



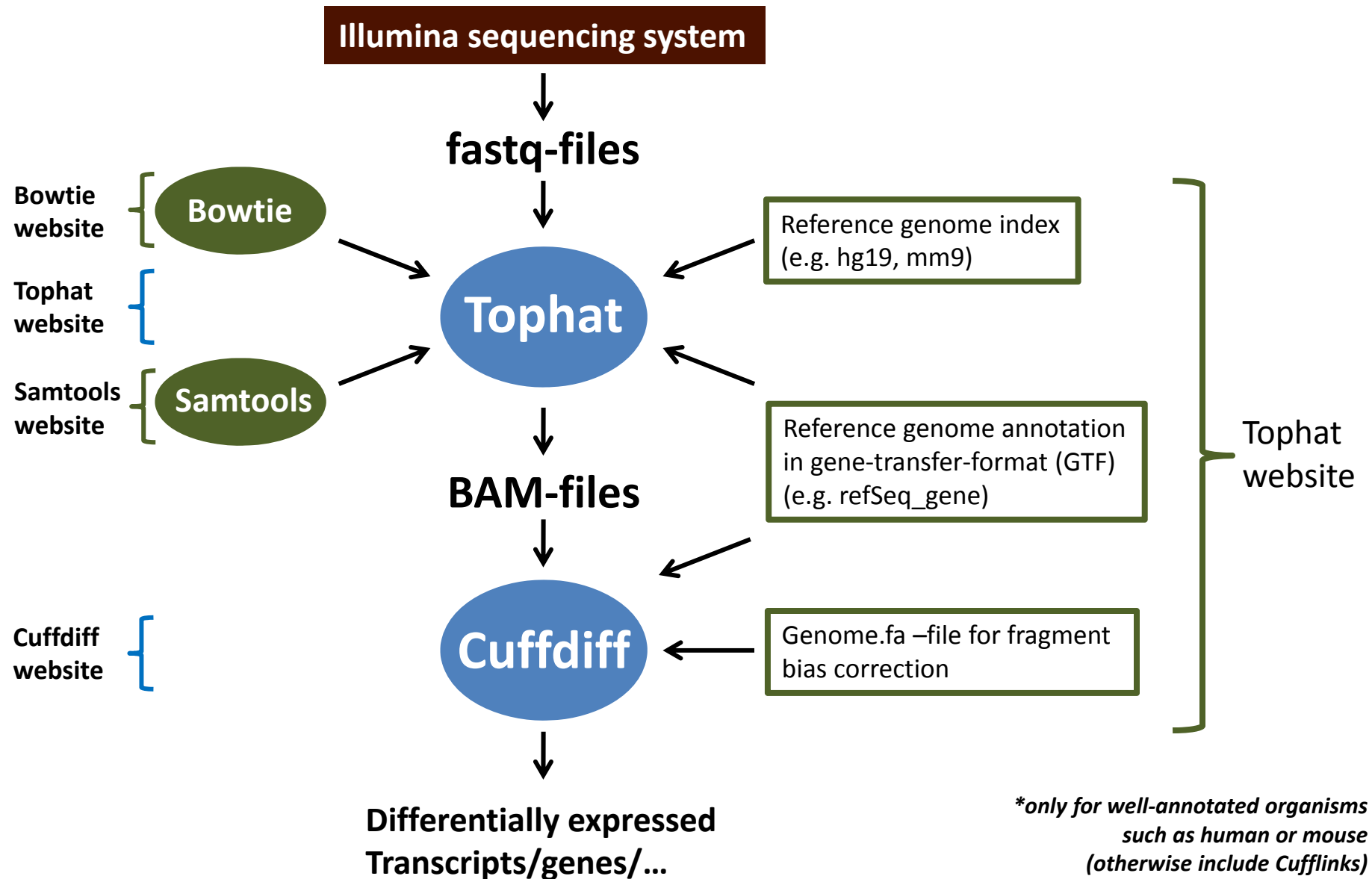
Icahn
School of
Medicine at
**Mount
Sinai**



Analysis of RNASeq data using Tophat and Cufflinks to identify differentially expressed genes

Jens Hansen (Ravi Iyengar lab)

