)

Create an Analysis

Analysis Name:

Description:

Present grouping:

Group Name	Dataset	operation	
TX	Sample5_FPKM	<input type="radio"/> modify	<input type="radio"/> delete
T0	Sample4_FPKM,Sample8_FPKM	<input type="radio"/> modify	<input type="radio"/> delete
T1	Sample7_FPKM	<input type="radio"/> modify	<input type="radio"/> delete
T2	Sample1_FPKM,Sample2_FPKM,Sample6_FPKM	<input type="radio"/> modify	<input type="radio"/> delete
T3	Sample3_FPKM	<input type="radio"/> modify	<input type="radio"/> delete

Add new a group to this Analysis

Group Name:

Selecting Dataset:

Sample1_FPKM Sample2_FPKM Sample3_FPKM Sample4_FPKM [Add to List](#)

Sample5_FPKM Sample6_FPKM Sample7_FPKM Sample8_FPKM [Reset](#)

[Add this Analysis Scheme](#) [Clear All](#)

Clustering Results

Clustering Result

View by group name
 View by sample name

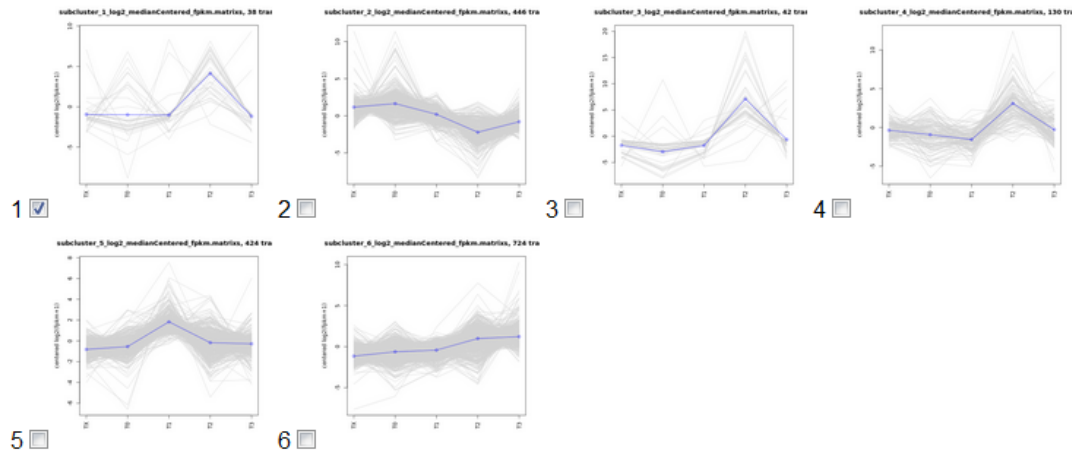
[view one](#)

time: 2014-01-03 16:36:51 Job note:

TX: Sample5_FPKM
T0: Sample4_FPKM, Sample8_FPKM
Group: T1: Sample7_FPKM
T2: Sample1_FPKM, Sample2_FPKM, Sample6_FPKM
T3: Sample3_FPKM

p: 0.001

fold: 4 k: 6



Select Analytic Approach:

- Show Gene List
- Functional enrichment KEGG
- GO

[send](#) [reset](#)

KEGG Pathway

3. Select Analytic Approach:

- Show Gene List
- Calculate GO term enrichment
- Calculate KEGG pathway enrichment
- Draw heatmap with 2D clustering

send reset

Total:2348 input gene symbol. hit:843 used. nohit:1505 excluded.

