

Early Lung Cancer Detection using ctDNA shedding model

Harinaryana Mellacheruvu¹, Alex Gorelick², and Johannes Reiter

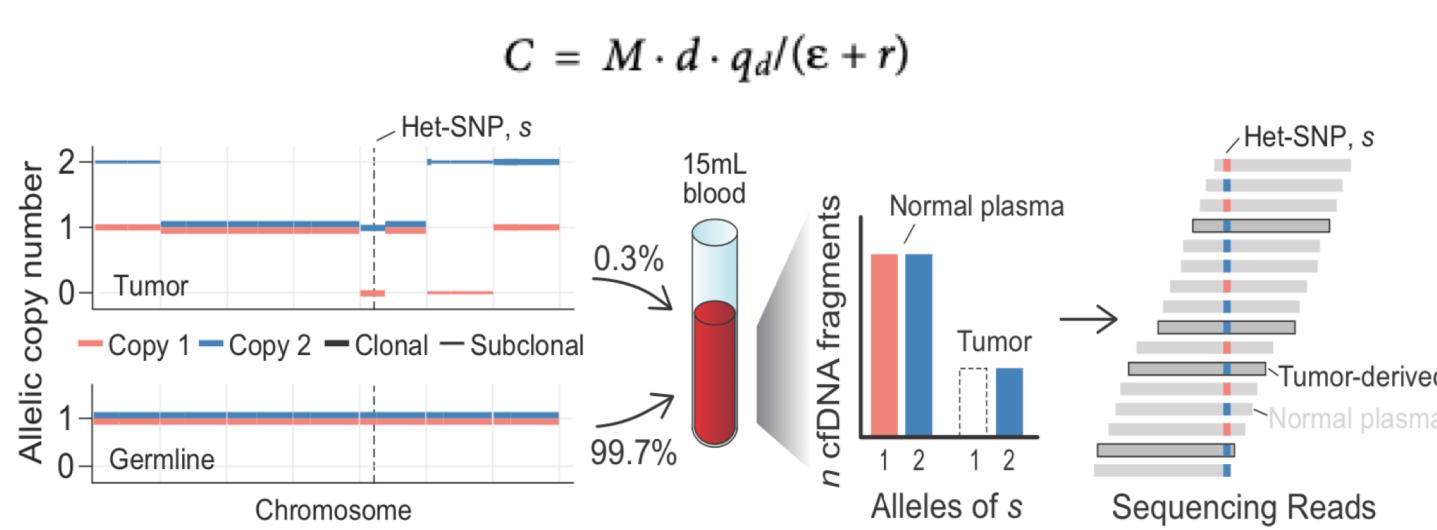
¹Santa Clara University, Santa Clara, CA; ²Stanford University, Stanford, CA

Introduction

- Over 50% of lung cancer cases are diagnosed late stage
- Circulating Tumor DNA (ctDNA) is a promising biomarker that can be obtained non-invasively
- Mutations found through sequencing ctDNA highly correlate with those found in primary tumor
- Sequencing also allows detection of copy number alterations which are prevalent across numerous cancers
- Quantity of ctDNA is indicative of cancer stage, progression, treatment response, and disease recurrence

