Investigation of Cell-Free DNA Methylation as a Biomarker of Malignant vs. Benign Lung Nodules

Jennifer Guo (CAS 2025)[1]; Jacob E. Till, MD, PhD[1]; Thrusha Puttaraju[1]; Aseel Abdalla[1]; Zhuoyang Wang[1]; Danielle Shoshany[1]; Wanding Zhou, PhD[2]; Jeffrey Thompson, MD[3*]; Erica L. Carpenter, MBA, PhD[1*];

[1]Division of Hematology-Oncology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania; [2]Center for Computational and Genomic Medicine, The Children's Hospital of Philadelphia;
[3]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania; *Co-Senior Author;

Background

- Many potentially malignant lung nodules (radiographic opacities ≤3 cm in diameter) are discovered incidentally or via screening.
- Smaller nodules lacking distinctly malignant features are classified as "indeterminate" and require repeat scans for monitoring.
- Evaluating these indeterminate lung nodules is a complex process with a high false positive rate, leading to unnecessary follow-up that may include repeat imaging or tissue biopsy.[1]
- Large-scale studies demonstrating that early detection of lung cancer via screening significantly reduced mortality have led to more low-dose CT screening and an increase in lung nodules being discovered (3.9 to 6.6 per 1000 going from 2008 to 2012).[2]

ccfDNA and ctDNA:

- DNA circulating in the blood, called circulating cell-free DNA (ccfDNA), is found even in healthy individuals.
- In cancer patients, a fraction of ccfDNA is tumor-derived, called circulating tumor DNA (ctDNA).
- Evaluation of ccfDNA and/or ctDNA represents a non-invasive method of detecting cancer in comparison to tissue biopsy.

Application of ccfDNA Methylation in Lung Nodule Evaluation:

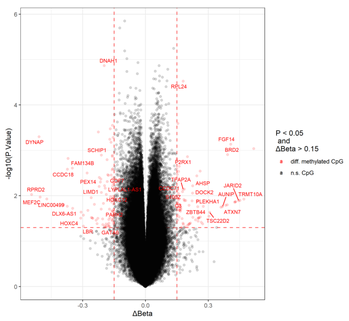
- Methylation is a type of epigenetic modification generally associated with gene silencing. At CpG sites, a methyl group can be added to a cytosine nucleotide.
- Methylation profiling of ccfDNA may be used to distinguish malignant from benign lung nodules. For example, hypermethylation at tumor suppressor gene promoters could indicate malignancy.[3]
- ccfDNA is a mixture of DNA derived from a variety of cells. The percentage of DNA originating from each type of cell can be estimated by referencing the unique methylation signatures of different cell types, in a process known as deconvolution.
- The results of deconvolution could be used as another method to predict if patients have lung cancer. A hypothesis could be that the increased apoptosis and necrosis associated with lung cancer could result in a larger proportion of ccfDNA originating from the lung.

Methods

- Blood samples were collected from patients with lung nodules (n = 10 benign, n = 14 malignant) under IRB protocol #824357.
- ccfDNA was extracted from plasma using QIAamp MinElute ccfDNA Mini Kit (#55204) and quantified with SYBR Green-based qPCR for the ALU115 amplicon.[4]
- 10 ng of extracted ccfDNA was converted and amplified using the NEBNext® Enzymatic Methyl-Seq Kit (#E7120).
- 500 ng of this DNA was sent for methylation profiling using the MethylationEPIC BeadChip (Infinium) 850K microarray (at University of Minnesota Genomics Center).
- Methylation data for determining signatures of reference cell types retrieved from the Genome Expression Omnibus (GEO).[5,6,7,8]
- Methylation data analysis performed in R with Sensible Step-wise Analysis of DNA MEthylation BeadChips (SeSAMe) package.[9]
- Copy number calculation from deconvolution percentages: [(ng DNA/mL plasma) * % from cell type]/(3.3 pg/haploid genome)

