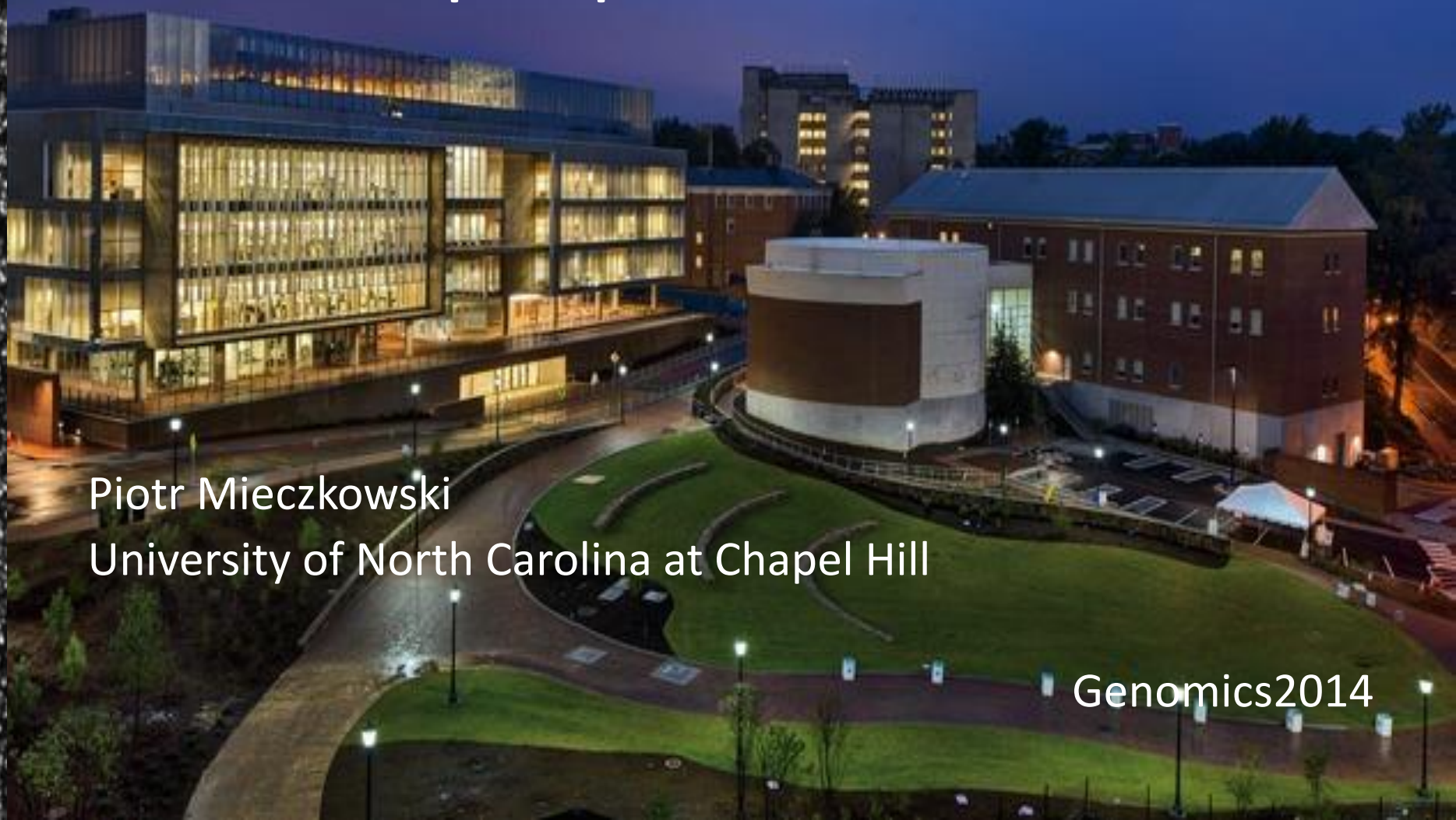
Analysis of circulating free DNA in peripheral blood



Piotr Mieczkowski
University of North Carolina at Chapel Hill

Genomics2014

Next Generation Sequencing for Clinical Care of Cancer Patients - UNCseq

Production Summary Report

Tue Apr 15 23:39:36 2014

requested by dnhayes

UNCseq Status Summary

1064 consented

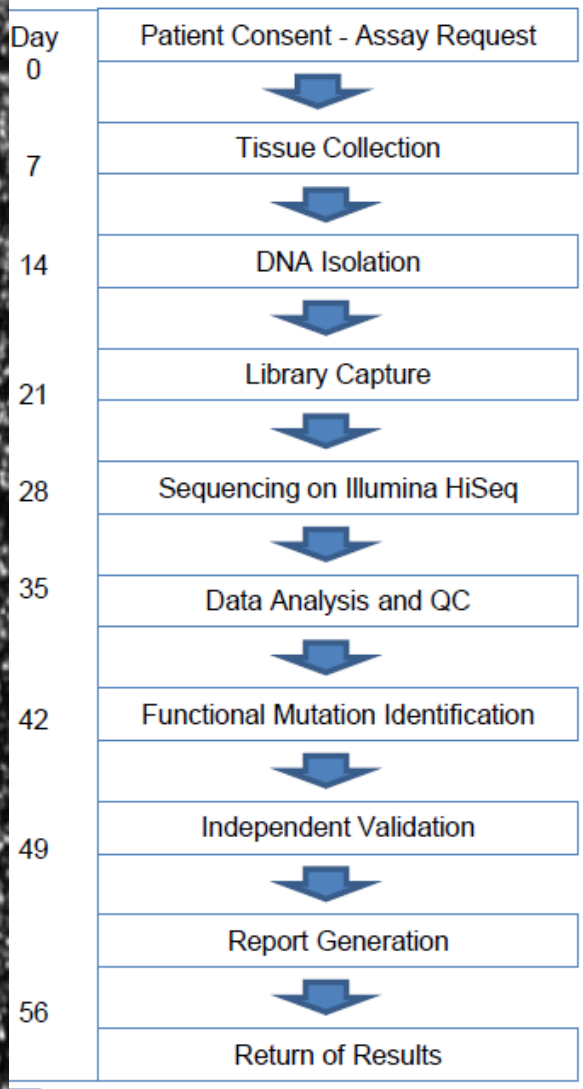
300 active
113 sample failure

436 completed
248 mutations reported
188 no reportable mutations

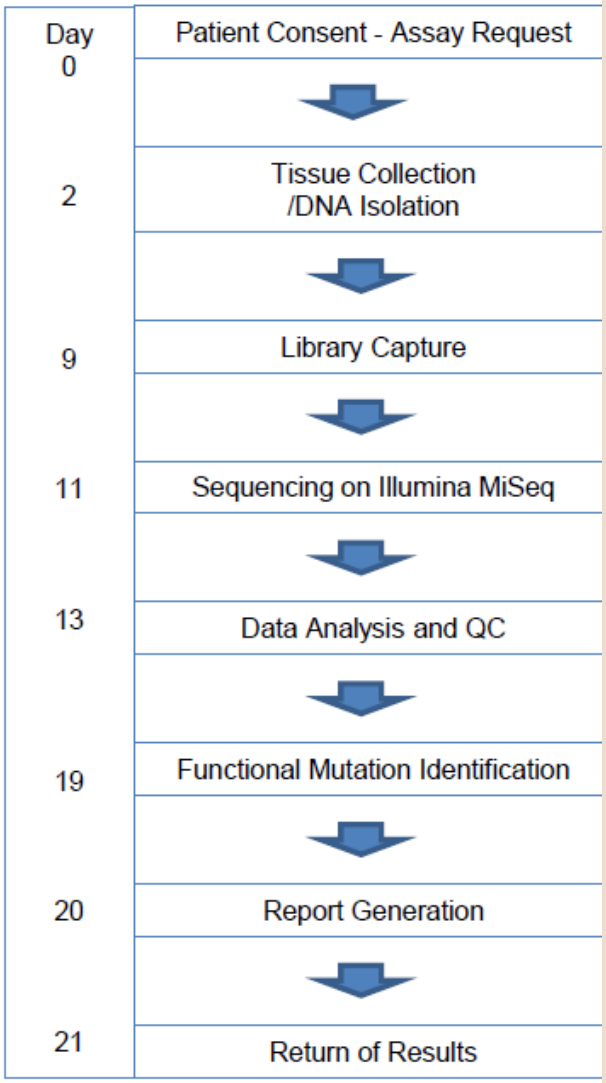
Sample Distribution

Tumor	Normal
215/300 distributed	227/300 distributed
35/300 pending distribution	4/300 pending distribution
49/300 not yet collected	26/300 not yet collected
0/300 not initiated	43/300 not initiated

Ideal Schematic Production Timeline



A **Current UNCSeq Assay Time Requirement = 56 days**



B **Improved UNCSeq Assay Time Requirement = 21 days**

Time to Completion Stats

Early Study

Total patients = 400

Physician Perception
Consent to Phone Message time (344)
Range: .4 - 16months
Average: 6months

Consent to Discussion (344)
Range: 1.9 - 12.5months
Average: 4.6months

Sample Collection to Phone Message time (344)
Range: -.5 - 12.7months
Average: 4.6months

Processing Time
Sample Collection to Discussion (344)
Range: -.8 - 11.9months
Average: 3.3months

Active Study

Total patients = 539

Physician Perception
Consent to Phone Message time (180)
Range: 1.6 - 6.5months
Average: 3.6months

Consent to Discussion (180)
Range: 1.6 - 5.8months
Average: 2.9months

Sample Collection to Phone Message time (180)
Range: 1.1 - 5.7months
Average: 2.8months

Processing Time
Sample Collection to Discussion (180)
Range: .5 - 4months
Average: 2.1months

How to monitor progress of treatment?

How to monitor patients after treatment?

How to screen population to detect early stages of cancer?

