Repeated observation of immune gene sets enrichment in women with non-small cell lung cancer

Jhajaira M. Araujo¹, Alexandra Prado¹, Nadezhda K. Cardenas², Mayer Zaharia¹, Richard Dyer³, Franco Doimi³, Leny Bravo², Luis Pinillos¹, Zaida Morante⁴, Alfredo Aguilar¹, Luis A. Mas¹, Henry L. Gomez¹, Carlos S. Vallejos¹, Christian Rolfo⁵, Joseph A. Pinto¹

¹Unidad de Investigación Básica y Traslacional, Oncosalud-AUNA, San Borja, Lima 41, Peru
²Escuela de Medicina Humana, Universidad Privada San Juan Bautista, Chorrillos, Lima 09, Peru
³Departamento de Patología, Oncosalud-AUNA, San Borja, Lima 41, Peru
⁴Departamento de Medicina Oncológica, Instituto Nacional de Enfermedades Neoplásicas, Surquillo, Lima 41, Peru
⁵Phase I – Early Clinical Trials Unit, Antwerp University Hospital, Antwerp, Edegem 2650, Belgium

Correspondence to: Christian Rolfo, e-mail: christian.rolfo@uza.be

Keywords: non-small cell lung cancer, gender, GSEA, CIBERSORT, immune gene sets

Received: November 19, 2015  Accepted: February 11, 2016  Published: March 06, 2016

ABSTRACT

There are different biological and clinical patterns of lung cancer between genders indicating intrinsic differences leading to increased sensitivity to cigarette smoke-induced DNA damage, mutational patterns of KRAS and better clinical outcomes in women while differences between genders at gene-expression levels was not previously reported. Here we show an enrichment of immune genes in NSCLC in women compared to men. We found in a GSEA analysis (by biological processes annotated from Gene Ontology) of six public datasets a repeated observation of immune gene sets enrichment in women. “Immune system process”, “immune response”, “defense response”, “cellular defense response” and “regulation of immune system process” were the gene sets most over-represented while APOBEC3G, APOBEC3F, LAT, CD1D and CCL5 represented the top-five core genes. Characterization of immune cell composition with the platform CIBERSORT showed no differences between genders; however, there were differences when tumor tissues were compared to normal tissues. Our results suggest different immune responses in NSCLC between genders that could be related with the different clinical outcome.

INTRODUCTION

Lung cancer is the most common malignancy and the leading cause of death worldwide where tobacco smoking is the main risk factor for the development of this cancer [1, 2].

It is well known that lung cancer shows different patterns of clinical characteristics and outcomes according to sex. Women have higher susceptibility to cigarette smoke-induced DNA damage and an increased risk for lung cancer, including higher levels of DNA adducts and higher frequency of the KRAS G12C mutation than men [3, 4]. In addition, female patients tend to present lung cancer at a younger age and in more advanced stages; however, women have better prognostic than men [5, 6].

Due to different susceptibility to DNA damage, according to sex, our research group previously evaluated differences between genders in gene expression and mutational status in genes involved in DNA repair without observing differences. In contrast, in a global evaluation, some gene sets were differentially enriched in women, including immune gene sets [7]. Despite the great advances in the knowledge of the genomic landscape of lung cancer, transcriptional differences between genders have not been previously explored.

With the aim of exploring transcriptional differences of Non-small cell lung cancer (NSCLC) between the genders, we performed a comprehensive analysis of differentially enriched immune gene sets annotated by Gene Ontology Biological Process (C5 BP).
RESULTS

Gene set enrichment analysis (GSEA) reveals enrichment of immune gene sets in NSCLC in women

The structure of the data for GSEA analysis is shown in Table 1. Immune gene sets enrichment was not observed in men. In contrast, we found 40 immune gene sets enriched in women where the most over-represented where “Immune Response” ([GO:0006955]; p-values < 0.00000001; FDR’s 0.4%–24.2%) and “Immune System Process” ([GO:0002376]; p-values < 0.00000001; FDR’s: 0.4%–24.3%) in 06 subsets; “Defence Response” ([GO:0006952]; p-values < 0.00000001; FDR’s: 0.3%–13.6%), “Cellular Defense Response” ([GO:0006968]; p-values between < 0.00000001 to 0.00784929) and “Regulation of Immune System Process” ([GO:0002682]; p-values between 0.00176991 to 0.0060423) in 05 subsets (Figure 1 and Figure 2).