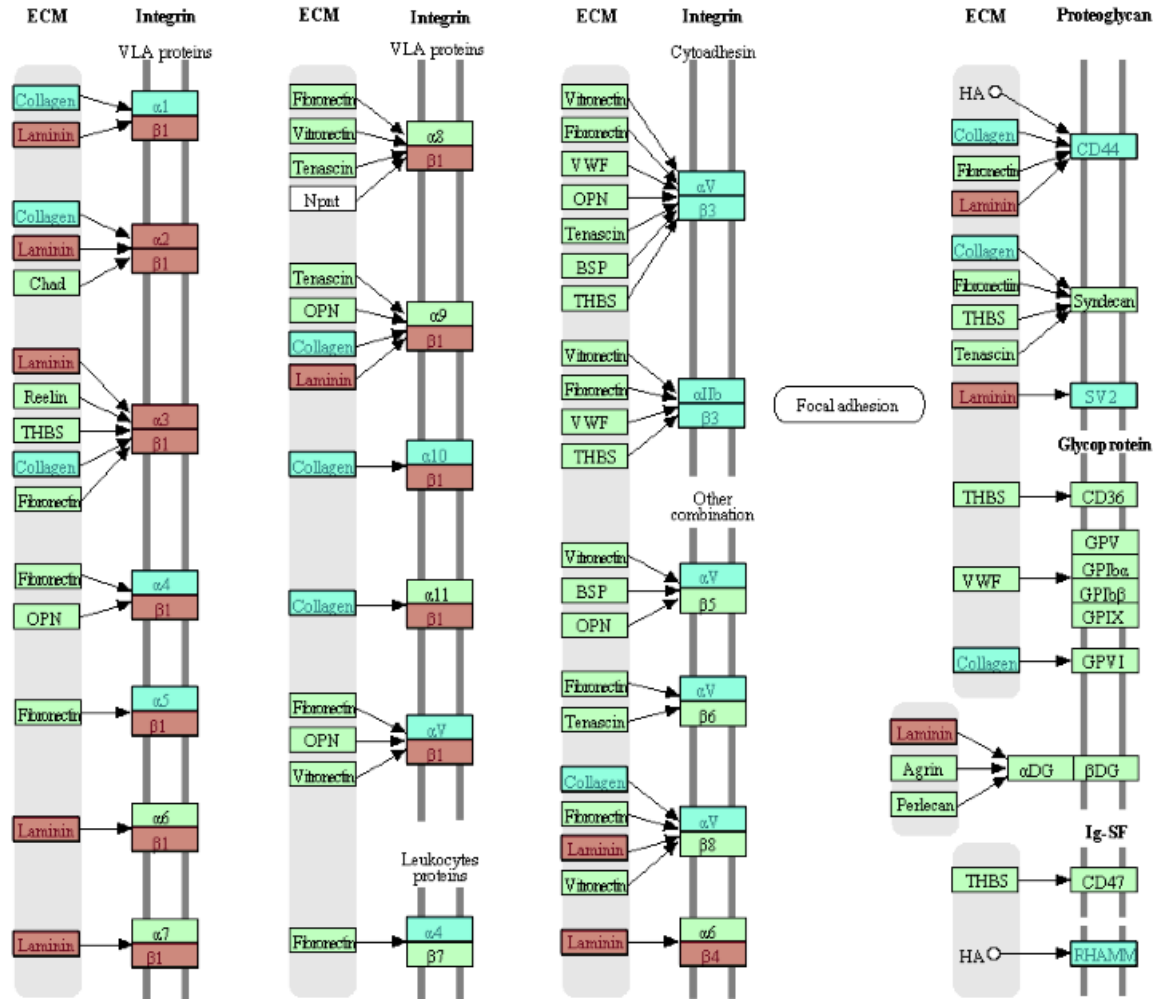
Show entries

Search:

Pathway name	Knnumbers frequency	Background frequency	P-value	transcriptid associated to the term
PI3K-Akt signaling pathway	55 out of 588 knumbers	243 out of 8706 knumbers	4.50e-16	ENST00000205386 ENST00000222139
Pathways in cancer	62 out of 588 knumbers	322 out of 8706 knumbers	1.89e-14	ENST00000205386 ENST00000222462
ECM-receptor interaction	25 out of 588 knumbers	64 out of 8706 knumbers	1.10e-13	ENST00000205386 ENST00000225964
Transcriptional misregulation in cancers	40 out of 588 knumbers	158 out of 8706 knumbers	1.10e-13	ENST00000222462 ENST00000247182
Focal adhesion	39 out of 588 knumbers	151 out of 8706 knumbers	1.12e-13	ENST00000205386 ENST00000225964
Regulation of actin cytoskeleton	40 out of 588 knumbers	162 out of 8706 knumbers	2.67e-13	ENST00000257290 ENST00000261799

KEGG Pathway

ECM-RECEPTOR INTERACTION



Enrichment Analysis

Insert a list of interesting genes to see which pathway they are involved.