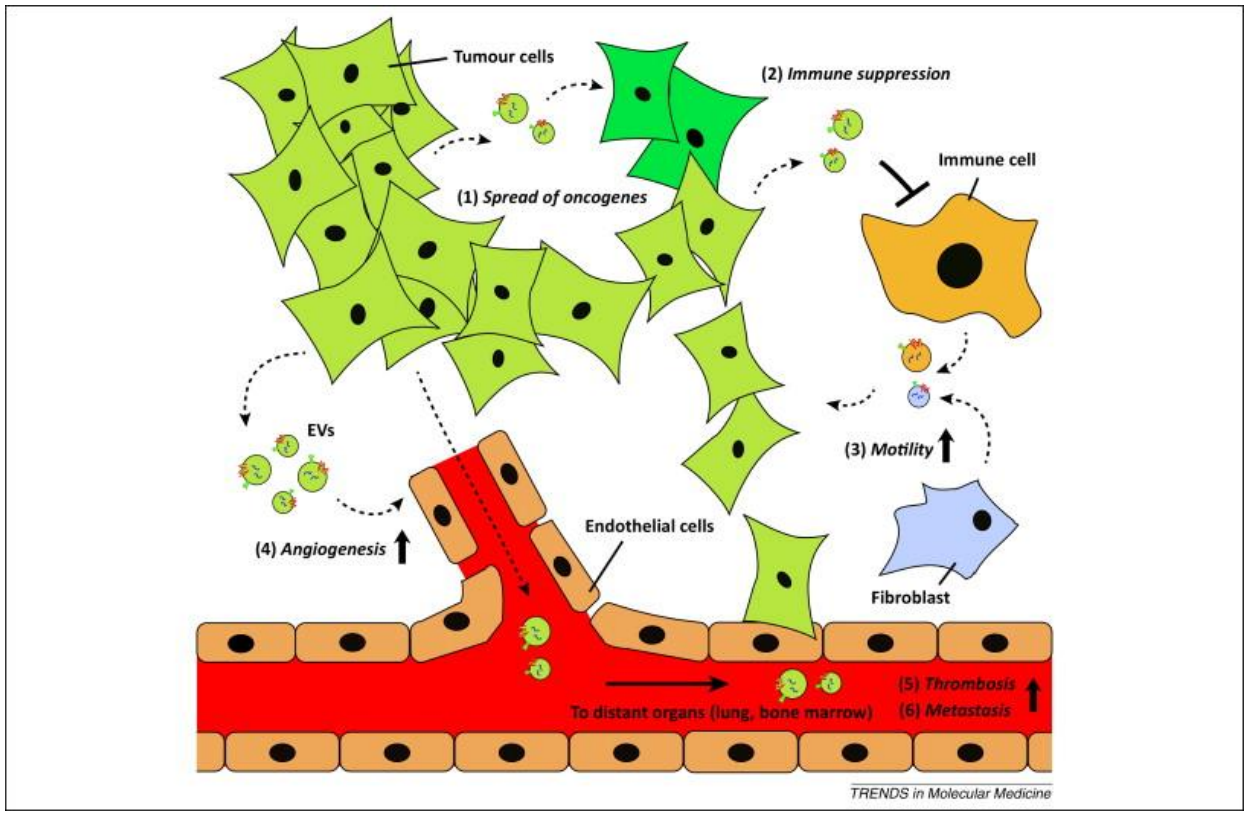
Liquid Biopsies

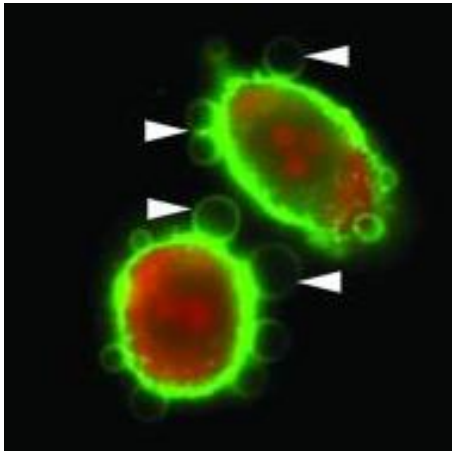
Ceppellini, R., E. Polli, and F. Celada. A DNA reacting factor in serum of a patient with lupus erythematosus diffusus. Proc. Soc. exp. Biol. (N .Y.) 1957, 96, 572.

Tan, E. M., P. H. Schur, R. I. Carr, and H. G. Kunkel. 1966. Deoxyribonucleic acid (DNA) and antibodies to DNA in the serum of patients with systemic lupus erythematosus. J. Clin. Invest. 45: 1732-1740.

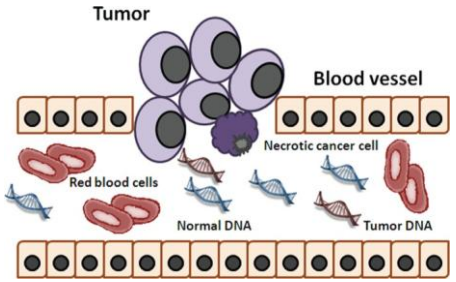
Steinman CR. Free DNA in serum and plasma from normal adults. J Clin Invest. 1975 Aug;56(2):512-5. PubMed PMID: 1150882; PubMed Central PMCID: PMC436612.

Source of circulating tumor DNA

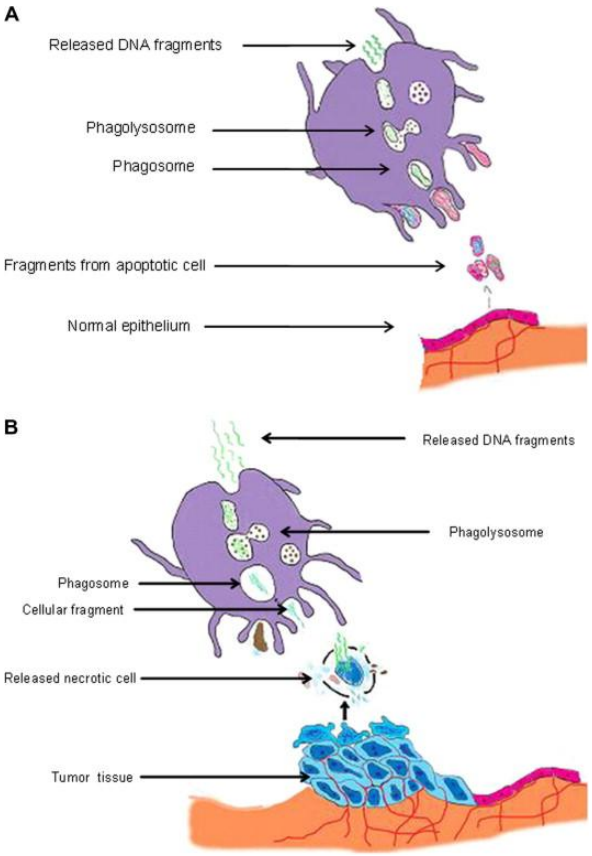




- **DNA released in oncoexosomes**
- **DNA released to plasma after necrosis or from apoptotic cells**



Alternative mechanisms of cfDNA release during phagocytosis.



Unequally sized DNA fragments result from phagocytosis of a necrotic cell (A), whereas uniformly sized DNA fragments are released by macrophage from apoptotic cell (B).

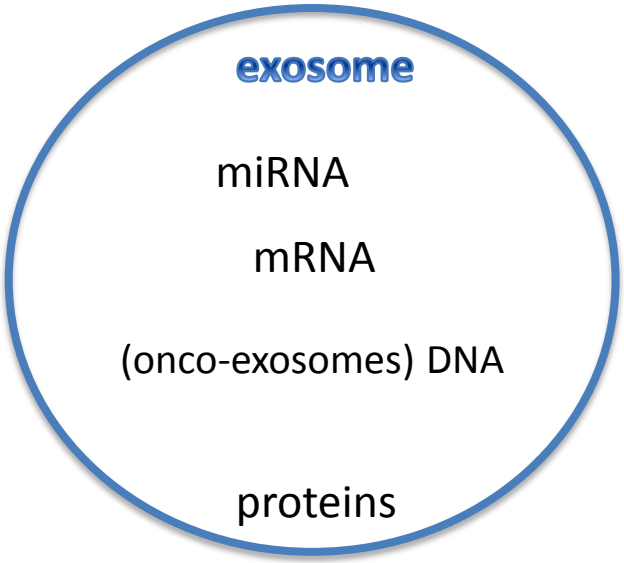
L. Benesova , B. Belsanova , S. Suchanek , M. Kopeckova , P. Minarikova , L. Lipska , M. Levy , V. Visokai , M. ...

Mutation-based detection and monitoring of cell-free tumor DNA in peripheral blood of cancer patients

Analytical Biochemistry, Volume 433, Issue 2, 2013, 227 - 234

http://dx.doi.org/10.1016/j.ab.2012.06.018

Plasma



Cell

Source of DNA and RNA contamination for tests

miRNA

proteins

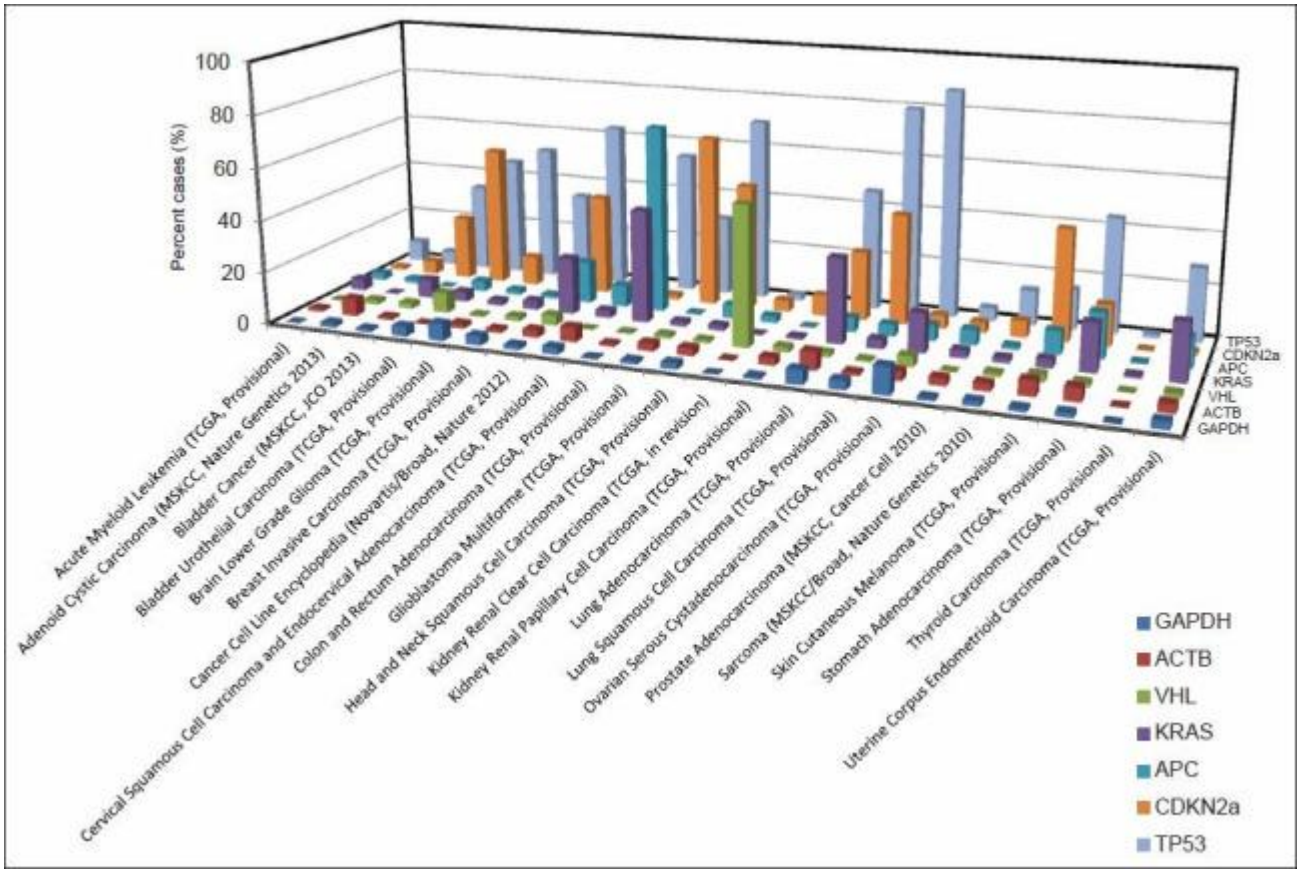
DNA

DNA from 7-14 mln cells is circulating in our bodies now (50-100ug).