Materials & Methods

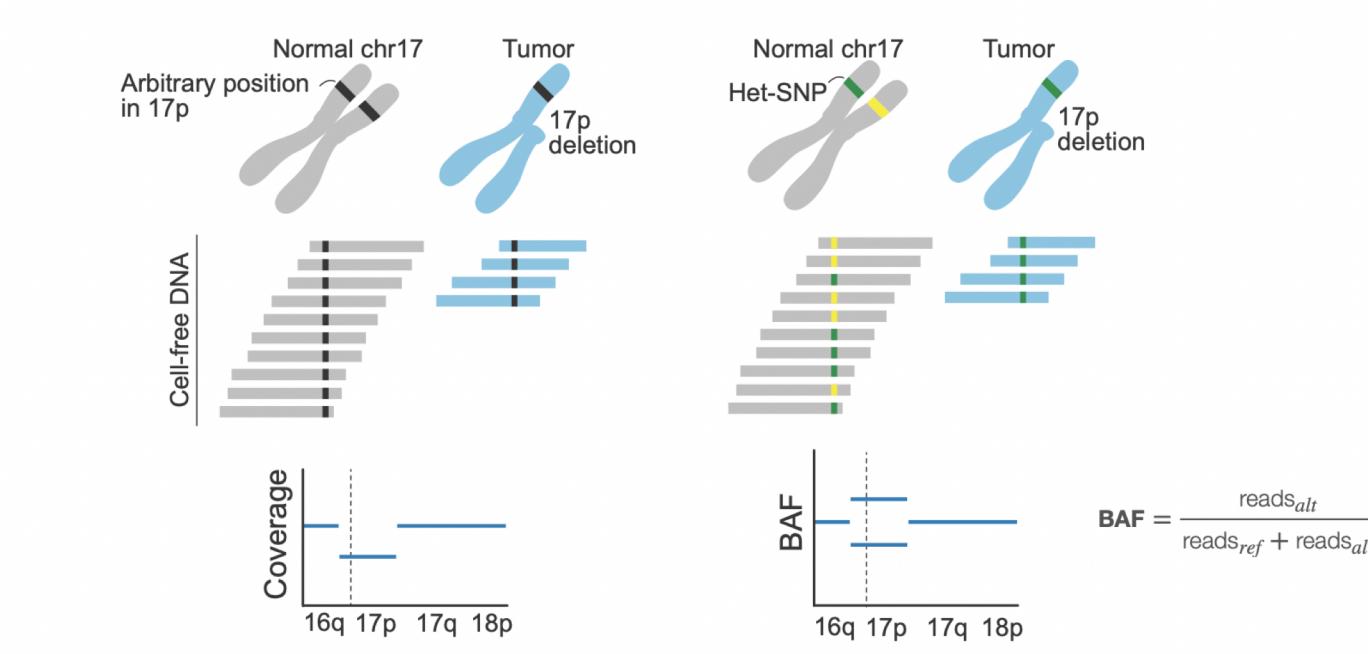
Intuition for detecting CNVs from cfDNA reads at heterozygous SNPs



Step 1: ctDNA read simulation at heterozygous SNPs. Copy Number Profiles from TCGA for lung adenocarcinoma patients in conjunction with whole genome from 1000 Genomes cohort used to simulate tumor reads for alleles 1 and 2 as according to tumor parameter input: Tumor size, diameter, tumor fraction. Reads simulated used multinomial distribution.

