Liquid biopsy-based multiomics profiling using low-pass whole genome sequencing and proteomics with computational modeling reveals molecular features of disease severity in EGFR/ALK wildtype NSCLC patients

Abstract ID. 5113

Zheng Feng[1], Danyi Wang[1], Francesco Vallania[2], Neil Smith[1], Anupriya Tripathi[2], Juergen Scheuenpflug[3#]

[1]Global Clinical Biomarkers and Companion Diagnostics, Translational Medicine, Global Development, EMD Serono Research and Development Institute, Inc., Billerica, MA, USA , A business of Merck KGaA, Darmstadt, Germany; [2]Freenome Holdings, Inc. South San Francisco, CA, USA;
[3]Global Clinical Biomarkers and Companion Diagnostics, Translational Medicine, Global Development, Merck KGaA, Darmstadt, Germany

[#]Corresponding author

BACKGROUND

- Lung cancer is the leading cause of cancer-associated deaths in the United States, with the vast majority being non-small cell lung cancer (NSCLC)[1]
- Non-invasive discovery of biomarkers associated with NSCLC prognosis and disease severity can enable precision oncology efforts[2]
- Liquid biopsy based multiomics profiling of plasma provides a promising and robust approach to interrogate both tumor- and non-tumor-derived signals

OBJECTIVE

The objective of this study was to investigate the potential of plasma-derived cell-free DNA (cfDNA) and circulating proteins as biomarkers of cancer prognosis in a cohort of treatment naive ALK/EGFR wild type patients with late-stage non-small cell lung cancer (NSCLC)

METHODS

- Cell free DNA (cfDNA) was extracted from 34 treatment-naïve stage IV EGFR/ALK wildtype patients
- Low pass whole-genome sequencing was performed to characterize cfDNA fragments, which reflect nucleosome protection and chromatin state
- Gene activation for protein-coding genes was modeled from fragment distribution around transcription start sites[3] (a schematic of our approach is shown in the diagram below)
  - The level of IL1RN, a previously discovered marker negatively associated with progression-free survival, was inferred using gene activation probability scores[2]
- The abundances of 644 plasma proteins including immune response markers, inflammation markers, cancer associated markers and DNA repair associated markers were measured
- Partial least-squares discriminant analysis (PLS-DA) was used for marker selection on both genomics and proteomics data. Number of features were selected for each analyte for the first two components using permutation based testing. Top 20 features for each component are shown with their group association.
- Gene Set Enrichment Analysis (GSEA) was applied to the MSigDB gene sets. *P*-values were corrected for multiple hypothesis testing using FDR.

Figure 1. Multiomics scope of the study

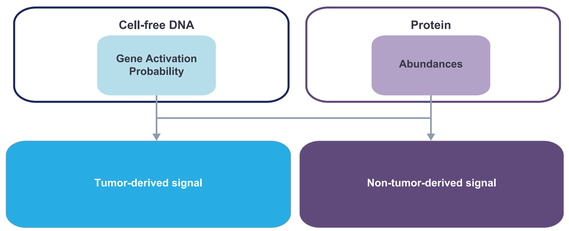
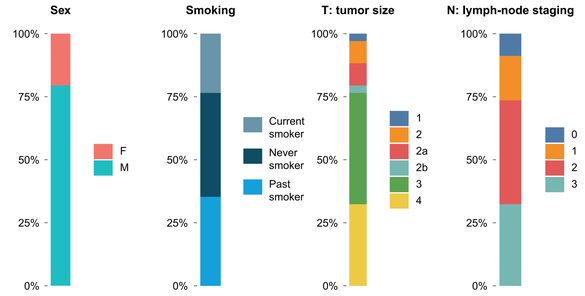
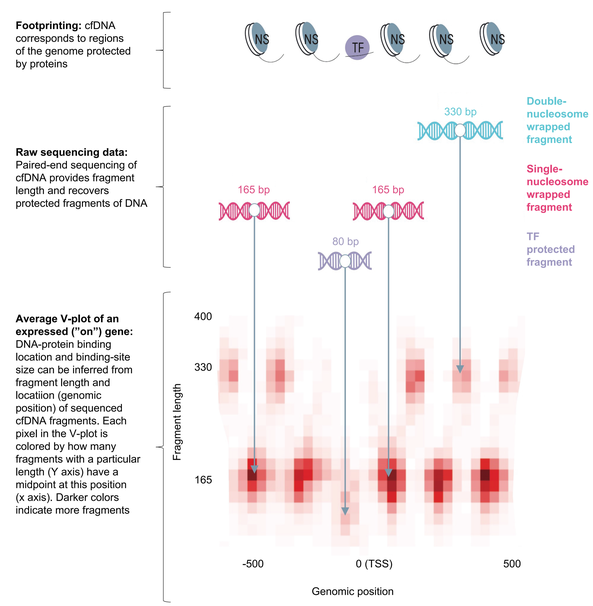