Why we are interested in analysis of circulating free DNA from plasma?

- **Cancer research**

1. Abundance of mutated genes corresponds to tumor burden.



Mining genome sequencing data to identify the genomic features linked to breast cancer histopathology
Zheng Ping, Gene P. Siegal, Jonas S. Almeida, Stuart J. Schnitt, Dejun Shen J Pathol Inform. 2014; 5: 3. Published online 2014 January 31.doi: 10.4103/2153-3539.126147 PMCID: PMC3952399

- **Prenatal testing**

1. Circulation of fetal DNA gives opportunity for easy screening for chromosomal aberrations.

- **Other**

Overview of techniques used for detection of cfDNA in plasma of cancer patients.

Method	Sample costa	Instrumentation	Percentage of CRC tumors with KRAS mutation	Estimated detectable fraction of mutated alleles (%)
DCE (denaturing capillary electrophoresis)	low	PCR cycler and capillary sequencer	30–35% (tissue) 28% (plasma)	0.30%
ME-PCR (mutant-enriched PCR)	low	(real-time) PCR cycler + arbitrary method to detect amplicons	35% (tissue) 39% (plasma) 62% (stool)	0.10%
Cold-PCR (co-amplification at lower denaturation temperature-PCR)	low	(real-time) PCR cycler + arbitrary method to detect amplicons	37.5% (tissue)	0.8–2.5%
BEAMing (beads emulsion amplification and magnetics)	high	PCR and flow cytometry systems	32% (tissue) 52% (plasma) 63% (stool)	0001%–1.7%
Digital PCR	high	Dedicated digital PCR system	38% (tissue) 47% (stool)	0005%–0.1%
Single-molecule DNA sequencing	high	Single-molecule sequencer	18–30% (tissue)	0015%–3%
Targeted (tagged-amplicon) deep sequencing	medium	Next-generation sequencer	59% (tissue)	2%

L. Benesova , B. Belsanova , S. Suchanek , M. Kopeckova , P. Minarikova , L. Lipska , M. Levy , V. Visokai , M. ...

Mutation-based detection and monitoring of cell-free tumor DNA in peripheral blood of cancer patients

Analytical Biochemistry, Volume 433, Issue 2, 2013, 227 - 234

http://dx.doi.org/10.1016/j.ab.2012.06.018

Calculation of assay sensitivity

10ng of human DNA – 3000 single haploid genomes (C-value – 3.3pg per haploid genome)

We would like to have 60-80% efficiency of tagging what corresponds – around 2000 genomes.

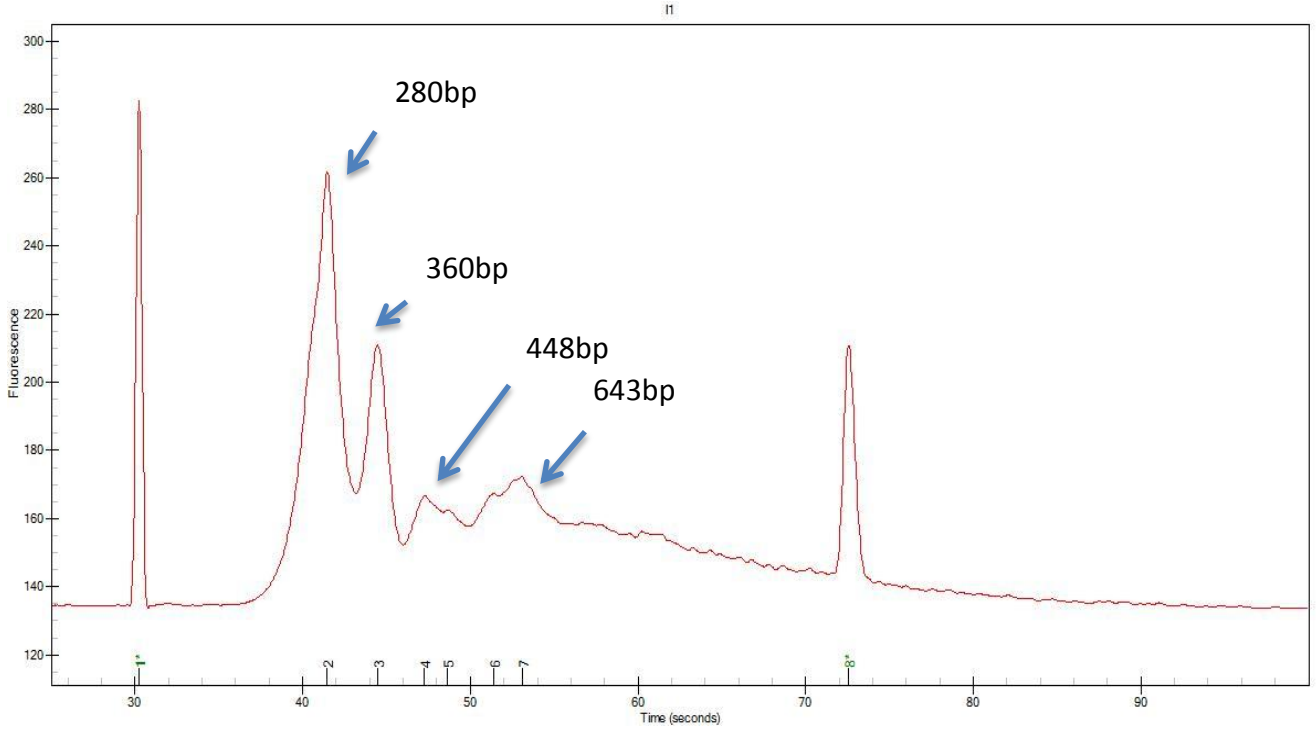
We need depth of sequencing 10,000 -15,000x to saturate system- each molecule must have around 5 copies .

Exome Capture
If our panel has 3 Mb of sequence – we need 45 Gb of sequence. We should get around 30-40 Gb from one/two lanes PE 2x100. We need to use HiSeq2500 for exome capture project.

Amplicon
MiSeq can be use for amplicons – 15mln reads – 500 amplicons

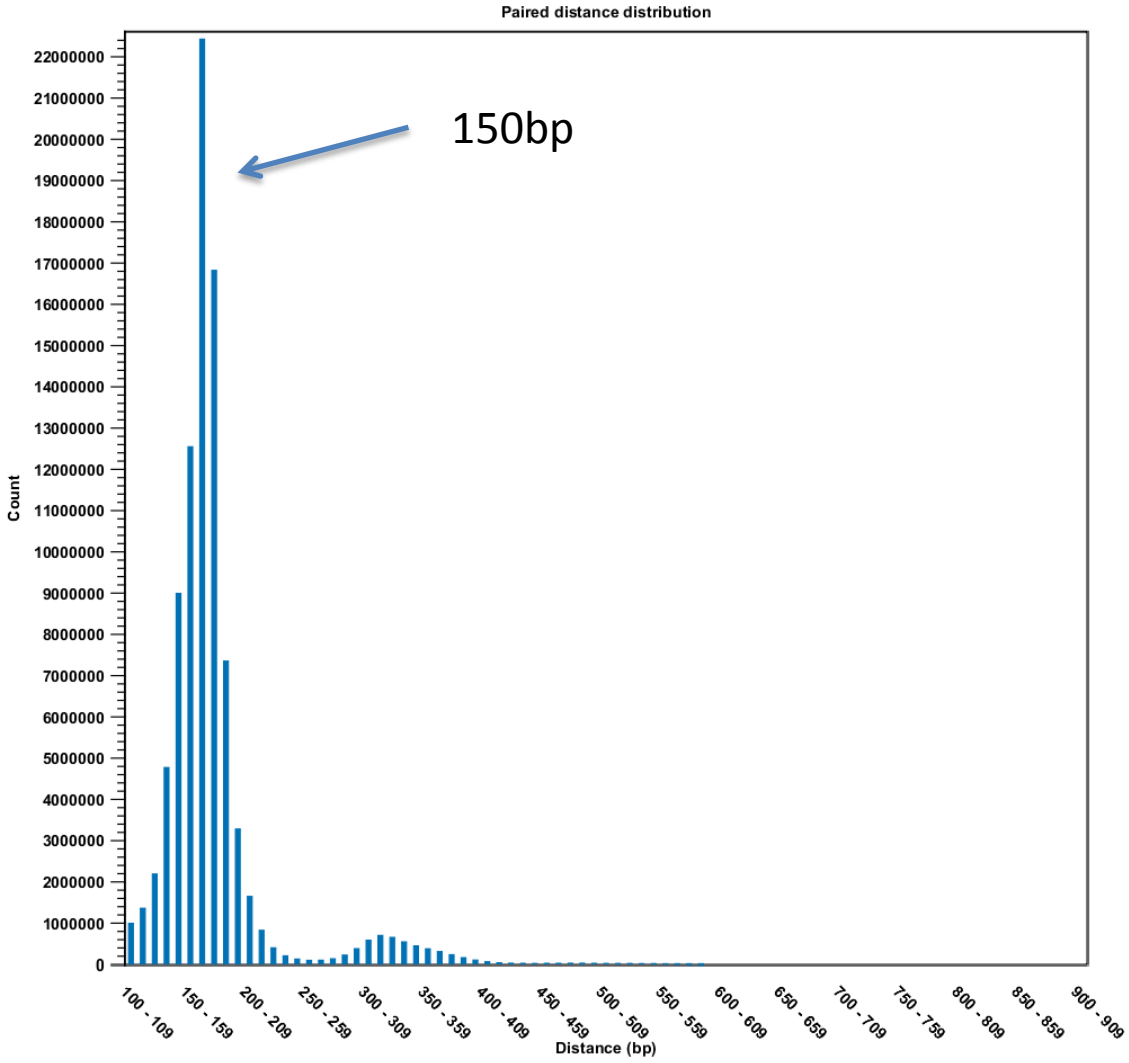
Pattern of DNA size suggests that Histones are involved in DNA protection

