Cancer Center
NCI-DESIGNATED COMPREHENSIVE CANCER CENTER

WES and WGS Planning and Best Practices

General considerations to increase the likelihood of high quality data

Consult Analyst
Engage with an analyst while planning the experiment

Discuss goals available samples, batches, etc.

Define Experimental Variables
Define experimental question and design experiments to properly test while avoiding confounders

Confirm your experimental groups (e.g. positive controls, genetic change)

Prevent Batch Effects
Avoid uncontrollable technical variables (e.g. do not process your control and experimental groups with different kits)

Avoid mixing fresh/frozen with FFPE samples

Conduct Pilot Studies
Perform preliminary sequencing to confirm quality and positive/ negative controls

Pilot studies can save money!

Include Proper Controls
Identification of somatic alterations requires a "control" sample (e.g. blood, healthy tissue) from each donor

Consult Analyst
Review raw data and results

Discuss data management plan

Consult Data Generating Facility
Ensure compliance with sample input policies and understand the deliverables for the data generating facility (e.g. buffers, concentrations)
Use Genomics.Core@CUAnschutz.edu for the Genomics Shared Resource

Ensure Sufficient Sample Size
Verify adequate sample numbers for analysis

Avoid Sample Pooling
Pooling can introduce unwanted variability and mask variability across replicates

Collect Metadata
Gather necessary sample information (e.g. sex, time point, any batch)

Perform Quality Control
Conduct quality checks throughout data generation (e.g. concentration and integrity)
Go/no-go decisions

Technology comparison

	Whole-exome-seq (WES)	Whole-genome-seq (WGS)
Intended capture	Exons (~2% of genome) Requires an exon capture kit	Whole-genome
Types of calls	Germline and somatic (~5% detection limit of somatic)	Germline and somatic (~10-15% detection limit of somatic)
Typical coverage	100-300X, depth is highly variable	30X, typically fairly uniform
Intended variants	SNPs and indels within exons	CNA, gene fusions, SNPs and indels in high frequency
Cost	Lower	More costly than WES because of sequencing needed

Alternative technologies

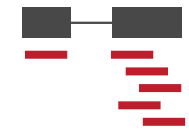
Duplex-seq
Tag-based error correction method to improve sequencing accuracy of select target regions
Capable of detecting mutations <0.1%

Genotyping arrays or low-pass WGS
Lower-cost alternative if interested in only germline alterations
Does not cover every base

Custom gene panels
Useful if only specific regions are of interest

Long-read sequencing
Provides better resolution of more challenging genomic regions and stuctural variants

Designing a WES or WGS experiment



Read general considerations (above)

Sample replicates
Typically not necessary for WES/WGS

Control samples
Identification of somatic alterations requires a "control" sample (e.g. blood, healthy tissue) from each donor

Avoid FFPE samples
Degradation typically results in low coverage & artifacts

Fragment size ~300-650 bp
Small fragments results in wasted sequencing costs (check kit specs)

Sample number
Varies by experimental design but can be high to overcome population variability

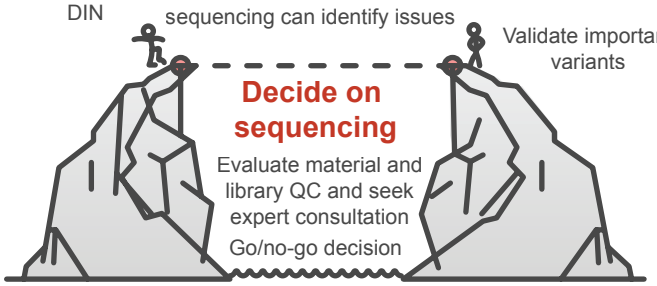
Quality control

Material evaluation
Assess concentration, fragment size, DIN

Library evaluation
Check size and quality (e.g. bioanalyzer, tapestation)
Pilot studies and shallow sequencing can identify issues

Data evaluation
Check QC metrics (e.g. depth, contamination, coverage regions)

Validate important variants



Contact the Biostatistics & Bioinformatics Shared Resource (BBSR) for cancer-related analysis

James.Costello@CUAnschutz.edu, Bioinformatics@CUAnschutz.edu

Contact CUBE for all non-cancer-related projects