Results

A. Volcano Plot (159 DMRs)

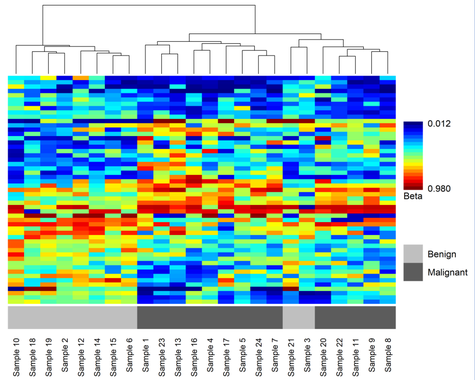


B. Differentially Methylated Loci in Named Genes

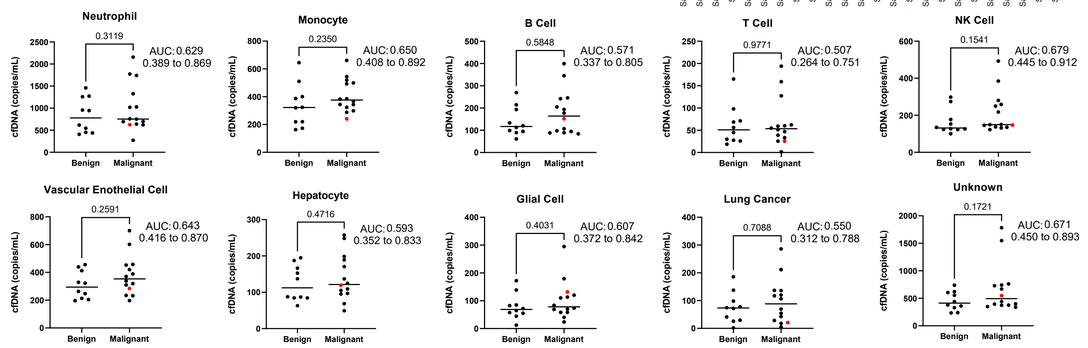
Hypomethylated	
RPRD2	KIF19
DYNAP	PEX14
MEF2C	LIMD1
LINC00499	SCHIP1
DLX6-AS1	DNAH1
CCDC18	LYPLAL1-AS1
FAM134B	ATP6V1G3
HOXC4	SMC4
LBR	GATA6
RASGRP3	APELA
CBX1	HOXC12
RNASEK	GDF7
DGKB	PARK2

Hypermethylated	
LOC101927924	AHSP
CCDC71	MBNL1
TFAP2A	CUL3
RPL24	DOCK2
FBXO47	HLA-DRB1
S100Z	PLEKHA1
RPRD1A	TSC22D2
C3	ATXN7
P2RX1	AUNIP
RANBP17	ZFP36L1
GLRA1	BRD2
ZBTB44	FGF14
CLTC	KDM5B
CAMK2D	CHURC1
MGC45800	JARID2
TRMT10A	

C. Supervised Clustering



D. Deconvolution



A. Volcano plot showing 159 differentially methylated regions (DMRs) from comparing the benign versus malignant subsets. Beta values range from 0 to 1 with higher values indicating that most copies of the loci are methylated within the sample. These loci had a large enough difference in methylation between subsets (quantified as a difference in the beta value of at least 0.15) and an unadjusted p-value of less than 0.05. Horizontal line represents the p-value cutoff and vertical lines represent the Δbeta cutoff.
B. A set of the differentially methylated loci that had a Δbeta of at least 0.175 and were located in named genes. 26 of these loci were hypomethylated and 31 were hypermethylated in the malignant subset compared to the benign subset.
C. Heatmap based on supervised clustering using the DMRs found. The set of loci among the 159 DMRs which were successfully read for all samples are displayed on the heatmap.
D. Deconvolution results showing the genomic copies per mL of plasma coming from each of the analyzed cell types for individual samples. Area Under Curve (AUC) and 95% confidence interval from Receiver Operator Curve (ROC) analysis, and Mann-Whitney test p-value noted on graphs.
*Red point represents the outlier which may be excluded (Sample 13). This sample had a greater number of failed methylation reads at certain loci.