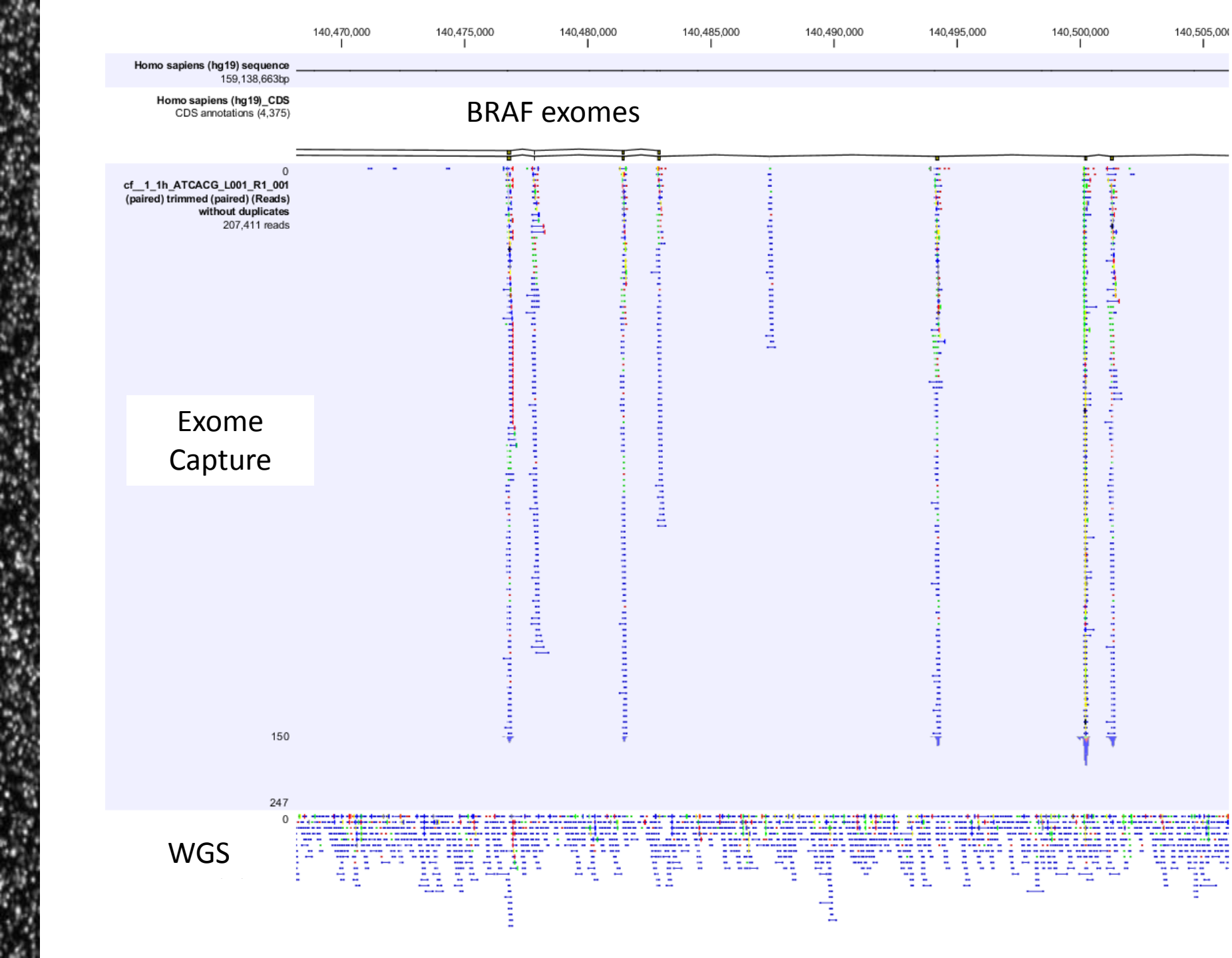
Experion 15K chip

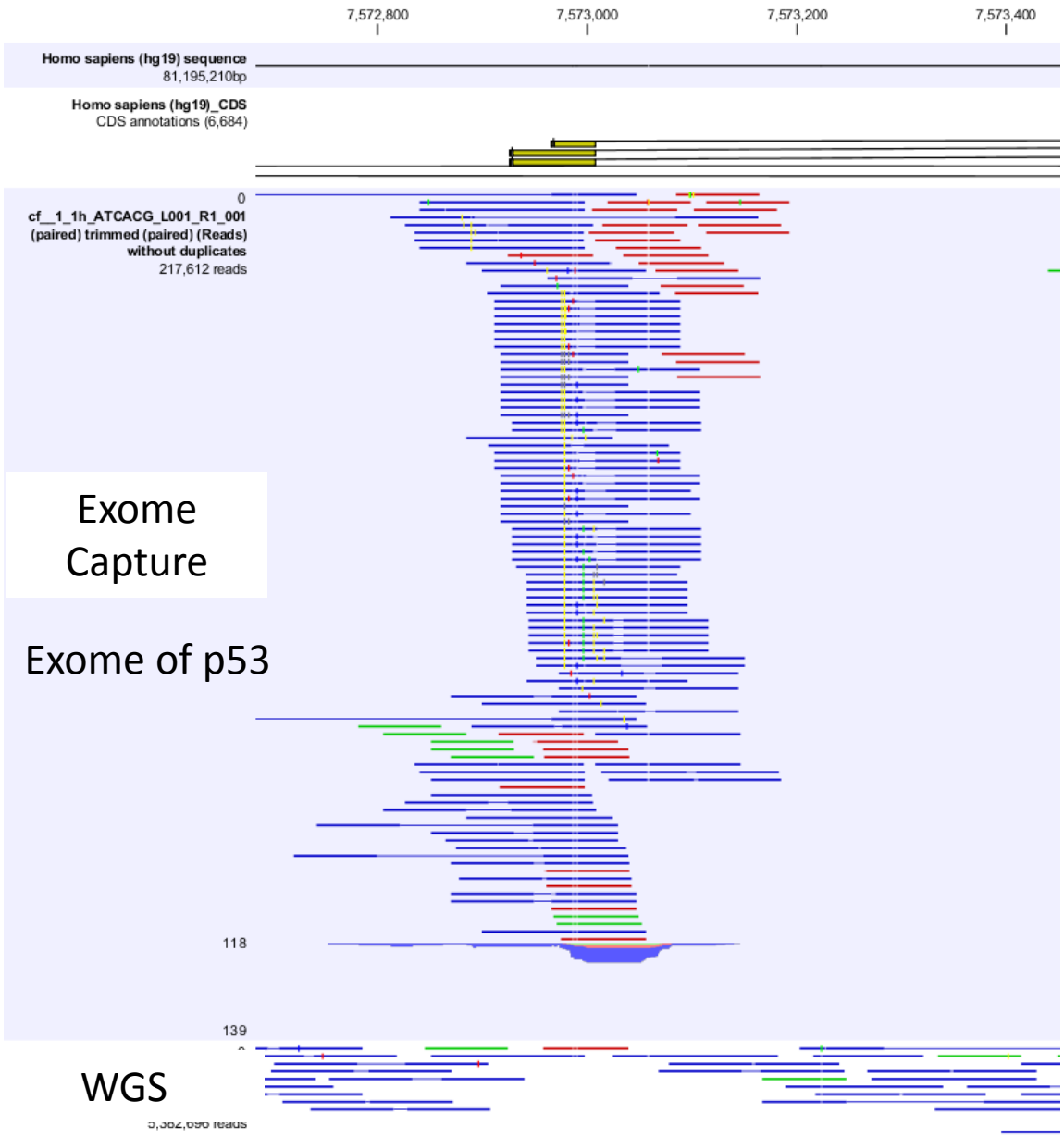


DNA library for Illumina sequencing was prepared from **2 ng** cfDNA**.**
We used Rubicon Genomics ThruPLEX kit for library prep.
The Bioanalyzer traces suggest substantial fragmentation of the circulating DNA.
The size of the dominant pick is similar to the size of DNA wrapped around histone proteins.
Therefore, our working hypothesis predicts the release of chromatin from dead cells and fragmentation by circulating nucleases in the plasma. Fragments of DNA interacting with histones are protected.

Prepared cf_1 library was subject for Illumina Pair End 2x100 cycles sequencing. Analysis of the sequencing data was performed using CLC Genomic Workbench 6.01. Insert size distribution confirmed previous observation.

