Ultra deep targeted sequencing approaches

March 28th 2022, Aarhus, Denmark

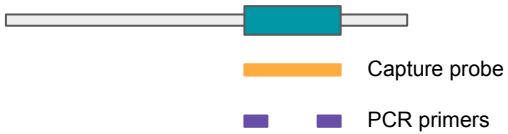
Mads Heilskov Rasmussen, PhD
Emil Christensen, PhD

Agenda

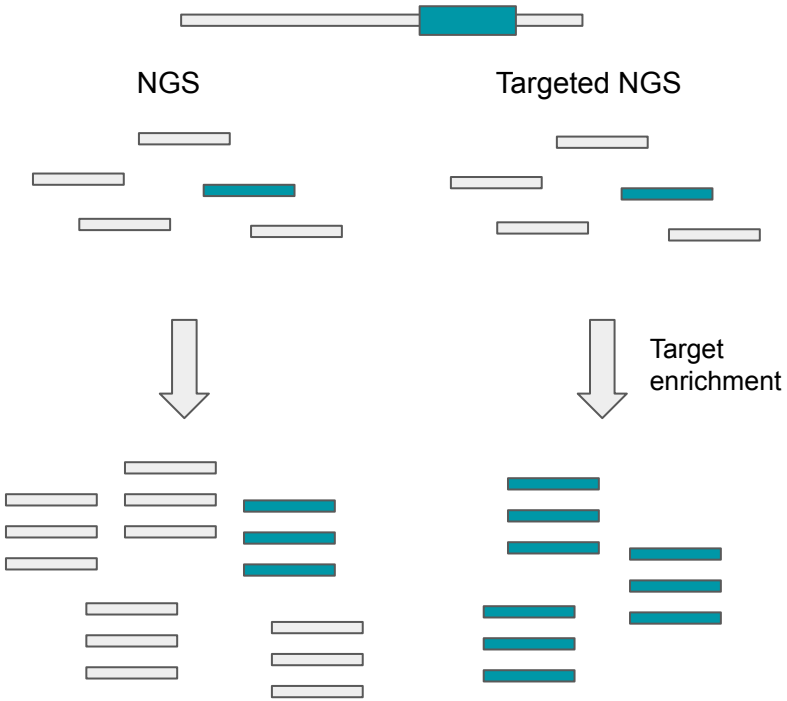
- Target enrichment
- Why use ultra deep targeted sequencing?
- Target enrichment design
- Introduction to ultra deep target enrichment technologies
- Sources of noise in NGS
- Unique molecular identifiers
- Examples of ultra deep target enrichment technologies
- Cases

Target enrichment

- Focus on genomic regions of interest
- Method for enrichment
  - Hybridization-based
  - Amplicon-based

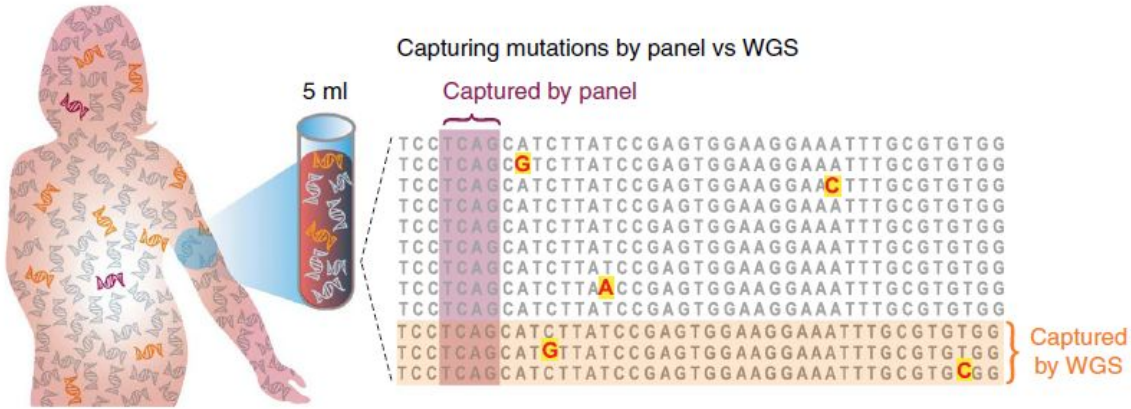


- Concentrate sequencing data



Why do we do ultra deep targeted sequencing?

