Copy number analysis of circulating cell-free DNA

Venkatraman E. Seshan

Nicholas Socci Dana Tsui Julie Yang

Thanks: Ronglai Shen, Adam Olshen

MSKCC

November 5, 2018

BIRS 18w5202, Oaxaca

Cell-free DNA

- Cell free DNA (cfDNA) are fragmented DNA found in serum
- Nucleosome bound; typically $\sim$166 bases long
- These originate from cells that underwent apoptosis
- Doesn't stay in the bloodstream for long
- Could have originated from malignant tumor *i.e.* cancer
- Potential for clinical utility and hence "liquid biopsy"

cfDNA in Cancer

- Cancers are characterized by genetic changes
- Signature somatic mutations and copy number changes
  - mutations: APC in colon, VHL in kidney
  - copy number: ERBB2 in breast, MYC in neuroblastoma
- Ability to detect low frequency events: ddPCR, NGS
- However DNA from a cell is approximately 7.2 picograms
- 10 nanogram assay input has contribution from 1500 cells
- Median 134 pg/μL from 38 metastatic melanoma patients

  Volpone *et al* (2018)

  `https://doi.org/10.1016/j.ejca.2017.10.029`

Shallow Whole Genome Sequencing

- Paired end sequencing of cfDNA and aligned reference genome
- Shallow i.e. target 10 million read-pairs per sample
- Human genome is 3 billion bases so $\approx$3 fragments per Kb

Why? - Low cost. Estimate tumor fraction (TF).

- Very low TF - ddPCR
- Low TF - Deep sequencing of highly select gene panel
- Higher TF - Moderate sequencing of broader gene panel

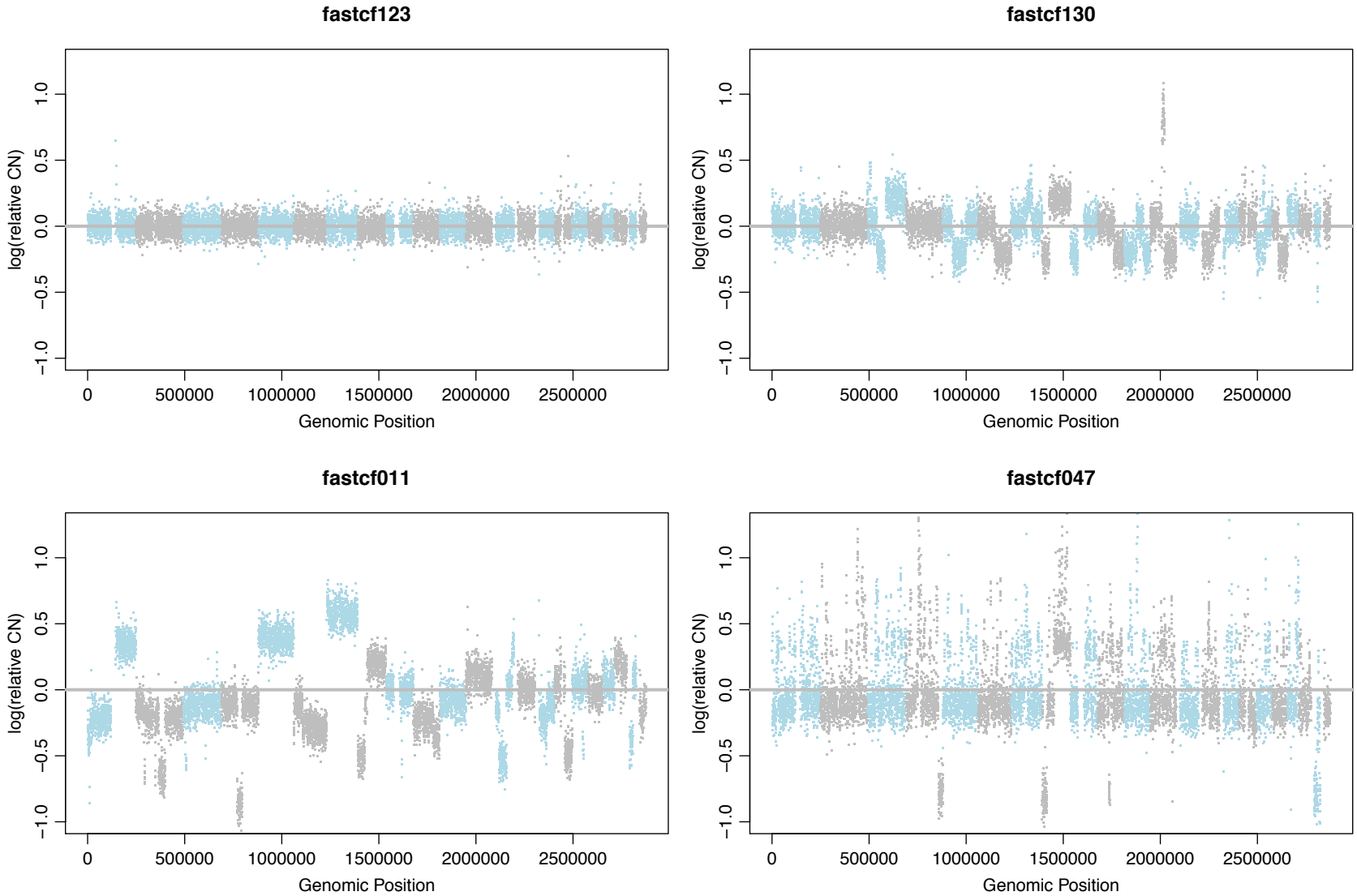
Copy Number Estimation from WGS

Solid tumors have large (whole chromsome or arm) copy number changes

- Read (fragment) counts overlapping bins obtained
- Counts are normalized for GC content and mappability
- Log-ratio of tumor (test) to normal (reference) computed
- Data are segmented (CBS, GLAD, HMM etc.)
- Regions of constant (relative) copy numbers