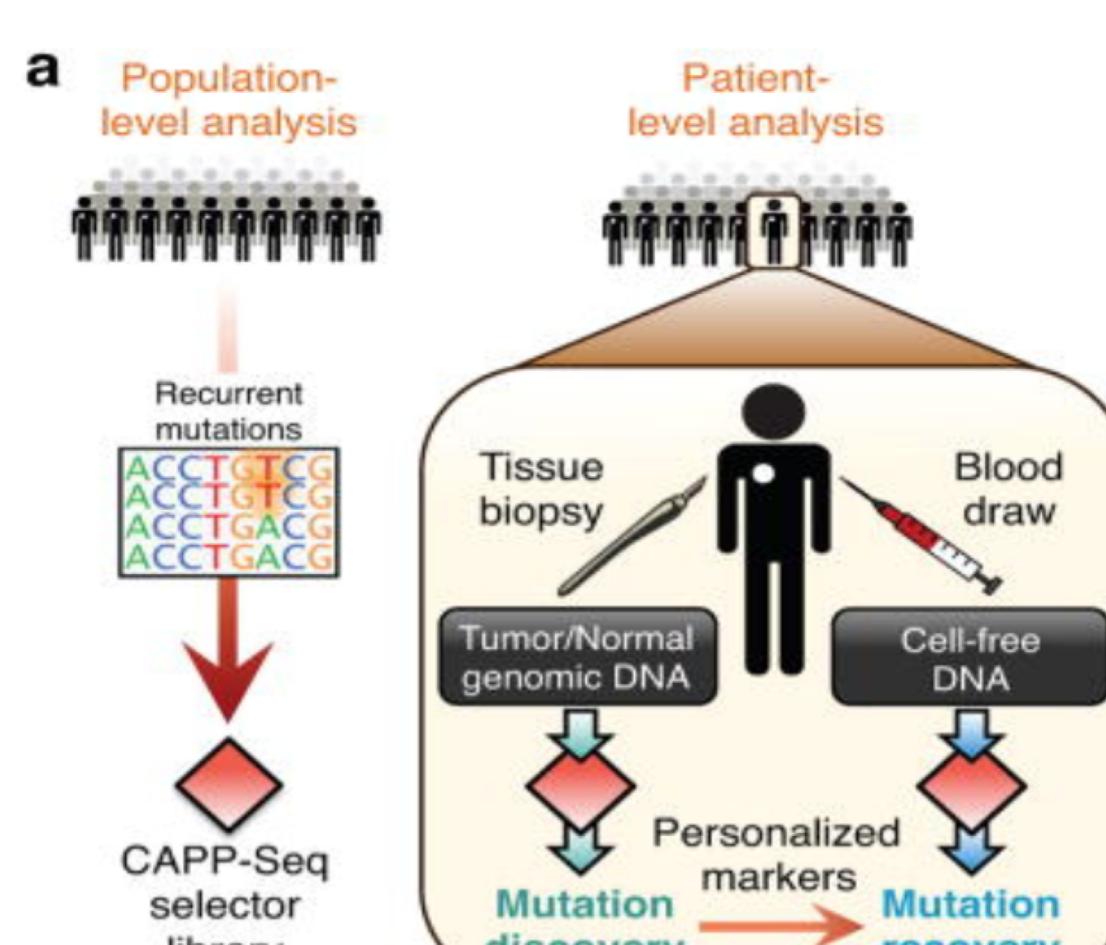
Step 2: Computational Phasing of Het-SNPs into haplotypes 1 and 2.

At heterozygous SNPs, copy-number variants provide cancer-specific signals detectable in cfDNA



Step 3: Calculation of B-Allele frequency (BAF), which is ratio of allele frequencies. Reads summed across SNPs found on chromosome arm to calculate BAF across the chromosome arm. Significant difference in BAF from 0.5 indicative of copy-number alterations. BAF calculated for chromosomes 1-22.

Step 4: Perform Steps 1 and 2 for patients from which copy number profiles obtained to generate dataset with B-allele frequencies



Step 5: ctDNA shedding model used to determine the likelihood of mutant DNA fragment shed into plasma using regions covered by CAPP-Seq panel and average number of mutations in these regions (6 somatic mutations).