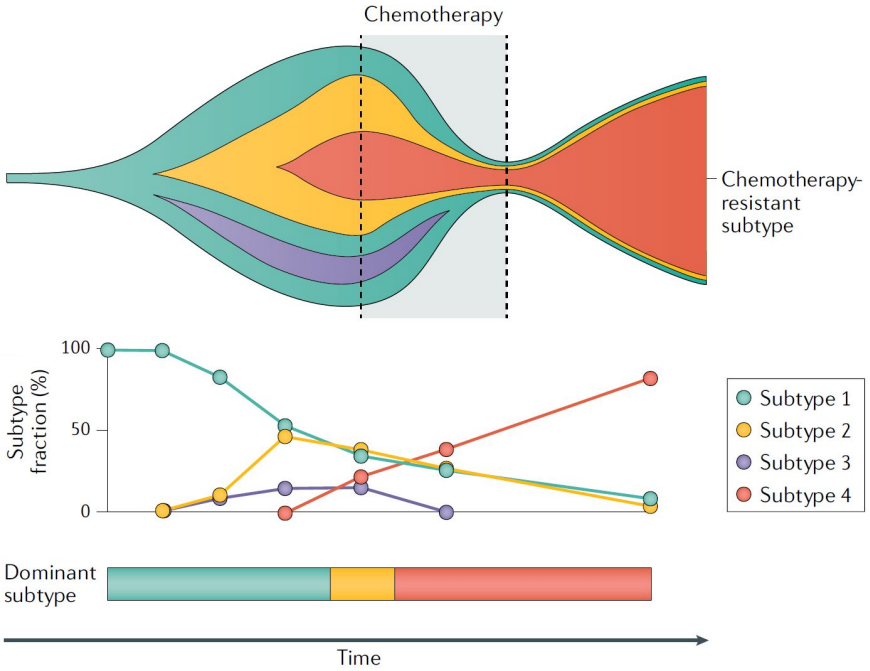
- Consider a blood sample
  - Limited amount of ctDNA
  - Maximize the chance of finding ctDNA
- Exhaustive assessment of regions of interest
- Relatively low sequencing costs



Zviran et al., Nature Medicine, 2020

What's the catch?

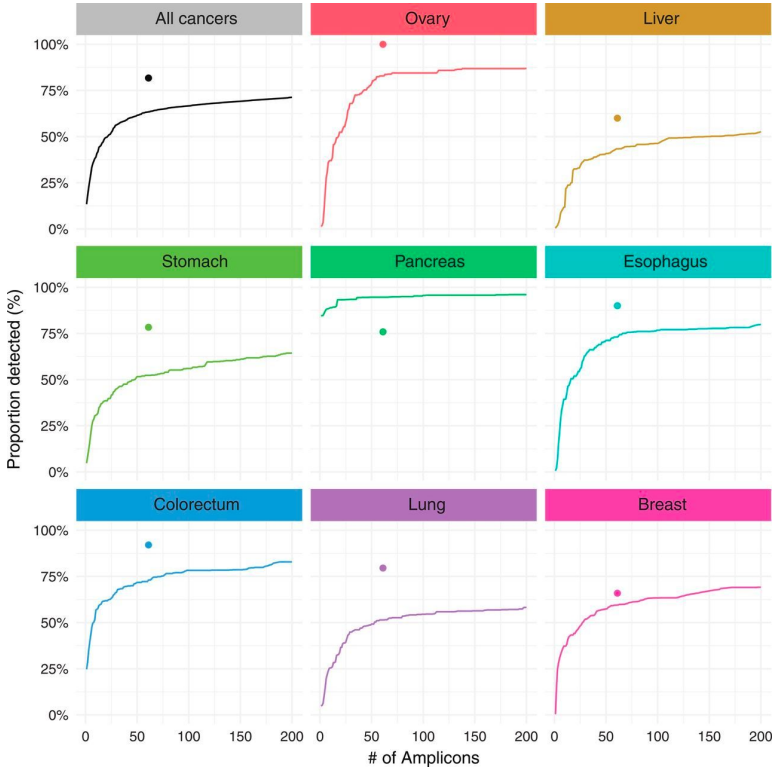
- What if the enrichment is not optimal?
  - Tumor heterogeneity
  - Tumor evolution
- New critical mutations may arise
- What if no ctDNA fragments originating from the regions of interest are present?



Meeks et al., Nature Reviews Urology, 2020

Enrichment design

- Tumor agnostic
- Tumor informed
  - Non-personalized



Cohen et al., Science, 2018

Enrichment design

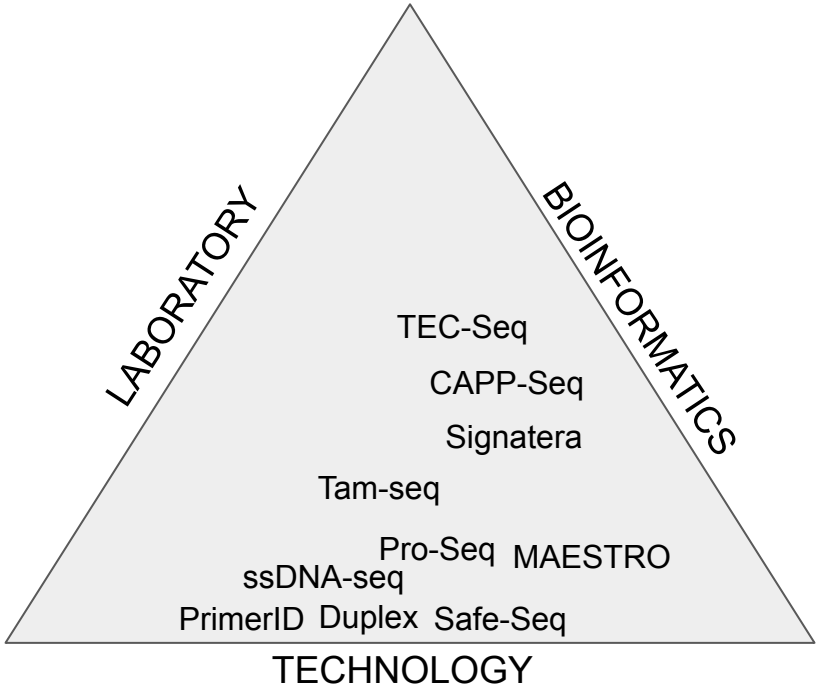
- Feasibility of tumor agnostic or tumor informed non-personalized approaches?
- Tumor informed and personalized

