Final Report: Year 4 (Reporting Year: 2017-2021)

NKFIH OTKA FK# 124652

Title: Evaluation of new and emerging biomarkers in lung adenocarcinoma

Background, Research Proposal

The etiology of lung cancer has been linked to numerous factors, including gut microflora. By definition, the human metagenome is the collective genome of all bacteria, archaea, fungi, protists, and viruses in the human body. The microbiome can control epithelial cell proliferation/differentiation, nutrition, detoxification, metabolism, and hormonal homeostasis. Certain bacterial species can direct immune cell trafficking, especially inflammatory or tolerant immune responses that can emerge or evolve. Lung cancer and chronic obstructive pulmonary disease (COPD) have been linked to microbiota changes and frequently follow gastrointestinal (GI) disorders. (R. Bingula et al., Journal of Oncology, 6, 1-15) The gut-lung axis can be linked to bacterial and tumor antigen mimicry or cross-reactivity. (L. Zitvogel, et al. Cell, 165, 276-87). Bacterial antigens can stimulate cytokine and interferon production and eliciting an immune response. Quantitative computed tomography (CT) based texture analysis (QTA), a non-invasive histopathological tumor assessment, might deliver information on tumor characteristics, including tumor-infiltrating immune cells.

Recent data suggest that the gut-lung axis regulates systemic metabolic and immune functions, and microbiota might alter exercise tolerance. Cachexia is associated with decreased survival in cancer patients and has a prevalence of up to 80%. The etiology of cachexia is poorly understood, and limited treatment options exist. Therefore, we aimed to analyze the associations of the human gut microbiome and cachexia. There is a lack in our understanding of pathogenesis and mechanisms accounting for the large variability in patients' response to oncotherapy. The predictive and prognostic role of the bacterial microbiome in lung cancer has not been explored.

Aim of the study

We aim to develop innovative methodological strategies for establishing associations between the sample's bacterial microbiome and lung cancer patients' outcomes, oncotherapy efficacy, and toxicity.

The next part of our study aims to combine a QTA and microbiome-based biomarker signature for the early prediction of survival and prognosis of NSCLC patients.

We aim to characterize the gut microbiome associations with malnutrition and exercise intolerance.

Progress and Results of Research (Year 1-4)

As proposed in the original Work Plan of our research proposal, we made progress in the grant project and completed the patient recruitment, sample collection, and sequencing. The results were published or under review in peer-reviewed open access journals or presented in world conferences (ERS, WCLC). We analyzed the clinical relevance of the bacterial microbiome in lung cancer.

Clinical Samples and Clinicopathological Data

This study is conducted following the guidelines of the Helsinki Declaration of the World Medical Association and has the approval of the national level ethics committee. We obtained individual informed consent from all patients for this study. Recruitment started in 2017 and, as planned, finished in 2019, includes a total of (n=362) patients in the study according to selection criteria at diagnosis as described in the original research plan. Patients with lung nodules identified on CT scans and NSCLC patients at diagnosis receiving standard of care treatments including chemotherapy, targeted therapy, and immunotherapy were included in our study. Samples were collected from lung cancer patients at diagnosis, before the initiation of oncotherapy (baseline), and at follow-up (12 weeks or after surgery). Accordingly, we set up an extensive database on patients' clinicopathological, survival, and microbiome sequencing data to enhance future research and collaborations on microbiome and metabolome in lung cancer. The below-described cohort of patients with available samples and clinical data were analyzed. Patients who proceeded to further successful sequencing steps were included in subsequent biomarker studies.

In **cohort #1** (n= 98 patients), the microbiome sequencing of the stool samples of advanced-stage and chemotherapy-treated patients (n=78) and early-stage resected patients (n=20) were performed.

Cohort#2, we included advanced stage (n=129) patients and analyzed sequencing data of stool bacterial species, and we evaluated Quantitative computed tomography (CT) based texture analysis (QTA).

Next, we analyzed the relevance of microbiome in malnutrition in predominantly advanced-stage **cohort#3** (n=31).

Next, we investigated the role of exercise tolerance and disease recurrence in **cohort #4** (n=15) of early-stage lung cancer patients with samples collected before and after surgical resection.