relative copy numbers $\longrightarrow$ integer copy number and tumor fraction

Some samples



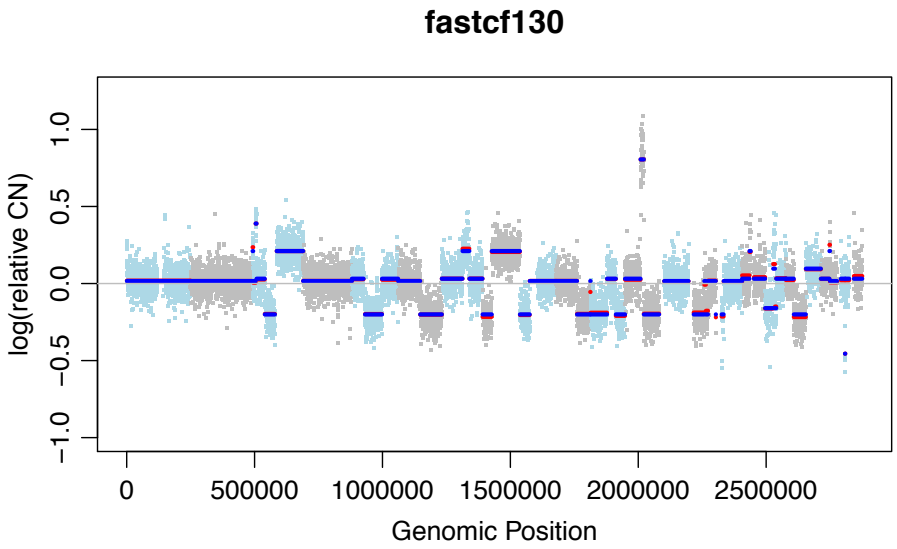
Data and Processing

- cfDNA from 118 samples were subject to shallow WGS
  - Bladder (45), Prostate (35), Germ-cell (10),
  - Breast (15), Lung (11), Other (2)
  - All had Stage 4 metastatic disease
- Paired end sequencing with a target of 10M read-pairs
- Unmatched cfDNA normal (5XX & 5XY) 2.5M read-pairs
- Data binned (100b), GC normalized, coarser (100-250k) bin
- Segmented using Circular Binary Segmentation

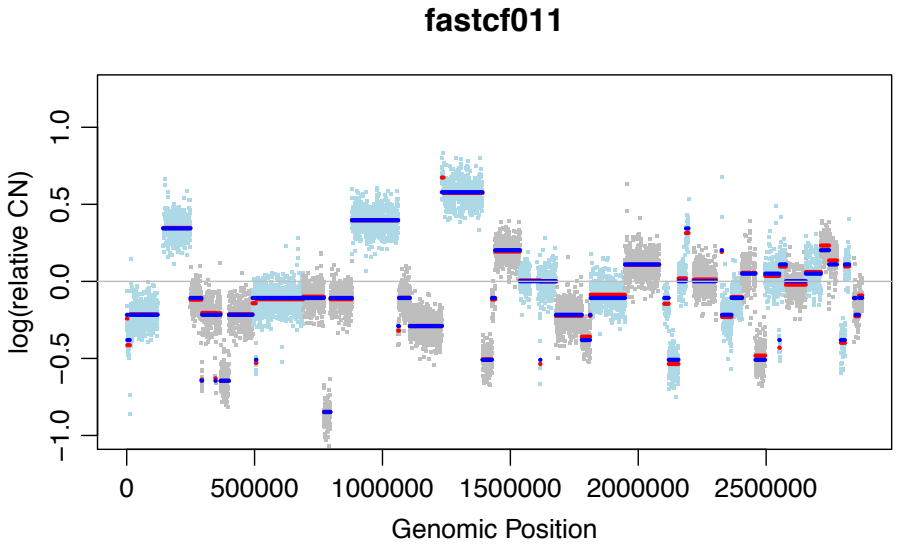
  Olshen *et al* (2004), Venkatraman and Olshen (2007)
- Modified BIC criterion to obtain distinct segment mean levels

  Zhang and Siegmund (2007)

Examples



fastcf130 has 55 segments; mbic reduces it to 6 (and 3 focal) distinct levels



fastcf011 has 56 segments; mbic reduces it to 14 distinct levels

Estimation Procedure

- Let $\mathcal{M}_1, \ldots, \mathcal{M}_k$ be the distinct segment mean levels
- Remove narrow (< 15Mb) segments with low and high means

  homozygous deletions and high level amplifications
- Let $\mathcal{M}_0$ be diploid level (some segments correspond to it)
- Let ρ_c be the clonal tumor fraction
- Then $\mathrm{E}(\mathcal{M}_i - \mathcal{M}_0)$ is $\log_2[\rho_c m_i + 2(1-\rho_c)] - 1$ where m_i is the integer copy number of the segment (between 1 and 5)
- Get $\mathcal{M}_0$ and ρ_c that minimize absolute deviations
- Fit a model with a single subclone fraction ρ_s ($< \rho_c$) for single copy gain or loss to allow for subclonal structure

Results

Estimated tumor fratcion below 5% in 68 samples, between 5% and 25% in 22 samples and above 25% in 28 samples

TF	Bladder	Prostate	GermCell	Breast	Lung	Other
< 0.05	29	14	3	10	10	2
$0.05 - 0.25$	7	7	4	4	0	0
> 0.25	9	14	3	1	1	0

Caveats: samples may not be representative of patient population

Over-representation of genitourinary cancer

Other methods

Plasma-seq Heitzer *et al* Genome medicine (2013)

`https://dx.doi.org/10.1186/gm434`

Computes a genomewide Z-score a mesure of how much local average of tumor read counts differs from normal