Catalogue somatic mutations

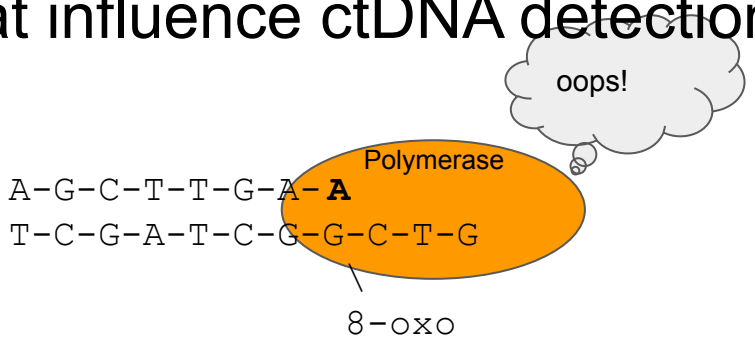
Enrichment of mutations for all patients

- Time ⇧
- Effort ⇧
- Consistency and reproducibility?
- Retrospective/prospective analysis setup

Development of novel ultra deep NGS technologies ... and their methods



Sources of errors in NGS that influence ctDNA detection

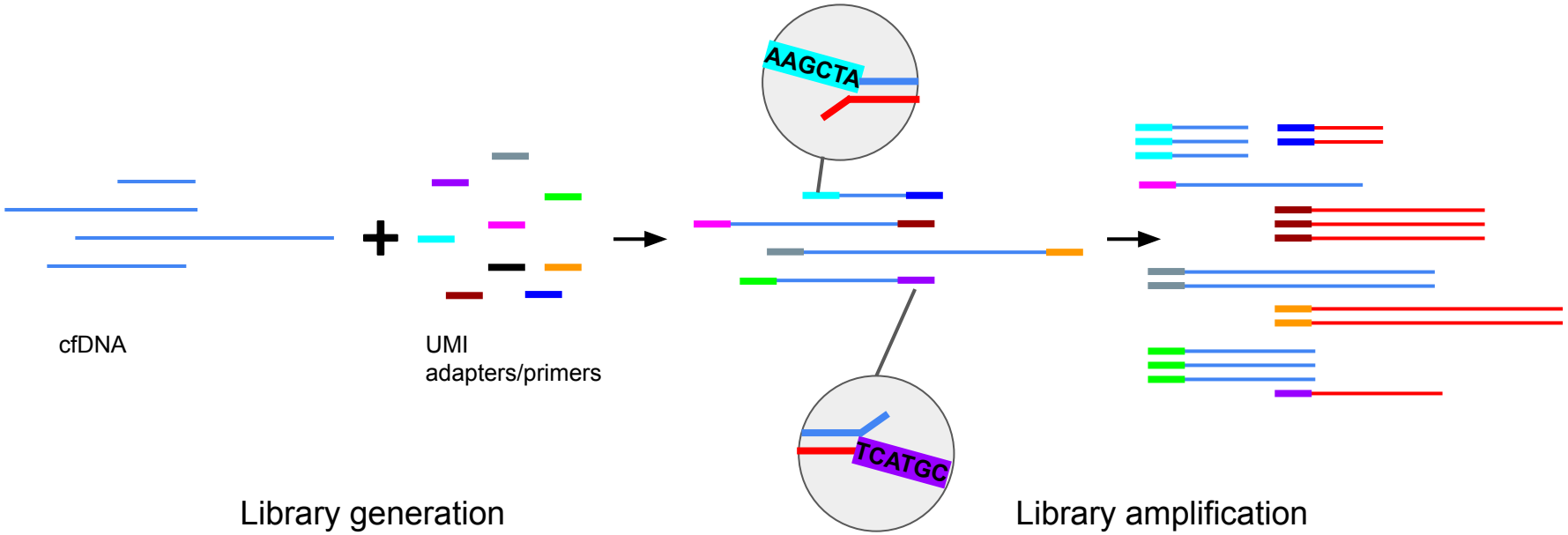


1. PCR errors
2. PCR "errors" on abasic nucleotides (sample storage and handling)
   a. Oxidation of guanine **G:C** (-> 8-oxo-G:C) **-> T:A** mutations
   b. Deamination of cytosine: **C:G** (-> U:G) **-> T:A**
   c. Others
3. Image acquisition and interpretation (sequenator)
4. Biological noise: mutational signal from non-cancerous cells (e.g. "clonal hematopoietic expansion of unknown potential" = CHIP)