Results: Statistical Significance

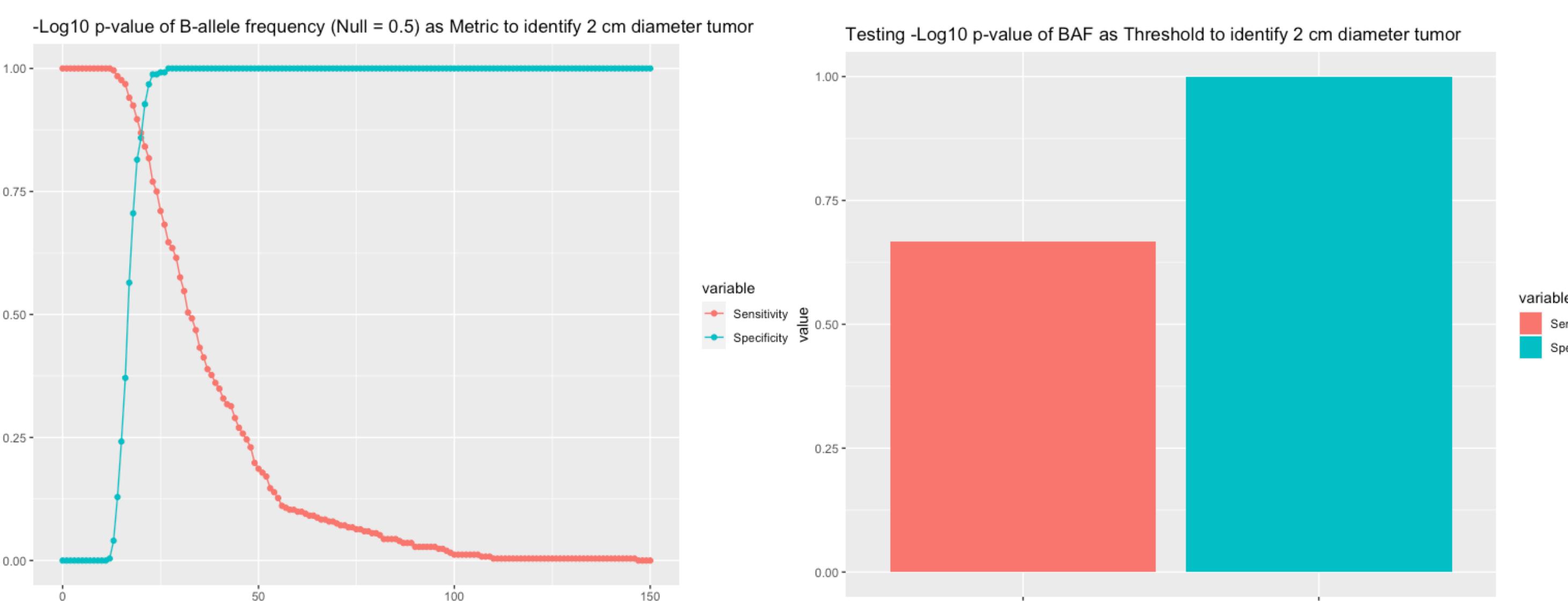


Figure 1: A Crude Metric to identify 2 cm diameter tumors. Binomial test performed comparing number of reads for haplotype 1 for chromosome arm to number of reads for haplotype 1 + 2. Significant difference from 0.5 indicates CNA. $-\log_{10}$ of p-value obtained for each chromosome arm and summed across chromosome arms to obtain metric. Metrics in range tested to identify which value identifies tumors at 99% specificity

Figure 2: Testing of optimal metric on independent 2 cm diameter tumor dataset to assess performance

Conclusions

- $-\log_{10}$ Metric based on statistical significance alone provides sensitive and specific detection of 4-billion cell tumors
- Machine learning methods such as the random forest classifier which are more pattern recognition based and go beyond statistical significance also provide highly sensitive and specific detection of 4-billion cell tumors
- Comparing the two detection methods of using copy number alterations versus mutation calling, we can more sensitively identify tumors through copy number alterations for tumors below 2 cm in diameter