Figure 2. Cohort description and profiling



RESULTS

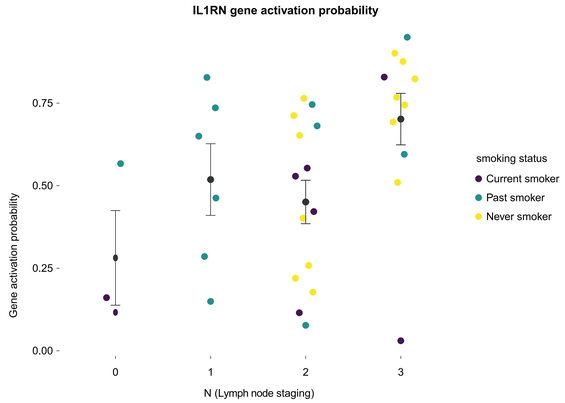
Figure 3. V-plots at the transcriptional start sites capture promoter activation states



TF = Transcription Factor (small region protected); NS = Nucleosome (large region protected)

- IL1RN levels have been previously associated with decreased survival time and increased progression in NSCLC[3]
- IL1RN gene activation probability scores are significantly associated with worse lymph-node staging (N) (P = 6E-3) in the current Stage IV EGFR wt/ALK wt NSCLC cohort, which is indicative of worse prognosis.

Figure 4. IL1RN is significantly associated with worse prognosis based on lymph-node staging



- Exploratory pathway analysis was performed comparing samples with high vs low estimated activation levels of IL1RN. Samples with high and low levels of IL1RN were separated by fitting a mixture model of two normal distributions
- 102 out of 32284 pathways were significantly enriched in patients with high levels of IL1RN (FDR < 5%)
- Top 10 most enriched pathways (ranked by significance and effect size) show inflammatory response and leukocyte activity signatures

Figure 5. Immune activity gene sets are enriched in patients with high IL1RN activation levels

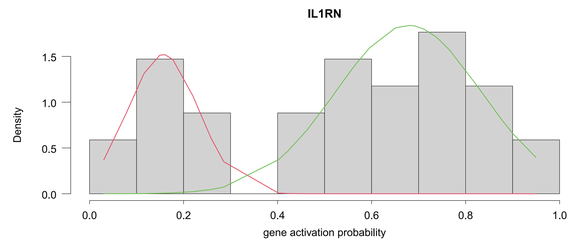
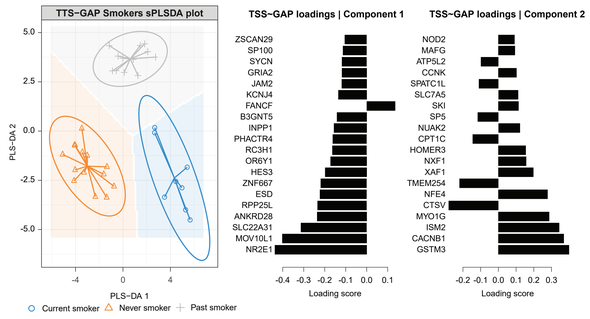


Table 1. Enrichment analysis identifies multiple gene sets enriched in patients with high IL1RN estimated activity

Pathway	*p*-value	ES	Size
C8||DESCARTES_FETAL_CEREBELLUM_MICROGLIA	< 1E-5	0.42360679	476
C4||MODULE_84	< 1E-5	0.41838963	486
C5||GOBP_INFLAMMATORY_RESPONSE	< 1E-5	0.38963768	642
C8||MANNO_MIDBRAIN_NEUROTYPES_HMGL	< 1E-5	0.37865283	517
C5||GOBP_LEUKOCYTE_MEDIATED_IMMUNITY	< 1E-5	0.3614014	683
C5||GOCC_CELL_SURFACE	< 1E-5	0.35916596	703
C5||GOMF_MOLECULAR_TRANSDUCER_ACTIVITY	< 1E-5	0.34934022	1285
C5||GOBP_CELL_ACTIVATION	< 1E-5	0.3481811	1242
C5||GOBP_DEFENSE_RESPONSE	< 1E-5	0.34688342	1458
C5||GOBP_DEFENSE_RESPONSE_TO_OTHER_ORGANISM	< 1E-5	0.34417198	986

- We performed an exploratory ANOVA analysis on these data to investigate associations between clinical variables
- We observe a significant association between smoking status and N (Lymph node staging) (*P* = 0.00795)
- This has been independently observed in a previous study[4]
- No observed association between tumor size and lymph node staging in this cohort, nor associations between tumor size and smoking status

Figure 6. Partial least square discriminant analysis identifies features associated with smoking status



