Cheatography

Experimental design Cheat Sheet

by rekins via cheatography.com/26597/cs/7480/

NGS Coverage

Coverage = Total raw output / Target sequencing size

50x for de novo sequencing

30x coverage for resequencing and transcriptomics

Target sequencing size estimation

GENOME	Number of samples x target genome size
AMPLICON	Number of molecules x average molecule size
TRANSCRIPTOME	N genes x average length (bp)

Genome Information

Species	Size	N genes
E. coli	5 Mbp	4600
S. cerevisiae	13 Mbp	6000
A. thaliana	130 Mbp	27000
H. sapiens	3 Gbp	25000

Variation Analysis

Technology	WGS	WES	RNA-seq
Mut, Indels	Y	coding	~coding
CNV	Y	coding	N
Subclonal	N	Y	N
Fusions	~Y	N	Y
Expression (variants)	N	N	Y

Filters for somatic variants detection: *Paired normal tissue, Minor Allele Frequency (MAF) >0.1% in dbSNP, 1000 genomes, Database of Genomic Variant (DGV).*

WES and WGS*: report bases not (sufficiently) covered to detect variants.*

INFO

Made as support material for the break-out Genomics session - Successful Planning of Large Data Generating Experiments (Leuven, 16 March)

Sequencing Depth

Required Sequencing Depth (D)

RNA quality	mRNA	total RNA
Good	D	2D
Degraded	2D	4D
Bad	4D	8D

Specifications (Illumina)

Platform	Max read length	Max N reads	Output
MiSeq	2x300 bp	25 M	15 Gb
NextSeq Medium	2x150 bp	130 M	39 Gb
NextSeq High	2x150 bp	400 M	120 Gb
HiSeq2500 Rapid	2x250 bp	600 M	300 Gb
HiSeq3000	2x150 bp	2500 M	750 Gb
HiSeq2500 High	2x125 bp	4000 M	1 Tb
HiSeq4000	2x150 bp	5000 M	1.5 Tb
HiSeq X Ten	2x150 bp	6000 M	1.8 Tb

Number of samples per application

Platform	TotalRNA (40M reads)	WES (>90% >10x)	WGS (30x)
MiSeq	NA	NA	NA
NextSeq Medium	3	3	NA
NextSeq High	10	10	1
HiSeq2500 Rapid	15	20	1
HiSeq3000	50	96	6
HiSeq2500 High	100	150	10
HiSeq4000	125	180	12
HiSeq X Ten	NA	NA	16

Costs per sample (Library prep + NGS)

Cost	Application
~€4000	WGS 30x tumor, 10x germline (2x125 bp)
~€2500	WES 200x tumor, 100x germline (2x125 bp)
~€750	Total RNA seq, 20-30 M reads, stranded library, rRNA depleted (2x125 bp)
~€300	mRNA seq, 15-20 M reads, stranded library (75 bp)

Authors

Rekin's Janky, rekins.janky@vib.be (VIB Nucleomics Core, BE)
Judith Boer, j.m.boer@erasmusmc.nl (Erasmus Medical Centre, NL)
Peter-Bram 't Hoen, P.A.C._t_Hoen@lumc.nl (Leiden University Medical Center, NL)
Wouter Van Delm, wouter.vandelm@vib.be (VIB Nucleomics Core, BE)

D = Total raw output / read size

Related to coverage, the sequencing depth computes the expected number of reads per sample.

For example, in human : 25,000 genes x 2,000bp x 30 coverage = 1.5Gb per sample. With a single read sequencing of 75 bp reads, this corresponds to 20 M reads per sample.

By **rekins**
cheatography.com/rekins/

Published 31st August, 2020.
Last updated 9th June, 2016.