Results: Machine Learning / Pattern Recognition

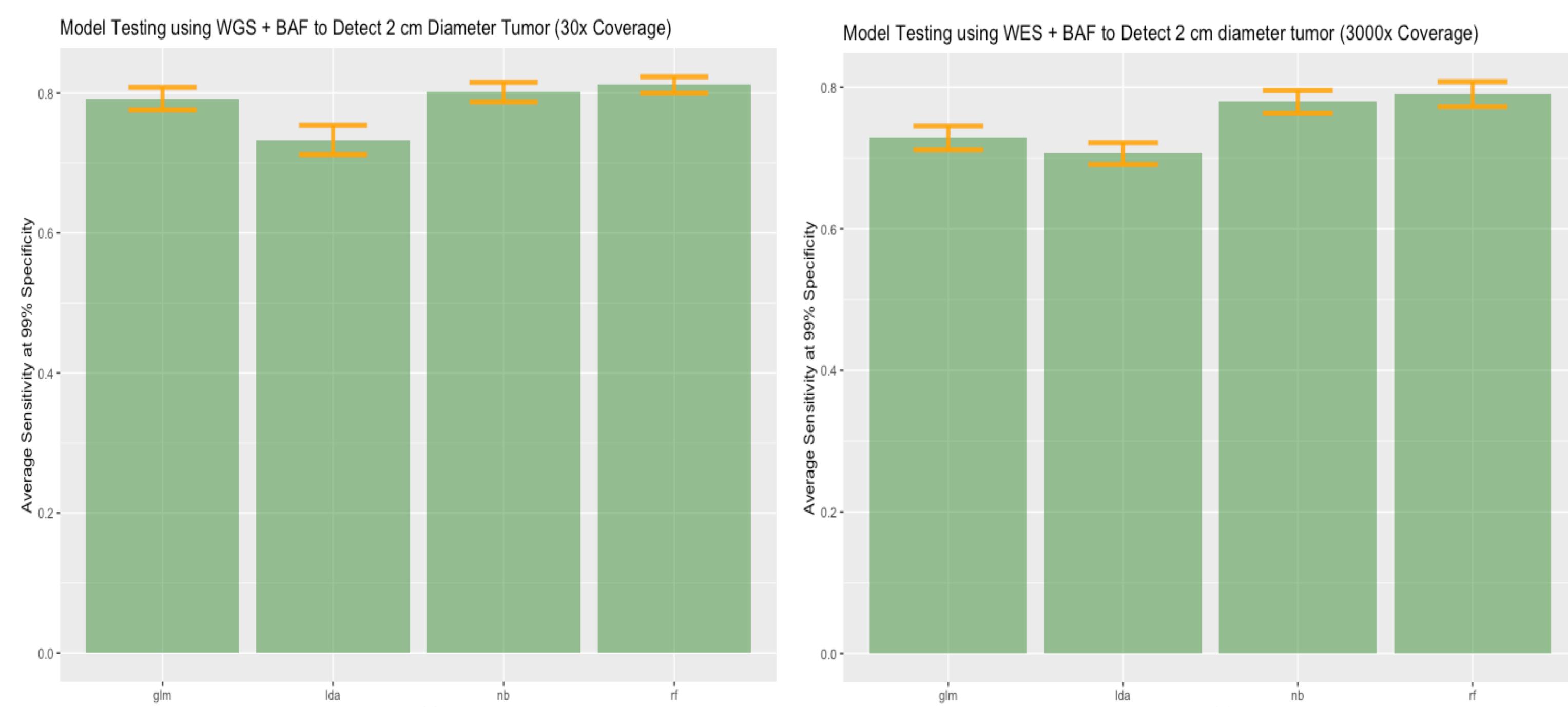


Figure 3: Sensitivity of logistic regression, linear discriminant analysis, naïve bayes, and random forest classifiers at 99% specificity for identifying 2 cm diameter tumors. B-allele frequency calculated from whole genome sequencing of cell-free DNA.

Figure 4: Sensitivity of logistic regression, linear discriminant analysis, naïve bayes, and random forest classifiers at 99% specificity for identifying 2 cm diameter tumors. B-allele frequency calculated from whole exome sequencing of cell-free DNA.