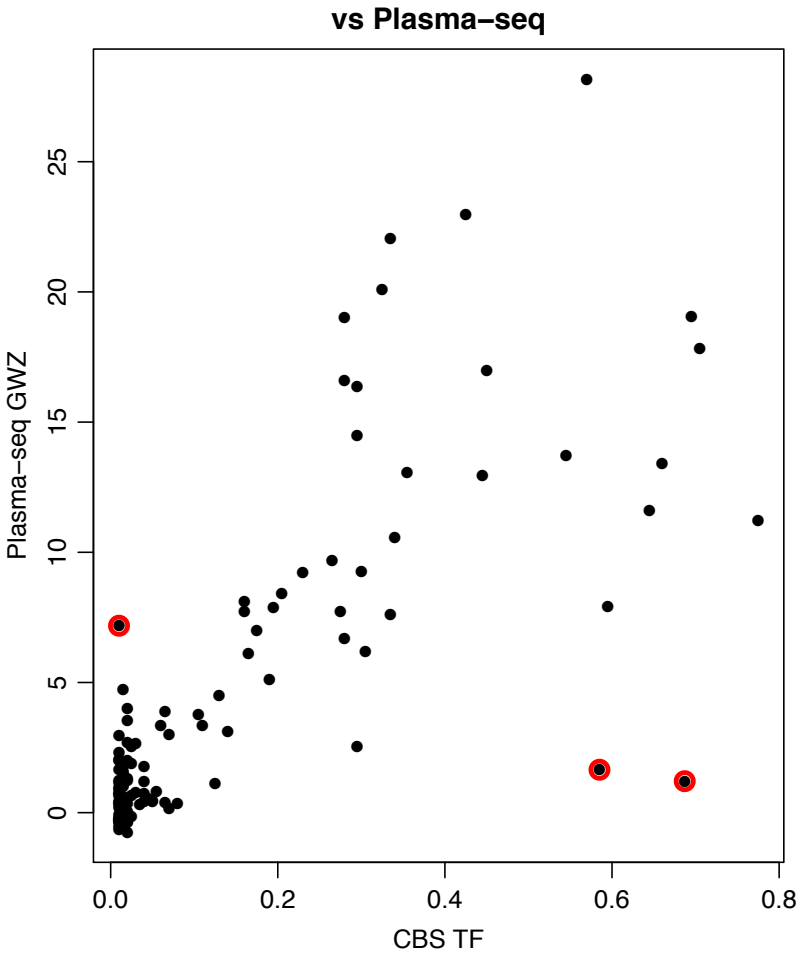
ichorCNA Adalsteinsson *et al* Nature communications (2017)

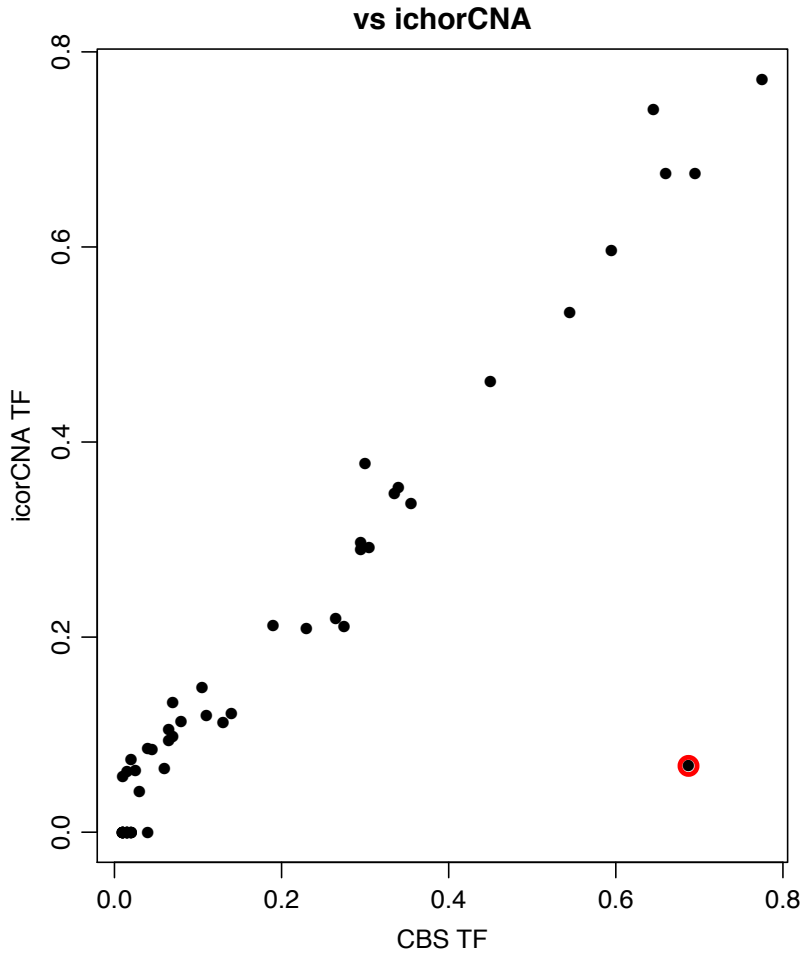
`https://www.nature.com/articles/s41467-017-00965-y`

Mixture model and HMM to segment and estimate copy number and tumor fraction simultaneously

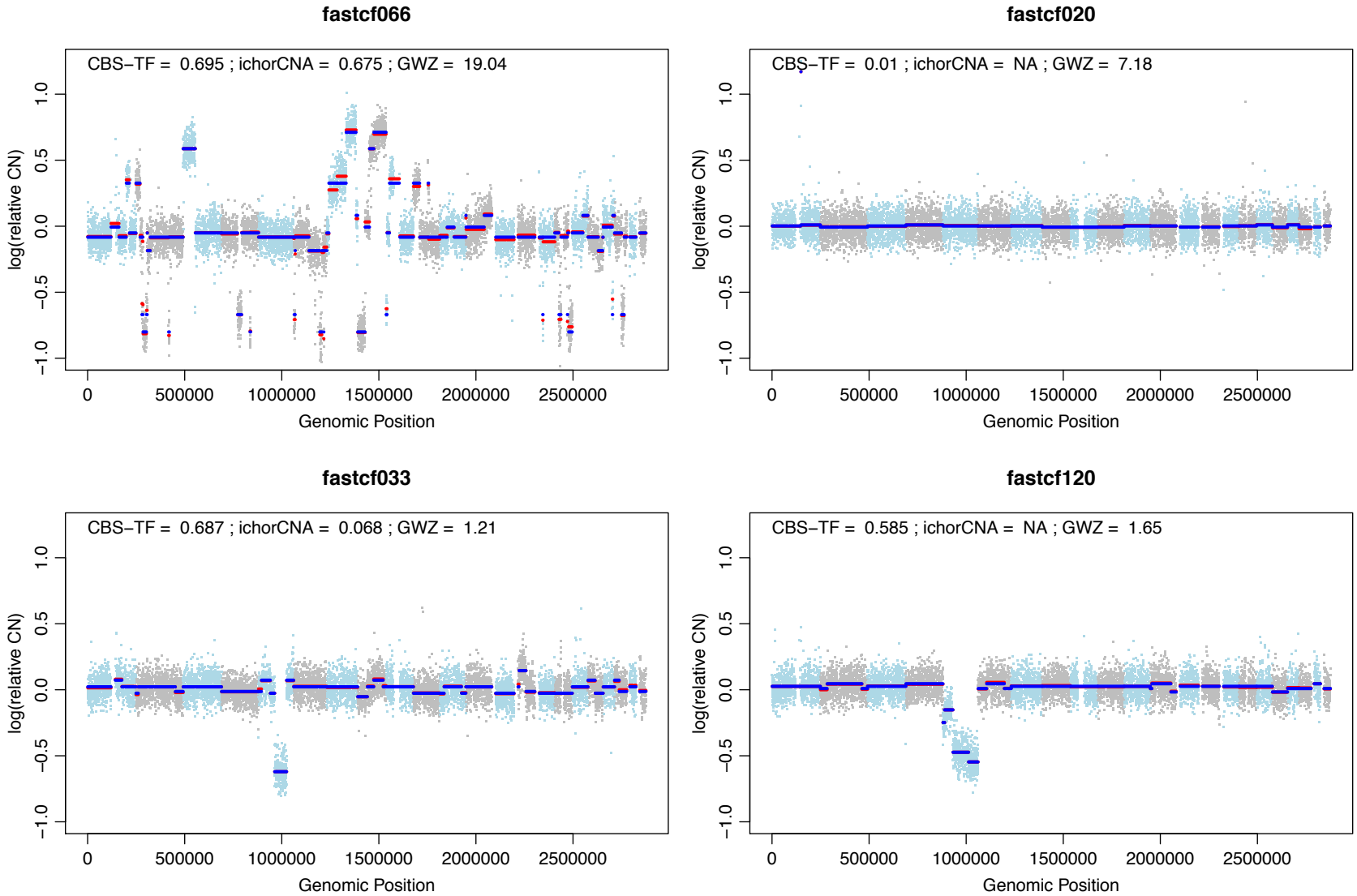
(only 55/118 were done)

