Airway Basal Cells of Healthy Smokers and Human Lung Adenocarcinomas Share a Common Embryonic Stem Cell-like Transcriptional Program

R. Shaykhiev, MD,PhD[1], R. Wang, M.D.,Ph.D[1], R.K. Zwick, BA[1], N.R. Hackett, Ph.D[1], R. Leung, MD[2], M. Moore, D.Phil[2], R.J. Downey, M.D.[2], Y. Strulovici-Barel, MSc[1], J. Salit, MS[1], R.G. Crystal, MD[1]
[1]Weill Cornell Medical College - New York, NY/US, [2]Memorial Sloan-Kettering Cancer Center - New York/US

Rationale: Various epithelial cancers, including lung adenocarcinomas, exhibit activation of a human embryonic stem cell (hESC)-like transcriptional program in association with their aggressiveness. We hypothesized that under chronic exposure to the pro-carcinogenic stress of cigarette smoking, airway basal cells, the stem/progenitor cell compartment of the large airway epithelium (LAE), acquire “cancerous” hESC-like molecular phenotype as an early step toward malignant tissue derangement.

Methods: Expression of 40 hESC-specific genes previously identified in a meta-analysis of the hESC transcriptome was assessed in freshly isolated LAE of healthy nonsmokers (n=21) and healthy smokers (n=31) and in the pure basal cell (BC) populations obtained from these samples (n=4 in both groups) using Affymetrix microarrays. BC transcriptomes were additionally analyzed by massive parallel sequencing (RNA-Seq). Expression patterns were then compared to those in primary human lung adenocarcinoma samples (n=193) and pure human lung adenocarcinoma cells propagated as xenografts in NOD/SCID/IL2R-null immunodeficient mice (n=4).

Results: Microarray analysis revealed low-level constitutive expression of a subset of hESC-specific genes in the normal healthy LAE, with higher expression in the BC population. Smoking dramatically and selectively up-regulated hESC-specific gene expression in BC. Remarkably, 35% of known hESC-specific genes were found up-regulated in BC of healthy smokers by both microarrays and RNA-Seq. The majority of these genes (71%) were not detected in nonsmokers’ BC indicative of their *de novo* induction by smoking. Strikingly, 82% of hESC-specific genes induced in BC of healthy smokers were also significantly up-regulated in both primary human lung adenocarcinomas and xenograft-derived human lung adenocarcinoma cells. Clustering analysis identified a subgroup of lung adenocarcinoma samples (n=32; 6%) among 193 primary lung adenocarcinoma samples based on the high expression of hESC-specific signature found up-regulated in BC of healthy smokers. This subset of adenocarcinoma subjects exhibited distinct clinical/pathological characteristics, such as longer smoking history ($p<0.0002$), lower lung function parameters FEV1 and DLCO (both $p<0.001$), higher COPD co-morbidity (44% *vs* 14% among other adenocarcinoma subjects), larger tumor size ($p<0.0002$), more advanced tumor stage, poorer differentiation grade, and

higher frequency of *TP53* gene mutations (50% *vs* 23% among other adenocarcinoma subjects).

Conclusions: Cigarette smoking reprograms airway BC *in vivo* toward a lung cancer-associated hESC-like molecular state prior to any clinical manifestation of lung cancer. The subset of lung adenocarcinomas highly expressing smoking-induced BC-hESC-like gene signature have a distinct, aggressive clinical phenotype.