Analysis pipeline*



FASTQ-files

Sequence identifier + optional description

Base-pair sequence of read

Base-pair quality values



```
@11V6WR1:102:C0NL8ACXX:2:1101:1131:2165 1:N:0:CGATGT
GAAAGATTCCTTGGGGCACATCCCGGGTGGGGTATGGGGCAATATTGCTTCTTTAGTCTCCCCCTCCCTCCTCTCAGNAGGCTTGGGGTCCAAGAA
+
@@@FFDEEHHHHHIIJJGDHGHJJJCFHHG7@FHIGGAHIHCAHHHHHHFFFFFDCCDDDD?CCCCBCDCCCCC#+2<@BCD<AB>BA:>@A3
@11V6WR1:102:C0NL8ACXX:2:1101:1231:2168 1:N:0:CGATGT
CATGAGCGGGGCTCAGAGCTTTGAAAGGGAGGGCACAGAACTAGGCAGAGCCAAATTGTCTTGTTTCAGATGAGACCAGGAAAACTATTTAGGACGA
+
CCCCFFFFHHHHHJJGIJJJJJHFFHHJJHJJHHHHHCFFFEEDDDDDDDDDDDDDDDDDDCDCDDDDCD>@CDCDDDBDCBDDDDDEDDDCDDDD
@11V6WR1:102:C0NL8ACXX:2:1101:1191:2235 1:N:0:CGATGT
GTGGGTACAGATGTGTAGGAATGCTAGGTGTGGTTGGTTTATTCTAGCGTCACTATTATCAGGCCTAGTTGGCTTGATGTAGAGAAGGCAATGATTTTT
+
BBCFFFFFHHHHHEHHIIJJJJJJFHICGFHIHIJJ?GHDFIGGHIJJGI>HJIIHJHEGHHHIECEHEFCDBA>CEFEED@CABDBBCC35>CDC
```

Tophat considers quality values during alignment

and assumes Sanger format encoding (Phred+33) as default

(in case of different encoding the corresponding format needs to be added when calling tophat)

(see Wikipedia: FASTQ format for more details)

Run Tophat from the command line

Add software folders to path:

```
$ PATH=$PATH:/Software/samtools-0.1.7_x86_64-linux:/Software/bowtie2-2.1.0:/Software/tophat-2.0.8:  
$ export PATH
```

Move to folder with fastq-files:

```
$ cd /Data/Experiment/RNASeq_fastq
```

Unzip compressed files:

```
$ gunzip Condition1_sample1.fastq.gz  
OR  
$ tar -xvzf Condition1_sample1.tar.gz
```

Run tophat for each fastq-file:

```
$ tophat -o '/Data/Experiment/Tophat_out/Condition1_sample1'  
--GTF /Data/reference_genomes/Mus_musculus/UCSC/mm9/Annotation/Genes/genes.gtf  
--no-novel-juncs* -p8  
/Data/reference_genomes/Mus_musculus/UCSC/mm9/Sequence/Bowtie2Index/genome  
condition1_sample1.fastq
```

(everything is separated by one space)

*(*approach does not consider reads that span novel unknown splice junctions)*

(if not Sanger format quality encoding, the right quality needs to be added)

Run Cuffdiff from the command line

Add software folders to path:

```
$ PATH=$PATH:/Software/cufflinks
```

```
$ export PATH
```

Move to folder with tophat output:

```
$ cd /Data/Experiment/Tophat_out
```

Run cuffdiff:

```
$ cuffdiff
```

```
-o ./cuffdiff_results
```

```
-L "Treatment","Control"
```

```
-p8
```

```
--multi-read-correct
```

```
-b /Data/reference_genomes/Mus_musculus/UCSC/mm9/WholeGenomeFasta/genome.fa
```

```
--upper-quartile-norm
```

```
/Data/reference_genomes/Mus_musculus/UCSC/mm9/Annotation/Genes/genes.gtf
```

```
Treatment_1_accepted_hits.bam,Treatment_2_accepted_hits.bam,Treatment_3_accepted_hits.bam
```

```
Control_1_accepted_hits.bam,Control_2_accepted_hits.bam,Control_3_accepted_hits.bam
```

(everything is separated by one space)

Cuffdiff output files

Among the cuffdiff output files are:

isoform_exp.DIFF

gene_exp.DIFF

gene_exp.DIFF:

gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
Acan	chr7:86198368-86259985	Treatment	Control	OK	45.4983	5.95898	-2.93267	-4.10656	5.00E-05	0.011509	yes
Actg2	chr6:83462902-83486245	Treatment	Control	OK	93.172	277.423	1.57412	2.51017	5.00E-05	0.011509	yes
Adcy1	chr11:6963491-7078508	Treatment	Control	OK	6.15244	0.215077	-4.83824	-3.74947	5.00E-05	0.011509	yes
Aif1	chr17:35307936-35312946	Treatment	Control	OK	23.1482	2.20051	-3.39499	-2.88255	5.00E-05	0.011509	yes
Asb2	chr12:104559351-104594211	Treatment	Control	OK	4.47988	26.6658	2.57346	2.73042	5.00E-05	0.011509	yes
Atp10b	chr11:42963378-43075787	Treatment	Control	OK	0.103141	1.84327	4.15957	3.31941	5.00E-05	0.011509	yes
Atp6v0d2	chr4:19803984-19849713	Treatment	Control	OK	6.11358	0.167771	-5.18745	-3.19626	5.00E-05	0.011509	yes
B4galnt3	chr6:120153827-120244577	Treatment	Control	OK	0.333629	3.19525	3.25962	2.52311	5.00E-05	0.011509	yes