Comparison


vs Plasma−seq
Plasma−seq GWZ
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
CBS TF


vs ichorCNA
icorCNA TF
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
CBS TF

Some samples



Targeted Sequencing

- Aimed at detecting somatic mutations
- High coverage at targeted regions 50x normal, $>$ 150x tumor
- Whole exome – all exons (coding part) of the genome

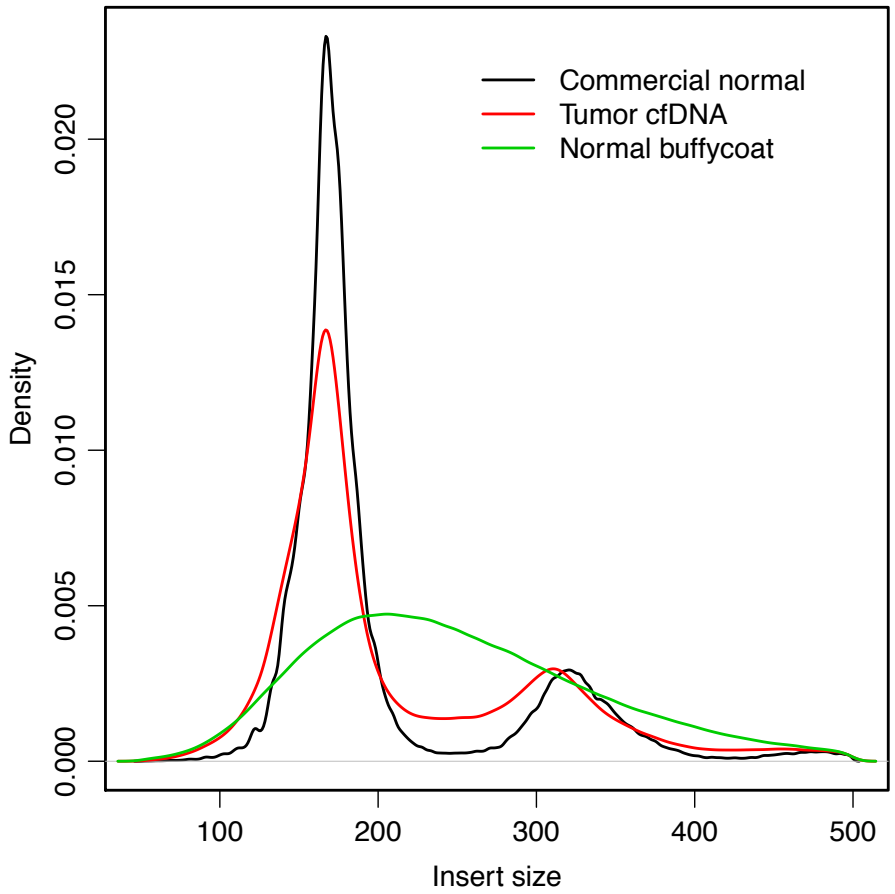
  targeted region cover $\sim$50 million bases
- Cancer gene panel
  - MSK-IMPACT $\sim$468 genes and 1 megabase
  - MSK-ACCESS $\sim$60 genes and 200 kilobases
  - GRAIL panel (don't have GUARDANT data)
- FACETS for allele specific copy number etsimation

FACETS

- ref and alt depths on a grid of SNPs and pseudo-SNPs
- log-ratio is the ratio of total depths
- log-odds-ratio for heterozygous loci (allelic imbalance)
- joint segmentation for regions of constant allelic copy number
- location of diploid state followed by copy number estimation
- tumor purity and ploidy