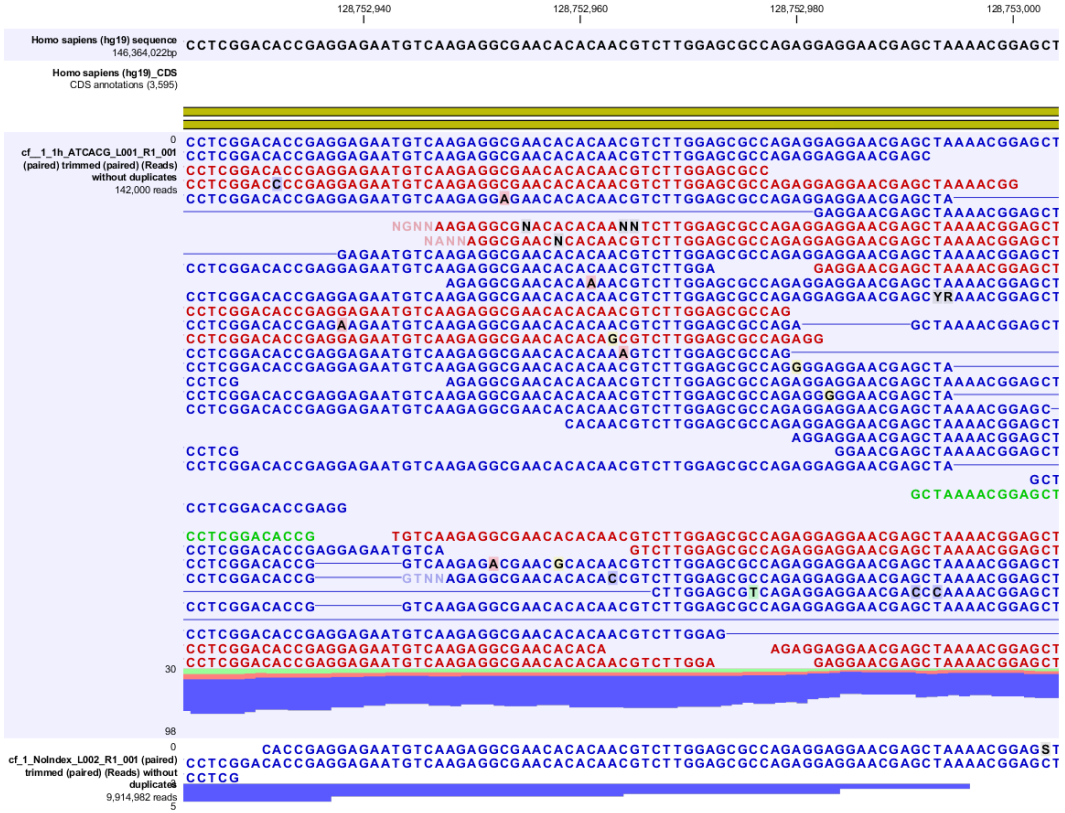


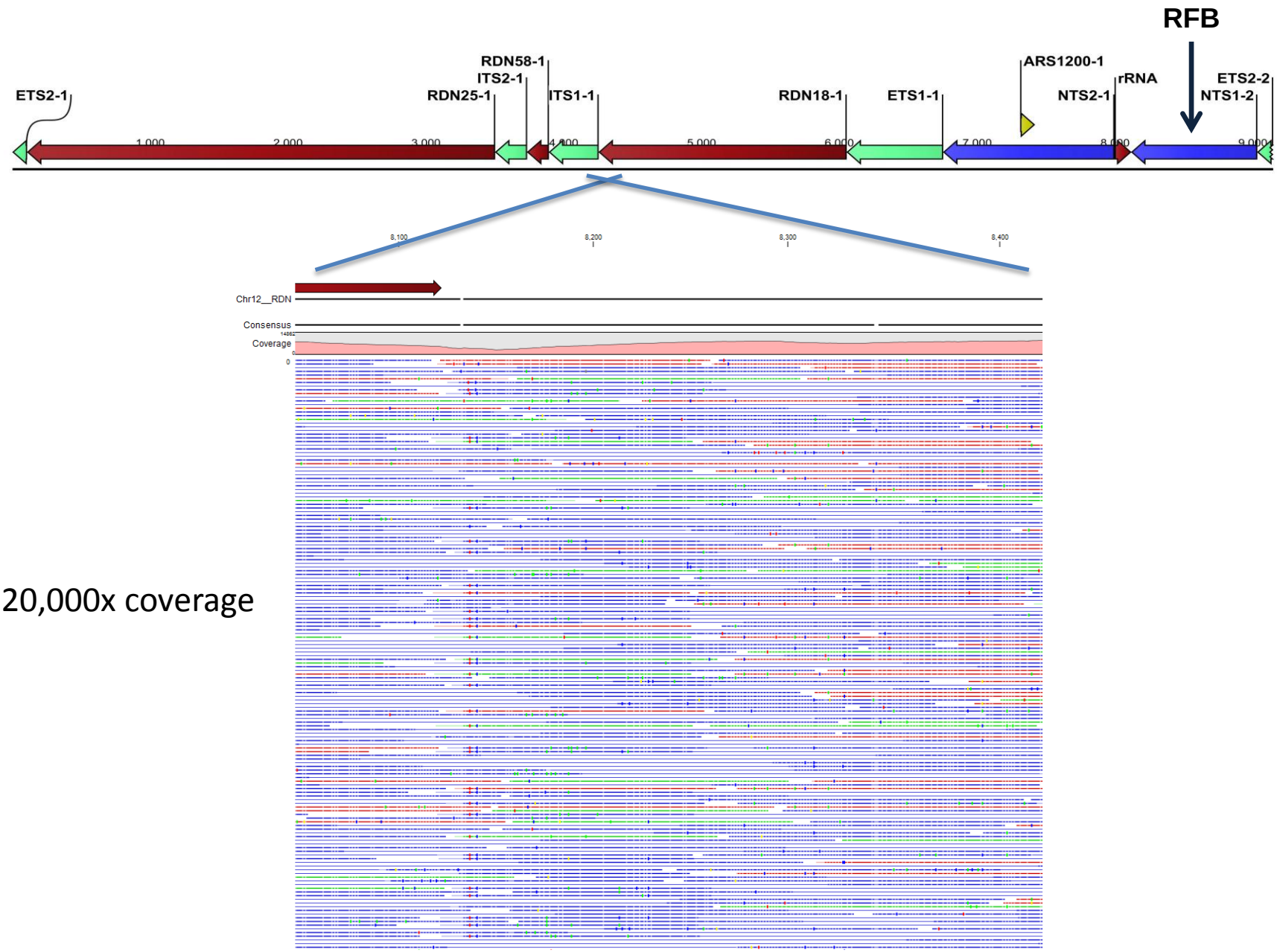

140,470,000
140,475,000
140,480,000
140,485,000
140,490,000
140,495,000
140,500,000
140,505,000
Homo sapiens (hg19) sequence
159,138,663bp
Homo sapiens (hg19)_CDS
CDS annotations (4,375)
BRAF exomes
0
cf__1_1h_ATCACG_L001_R1_001
(paired) trimmed (paired) (Reads)
without duplicates
207,411 reads
Exome
Capture
150
247
0
WGS


7,572,800
7,573,000
7,573,200
7,573,400
Homo sapiens (hg19) sequence
81,195,210bp
Homo sapiens (hg19)_CDS
CDS annotations (6,684)
0
cf__1_1h_ATCACG_L001_R1_001
(paired) trimmed (paired) (Reads)
without duplicates
217,612 reads
Exome
Capture
Exome of p53
118
139
WGS

Summary from capture:

- We have substantial number of duplicate reads after capture. We can increase input DNA for library prep from 2 to 10ng.
- We can improve quality of the reads by implementation of the Molecular Tags Molecular Unique Identifiers) into protocol and overlap sequencing and collapsing into consensus read.





20,000x coverage

Errors in the system

normal mutant

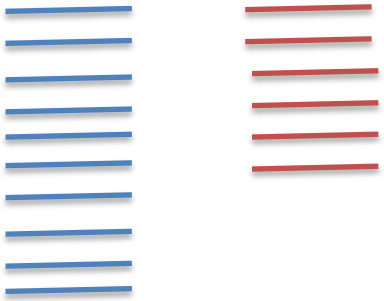


normal mutant



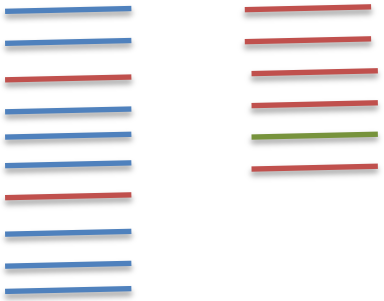
PCR bias

normal mutant

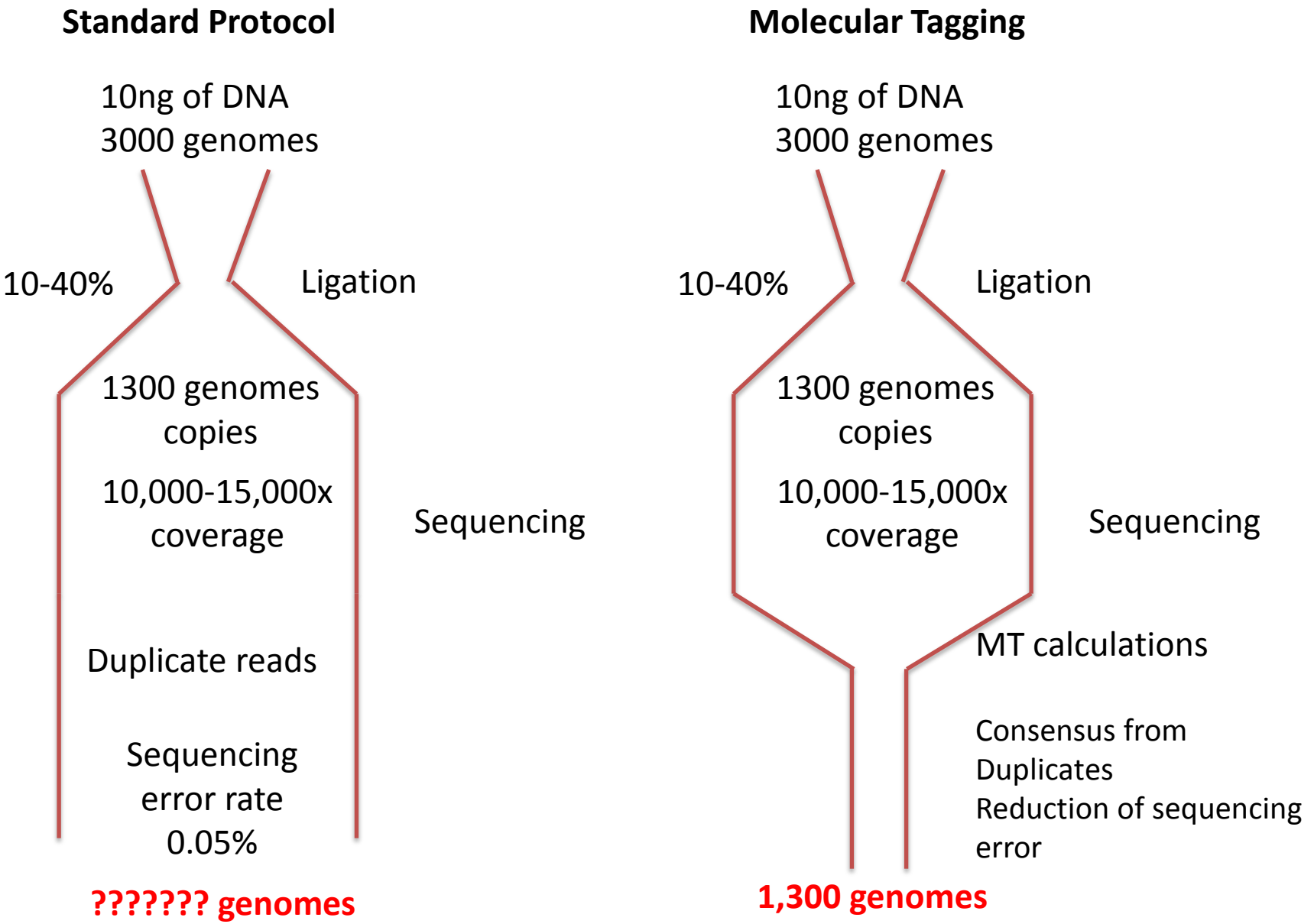


PCR errors

normal mutant



Bottlenecks



We can use a library containing Molecular Tags (Molecular Unique Identifiers) for both Exome Capture and Amplicon Sequencing to increase sensitivity of mutation detection