As proposed in the original Work Plan, we completed the clinicopathological data collection. We included gender, age, smoking, family history of cancer, previous cancer, stage, COPD Global Initiative for Chronic Obstructive Lung Disease (GOLD), body mass index, Global Subjective Assessment (SGA) ranking (cachexia score), and histopathological diagnosis, including molecular diagnostics. We divided patients into Responder (R) and non-Responders (NR) to oncotherapy according to RECIST 1.1 criteria. We also analyzed patients based on short-term (<6months) versus long-term overall survival [OS] (>6months) when stable disease (SD), partial response (PR), or complete response (CR) lasted less or more than six months, respectively.COPD patients are classified as Global Initiative for COPD (GOLD) 1-4 stage.

Microbiome Sequencing and Data Analysis

After successful DNA isolation, we completed the NGS sequencing of samples collected. The data analysis was conducted in R version 3.5.0 using in-house scripts. Library preparation and shotgun metagenomic sequencing were performed using Illumina HiSeq 4000 at an average depth of 6 Gb. We processed the sequenced reads with QC to remove the adapter regions, low-quality reads, and human DNA contaminations (bwa (version 0.7.4-r385) mem against human reference genome ucsc.hg19) 95% of the reads remained after the QC. following the previously described steps [Li J. et al. Proc Natl Acad Sci USA. 2016;113:E1306–1315]. We analyzed the overall bacterial load of the samples, and we clustered the sequences into operational taxonomic units (OTUs). The high-quality reads were taxonomically profiled using MetaPhlAn2 [Truong DT et al. Nat Methods. 2015;12:902–3.] with default settings. The differentially abundant taxa were identified using the Wald test implemented in the R package DESeq2 [Love MI et al. Genome Biol. 2014;15:550.] v1.22.2 on the unrarefied relative abundance data, and the statistical significance was filtered with $p < 0.05$ unless otherwise stated. Pathway analysis was done using the HUmann2 pipeline. We also compared microbiome diversity in our cancer patients with healthy controls from the Human Microbiome Project.

Bioinformatical analysis

R software (R 3.3.0) was used for statistical analysis. The Fisher test was used to compare categorical variables. The student's t-test was used for normally distributed clinical data and metabolites levels. Wilcoxon rank-sum test was used for continuous, not normally distributed data. We used Spearman correlation between variables and gut microbial species. Statistically different taxa were identified with the Wald test using R package DESeq2 (v1.22.2) on the rarefied relative abundance data. Two-tailed p-values < 0.05 were considered significant. FDR correction was applied for multiple testing corrections. (Benjamini Y, Hochberg Y. et al. 1995;57:289–300) We also compared microbiome diversity in our cancer patients with healthy controls from the Human Microbiome Project.A diversity index (Simpson's Diversity Index) was analyzed on phylum, genus, or species level. Diversity measures on phylogenetic relations among the individuals such as richness, divergence, or evenness were evaluated. We analyzed four measures of alpha diversity; Faith's PD (uses phylogenetic information), Shannon & Simpson (measures of complexity by probability), richness (just the pure number of different organisms). Chao1, a nonparametric method for estimating the number of species in a community, was used. Additionally, we analyzed on phylum, genus & species level the most common ranks. Statistics are performed based on the aims and the meaningful sample groups that are present in each setting. We used P-value <0.05 to assess significance for alpha and beta diversity. Beta diversity was assessed with a Non-Metric Multidimensional Scaling (NMDS) plot of different patient groups based on the gut microbial compositions using Bray-Curtis dissimilarities (ANOSIM). Principal Coordinates Analysis (PCoA, = Multidimensional scaling, MDS) was used to explore and visualize beta diversity and similarities

or dissimilarities of data. Due to the complex relationships present in the gut ecological community, we applied machine learning approaches, including the Extreme Gradient Boosting, a decision tree-based algorithm to the microbiome data to reveal the signature taxa of the responder and non-responder patient groups.

Results, Conclusions, and Future Perspectives

In cohort #1, we analyzed the associations of gut microbiota and Chronic Obstructive Pulmonary Disease Assessment Test (CAT) score in chemotherapy-treated lung cancer patients.

Patients with lung cancer have a high prevalence of chronic obstructive pulmonary disease (COPD). CAT score is a symptomatic questionnaire that has been shown to correlate with mortality and airflow limitation severity. COPD is a systemic inflammatory disease, and the role of the gastrointestinal (GI) microbiome in the pathogenesis of COPD has not been explored. Recent data suggest that the gut-lung axis regulates systemic immune function, and the GI microbiome composition in early life influences asthma risk. (Abrahamsson TR, et al. Clin Exp Allergy 2014;44:842–850).

Here, we aimed to evaluate the associations of GI microbiota and COPD CAT score in chemotherapy-treated lung cancer patients.