Shen and Seshan (2014) Nucleic Acids Research

Insert Size



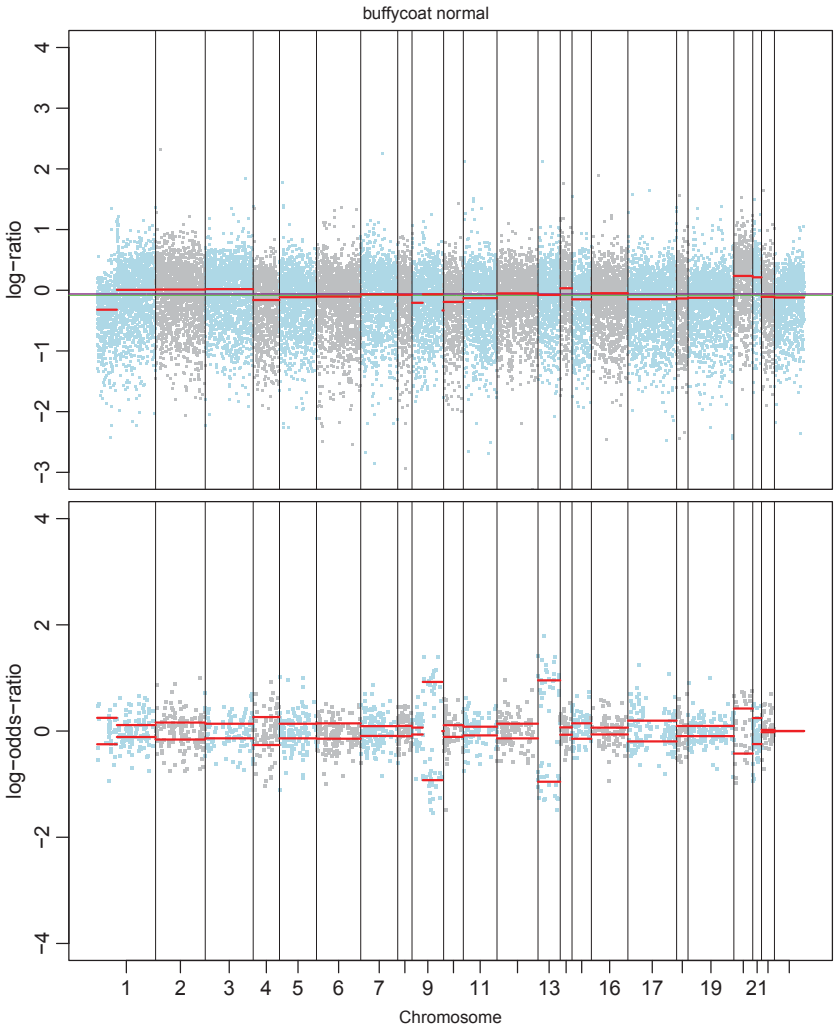
cfDNA expected to have two sharp peaks

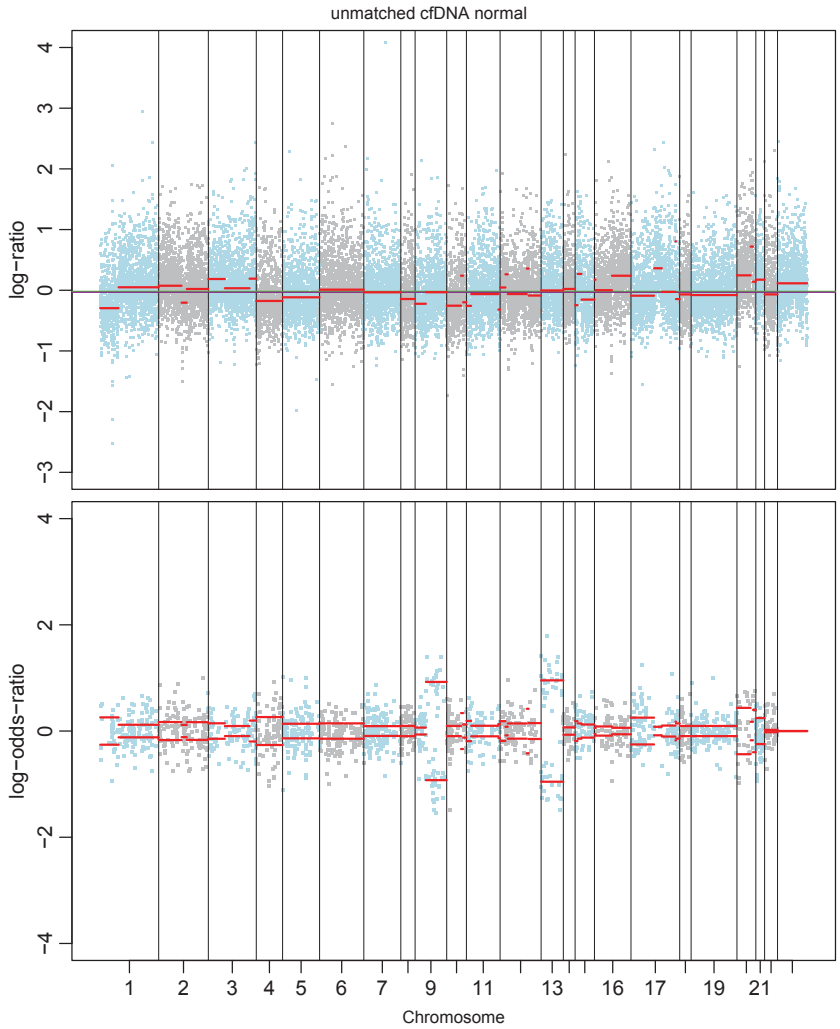
Genomic DNA is sheared and is more spread out

tumor cfDNA seems like a mixture

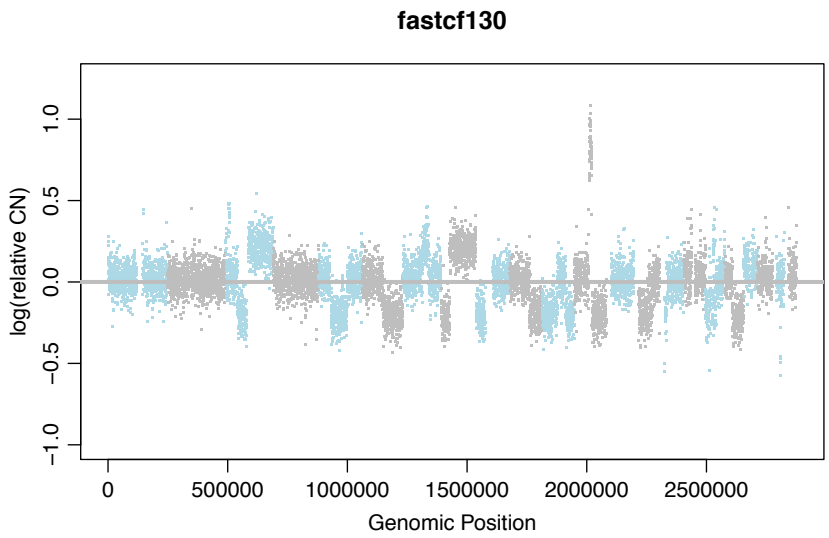
Insert size can affect copy number log-ratio