Prepare MT Sequencing Libraries



Duplex adapters *

Modified TruSeq adapters

Single stranded DNA assay

Amplicon sequencing of selected targets

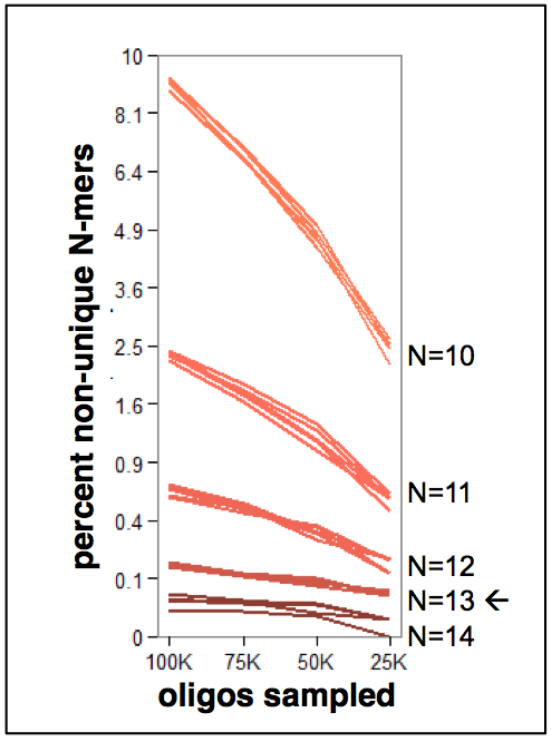
Detection of ultra-rare mutations by next-generation sequencing
Michael W. Schmitt, Scott R. Kennedy, Jesse J. Salk, Edward J. Fox, Joseph B. Hiatt, Lawrence A. Loeb
Proc Natl Acad Sci U S A. 2012 September 4; 109(36): 14508–14513. Published online 2012 August 1. doi: 10.1073/pnas.1208715109
PMCID: PMC3437896

Molecular Tags (MT) – Amplicon Strategy

```
                sequencing primer                          MT/FS                    515F
1) GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG NNNNNNNN  GAGTGCCAGCMGCCGCGGTAA  →

                                  806R                  MT/FS              sequencing primer
      1) ←  TAATCTWTGGGVHCATCAGGCA      NNNNN TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG
```

Duplicate tags



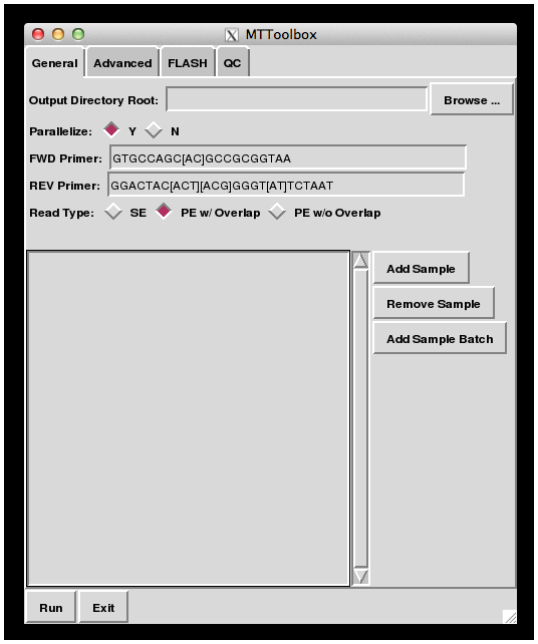
- **Tag individual DNA templates with a random oligo before PCR and sequencing**
- **Reduces sequencing errors and PCR bias**

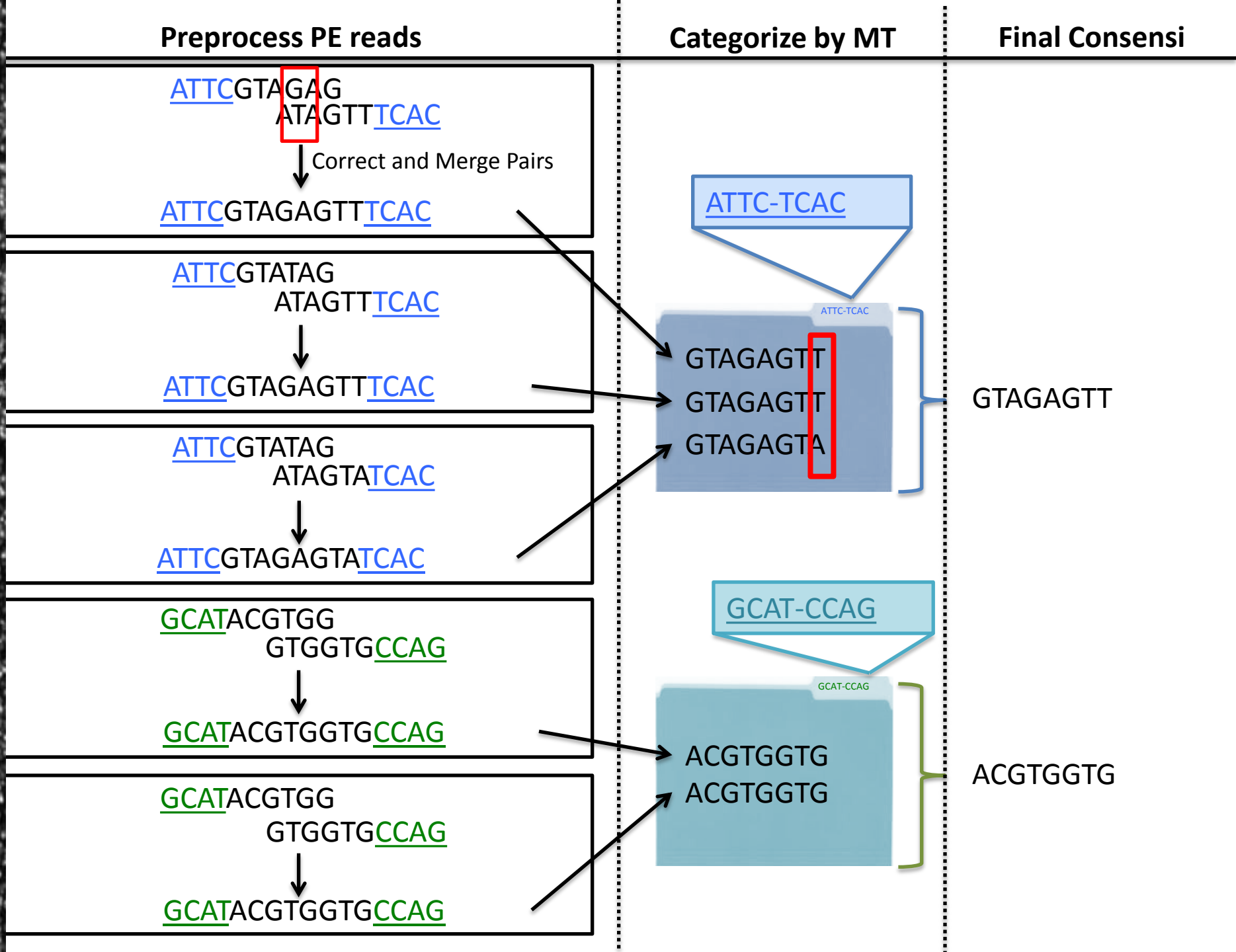
GGTC ATAGGTCAGAT
GGTC ATAGGTCGCAT
GGTC ATAGGTCAGAT
GGTC ATAGGTCAGAT

The **birthday paradox** concerns the probability that, in a set of *n* randomly chosen people, some pair of them will have the same birthday. The probability reaches 100% when the number of people reaches 367 (since there are 366 possible birthdays, including February 29). However, 99% probability is reached with just 57 people, and 50% probability with 23 people.

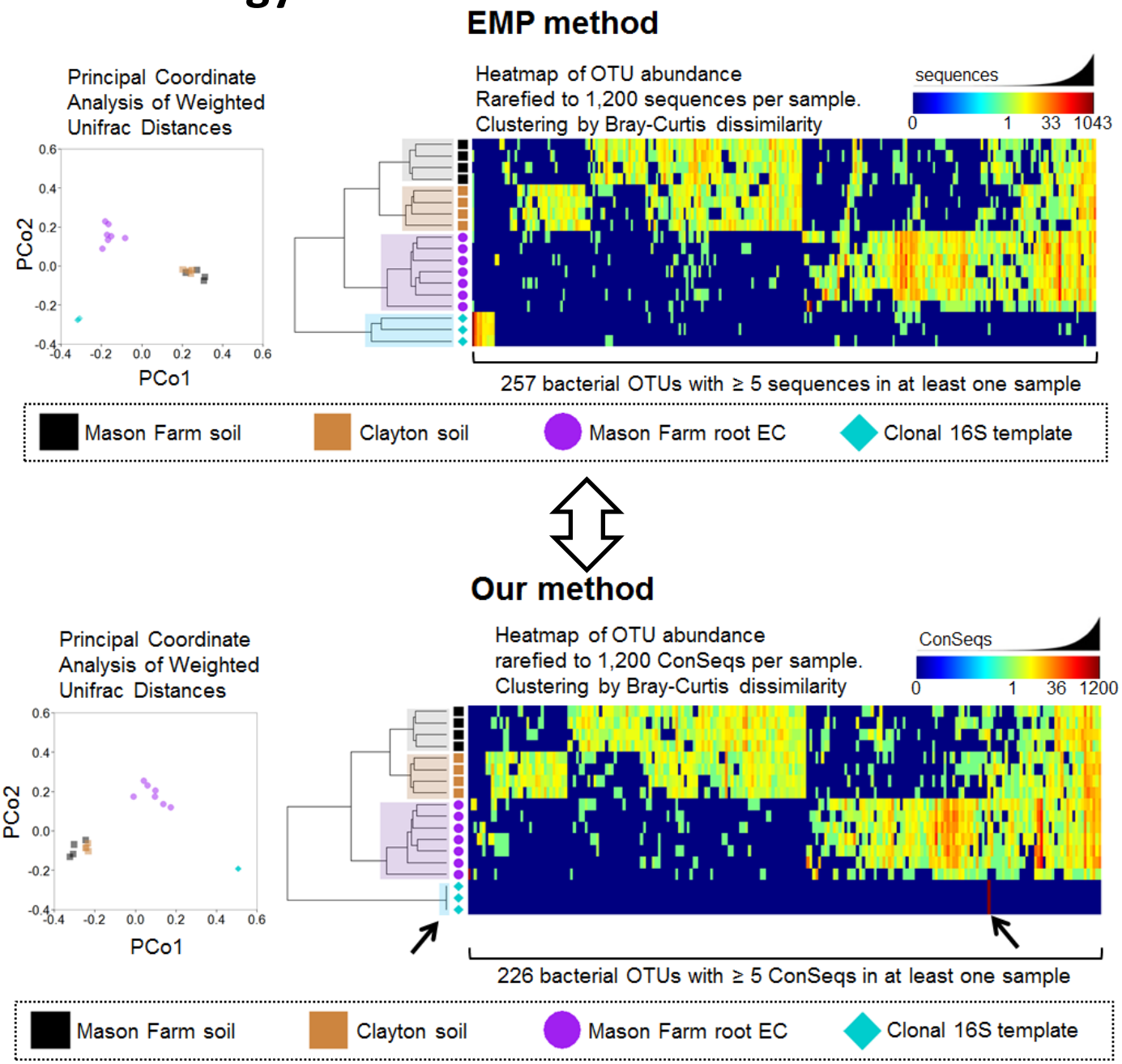
MTToolbox

- https://sites.google.com/site/moleculetagtoolbox/
- Parallelizable
  - A 96 sample run takes ~1 hour
- GUI
- Extended Edition
  - Build OTUs (OTUpipe)
  - Remove host contaminants (BLAST+)
  - OTU taxonomy assignments (RDP Classifier/QIIME)