We included 98 lung cancer patients (adenocarcinoma (n=47), squamous cell cancer (n=21), other (n=30)). We surveyed stool samples from patients with COPD (n=59) and non-COPD (n=39). We analyzed the gut microbiome according to CAT score and treatment response to chemotherapy (CHT) using high-throughput sequencing.

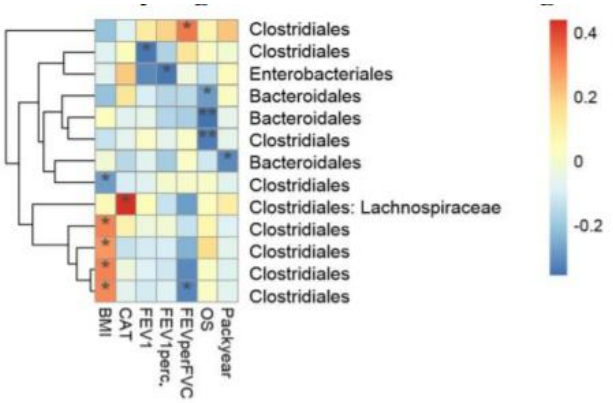

Figure 1 shows correlations of BMI, CAT, lung function parameters (forced expiratory volume in 1 second (FEV1), and forced vital capacity (FVC)), overall survival (OS), smoking in pack-year, and bacteria. Spearman correlation is represented on a scale (r-value), Significance * p <0,05, ** p <0,01 (with FDR correction).

Lachnospiraceae bacterium showed a positive correlation with CAT score. Acidaminococcus intestini and Eubacterium dolichum had negative correlations with CAT scores. Non-responders had significantly higher CAT scores (vs. responders, median 23 vs 13, respectively ($p<0.05$). There was a trend towards longer OS in non-COPD patients (vs. COPD, median OS 11 vs 7.3, months, respectively). In cohort#1, we found associations with certain gut bacteria species and CAT scores. A higher CAT score predicts poor response to CHT during the treatment of lung cancer patients. The data were presented at the European Respiratory Society Congress.

Next, we analyzed cohort #2. As described in the original research plan, we next analyzed CT QTA and microbiome. This part of the study aims to combine a QTA and microbiome-based biomarker signature in NSCLC to predict patients' survival, including anti-PD immunotherapy (ICI)-treated patients.

We included (n=129) patients and analyzed stool shotgun metagenomics (MG) and CT scans of advanced-stage NSCLC patients treated with ICI. QTA was applied to primary tumors obtaining 103 continuous CT parameters that describe tumor pixel distribution. We analyzed the abundance of bacterial species. QTA was applied to CT images of primary tumors. Three-dimensional tumor segmentation was performed using the 4.10 version of 3D Slicer, and a total of 103 CT parameters from each CT image were obtained. We used the Sklearn machine learning library in Python for data preprocessing and standardization, reducing the number of CT parameters by Principal Component Analysis (PCA). The components thus obtained were further analyzed with hierarchical cluster analysis. Responders versus non-responders and patients were analyzed based on naïve Bayes and k-means clustering. We verified the accurateness of the machine learning (AI) algorithms with leave-one-out cross-validation. AI-based on Extreme Gradient Boosting, a decision tree-based algorithm evaluated associations between outcomes and various clinicopathological parameters, including COPD and drug toxicity.