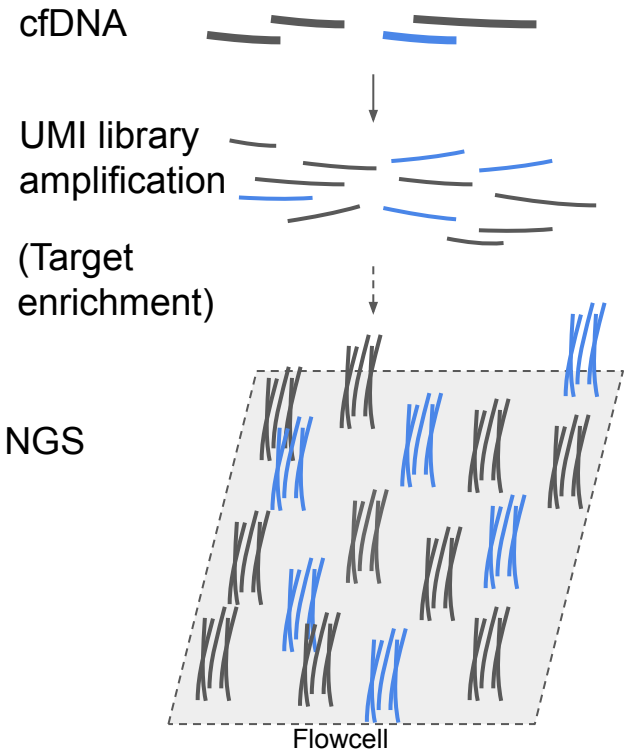
False positive signal is a critical limitation for ultra-sensitive ctDNA detection methods
-> solution (1., 2. and 3.): UMI technology

How UMI mediated noise reduction works

UMI: Unique Molecular Identifier, a molecular **BARCODE**



How UMI mediated noise reduction works



Reference: AGTCAGTTCGCTG

PCR duplicates:
AGTCAGTGCGCAG
AGTCAGTTCGCAG
AGTCCGTTCGCAG
AGTCAGTTCGCAG
AGTCAGTTTTCAG
AGTCAGTTCGCAG
AGTCAGTTCTCAG

UMI family (same UMI + genomic pos)

Consensus read: AGTCAGTTCGC**A**G

Calls: T>A

Principal workflow of UMI mediated noise reduction

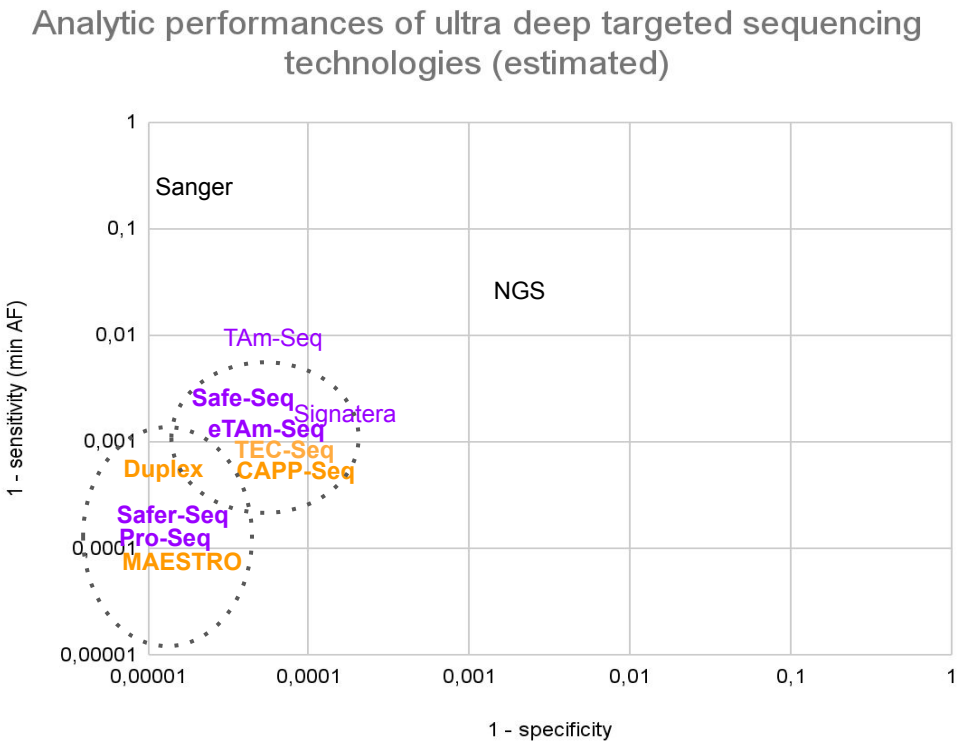
1. Library prep
2. Target enrichment
3. **Redundant sequencing**