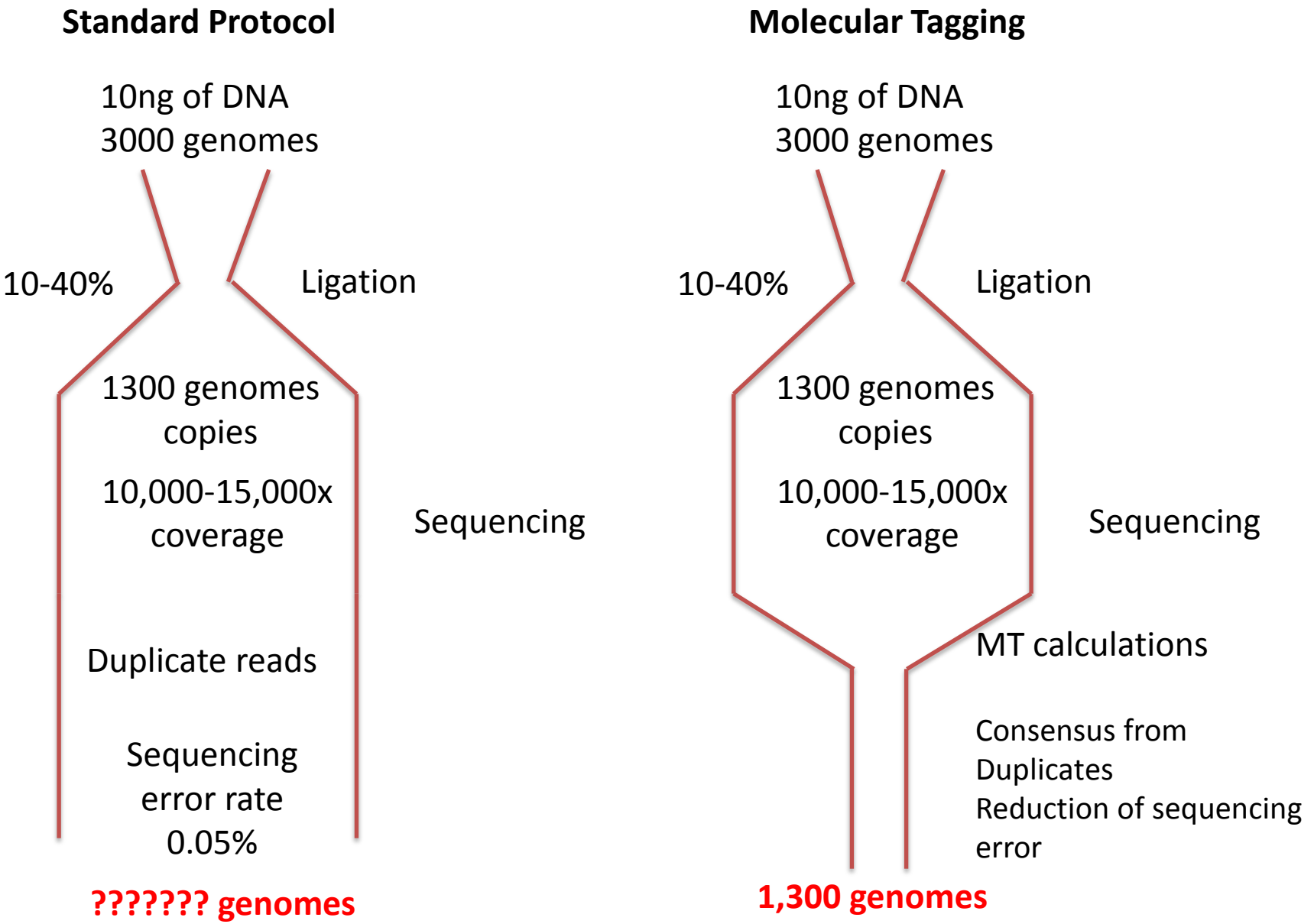



Preprocess PE reads
Categorize by MT
Final Consensi
ATTCGTAGAG
ATAGTTTCAC
Correct and Merge Pairs
ATTCGTAGAGTTTCAC
ATTCGTATAG
ATAGTTTCAC
ATTCGTAGAGTTTCAC
ATTCGTATAG
ATAGTATCAC
ATTCGTAGAGTATCAC
GCATACGTGG
GTGGTGCCAG
GCATACGTGGTGCCAG
GCATACGTGG
GTGGTGCCAG
GCATACGTGGTGCCAG
ATTC-TCAC
ATTC-TCAC
GTAGAGTT
GTAGAGTT
GTAGAGTA
GTAGAGTT
GCAT-CCAG
GCAT-CCAG
ACGTGGTG
ACGTGGTG
ACGTGGTG

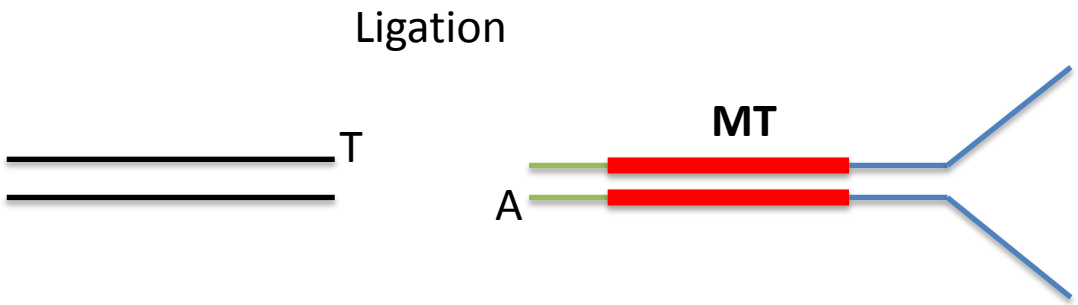
Amplicon Strategy



Bottlenecks - Exome Capture Strategy

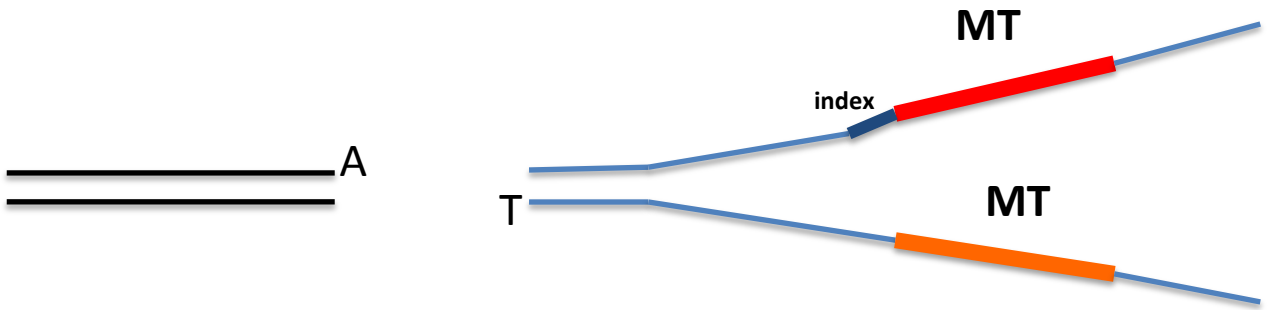


Adapters containing Molecular Tags (MT) used for experiment

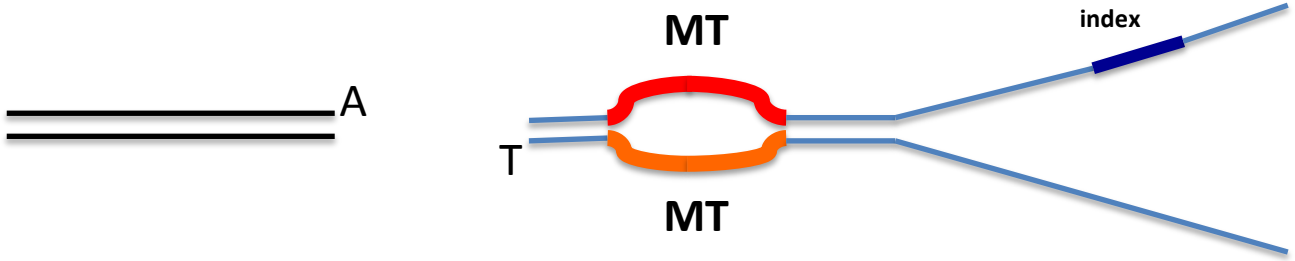


Duplex MT Adapter

Detection of ultra-rare mutations by next-generation sequencing
Michael W. Schmitt, Scott R. Kennedy, Jesse J. Salk, Edward J. Fox, Joseph B. Hiatt, Lawrence A. Loeb
Proc Natl Acad Sci U S A. 2012 September 4; 109(36): 14508–14513. Published online 2012 August 1. doi: 10.1073/pnas.1208715109
PMCID: PMC3437896

