Pilot Analysis: CC16 (SCGB1A1 / Uteroglobin) Signal in Bronchial Epithelial Samples

Public Dataset: GSE37147 (Affymetrix Human Gene 1.0 ST, GPL13243) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37147**)**

Background

Club Cell Secretory Protein (CC16), encoded by *SCGB1A1* and historically referred to as uteroglobin, is a well-established airway epithelial biomarker implicated in epithelial integrity, smoking-related injury, and chronic obstructive pulmonary disease (COPD). Given its relevance to airway biology and population-based respiratory research, a focused pilot analysis was conducted to evaluate CC16 behavior at the transcriptomic level in bronchial epithelial samples and to assess its integration with available clinical and exposure-related phenotypes. The primary goal of this work was methodological and exploratory: to validate CC16 probe mapping on the platform, characterize expression behavior, and examine whether CC16 transcript levels demonstrate biologically coherent relationships with smoking exposure and lung-function severity.

Data and Methods

Publicly available gene-expression data from GSE37147 were analyzed. Expression profiles were generated from bronchial epithelial brushings using the Affymetrix Human Gene 1.0 ST microarray (GPL13243). The series matrix contained 19,793 probe rows across 269 samples. Platform annotation was explicitly downloaded and parsed to confirm probe-to-gene mapping. CC16 was identified as *SCGB1A1 (uteroglobin)* and uniquely mapped to probe **7356_at**. Expression values corresponding to this probe were extracted across all samples marked as usable in the original dataset.

Quality control focused on distributional assessment and verification of signal stability rather than aggressive filtering. Sample-level CC16 expression was subsequently merged with clinical metadata extracted from the GEO series header, including smoking status, COPD status, and lung-function measures such as FEV1% predicted.

CC16 Expression Characteristics

Across 269 bronchial epithelial samples, CC16 expression demonstrated a stable and biologically plausible distribution on the log scale. The mean expression value was **12.28**, with a standard deviation of **0.39** and a median of **12.40**. Observed values ranged from **9.68 to 12.65**. Interquartile range-based screening identified approximately **30 samples** at the distribution extremes. These samples were retained to preserve potential biological variability rather than excluded as technical outliers.

Overall, the distribution was unimodal and right-skewed, consistent with expected airway epithelial expression patterns for CC16.

Integration with Clinical and Exposure Variables

Smoking status was available for **238 samples**, including **99 current smokers** and **139 ex-smokers**. CC16 expression differed significantly between these groups. Current smokers exhibited a mean CC16 expression of **12.14** (SD **0.54**), whereas ex-smokers showed a higher mean expression of **12.39** (SD **0.18**). A Welch two-sample t-test confirmed this difference to be statistically significant ($\mathbf{p = 2.32 \times 10^{-5}}$), with a moderate-to-large effect size (**Cohen's d = −0.66**). These findings are consistent with smoking-related suppression of CC16 expression in airway epithelial cells.

COPD status was also available for **238 samples**, comprising **87 COPD cases** and **151 controls**. Mean CC16 expression was **12.24** among individuals with COPD and **12.32** among controls. This difference did not reach statistical significance (**p = 0.14**), indicating that CC16 transcript levels alone may not strongly discriminate COPD status without additional contextual variables.

Lung-function data (FEV1% predicted) were available for **238 samples**. Across all samples, CC16 expression showed a modest but statistically significant positive correlation with FEV1% predicted ($\mathbf{r = 0.24, p = 1.98 \times 10^{-4}}$). Stratified analyses indicated that this association was stronger among current smokers (**r ≈ 0.37**), suggesting effect modification by active smoking exposure and supporting the relevance of CC16 as a severity-linked biomarker in exposed populations.

Interpretation and Relevance

This pilot analysis confirms that CC16 is robustly represented at the transcript level in bronchial epithelial samples and exhibits biologically coherent relationships with smoking exposure and lung-function severity. While CC16 differences by COPD status alone were modest, its integration with exposure and physiological measures revealed meaningful patterns consistent with airway injury and epithelial dysfunction. Methodologically, this work demonstrates a reproducible framework for probe validation, expression quality assessment, and phenotype integration that is directly extensible to larger cohort analyses, longitudinal modeling, and multi-biomarker workflows.

Summary

In summary, CC16 (*SCGB1A1*) expression in bronchial epithelial samples from GSE37147 displays stable signal behavior, significant suppression among current smokers, and a positive association with lung-function measures. These results support CC16's utility as a contextual airway biomarker and provide a clean, validated foundation for broader population-level respiratory disease analyses.