In lab

4. Mapping of raw fastq reads
5. **Grouping of reads sharing UMI barcode and genomic position into "families"**
6. **Consensus sequence generation within UMI families**
7. **Mapping of consensus reads**
8. Variant calling

On computer

Technology performances

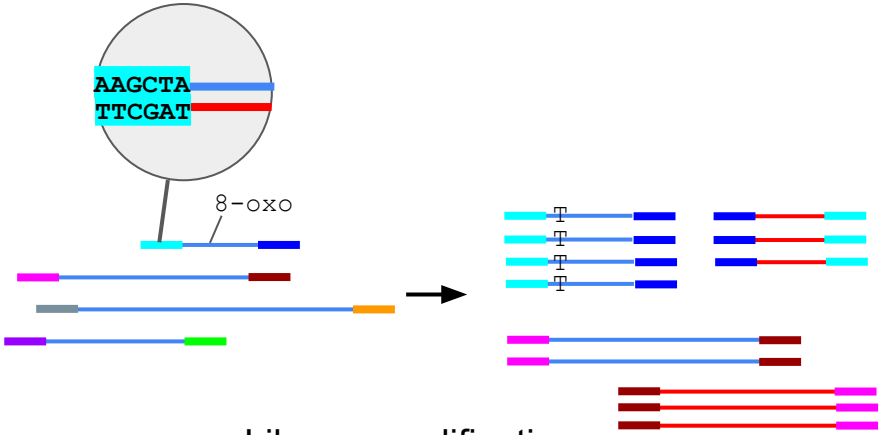


Targeting method

Non UMI-based
UMI-based

Approximate estimates for illustrative purpose only

Duplex sequencing



Library amplification

⇧ Duplex Seq increases specificity 10-500x by utilizing information from both strands to eliminate abasic and "jackpot" PCR errors.

⇩ Requires sampling of both strands effectively increasing NGS data cost and lowering sensitivity.

Reference: AGTCAGTTCGCTG

Sense UMI family:
AGTCATTGCGCAG
AGTCATTTCGCAG
AGTCCTTTCGCAG
AGTCATTTCGCAG

Anti-sense UMI family:
AGTCAGTTTTCAG
AGTCAGTTCGCAG
AGTCAGTTCTCAG

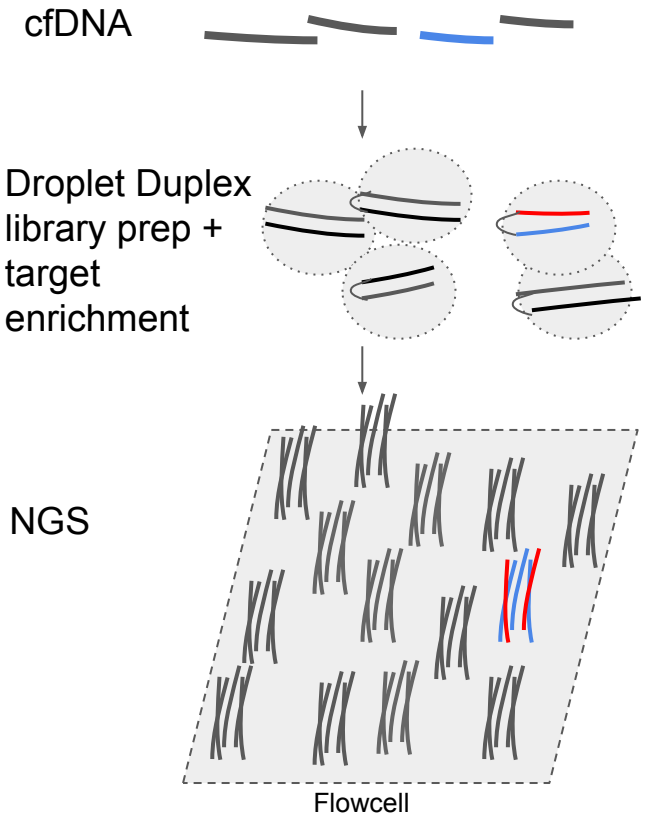
Duplex family

Consensus read: AGTCAGTTCGC**A**G

↓

Calls: T>A

Proximity sequencing



Reference: AGTCAGTTCGCTG

Flowcell cluster duplicates:

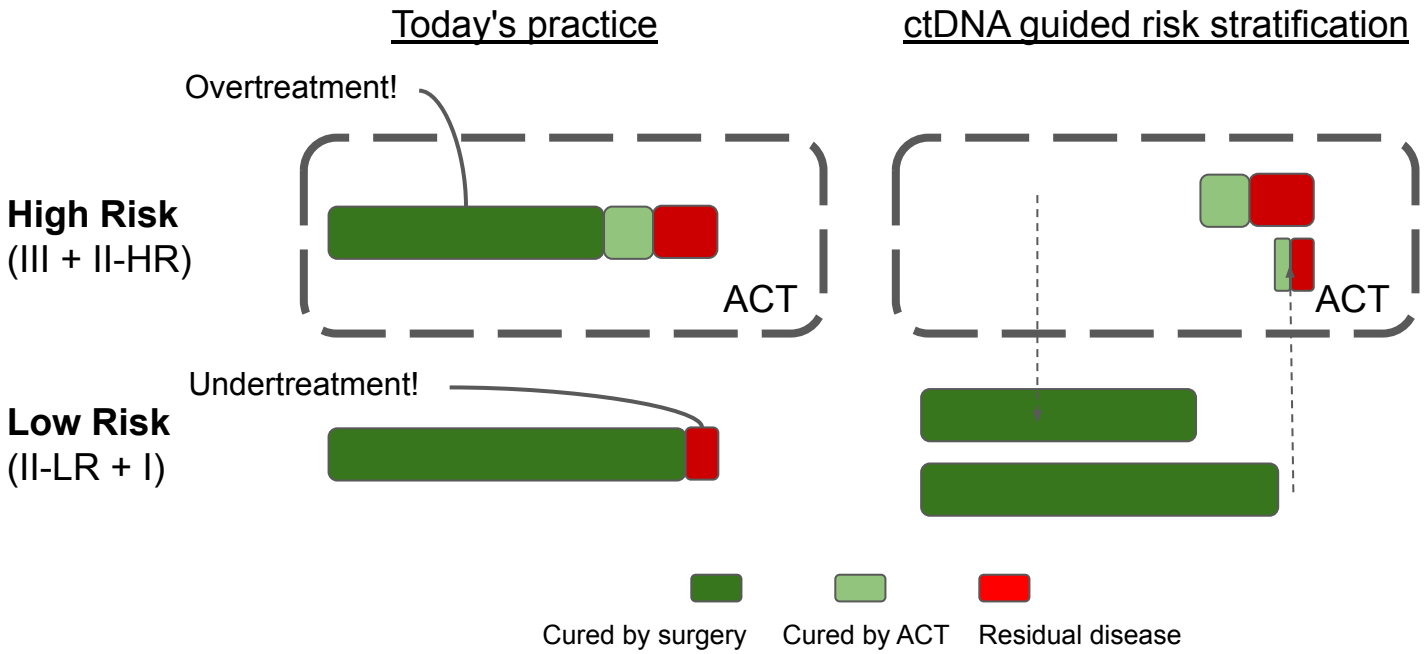
AGTCAGT**G**CGC**A**G
AGTC**CT**TTCGC**A**G
AGTCAGTTCGC**A**G
AGTCA**T**TTCGC**A**G
AGTCA**T**TTCGC**A**G
AGTCAGTTCGC**A**G

Flowcell consensus: AGTCANTTCGC**A**G

Calls: T>A

Pro Seq has Duplex level-noise reduction using 50x less data effectively increasing sensitivity

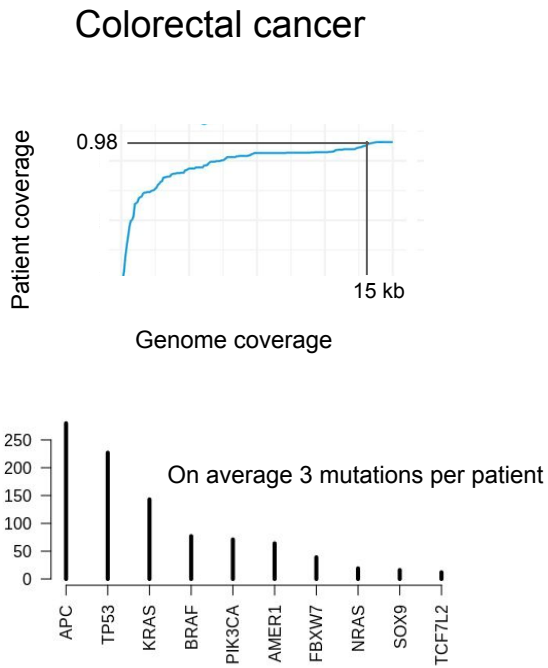
Case I: can ctDNA improve the post-OP treatment of CRC ?