List of all datasets with immune gene sets enrichment in ≥ 2 subsets are shown in Tables S1 and S2. We were not able to find differences in the dataset GSE7670 (without information about smoking status). There were no differences in relative frequencies of immune cell composition in both genders according to the LM22 signature (Figure 3). An important difference was seen between healthy lung tissues vs NSCLC, regardless the smoking status or gender. The main component in tumors was plasma B cells and macrophages. In contrast, healthy lung tissues have a higher proportion of T cells CD8, mast cells and a lower proportion of plasma cells, compared to lung tumors. Detailed values of immune cell composition are described in the Table S3.

Immune cell composition inferred from the transcriptional background

There are few reports describing the value of tumor-infiltrating lymphocytes in NSCLC where a more significant prognostic factor is the presence of tumor-induced bronchus-associated lymphoid tissue (Ti-BALT) composed by mature dendritic cell (DC)/T-cell clusters adjacent to B-cell follicles [18]. Despite of the controversial value of tumor-infiltrating lymphocytes in NSCLC, peripheral leukocytes have more prognostic relevance. A recent meta-analysis study described that a high neutrophils/lymphocyte ratio in peripheral blood is related with a poor prognosis [19]. Also, gene expression patterns of peripheral blood mononuclear cells are highly correlated with the tumor burden in NSCLC and this biological signal could disappear after tumor resection [20, 21].

A different composition of tumor infiltrating immune cells between genders might be a good explanation for our findings in the GSEA analyses and could be more easily linked to the outcomes in women patients with NSCLC; however, the similar composition of immune cells suggests differences in activation of immune pathways and interaction between immune and tumoral cells.

There are several methodologies for the analysis of various datasets to minimize the batch effect [22]. In this work we preferred to construct subsets and perform analyses in each subset individually instead of pooling data. An overall comparison in each dataset could lead to biased results due to smoking status is the main cofounder
and difficult to control. Although we decide not to use a more stringent statistics (FDR < 5%), the strength of our analysis strategy is based on the multiple validation that lead to be able to identify subtle differences in biological signals that could be masked with the batch of samples or more stringent statistics.

Our results are supported by a recent report by Qu et al. (2015), providing the first evidence of gender differences of immune regulation by elements that escape X chromosome inactivation and trigger regulatory networks and activation of genes with immune function in other autosomes in a study done in primary humans T cells [23]. Our results suggest different activities of immune gene sets regardless the immune cell composition in the tumor. Genes such as APOBEC3G and APOBEC3F seem to play and important role in the immune response in NSCLC, while over-representation of CCL5 (a gene widely studied in other malignancies), CD1D, LAT, TRAT1, IL32 and others, suggest different regulatory activities of T lymphocytes in NSCLC in women compared to men. Although “Immun Response” and “Immune System Process” “T-Cell Activation”, “Regulation of Cytokine Biosynthetic Process” (Figure 2).

Recently, immune checkpoints inhibitors such as nivolumab and pembrolizumab are included in the National Comprehensive Cancer Network (NCCN) guidelines for the treatment of metastatic NSCLC subsequent to first line of chemotherapy [24]. A phase III study comparing nivolumab with docetaxel in patients with advanced non-squamous NSCLC showed that nivolumab not improved the overall survival in the subgroup of women patients (HR:0.78, CI: 0.58–1.04) [25]. Likewise, the results of KEYNOTE-010 trial comparing pembrolizumab versus docetaxel in patients previously treated with platinum-doublet chemotherapy showed that pembrolizumab does not increase the progression free survival in women (HR: 1.02, CI: 0.78–1.32). Conversely, both of these immune checkpoint inhibitors improved the outcome of the subgroup of men patients [26]. This contrasting results might be explained by differences in expression of immune genes.