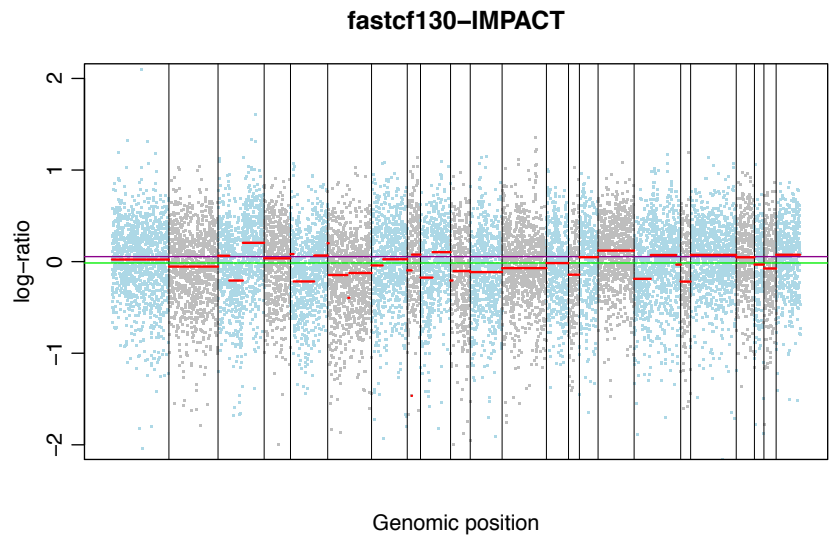
Log-ratio

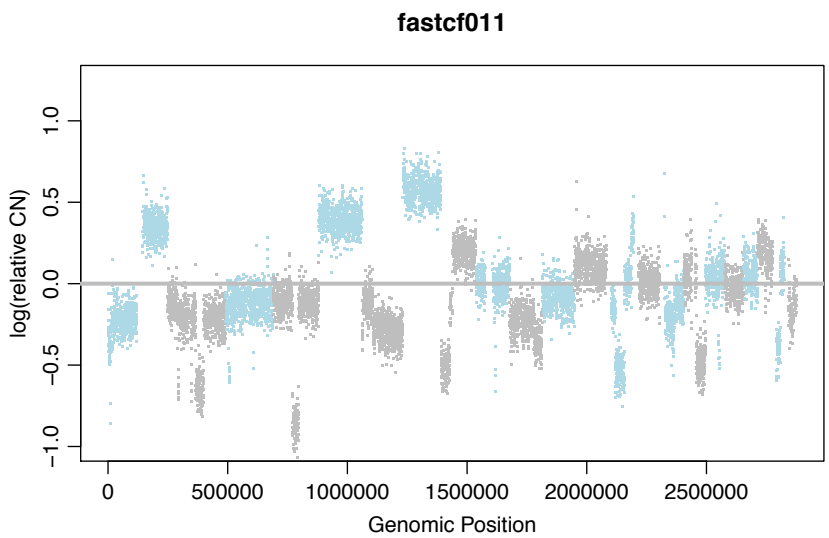

buffycoat normal
log−ratio
log−odds−ratio
Chromosome
1 2 3 4 5 6 7 9 11 13 16 17 19 21

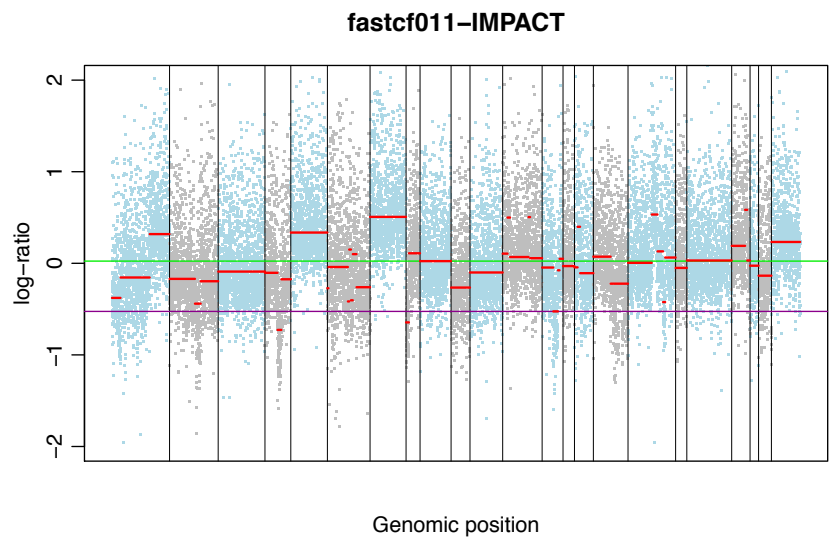

unmatched cfDNA normal
log−ratio
log−odds−ratio
Chromosome
1 2 3 4 5 6 7 9 11 13 16 17 19 21