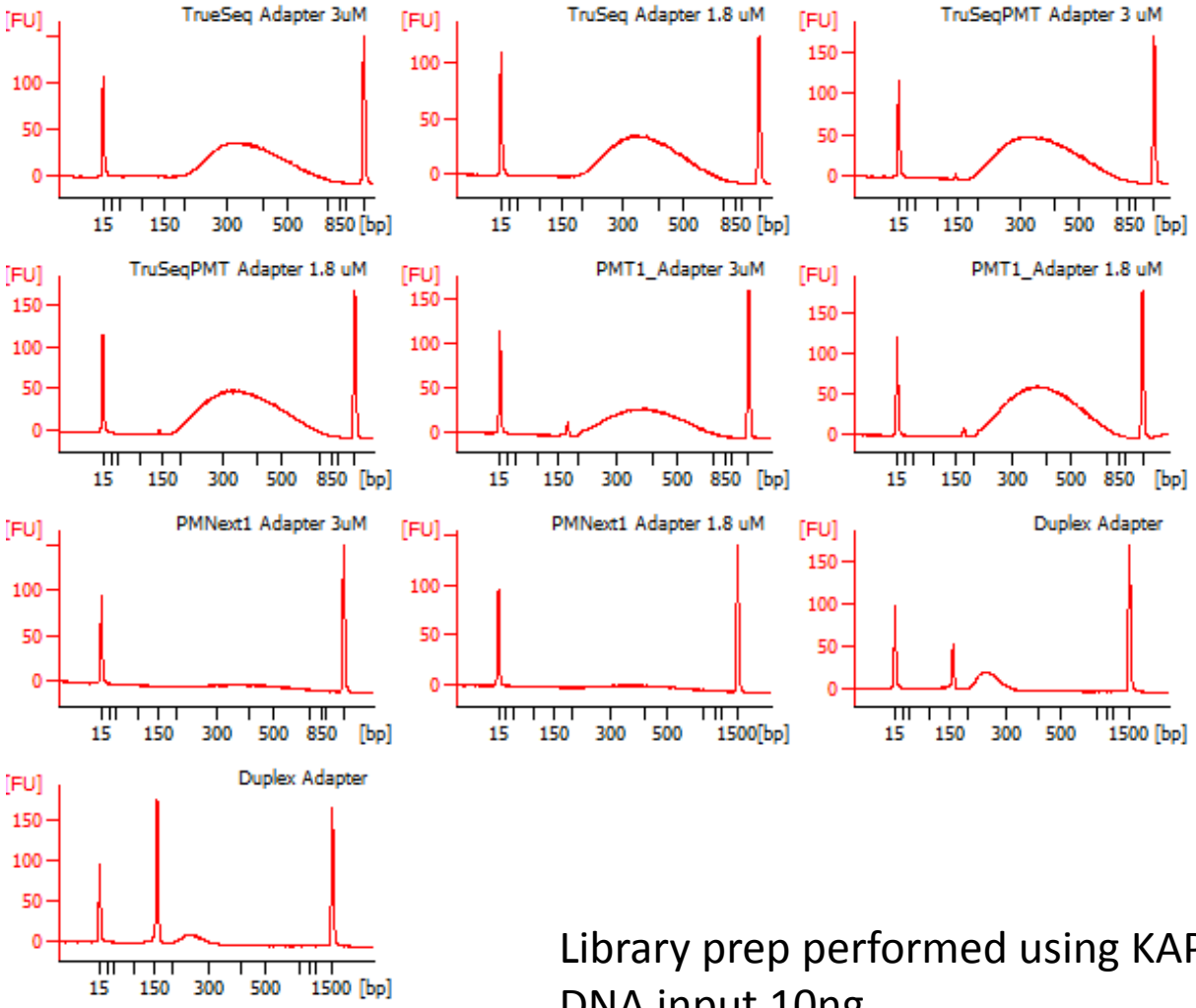


TruSeqPMT Adapter



PMT1 Adapter

Efficiency of Library construction using different types of adapters



Adapter		Library (ng/uL)	
TruSeq	3 uM	73	*
TruSeq	1.8 uM	65	*
TruSeq_PMT	3 uM	110	*
TruSeq_PMT	1.8 uM	108	*
PMT1	3 uM	39.8	*
PMT1	1.8 uM	39.2	*
PMNext1	3 uM	4.52	
PMNext1	1.8 uM	3.76	
Duplex	3 uM	9.34	
Duplex	1.8 uM	9.14	

*30 ng loaded on Bioanalyzer

Library prep performed using KAPA Hyper Library prep kit
DNA input 10ng

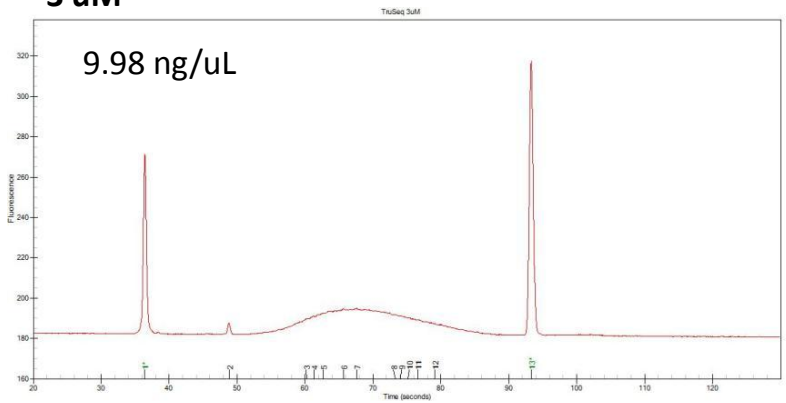
Input – 1 ng

3 uM

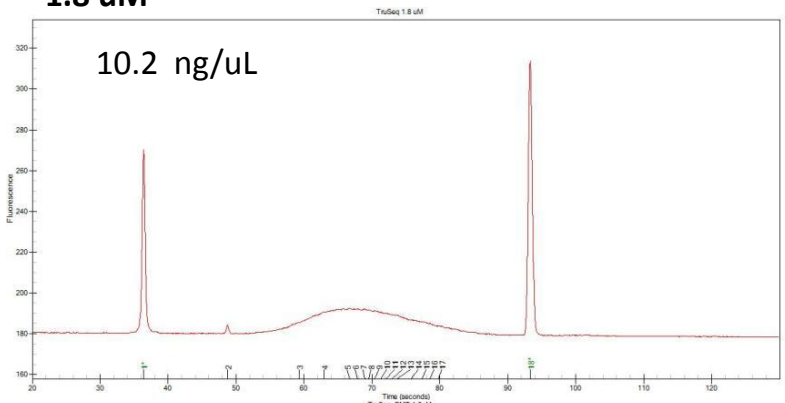
1.8 uM

TruSeq

9.98 ng/uL

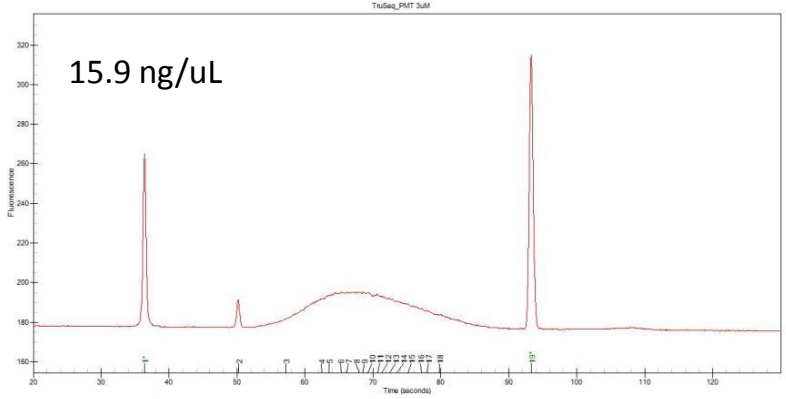


10.2 ng/uL

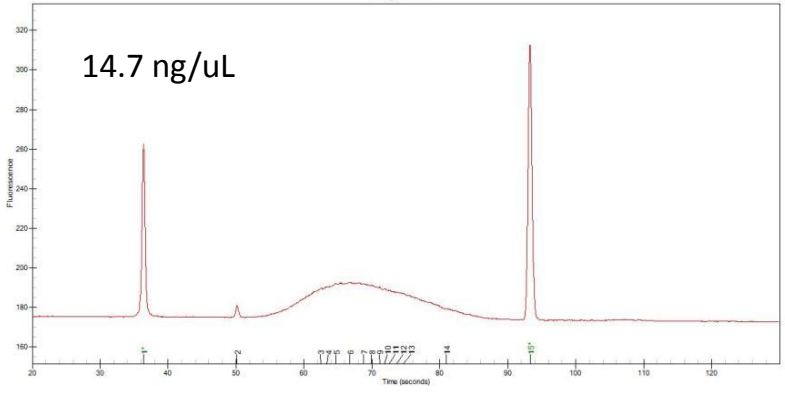


TruSeq_PMT

15.9 ng/uL

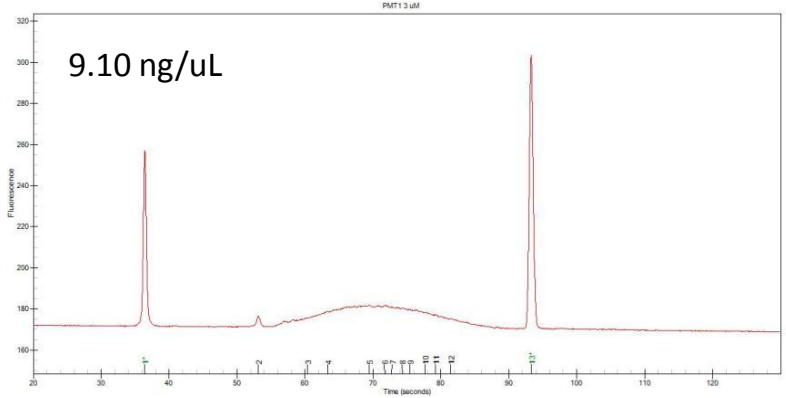


14.7 ng/uL

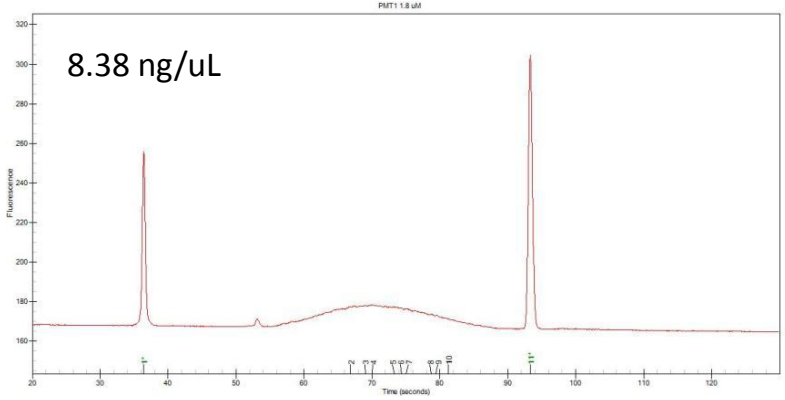


PMT1

9.10 ng/uL



8.38 ng/uL



Acknowledgements

Mieczkowski Lab + HTSF
Ewa Malc
Donghui Tan
Liz Sheffield
Maryam Clausen
Alicia Brandt
Nick Schuch
Uma Veluvolu
Scot Waring
Tara Skelly

Hemant Kelkar
Tristan De Buysscher

Corbin Jones

Margaret L. Gulley

Tomasz Kozlowski

The UNCseq Team

(Earp, Hayes, Sharpless, Grilley-Olson)

Over 30 individuals involved at LCCC

NATURE METHODS

Practical innovations for high-throughput amplicon sequencing

Derek S Lundberg[1,2,9], Scott Yourstone[1,3,9], Piotr Mieczkowski[4–6], Corbin D Jones[1–3,5,6] & Jeffery L Dangl[1,2,6–8]