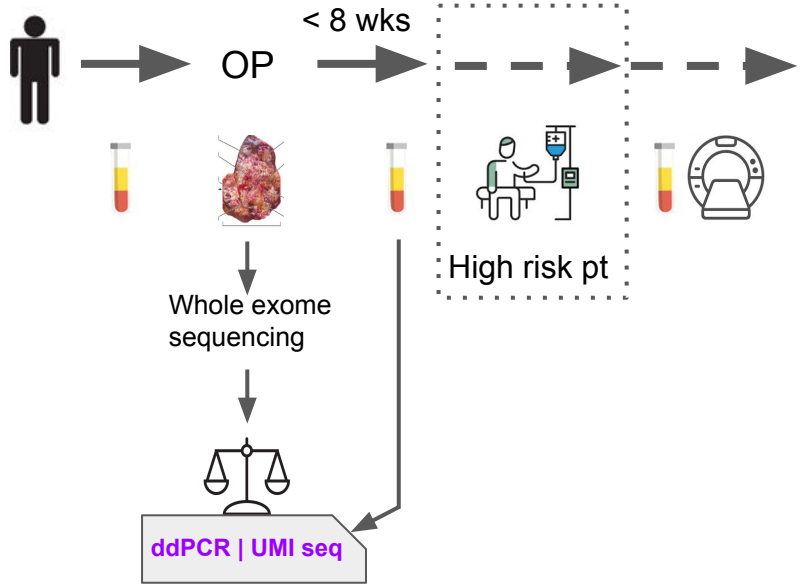


Clinicians need a postOP response - ctDNA positive or negative ?

@ Modified from Christina Demuth

Case I: can ctDNA improve the post-OP treatment of CRC ?

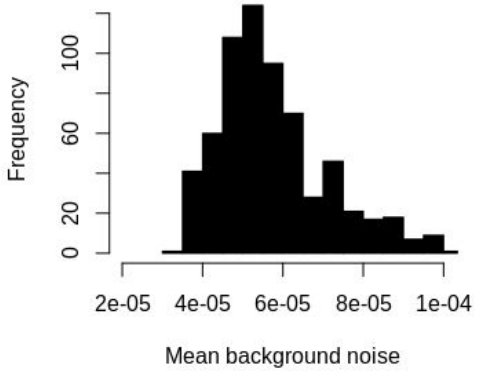




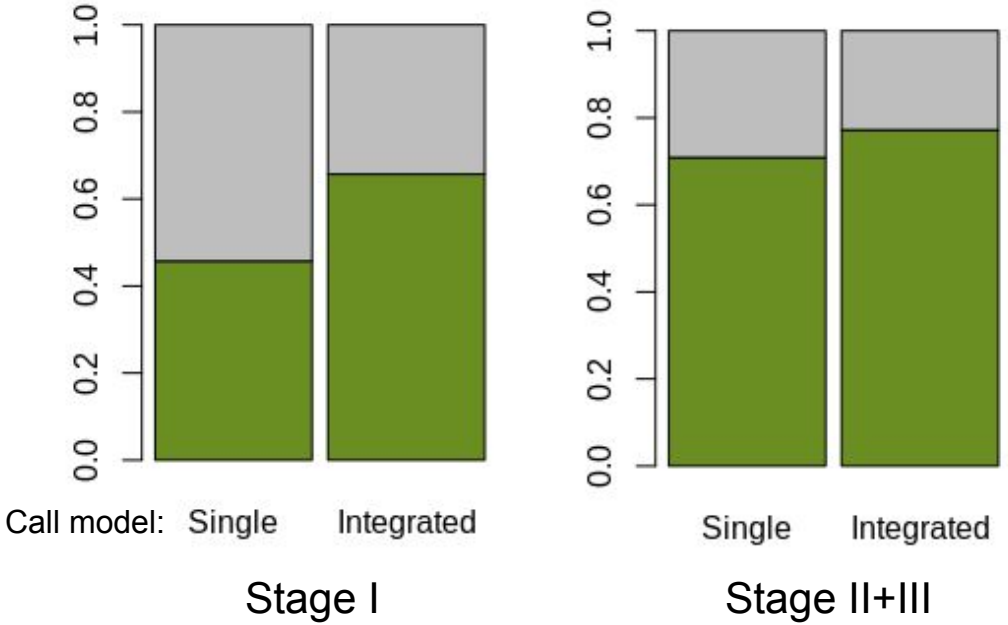
Case I: can ctDNA improve the post-OP treatment of CRC ?

Why is it not perfect ?

- Low shedding (T1 tumors)
- cfDNA -> NGS efficiency
- ctDNA sampling effects
- Mean LOD 0.032 % (plasma)