Home Full-text search on Annotation tables Library Compare **Enrichment Analysis** Clustering

KEGG GlobalView

Enrichment Analysis

1. Enter genesymbol:(one id per line)

TRPA1
VIL1
VTCN1
WT1
ZFP57

Or upload from file:
 No file chosen [download example](#)

Save to file:1389348774137 done!

2. Select Analytic Approach:

KEGG
 GO

KEGG Global View

KEGG Global View provide an overview picture of KEGG pathway of human (hg19) and mouse (mm10) organisms. You can investigate specific metabolic pathway by exploring each category.

Home Full-text search on Annotation tables Library Compare Enrichment Analysis Clustering

KEGG GlobalView

BRITE hierarchies (33338)

Network hierarchy (22563)

- [KEGG Orthology \(KO\) \(20622\)](#)
- [KEGG modules \(1941\)](#)

Protein families: metabolism (5182)

- [Enzymes \(3786\)](#)
- [Protein kinases \(484\)](#)
- [Peptidases \(494\)](#)
- [Glycosyltransferases \(214\)](#)
- [Lipid biosynthesis proteins \(73\)](#)
- [Prenyltransferases \(16\)](#)
- [Amino acid related enzymes \(60\)](#)
- [Cytochrome P450 \(55\)](#)

Protein families: genetic information processing (2696)

- [Transcription factors \(1046\)](#)
- [Transcription Machinery \(280\)](#)
- [Spliceosome \(492\)](#)
- [Ribosome \(199\)](#)
- [Ribosome biogenesis \(9\)](#)
- [Transfer RNA biogenesis \(203\)](#)
- [Translation factors \(51\)](#)
- [Chaperones and folding catalysts \(44\)](#)
- [SNAREs \(43\)](#)
- [Ubiquitin system \(283\)](#)
- [Proteasome \(21\)](#)
- [DNA replication proteins \(25\)](#)

Protein families: signaling and cellular processes (2897)

- [Transporters \(371\)](#)
- [Secretion system proteins \(17\)](#)
- [G Protein-Coupled Receptors \(778\)](#)
- [Enzyme-linked receptors \(66\)](#)
- [Cytokine receptors \(89\)](#)
- [Nuclear receptors \(48\)](#)
- [Ion Channels \(284\)](#)
- [GTP-binding proteins \(184\)](#)
- [Cytokines \(12\)](#)
- [CD molecules \(794\)](#)
- [Proteoglycans \(15\)](#)
- [Heparan sulfate/heparin binding proteins \(186\)](#)
- [Glycan Binding Proteins \(53\)](#)

Show

Pathway name	frequency
Glycolysis / Gluconeogenesis	36 / 90
Citrate cycle (TCA cycle)	22 / 54
Pentose phosphate pathway	18 / 57
Pentose and glucuronate interconversions	12 / 56
Fructose and mannose metabolism	18 / 79
Galactose metabolism	21 / 64
Ascorbate and aldarate metabolism	7 / 37
Fatty acid biosynthesis	5 / 30
Fatty acid elongation	18 / 21
Fatty acid metabolism	29 / 49

Showing 1 to 10 of 317 entries

◀ Previous Next ▶

What to do if you have no replicates?