sWGS vs IMPACT









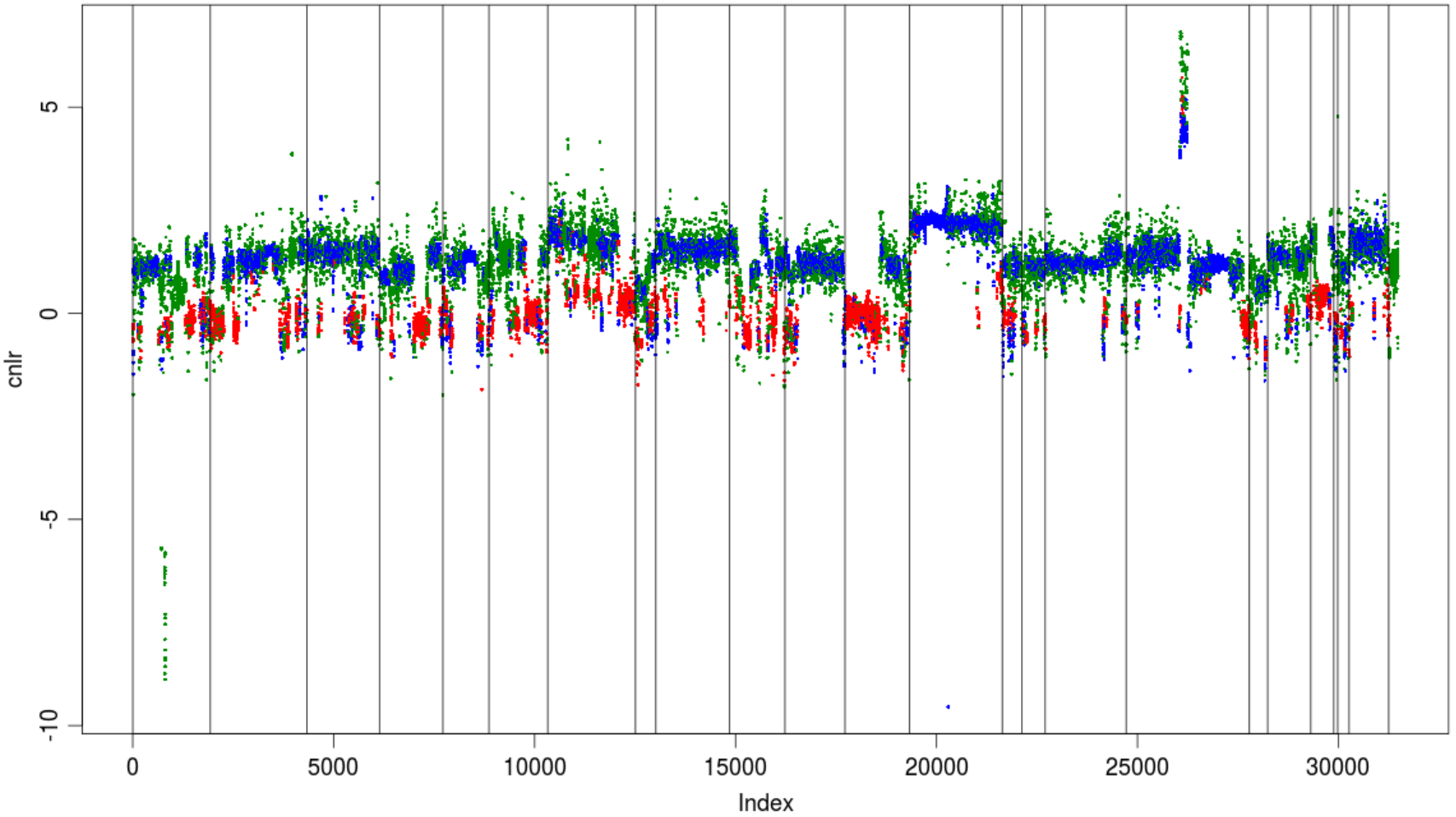
Ultra-deep Sequencing

- Goal to detect mutations with very low frequency, say 0.1%
- Sequencing with unique molecular identifiers (UMI)
- Typically 20-50k depth - fewer unique molecules
- 1500-5000 unique - collapse duplicates
- Beware of mutations in normal tissue (esophagus epithelium)

DOI: 10.1158/2159-8290.CD-RW2018-187

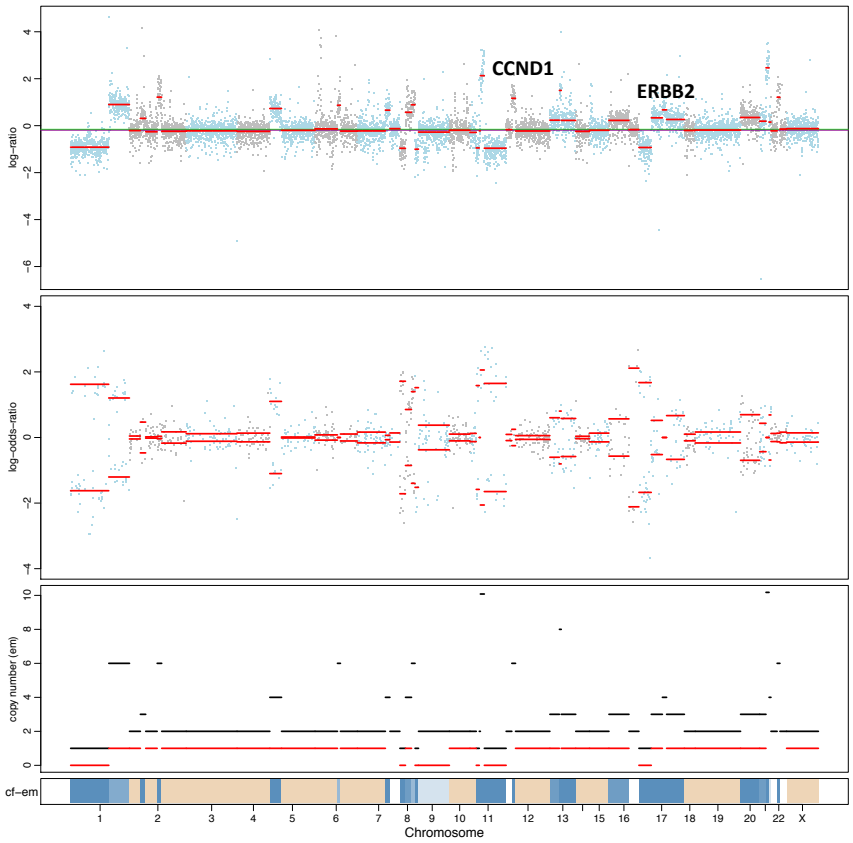
- MSK-ACCESS: targets with both 20k and 1k coverage