Literature Review

Author (Year)	Loci	Test Set (n)	AUC for Test Set (95% CI)	Summary
Chen, et al. (2020)[10]	CDO1, SOX17, HOXA7	246	1: 88 (.84-.93)	• 8 lung cancer-specific genes tested, best performing panel identified • 246 patients with nodules, (163 M, 83 B)
Hulbert, et al. (2016)[11]	CDO1, SOX17, TAC1	210	.77 (.68-.86)	• Tested 6 cancer-specific genes from The Cancer Genome Atlas • 210 patients with nodules (150 stages I/IIA, 60 B)
Qi, et al. (2021)[12]	Top 300 DMRs	97	.96 (.96-.97)	• Test utilized top 300 differentially methylated regions (DMRs) from whole genome cfMeDIP-seq • 7 patients without nodules, 23 benign nodules, 35 malignant nodules (≤3 cm), 32 tumors (>3 cm)
Wielscher, et al. (2015)[13]	HOXD10, PAX9, PTPRN2, STAG3	46	.85 (.72-.95)	• Illumina methylation array identified DMRs of interest, 64 passed QC and were used to generate 4-gene model in a 204-patient cohort • 4-gene model was tested in a 46-patient cohort (23 M, 23 healthy)
Huang, et al. (2020)[14]	SHOX2, PTGER4	140, 30	.86 (.80-.92)	• Methylation of SHOX2 and PTGER4 shown to be biomarkers of lung cancer in literature • 140 patients with nodules in first set (104 M, 36 B) • 30 patients in validation set (19 M, 11 B)
Liang, et al. (2019)[15]	9 DMRs	66	.82 (.70-.93)	• Tissue-derived cancer-specific methylation markers from 230 samples, 9 markers selected with training set of 66 plasma samples • Independent test set of another 66 plasma samples (39 M, 27 B)
Gao, et al. (2015)[16]	APC, RASSF1A	89	.81 (N/A)	• Tumor suppressor genes which are commonly hypermethylated in cancer patients • 89 patients with nodules (58 M, 31 B)
Vrba, et al. (2020)[17]	10 Marker Set	65	.96 (.91-1.0)	• MIR129-2, LINC01158, CCDC181, PRKCB, TBR1, ZNF781, MARCH11, VWC2, SLC9A3, HOXA7 • Subset best for NSCLC selected from 1,250 biomarkers discovered *in silico* designed to detect 10 carcinoma types • Panel tested on 65 patients (18 M, 47 healthy)

Discussion and Future Directions

- Supervised clustering using the DMRs worked fairly well to separate benign versus malignant cases, indicating that the two subsets had distinct methylation patterns at those loci. It is possible that testing for methylation at a panel of DMRs identified in this discovery set could be useful in lung nodule diagnosis.
- Unadjusted p-value was used to select for DMRs instead of adjusted p-value since the sample size was not large enough for the adjusted p-value to be significant for any of the loci. Larger cohorts are necessary for statistically significant results especially since there are hundreds of thousands of loci being assessed for each patient.
  - The performance of a subset of DMRs with the greatest predictive value could be tested in a larger cohort.
- Deconvoluted percentages and copy numbers were not predictive of diagnosis, demonstrated by the high p-values and the AUC values close to 0.5. However, the difference in average percentage and copy number for benign versus malignant subsets was greatest for lung cancer-derived ccfDNA in comparison to other cell types assessed.

Acknowledgements

I would like to thank Dr. Till, Aseel, and everyone else at the lab for being extremely helpful and supportive throughout the learning process. I would like to specifically thank Dr. Carpenter and Dr. Thompson for providing the opportunity to get involved with this project. Thank you to CURF for providing funding through the Penn Undergraduate Research Mentoring Program.

