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

Zheng Feng1, Danyi Wang1, Francesco Vallania1, Neil Smith1, Anupriya Tripathi2, Juergen Scheuenpflug3
1. Global Clinical Biomarkers and Companion Diagnostics, Translational Medicine, Global Development, EMD Serono Research and Development Institute, Inc., Morristown, MA, USA. 2. A business of Merck KGaA, Darmstadt, Germany. 3. Freenome Holdings, Inc. South San Francisco, CA, USA.

Corresponding author.

ABSTRACT

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

METHODS

- Cell-free DNA (cfDNA) was extracted from 36 blood samples of treatment-naive stage IV NSCLC mutation positive patients.
- Low-pass whole-genome sequencing was performed to characterize cfDNA fragments, which reflect nucleosome protection and transcription start sites.
- Gene activation for protein-coding genes was modeled from fragment distribution around transcription start sites (a schematic of our approach is shown in the diagrams below).
- The distribution of fragment length and number of fragments per particular size (60-80 bp) was normalized.

RESULTS

- The level of IL1RN in plasma was negatively associated with progression-free survival (PFS) and overall survival (OS).
- Gene Set Enrichment Analysis (GSEA) was applied to the MSigDB gene sets.
- Among all significant sets, we identified multiple cell-type signatures of myeloid cells to be significantly enriched in never smokers (see Table 2).
- GSEA-identified significant gene sets included inflammatory response genes.

CONCLUSIONS

- Using our multispecies liquid biopsy platform, we characterized a non-invasive cohort of treatment-naive stage IV EGFR/ALK wild-type NSCLC patients.
- Gene activation scores of IL1RN, a previously discovered marker associated with progression in NSCLC, were associated with worse lymph-node (N) staging in this cohort, and with higher enrichment of inflammatory gene signatures.
- IL1RN staging was significantly associated with smoking status, where never-smokers were associated with higher IL1RN expression.
- Future applications of this multispecies platform provide promise for stratification of NSCLC patients for patient-centric oncology precision medicine.

REFERENCES


Table 1. Enrichment analysis identifies multiple gene sets enriched in never smokers

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

Table 3. Enrichment analysis identifies multiple gene sets enriched in never smokers

Figure 1. Multimics scope of the study

Figure 2. Cohort description and profiling

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

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

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

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

Figure 7. Exploratory pathway analysis was performed comparing patients with high vs low activated IL1RN levels of > 0.5. Samples with high and low levels of IL1RN were separated by tSNE in a mixture model of two normal distributions.

Figure 8. Permutation analysis was performed to determine significance of our enrichment features. TSS-GAP optimal feature (G) was compared (p < 0.01), with no significant differences in our permutation analysis (P < 0.05).

Figure 9. TSS-GAP identified 118 sets of out of 32,964 to be significantly enriched in never smokers (FDR < 0.05).

Figure 10. Among all significant sets, we identified multiple cell-type signatures of myeloid cells to be significantly enriched in never smokers (see Table 2).

Figure 11. Proteins

Figure 12. CFNA

Figure 13. Tumor derived signal

Figure 14. Non-tumor derived signal

Figure 15. cfDNA provides fragment footprinting of start sites and transcription factor recruitment.

Figure 16. Rapid sequencing of plasma-derived cfDNA fragments allows for a non-invasive monitoring of tumorigenesis.

Figure 17. cfDNA is fragmented in 60-80 bp fragments, which reflect transcriptional start sites and transcription factor binding sites.

Figure 18. V-plots at the transcriptional start site capture promoter activation states.