Differential Depth Issues



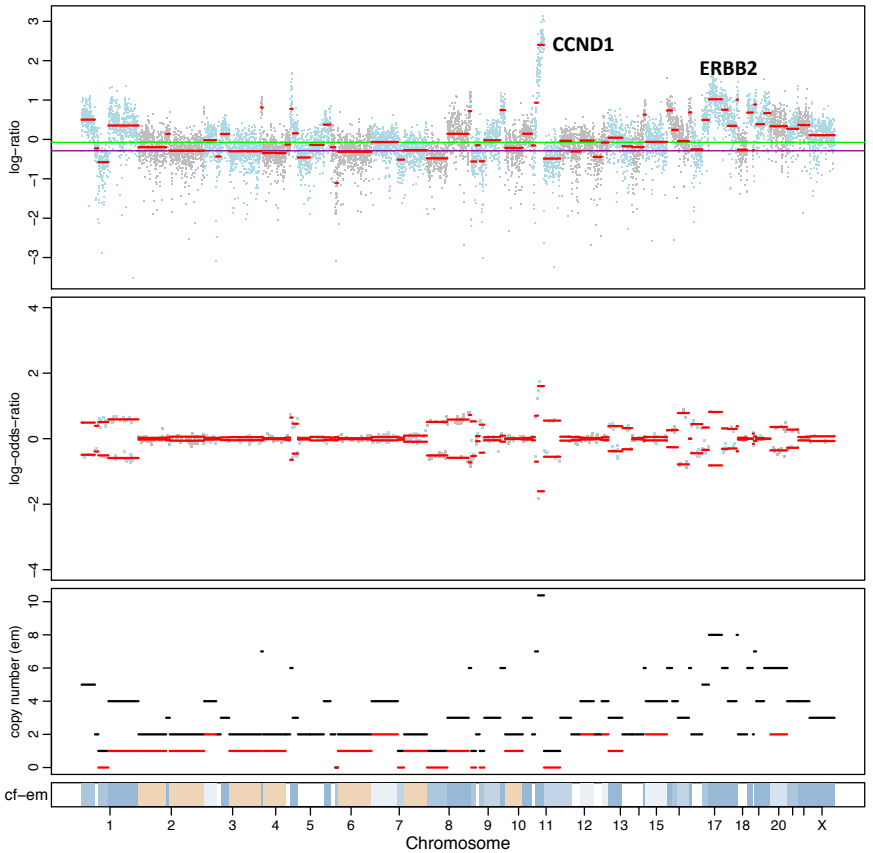
MSK-VB-0044

Tumor



Estimated tumor purity = 81%

cfDNA

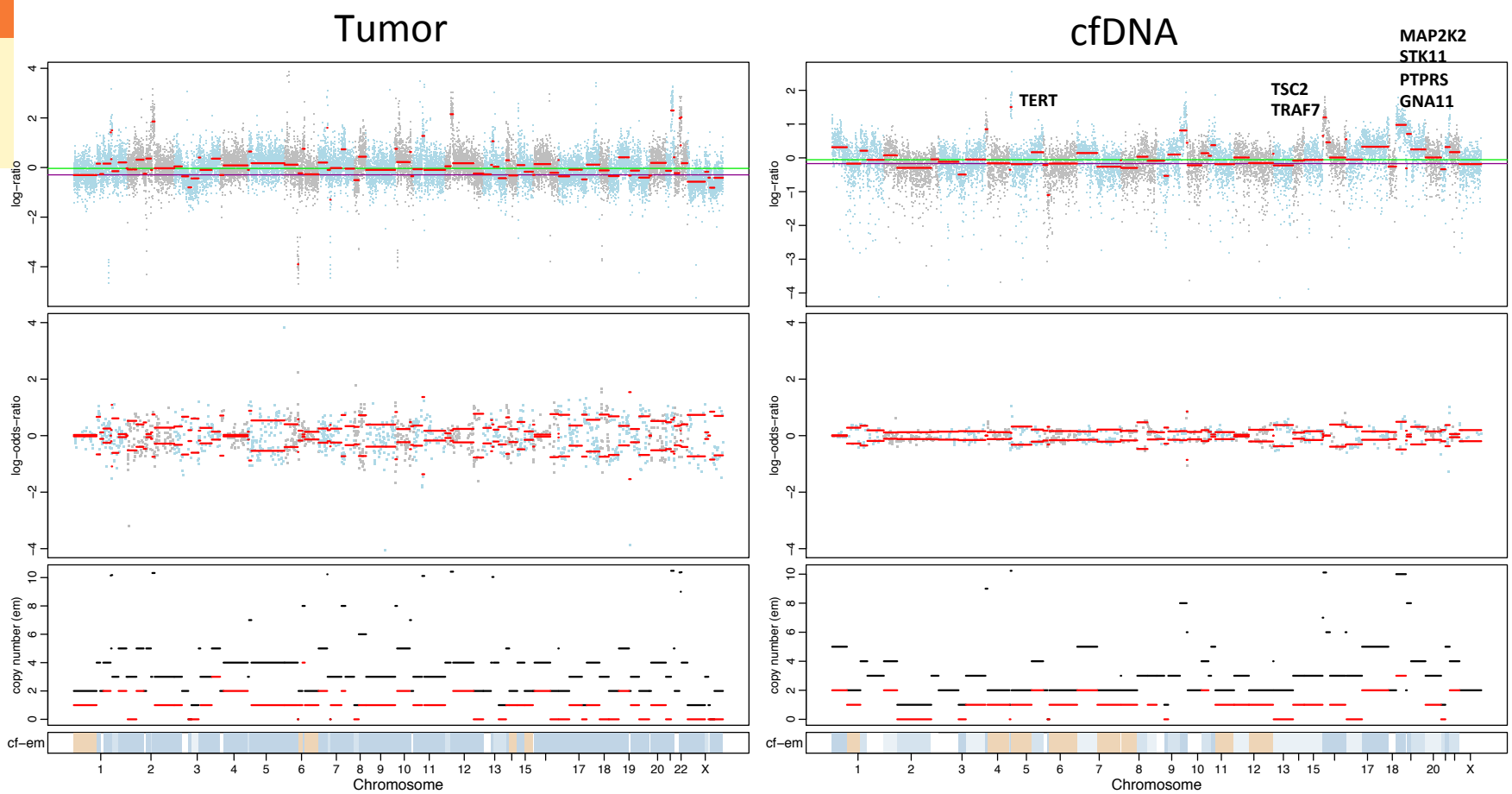


Estimated tumor purity = 52%

- ERBB2 higher level gain detected in cfDNA.
- Ploidy shift from 2.06 in the tumor to 2.58 in cfDNA

Memorial Sloan Kettering Cancer Center

MSK-VB-0029



Several focal amp in cfDNA not detected in the tumor:

- 5p (TERT) likely subclonal in the tumor
- 16p (TSC2, TRAF7)
- 19p (MAP2K2, STK11, PRPRS, GNA11)

Memorial Sloan Kettering Cancer Center

Summary

- cfDNA presents a variety of challenges
- Method to estimate tumor fraction from shallow WGS that Compares well to existing methods
- Helps in followup assay to use as tumor fraction is varied even in metastatic cancer patients
- R package work in progress

Ackowledgements: Margaret Steiner (2018 QSURE intern)

Erica Gedvilaite and Center for Molecular Oncology collaborators

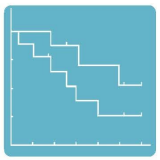
Memorial Sloan Kettering Cancer Center QSURE
Quantitative Sciences Undergraduate Research Experience

SUMMER INTERNSHIP

MSKCC| New York, NY| Summer 2019

Hands-on quantitative research experience in cancer through
Biostatistics, Epidemiology and Health Outcomes Research

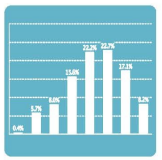
POTENTIAL RESEARCH AREAS


Survival Analysis


Prediction Modeling


Clinical Trials


Genetics & Genomics


Cancer Epidemiology

For details:
https://www.mskcc.org/departments/epidemiology-biostatistics/student-intern

Supported by the National Cancer Institute,
R25 CA214255