Genomic Landscape of Patients With Non-Small Cell Lung Cancer (NSCLC)

Neeraj Kumar Singh[1], Kamal S. Saini[2], Sangavai Chakkrapani[1], Eric Severson[2], Nahush Nagaraj[1], Shakti Ramkissoon[2], Isagani Chico[2], Ariel Aguilo[2], Rohini George[1], Laura Vidal[2], Smita Agrawal[1]

1 - ConcertAI, Bengaluru, India; 2 - Labcorp Drug Development Inc, Princeton, NJ

**Contact Info.: nesingh@concertai.com*

Abstract #923

Background

- Non-small cell lung cancer (NCSLC) is the leading cause of cancer deaths in the United States with a 5-year survival rate of 22.9% (7% in those with metastases).
- Targeted therapies have shown promise but a better understanding of the genomic landscape of the tumor may help further development of personalised medicines and combination therapies which in turn could lead to better survival.
- We have developed a deeply curated real-world clinico-genomics database of >11k NSCLC patients and here, we present the genomic landscape and key characteristics of these NCSLC patients.

Methods

- The ConcertAI Genome360™ NCSLC dataset was created by curating >100 clinical variables from unstructured clinical data from EMRs combined with extraction of detailed genomics and molecular pathology (MP) data from original NGS and other MP reports. This currently contains data for 11349 NSCLC patients from across the USA who have undergone a next generation sequencing test of their tumor and covers >1000 genes.
- Here we present the demographic, clinical and genomic characteristics of this cohort.
- The top 20 genes with pathogenic mutations are presented and co-mutational analysis was performed to identify the most significant genes that were positively or negatively correlated to these genes using chi-squared test with Bonferroni correction.
- Other clinically relevant biomarkers such as tumor mutation burden (TMB), microsatellite instability (MSI) and Programmed death-ligand 1 (PD-L1) status are also characterised.

Results

Table 1: Distribution of key demographic and clinical features

Demographics/Clinical Features		Sample Size (%)
Cohort		11349
Age	Mean (Min-Max) in years	70 (21-89)
Gender	Male	5646 (49.8)
	Female	5703 (50.2)
Race	White	8915 (78.55)
	Black or African American	1468 (12.93)
	Other or Unknown Race	966 (8.51)
Tumor Stage	Stage 1	1140 (10.04)
	Stage 2	696 (6.13)
	Stage 3	2179 (19.19)
	Stage 4	6450 (56.83)
	Unknown	884 (7.79)
Histology(*)	Adenocarcinoma, NOS	7831 (69)
	Squamous cell carcinoma, NOS	2771 (24.42)
	Non-small cell carcinoma	770 (6.78)
	Acinar cell carcinoma	331 (2.92)
	Others	712 (6.27)

(*) Some patients have more than one histology,

Table 2: Distribution of key immune biomarkers

Biomarkers	Tested	High	Prevalence (%)
PD-L1	7167	2118	29.51
TMB	6528	1866	28.42
MSI	7841	63	~1

- In patients with metastatic disease, brain, bone, and liver were the most prevalent metastatic sites.
- The 20 most frequently mutated genes along with their top 3 most significant (p-value < 3e-5) positively and negatively correlated genes are listed in Table 3 .
- For example: Co-mutational analysis shows that KRAS mutations are positively correlated with STK11 mutations but negatively correlated with EGFR mutations [1]

Table 3: Top 20 mutated genes and associated positively/negatively correlated genes

Biomarker	Prevalence (%)	Positively Correlated Genes	Negatively Correlated Genes
TP53	65	SOX2, NFE2L2,FGFR1	KRAS, MDM2, STK11
KRAS	28	STK11, RBM10, ATM	EGFR, KMT2D, SOX2
CDKN2	19	TP53, CCND1, SOX2	RB1
EGFR	13	MDM2, CDK4, CCNE1	STK11, KEAP1, KMT2D
STK11	12	KEAP1, KRAS, MYC	SOX2, PIK3CA, KMT2D
MTAP	9	CDKN2A	RB1
PIK3CA	7	SOX2, TP53, NFE2L2	KRAS, STK11
KEAP1	6	STK11, KRAS, CDKN2A	EGFR
SMARCA4	6	STK11, KEAP1, APC	EGFR
NF1	6	TP53, RASA1, PTPN11	KRAS
RBM10	5	KRAS, IL7R, ATM	TP53
KMT2D	7	TP53, NFE2L2, NOTCH1	KRAS, EGFR, STK11
DNMT3A	6	TET2, ASXL1, CHEK2	-
RB1	7	TP53, PTEN, CCNE1	MTAP, KRAS, CDKN2A
ARID1A	7	CDKN2A, TP53, FAT1	EGFR
ATM	4	KRAS, NKX2-1, NFKBIA	TP53
PTEN	7	RB1, TP53, KMT2D	KRAS
BRAF	11	SETD2, CDK6	KRAS
MET	7	CDK6, MDM2, ROS1	KRAS
TERT	4	MITF, PTCH1, EP300	-

Conclusions

- We have created a real-world clinico-genomics dataset of NSCLC patients from across the USA which is EMR and NGS testing lab agnostic and provides a comprehensive view of the patients' clinical journey with complete molecular pathology data.
- This analysis provides a deeper understanding of the genomic landscape of NSCLC.
- The co-mutational analysis provides insights into pathways that are perturbed together providing opportunities to develop treatments to overcome such co-mutations.
- Using this ConcertAI Genome360™ NSCLC dataset, we have previously identified genomic biomarkers for sensitivity/ resistance to PD-L1/PD-1 check point inhibitors in NSCLC [2].

References

1. Choughule, A et al. doi:10.1038/bjc.2014.401.
2. Singh, N et al. doi: 10.1136/jitc-2022-SITC2022.0531