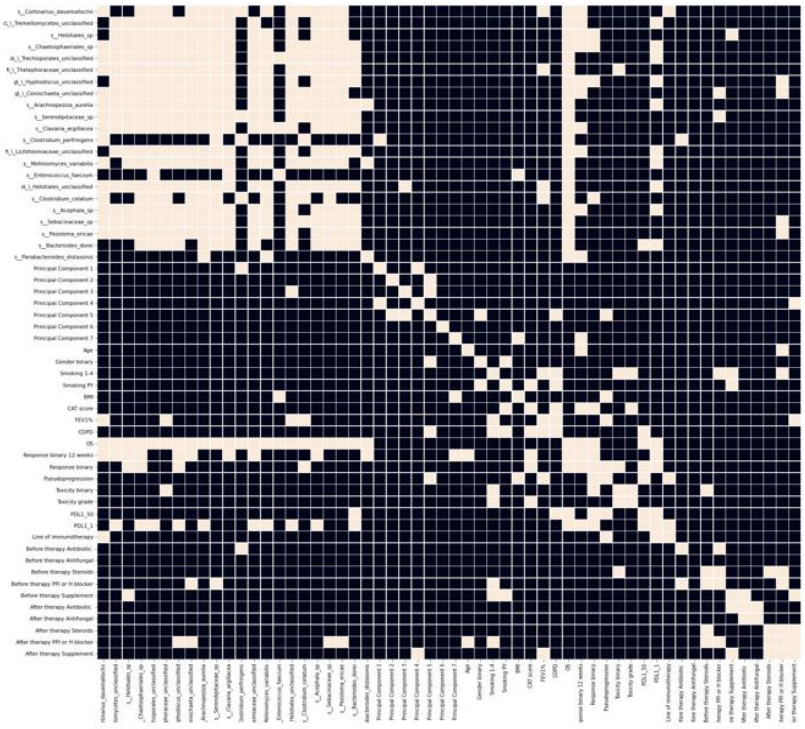

Figure 2. Correlation of QTA, microbiome,and clinicopathological parameters. Positive correlations (r > 0.3) with a p<0.05 threshold is represented with bright color. (Spearman correlation.). Not published data

Principal component analysis (PCA) was performed on CT-derived parameters, identifying seven components that explained 80% of the data variation. We found that two out of 7 primary components significantly differed in patients with COPD (p=0.045) and response to therapy (p=0.0015). Bacteroides sp. and Parabacteroides sp. were associated with long OS (>6 mo), and Thelephoraceae and Lachnospiraceae bacterium with treatment toxicity. AI identified MG

signature for patients with favorable response to ICI and high PD-L1 expression (≥50%) with 84% and 79% accuracy, respectively. From QTA parameters and MG combined, positive predictive value (PPV) for response to therapy and OS were 90% for both.

In cohort# 2, we found distinct gut bacterial microbiome communities together with CT QTA signatures might help selecting NSCLC patients for therapy. These data are under review.

Next, we analyzed the interaction of microbiome and malnutrition in **cohort#3.** (Ni Y, Lohinai Z, et al. ISME J. 2021 May 17.) The etiology of cachexia is poorly understood, and we have no effective therapy to prevent cachexia. Cancer cachexia is associated with worse performance status and frequently limits oncotherapy administration. We aimed to analyze the role of the human gut microbiome in cachexia by integrating shotgun metagenomics and plasma metabolomics of 31 lung cancer patients. The cachexia group showed significant differences in the gut microbial composition compared to non-cachectic patients. There was a significant difference in beta but not alpha diversity of phylum abundance between cachexia and non-cachexia patient groups. (Figure 3). We decided to use the plasma samples collected for metabolomic analysis. The performed chromatographic and mass spectrometric characteristics added more novelty and validation aspects of microbiome functionality than circulating DNA and RNA that was extensively studied by others. (Wu TH et al., 2019 Aug;78:31-41, Scilla KA et al, 2019 Jun 15;20(7):61.) The annotation of metabolites was processed according to the chemical analysis working group metabolomics standards initiative. Our results showed that branched-chain amino acids (BCAAs), methylhistamine, and vitamins were significantly depleted in the plasma of cachexia patients, which was also reflected in the depletion of relevant gut microbiota functional pathways. BCAAs and 3-oxocholic acid enrichment in non-cachectic patients were positively correlated with gut microbial species Prevotella copri and Lactobacillus gasseri, respectively.

Furthermore, the gut microbiota capacity for lipopolysaccharides biosynthesis was significantly enriched in cachectic patients. The involvement of the gut microbiome in cachexia was further observed in a high-performance machine learning model using solely gut microbial features. Of note, the abundance of sputum and bronchoalveolar lavage microbiome bacterial density was low, increasing the risk of false-positive signals in the preliminary dataset, in line with other studies published. (Lee SH et al. 2016 Dec;102:89-95.) Additionally, the abundance and complexity of the interaction of gut microbiota with immunity seemed more relevant; therefore, we focused on the gut microbiome in-depth analysis in our work.