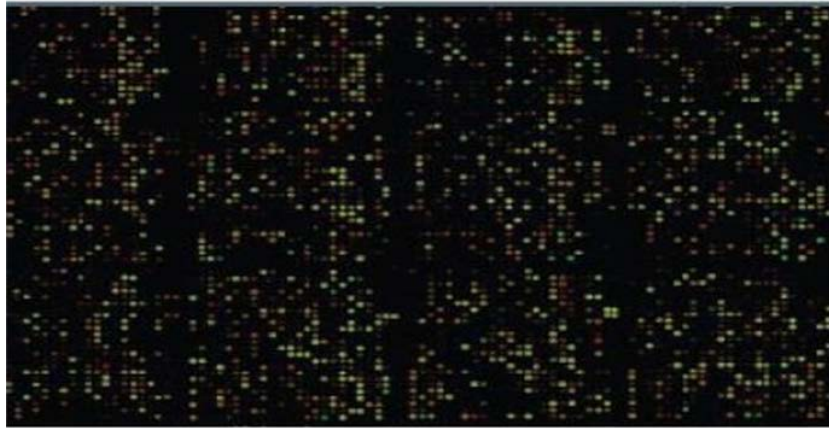
Suggestions from edgeR authors

- **Be satisfied with a descriptive analysis**, that might include an MDS plot and an analysis of fold changes. Do not attempt a significance analysis. This may be the best advice.
- Simply **pick a reasonable dispersion value**, based on your experience with similar data, and use that for DE detection
 - In edgeR (empirically):
 - 0.4 human data (genetically unidentical)
 - 0.1 for genetically identical model organisms
 - 0.01 for technical replicates
- **estimate dispersion** from dataset reducing one (less critical) experiment factor
- **estimate dispersion** from a sizeable number of control transcripts that should not be DE if there exists

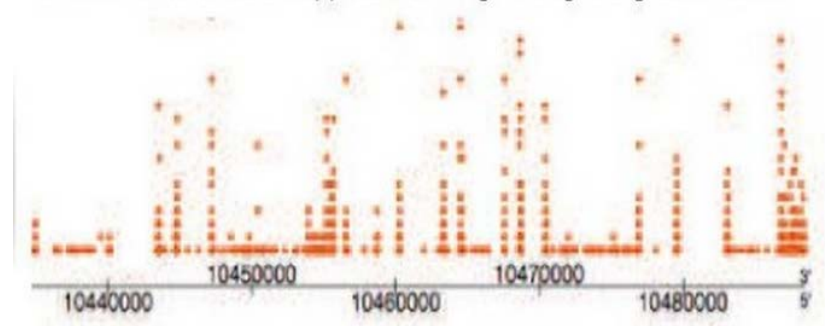
edgeR paper <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2796818/>

menu <http://www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf>

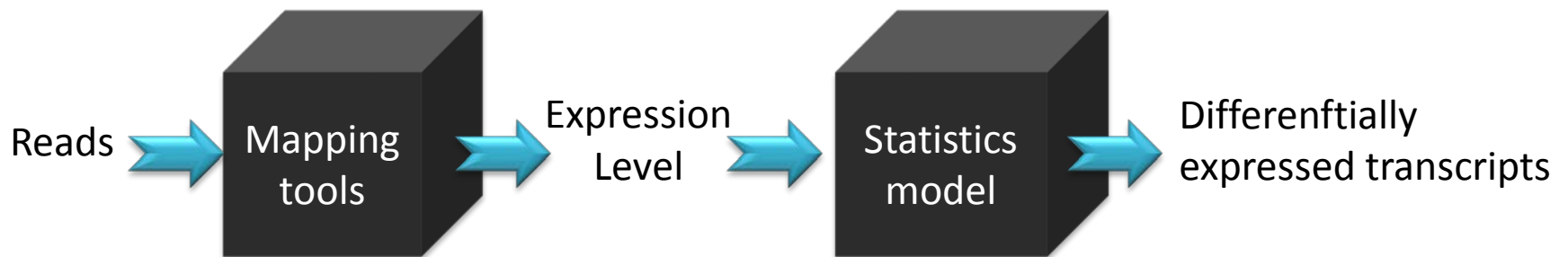
Analog signal vs Digital signal



A dot means a read mapped to the region beginning at the base



<http://www.slideshare.net/ueb52/uebuat-bioinformatics-course-session-23-vhir-barcelona>



Limitations

- Assumption of “Uniformity” of all expressed transcripts may not always be true
- Uncertain problems in mapping
 - Transcripts length issue
 - Redundant seq in genome
 - Reference is never a perfect match to the actual biological source of RNA being sequenced
- Reference & no Reference
- Lag in analytic tools.
- No single robotic analysis scheme fits all kinds of needs
- Cost !!





Thanks for your Attention