- Sparse partial least square discriminant analysis (sPLSDA) was applied to protein abundances and TSS~GAP gene activation scores
- Permutation analysis was performed to determine significance of our selected features. TSS~GAP optimal feature number was 50 genes ($p < 10^{-2}$). Proteomics data did not show significant differences in our permutation analysis ($p > 0.05$)
- Top 20 features (out of 50) per component ranked by loadings
- We performed an exploratory gene sets enrichment analysis measuring associations with smoking status on both TSS~GAP scores and proteomics data
- TSS~GAP GSEA identified 108 sets out of 32284 to be significantly enriched in never smoking patients (FDR < 5%)
- Among all significant sets, we identified multiple cell-type signatures of myeloid cells to be significantly enriched in never smokers (see **Table 2**)
- Proteomics GSEA identified no significant gene sets.

Table 2. Enrichment analysis identifies multiple gene sets enriched in never smokers

pathway	*p*-value	ES	Size
C8||DESCARTES_FETAL_CEREBELLUM_MICROGLIA	1.00E-05	-0.3816575	476
C8||DESCARTES_FETAL_CEREBRUM_MICROGLIA	1.00E-05	-0.4063936	349
C8||DESCARTES_FETAL_PANCREAS_MYELOID_CELLS	1.00E-05	-0.4414363	193
C8||DESCARTES_FETAL_KIDNEY_MYELOID_CELLS	1.00E-05	-0.4509753	155
C8||DESCARTES_FETAL_LUNG_MYELOID_CELLS	1.00E-05	-0.4622521	145
C8||DESCARTES_FETAL_PLACENTA_LYMPHOID_CELLS	1.00E-05	-0.4832378	115
C8||CUI_DEVELOPING_HEART_C8_MACROPHAGE	3.00E-05	-0.4244807	243
C8||HAY_BONE_MARROW_PLATELET	3.00E-05	-0.4253388	206
C8||DESCARTES_FETAL_ADRENAL_MYELOID_CELLS	3.00E-05	-0.4472849	151
C8||DESCARTES_FETAL_CEREBELLUM_VASCULAR_ENDOTHELIAL_CELLS	4.00E-05	-0.3685401	503
C8||GAO_LARGE_INTESTINE_24W_C11_PANETH_LIKE_CELL	4.00E-05	-0.403068	283
C8||GAO_LARGE_INTESTINE_ADULT_CJ_IMMUNE_CELLS	4.00E-05	-0.3791514	425
C8||DESCARTES_FETAL_CEREBRUM_VASCULAR_ENDOTHELIAL_CELLS	7.00E-05	-0.3720602	434
C8||FAN_EMBRYONIC_CTX_BRAIN_ENDOTHELIAL_1	8.00E-05	-0.3786333	398
C8||DESCARTES_FETAL_INTESTINE_MYELOID_CELLS	8.00E-05	-0.4183012	185
C8||DESCARTES_FETAL_ADRENAL_LYMPHOID_CELLS	8.00E-05	-0.4460872	130
C8||AIZARANI_LIVER_C25_KUPFFER_CELLS_4	9.00E-05	-0.4318588	154
C8||GAO_LARGE_INTESTINE_24W_C2_MKI67POS_PROGENITOR	0.00015023	-0.4618927	104

CONCLUSIONS

- Using our multiomics liquid biopsy platform, we characterized a retrospective cohort of treatment-naïve stage IV EGFR/ALK wildtype NSCLC patients
- Gene activation scores of IL1RN, a previously discovered marker associated with progression in NSCLC, are associated with worse lymph node (N) staging in this cohort, and with higher enrichment of inflammatory gene signatures
- N staging was significantly associated with smoking status, where never-smokers are associated with higher N
- All of our findings require an independent validation cohort for confirmation. Our results show the potential of our multiomics platform to non-invasively characterize NSCLC.
- Future applications of this multiomics platform provide promise for stratification of NSCLC patients for patient-centric precision oncology

REFERENCES

1. Siegel et al. *CA Cancer J Clin*. 2020. doi: 10.3322/caac.21590.
2. Vallania F, Warsinke H, Ulz P, et al. Multiomic plasma profiling identifies potential signatures of disease progression in early-stage NSCLC. Presented at: Annual Meeting of the American Association for Cancer Research. 2020; virtual.
3. St. John J, Gafni E, White B, et al. Predicting gene expression using cfDNA fragment length and position. Presented at: Annual Meeting of the American Association for Cancer Research. 2019; Altanta, Georgia.
4. Jiang F, Zhao H, Li X, et al. Association between smoking status and lymph node metastasis in Chinese patients with T1 non-small cell lung cancer: A real-world observational study, In *Journal of Clinical Oncology*. 2019. DOI: 10.1200/JCO.2019.37.15_suppl.e20012

Presented at American Association for Cancer Research (AACR) Annual Meeting | April 8-13, 2022. New Orleans, LA, USA