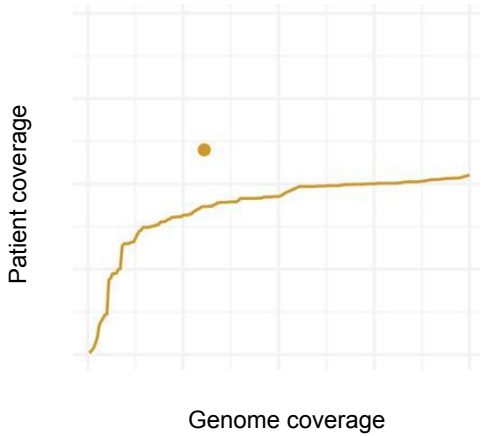
PreOP detection rates



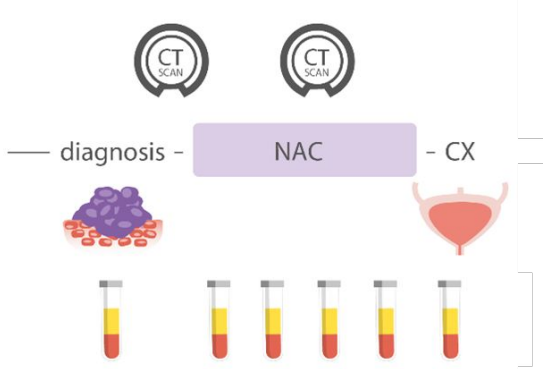
III

Case II

Bladder cancer



Retrospective setup



92 patients

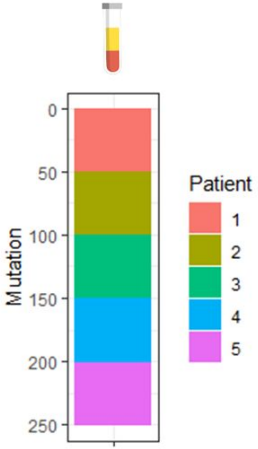
Plasma and urine samples

No response to be given to clinicians - proof of principle

Case II

Whole exome sequencing
Catalogue somatic mutations

50 mutations/patient

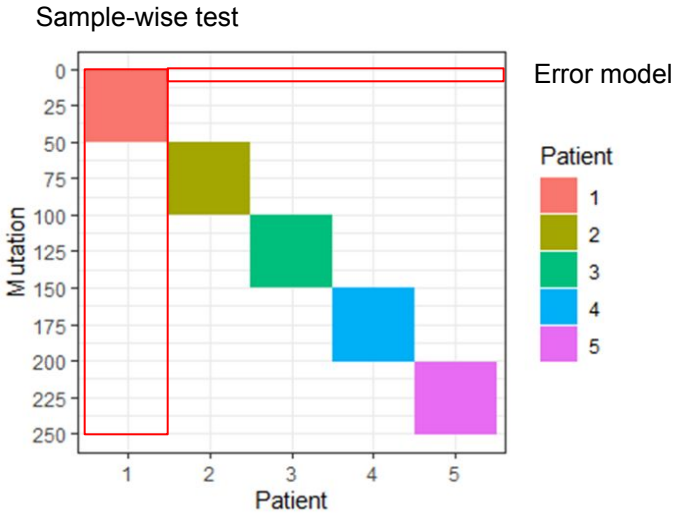


Ultra deep sequencing with UMIs

Case II

- Exploit cross patient data
- Single mutation calling
  - Shearwater algorithm[1,2]
    - Test vs. error model based on “normal samples”
- Sample level calling
  - Fisher’s method for target mutations
  - Bootstrapping of random non-target mutations

    Fisher’s method
  - Rank target mutation score in relation to non-target scores



[1] Gerstung et al., Bioinformatics, 2014

[2] Martincorena et al., Science, 2015

Key points: Ultra deep targeted sequencing

- Very high mean depth on relatively narrow genomic space
- The clinical situation and practical matters are important for the enrichment design
- False positive signals arise from 1) NGS image interpretation, and 2) PCR base misincorporation (especially on abasic bases)
- UMI directed strategies might be necessary to achieve sufficient sensitivity (for most clinical settings) due to false positive signal inherent to NGS

Selected references

NGS noise: Pfeiffer et al Sci Rep. 2018 Jul 19;8(1):10950

NGS specs: Loman et al Nat Biotechnol. 2012;30(5):434–9, Fox et al Next Gener Seq Appl. 2014;1,

Artifactual noise: Shibutani et al Nature. 1991;349:431–434; Stiller et al Proc Natl Acad Sci U S A. 2006;103:13578–13584; Chen et al Science. 2017 Feb 17;355(6326):752-756; Costello et al Nucleic Acids Res. 2013 Apr 1;41(6):e67

PrimerID: Jabara et al. Proc Natl Acad Sci U S A. 2011;108(50):20166–71

Safe-Seq: Kinde et al Proc Natl Acad Sci U S A 2011;108(23):9530–5, Cohen et al 2021 Nature Biotechnology 39(10):1-8

Tam-Seq: Forshew et al Sci Transl Med 2012 May 30;4(136):136ra68; Gale et al PLoS One. 2018 Mar 16;13(3):e0194630

Duplex: Schmitt et al. Proc Natl Acad Sci U S A. 2012;109:14508–14513

CAPP-Seq/iDES: (Newman et al Nat Med 20, 548 (2014), Newman et al Nat Biotechnol 34, 547 (2016))

Tec-Seq: Phallen et al Sci Transl Med. 2017 Aug 16; 9(403)

Pro-seq: Pel et al PLoS One. 2018 Oct 2;13(10):e0204265

MAESTRO: Gydush et al Nat Biomed Eng. 2022 Mar;6(3):257-266

ssDNA-seq: Gansauge et al Nat. Protoc. 2013 8, 737–748.; Burnham, et al Sci. Rep. 6, 27859 (2016); Snyder et al Cell 164, 57–68 (2016).