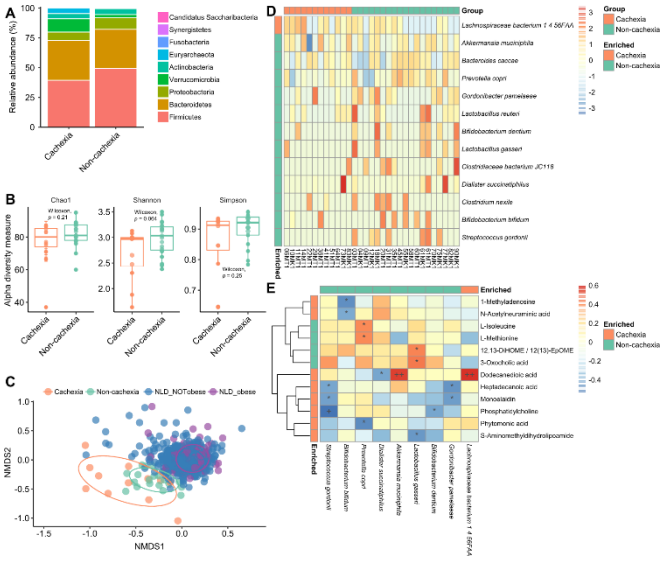

Figure 3. (A) Phylum abundance comparison between cachexia and non-cachexia patient groups. (B) Comparison of microbial alpha diversity: Chao1 ($p = 0.21$, Wilcoxon rank-sum test), Shannon index ($p = 0.064$, Wilcoxon rank-sum test), Simpson index ($p = 0.25$, Wilcoxon rank-sum test). (C) Non-metric multidimensional scaling (NMDS) plot comparing cachexia and non-cachexia patient groups together with a healthy Dutch cohort (NLD) of 471 subjects, based on the gut bacterial species compositions using Bray-Curtis dissimilarities ($p = 0.001$, $r = 0.212$, ANOSIM). The BMI cutoff of 25 was used to group NLD samples into "NLD_NOT obese" and "NLD_obese". (D) Heatmap of differentially abundant bacterial species ($p < 0.05$, prevalence higher than 20%). Color scale represents the row-scaled log-transformed relative abundances of species. (E) Potential mechanistic links between cachexia-associated gut microbiota species and serum metabolites. Spearman's rank correlations were calculated between differentially abundant species and differentially abundant metabolites (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, +FDR < 0.1, ++FDR < 0.05, +++FDR < 0.01, Spearman's rank correlation). (Ni Y, Lohinai Z, et al. ISME J. 2021 May 17.)

The cachexia group showed significant differences in the gut microbial composition, functional pathways of the metagenome, and the related plasma metabolites compared to non-cachectic patients. Our study demonstrates the links between cachectic host metabolism and specific gut microbial species and functions in a clinical setting, suggesting that the gut microbiota could influence cachexia with possible therapeutic applications.

Next, we analyzed the metabolic interactions of the microbiome and physical activity in cohort#4. Impaired exercise tolerance and lung function are markers for increased mortality in lung cancer patients undergoing lung resection surgery. Recent data suggest that the gut-lung axis regulates systemic metabolic and immune functions, and microbiota might alter exercise tolerance. Here, we aimed to evaluate the associations between gut microbiota and outcomes in lung cancer patients who underwent lung resection surgery. We analyzed stool samples from 15 early-stage lung cancer patients collected before and after surgical resection.

We analyzed microbiome associations with post-surgery lung function and cardiopulmonary exercise testing (CPET) to assess the maximum level of work achieved. There was a significant difference, between pre-and post-surgical resection samples, in microbial community functional profiles, and several species from Alistipes and Bacteroides genus associated with the production of short-chain fatty acids (SCFAs) increased significantly in abundance. Interestingly, an increase

in VO2 coincides with an increase in certain species and the "GABA shunt" pathway, suggesting that treatment outcome might improve by enriching butyrate-producing species. In order to predict tumor recurrence vs. non-recurrence, patients were accurately classified by a balance of three bacteria (AUC=95%). Here, we revealed associations between specific gut bacteria with the recovery of lung function and exercise capacity.