Biological differences of lung tumors among genders should be deeply explored in order to improve the immunotherapeutic approaches. Our study provides evidence of biological differences of NSCLC between genders and the basis for the distinct clinical outcome.

Table 1: Characteristics and composition of subsets included in the GSEA analysis

<table>
<thead>
<tr>
<th>GEO Accession</th>
<th>GSE10072</th>
<th>GSE32863</th>
<th>TCGA</th>
<th>GSE50081</th>
<th>GSE47115</th>
<th>GSE7670</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEO Accession</td>
<td>GSE10072</td>
<td>GSE32863</td>
<td>TCGA</td>
<td>GSE50081</td>
<td>GSE47115</td>
<td>GSE7670</td>
</tr>
<tr>
<td>Array</td>
<td>Affymetrix Human Genome U133A</td>
<td>Illumina HumanWG-6 v3.0</td>
<td>IlluminaHiSeq RNASeqV2</td>
<td>Affymetrix Human Genome U133 Plus 2.0</td>
<td>Illumina HumanHT-12 WG-DASL V4.0 R2</td>
<td>Affymetrix Human Genome U133A</td>
</tr>
<tr>
<td>Histology</td>
<td>Adenocarcinoma</td>
<td>Adenocarcinoma</td>
<td>Adenocarcinoma</td>
<td>NSCLC</td>
<td>NSCLC</td>
<td>Adenocarcinoma</td>
</tr>
<tr>
<td>Stage</td>
<td>IA–IV</td>
<td>IA–IV</td>
<td>IA–IIIB</td>
<td>IB–IIIB</td>
<td>IA–IV</td>
<td>IA–IV</td>
</tr>
<tr>
<td>Samples evaluated</td>
<td>n = 71</td>
<td>116</td>
<td>447</td>
<td>81</td>
<td>16</td>
<td>52</td>
</tr>
<tr>
<td>Subsets in dataset</td>
<td>-</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

| Never Smokers | Normal Tissue | 31 (43.7%) | 15 (48.4%) | 11 (73.3%) | 16 (51.6%) | 59 (50.1%) | 30 (50.8%) |
|               | Male | 4 (26.7%) | 7 (23.3%) |
|               | Female | 11 (73.3%) | 23 (76.7%) |
| Tumoral Tissue | 16 (51.6%) | 29 (49.2%) | 23 (79.3%) |
|               | Male | 13 (81.2%) | 17 (24.6%) |
|               | Female | 3 (18.8%) | 52 (75.4%) |
|               | 6 (20.7%) | 18 (75%) |
| Current Smokers | Normal Tissue | 16 (40%) | 28 (49.1%) | 6 (21.4%) | 22 (78.6%) |
|               | Male | 4 (25%) | 6 (21.4%) |
|               | Female | 12 (75%) | 22 (78.6%) |
| Tumoral Tissue | 24 (60%) | 29 (50.1%) | 29 (50.1%) | 378 (84.6%) | 57 (70.4%) | 16 (100%) |
|               | Male | 16 (66.7%) | 18 (62.3%) |
|               | Female | 8 (33.3%) | 19 (68.8%) |
|               | 7 (24.1%) | 9 (56.2%) |
| Unknown | Normal Tissue | 26 (50%) | 26 (50%) |
|               | Male | 5 (19.2%) | 5 (19.2%) |
|               | Female | 21 (80.8%) | 21 (80.8%) |
| Tumoral Tissue | 26 (50%) | 26 (50%) |
|               | Male | 5 (19.2%) | 5 (19.2%) |
|               | Female | 21 (80.8%) | 21 (80.8%) |
MATERIALS AND METHODS

NSCLC datasets

With the aim of explore differences in immune gene sets in NSCLC between genders in a functional approach, we retrieved normalized gene expression data of five NSCLC datasets (GSE10072, GSE32863, GSE50081, GSE47115, and GSE7670) from the NCBI GEO website (http://www.ncbi.nlm.nih.gov/geo/). One RNASeq Level 3 dataset of lung adenocarcinoma was downloaded from the TCGA website (https://tcga-data.nci.nih.gov/docs/publications/luad_2014/).