Results: Copy Number Alterations versus Mutations

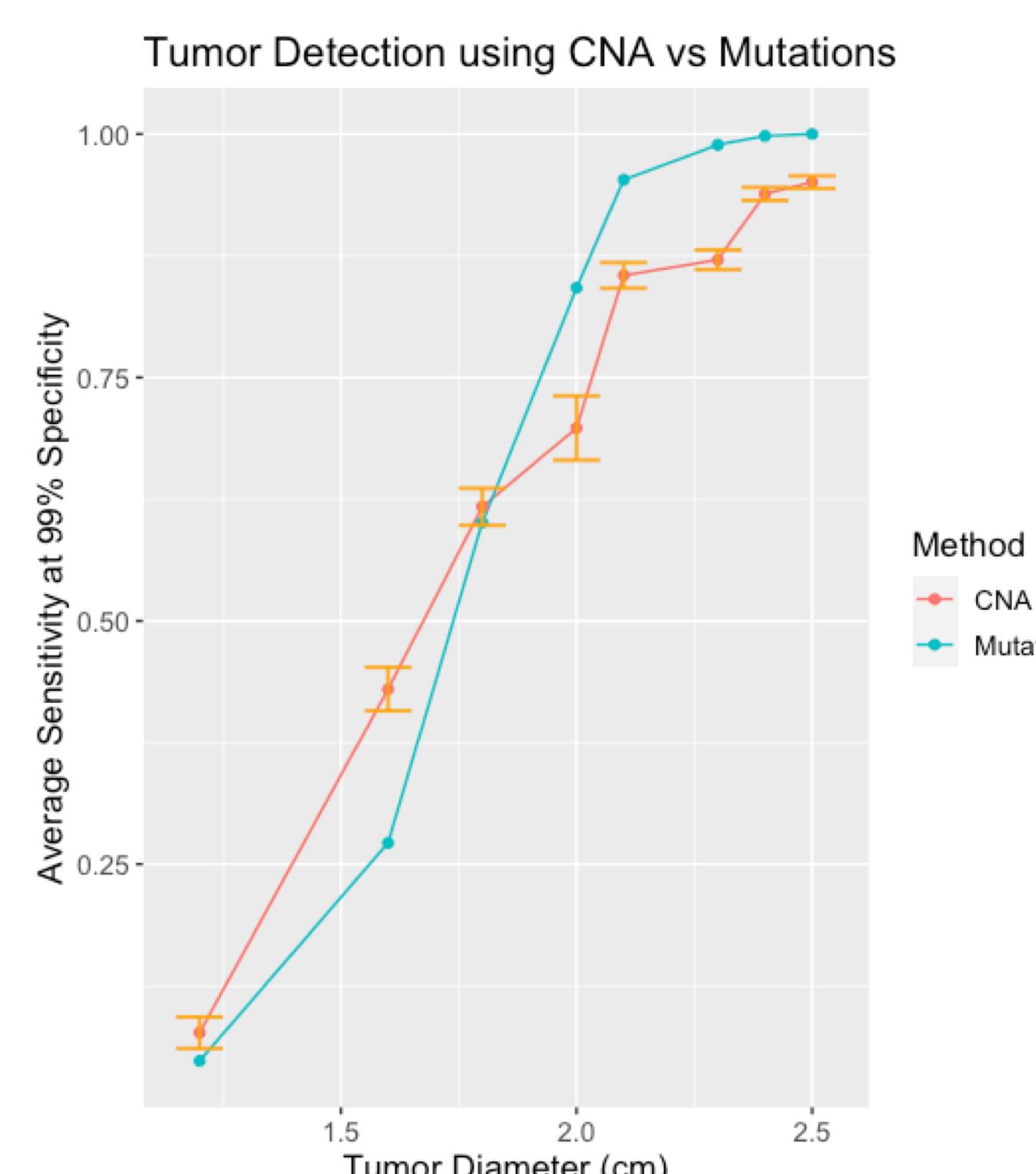


Figure 5: Tumor Detection for diameters in range 0-2.5 cm with average sensitivity of detection displayed at 99% specificity.

Literature Cited

- Newman AM, Bratman SV, To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*. 2014;20(5):548-554. doi:10.1038/nm.3519
- Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J, Lazar AJ; Cancer Genome Atlas Research Network, Cherniack AD, Beroukhim R, Meyerson M. Genomic and

Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell*. 2018 Apr 9;33(4):676-689.e3. doi: 10.1016/j.ccr.2018.03.007. Epub 2018 Apr 2. PMID: 29622463; PMCID: PMC6028190.

Acknowledgments

I'd like to thank my mentor Alex Gorelick, Principle Investigator Johannes Reiter, lab members Stefano and Giulia, and the Reiter Laboratory as a whole for supporting me throughout this project.