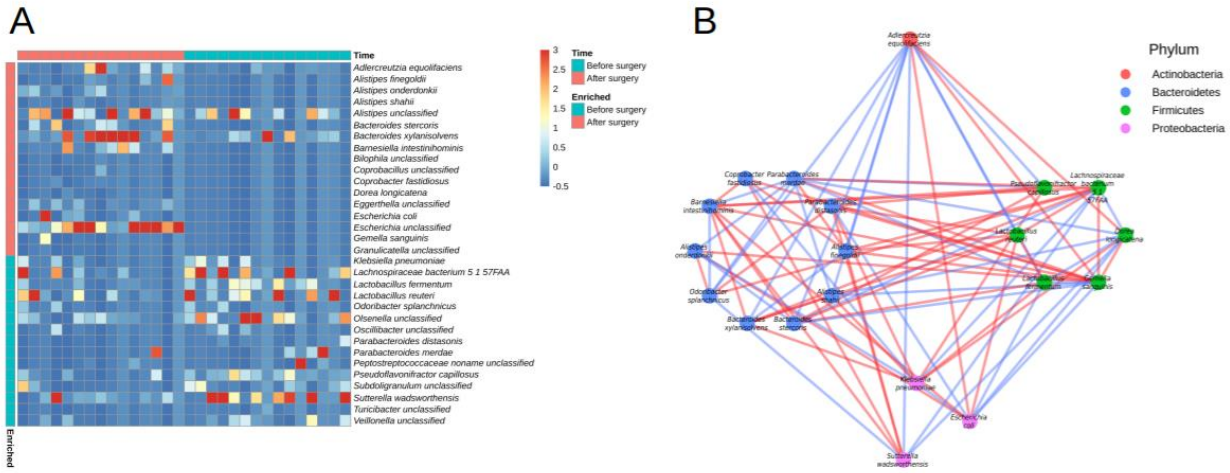

Figure 4. Taxonomic analysis of the gut microbiome according to surgical resection. (A) Heatmap of differentially abundant bacterial species (P<0.05, Wilcoxon signed-rank test) before and after surgical resection. Red and blue in the far-left column indicate increased and decreased relative abundance, respectively. (B) Co-abundance network of bacterial species using SparCC. (Rios-Covian, D et al.Front. Microbiol. 8, 376 (2017)) Only correlations between differentially abundant bacterial species (P<0.05, Wilcoxon signed-rank test) that changed direction were used for network construction. The nodes are colored based on their related phyla. Edge color indicates either correlation that changed from positive to negative (blue) or from negative to positive (red).

Conclusions

Our data suggest that the abundance and functionality of specific gut bacteria might substantially modulate lung cancer treatment outcomes. With the help of the current grant project, we had the opportunity to set up an extensive database on NSCLC patients' clinicopathological characteristics, survival, and microbiome sequencing data.

The present work conducted between 2017-2021 showed associations with gut microbiome functions with COPD and lung cancer, which may have potential therapeutic implications.

Additionally, we constructed a machine learning model that might predict high accuracy cancer treatment outcomes.

Based on our data, specific microbial species were associated with drug toxicity. Additionally, long-term compared to short-term survival in lung cancer patients show differences in microbiome composition and functionality. We used metabolomics to evaluate the role of bacterial interactions with cachexia. Next, changes in microbiota after lung resection surgery were evaluated. We showed associations of CPET parameters and metabolic relations of lung diseases and the microbiota.

With the help of the current NKFIH grant, we have not only had the opportunity to use the high-quality research tools and the state-of-the-art facilities, but we can accelerate our research to translate our data from bench to bedside as soon as possible. Our results are presented in world congresses and published or under review in high-impact journals.

Future directions

With the help of this grant, we are making data openly accessible and generate a large and diverse array of research data, such as: "omics" raw data like metagenomics, differentially abundant species, gene set enrichment, species-functions-metabolite correlations, and pathway analysis results. Final processed results of high value for the scientific community and data used for figures generated for publications and drafts for publications are published in a format that allows being easily reused by other scientists.
We plan to perform fecal microbiota transplantation, probiotics, or other methods to restore healthy or specific microbiota. That might also be a future approach to assess the clinical importance of gut microbiota in lung cancer.

Activities Related to this Grant

In the past 48-month period of the Research Project, our group presented data on our research at the World Conference on Lung Cancer and the European Respiratory Society International Congress. Also, we participated in several scholarly discussions with leading scientists involved in lung cancer research, diagnosis, and therapy. Results of the current project at this stage have made significant progress with publishing a paper on cachexia, and two manuscripts are under review. As requested by OTKA, we state in each one that we are a recipient of an OTKA grant.
With the help of the current OTKA grant, we had the opportunity to use high-quality research methods and network with participating clinicians and researchers Worldwide. Additionally, Dr. Zsolt Megyesfalvi, Ph.D. student and a researcher participant in the grant project, was awarded the first prize at the Semmelweis University Ph.D. Scientific Days 2019 Budapest, April 25–26, 2019 and at the Hungarian Medical Association of America annual congress, Balatonfüred Aug 30, 2019. Balazs Santa and Csilla Kugler medical students awarded prizes at the Students' Scientific Conference, Semmelweis University, and the National Students' Scientific Conference (OTDK 2019). In line with the original grant call/proposal, we successfully set up a young research group. The achieved results of the current project provide a reasonable basis and the opportunity to continue research in the field.