Data preprocessing

Values from datasets downloaded from the NCBI GEO were log₂ transformed and median centered. In the TCGA dataset, expression values of “zero” were set to the overall minimum value and all data were log₂ transformed and median centered.

GSEA analysis

We evaluated 825 gene sets in biological processes annotated from Gene Ontology (C5 BP) in the Molecular Signature Database (MSigDB; http://www.broad.

---

![Diagram showing gene expression data and -log10 (P-value)]
Figure 1: Immune gene sets enrichment more overrepresented in women. –log10 of P-values are shown in different subsets (A). Enrichment profile generated with GSEA in the gene set “immune process” comparing NSCLC in men vs NSCLC in women shown enrichment in female in smokers (B) and in non-smokers (C).
C  GSE32863 (Tumor/Smokers)

Enrichment plot:
REGULATION_OF_IMMUNE_SYSTEM_PROCESS

D  TCGA (Tumor/Non-Smokers)
Figure 2: Enrichment profile and heatmaps generated with comparing the biological process: “REGULATION_OF_IMMUNE_SYSTEM_PROCESS” between NSCLC in men vs NSCLC in women. This enrichment was observed in subsets of smokers (Figures A, B and C), non-smokers (D) and healthy tissue from non-smokers (E).
Because of the small sample size of the subsets, the GSEA was conducted with 1000 gene set permutations. The GSEA analyses were performed using the Java GSEA implementation downloaded from www.broad.mit.edu/gsea/msigdb/. Gene sets from samples of men vs women were compared. A gene set was considered enriched when it was included in the top 50 rank in at least two subsets with a \( p \)-value < 0.05 and a False Discovery Rate (FDR) < 25% and when was represented in only 1 gender.

To avoid the confounding effect of smoking and possible batch effects with the pooling of samples, we decided to divide each dataset into subsets according to smoking status (smokers vs no smokers) and type of tissue (normal vs tumoral) when these samples were available. Patients with unknown smoking status or former smokers were excluded from this analysis, except for patients from the dataset GSE7670 whose samples lack this information (Table 1).

**Analysis of immune cells composition from gene expression data**

To investigate if signal of enrichment of immune gene sets in women is related to a different background of immune cells in the samples, gene expression analysis with the online analytical platform CIBERSORT (https://cibersort.stanford.edu/) was done. CIBERSORT quantifies relative levels of the abundances of distinct cell types in a mixed cell population [27].

We evaluated our subsets profiled with Affymetrix platforms (GSE10072 and GSE50081) with the LM22 gene signature that is able to identify 22 immune cell types (LM22 signature is only validated for Affymetrix microarrays data). For this analysis, labels of affymetrix probes were replaced with gene names. Due to more than one affymetrix probe could represent a gene, genes were collapsed to the highest value. After this step, the data were quintile normalized. Analyses were done with 100 permutations with default statistical parameters. The results were filtered by a maximum \( p \)-value of 0.05.

![Figure 3: Relative leukocyte fractions evaluated by CIBERSORT in Affymetrix datasets to infer relative RNA fractions from 22 leukocyte subsets (LM22 signature) in each sample. Shown are the average fractions in each dataset.](image-url)
**Authors’ contributions**

J.A.P., J.M.A., H.L.G. and CV. were responsible for the design, execution of the study, data interpretation and final manuscript writing; J.M.A., A.P., N.K.C., M.Z., R.D., F.D. and LB assisted with the collation of study materials, data analysis and data interpretation; J.M.A., A.P., N.K.C. and L.B., assembly and verification of the data; L.P., M.Z., Z.M., A.A., H.L.G AND C.R., contributed to the interpretation of the results and participated to the manuscript writing. J.A.P., C.V. and C.R., supervised the study. All authors read and approved the final manuscript.

**ACKNOWLEDGMENTS AND FUNDING**

This work was funded by a Research Grant of AUNA.

**CONFLICTS OF INTEREST**

The authors declare they have not potential conflicts of interests with this research.

**REFERENCES**