References

1. Loverdos K, Fotiadis A, Kontogianni C, Iliopoulou M, Gaga M. Lung nodules: A comprehensive review on current approach and management. *Ann Thorac Med*. 2019;14(4):226-238. doi:10.4103/atm.ATM_110_19
2. Gould MK, Tang T, Liu IL, et al. Recent Trends in the Identification of Incidental Pulmonary Nodules. *Am J Respir Crit Care Med*. 2015;192(10):1208-1214. doi:10.1164/rccm.201505-0990OC
3. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene 21, 5427–5440 (2002). https://doi.org/10.1038/sj.onc.1205600
4. Fawzy A, Sweify KM, El-Fayoumy HM, Nofal N. Quantitative analysis of plasma cell-free DNA and its DNA integrity in patients with metastatic prostate cancer using ALU sequence. *J Egypt Natl Canc Inst*. 2016;28(4):235-242. doi:10.1016/j.jnci.2016.08.003
5. Moss J, Magenheim J, Neiman D, Zemmour H et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018 Nov 29;9(1):5068. PMID: 30498206
6. Kular L, Ewing E, Needhamsen M, Pahlevan Kakhki M et al. DNA methylation changes in glial cells of the normal-appearing white matter in Multiple Sclerosis patients. Epigenetics 2022 Jan 30;1-20. PMID: 35094644
7. Cho JW, Hong MH, Ha SJ, Kim YJ et al. Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. Exp Mol Med 2020 Sep;52(9):1550-1563. PMID: 32879421
8. Dietz S, Lifshitz A, Kazdal D, Harms A et al. Global DNA methylation reflects spatial heterogeneity and molecular evolution of lung adenocarcinomas. Int J Cancer 2019 Mar 1;144(5):1061-1072. PMID: 30150867
9. Zhou W, Triche TJ, Laird PW, Shen H (2018). "SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions." Nucleic Acids Research, gky691. doi: 10.1093/nar/gky691
10. Chen C, Huang K, Yin W, et al. Ultrasensitive DNA hypermethylation detection using plasma for early detection of NSCLC: a study in Chinese patients with very small nodules. *Clin Epigenetics*. 2020;12(1):39. Published 2020 Mar 5. doi:10.1186/s13148-020-00828-2
11. Hulbert A, Jusue-Torres I, Stark A, et al. Early Detection of Lung Cancer Using DNA Promoter Hypermethylation in Plasma and Sputum. Clin Cancer Res. 2017;23(8):1998-2005. doi:10.1158/1078-0432.CCR-16-1371
12. Qi J, Hong B, Tao R, et al. Prediction model for malignant pulmonary nodules based on cfMeDIP-seq and machine learning. *Cancer Sci*. 2021;112(9):3918-3923. doi:10.1111/cas.15052
13. Wielscher M, Vierlinger K, Kegler U, Ziesche R, Gsur A, Weinhäusel A. Diagnostic Performance of Plasma DNA Methylation Profiles in Lung Cancer, Pulmonary Fibrosis and COPD. *EBioMedicine*. 2015;2(8):929-936. Published 2015 Jul 2. doi:10.1016/j.ebiom.2015.06.025
14. Huang W, Huang H, Zhang S, et al. A Novel Diagnosis Method Based on Methylation Analysis of SHOX2 and Serum Biomarker for Early Stage Lung Cancer. Cancer Control. 2020;27(1):1073274820969703. doi:10.1177/1073274820969703
15. Liang W, Zhao Y, Huang W, et al. Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). Theranostics. 2019;9(7):2056-2070. Published 2019 Apr 6. doi:10.7150/thno.28119
16. Gao L, Xie E, Yu T, et al. Methylated APC and RASSF1A in multiple specimens contribute to the differential diagnosis of patients with undetermined solitary pulmonary nodules. J Thorac Dis. 2015;7(3):422-432. doi:10.3978/j.issn.2072-1439.2015.01.24
17. Vrba L, Oshiro MM, Kim SS, et al. DNA methylation biomarkers discovered in silico detect cancer in liquid biopsies from non-small cell lung cancer patients. Epigenetics. 2020;15(4):419-430. doi:10.1080/15592294.2019.1695333