SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Statistics and Quantitative Methods
Division of Biostatistics, Epidemiology and Public Health
Laboratory of Clinical Biostatistics

PhD program in Public Health
Cycle XXXIII
Curriculum in Epidemiology and Research on the System of Prevention, Diagnosis and Treatment

Sex-based differences in cancer immunotherapy efficacy

Surname: PAGAN
Name: ELEONORA
Registration number: 735302

Tutor: Prof. Vincenzo Bagnardi
Coordinator: Prof. Guido Grassi

Academic year 2019/2020

Abstract

In the last decade lots of research efforts were put in the field of immunotherapy, a relatively new class of treatments that boost the body's natural defenses to fight against cancer.

In medical research, despite growing evidence that sex (i.e., differences between men and women at biological level) might be a disease treatment modifier, sex-based differences in the presentation, progression, adverse events as well as in the prophylactic or therapeutic treatment of diseases were rarely analyzed and reported.

Only in 2018, Conforti and colleagues found, for the first time, that considering patients affected by several types of advanced cancers, and treated with immune checkpoint inhibitors as monotherapy or as a combination therapy, males derive larger effect on OS then females do.

The overall aim of this doctoral thesis was to extend previous (limited) knowledge on sex-based differences in cancer immunotherapy efficacy and to further explore mechanisms at molecular level that regulate anticancer immune response in men and women.

First, we investigated whether, and confirmed that, women with advanced lung cancer derive larger benefit than men from the combination of chemotherapy to an anti-PD-1/PD-L1. We found also an interaction between patients' sex and the efficacy in OS of two therapeutic strategies (anti-PD-1/PD-L1 alone or in combination with chemotherapy compared to standard chemotherapy) with opposite direction of the effect in men and women: men derive larger benefit than women with an anti-PD-1 treatment alone, while women have better survival with anti-PD-1/PD-L1 plus chemotherapy.

Then we deeply studied the sex-based dimorphism of the response to lung cancer immunotherapy, conjecturing that the heterogeneity of response to different immunotherapeutic strategies might be due to differences in the molecular mechanisms that drive anticancer immune response in men and women. We observed a less efficient tumor recognition and infiltration by immune system in men compared to women. In particular, in men, we found a lower abundance of a number of immune cell types in the tumor microenvironment (TME), a significantly higher T-cell exclusion score, a smaller T-cell receptors repertoire diversity and a lower amount of ubiquitous expanded T-cell receptors. We found that such poorer immune infiltration of tumors in men may depend on a less efficient tumor neoantigens presentation to the immune-system, due to lower expression levels of human leukocyte antigen (HLA) class I and II molecules, higher frequency of HLA type I loss of heterozygosity and/or alterations in other component of the antigen presentation machinery. We also showed that, among the molecular pathways and biological

processes most significantly enriched in the TME of women, there were many directly related to the anticancer immune response. Contrary, none of the gene sets found significantly enriched in tumors arising in men were directly related to anticancer immune responses. Moreover, we provided a clear example of the potential clinical implications of our findings, showing significant differences in the association between tumor mutational burden and survival benefit observed in men and women treated with anti-PD-1/PD-L1 antibodies.

We implemented several statistical methods to answer the different questions depending on the aim of each study. We used meta-analyses to combine results from several studies and to produce estimates of the overall sex-effect of interest. We used cox proportional hazard regression model to analyze survival data and, as a mean to investigate departures from linearity, restricted cubic splines were applied to model the relationship between continuous covariates and the survival outcome of interest. Several bioinformatic tools were used to process the data. Moreover, we implemented the Gene Set Enrichment Analysis methodology in the statistical software SAS with an extension to meta-analysis.

Our results have several straightforward implications in the context of both translational and clinical research including the need to explore differential therapeutic approaches in men and women with cancers, to improve results for both.

List of publications

I. Conforti F, Pala L, Bagnardi V, Viale G, De Pas T, Pagan E, Gelber RD and Goldhirsch A. **Sex-based differences of the tumor mutational burden and T-cell inflammation of the tumor microenvironment.**
Ann Oncol. 2019;30(4):653-655. doi:10.1093/annonc/mdz034

II. Conforti F, Pala L, Bagnardi V, Viale G, De Pas T, Pagan E, Pennacchioli E, Cocorocchio E, Ferrucci PF, De Marinis F, Gelber RD, Goldhirsch A. **Sex-Based Heterogeneity in Response to Lung Cancer Immunotherapy: A Systematic Review and Meta-Analysis.**
J Natl Cancer Inst. 2019;111(8):772-781. doi:10.1093/jnci/djz094

III. Conforti F, Pala L, Pagan E, Bagnardi V, De Pas T, Tortora G, Bria E, Minucci S, Joffe H, Gelber RD, Viale G, Giaccone G, Goldhirsch A. **Molecular characterization of sex-based dimorphism of anticancer immune response and mechanisms of immune evasion, in patients with early non-small cell lung cancer.**
Work in progress.

Contents

List of abbreviations

aDC	Activated dendritic cells
ALK	Anaplastic lymphoma kinase gene
APC	Antigen presenting cell
CAF	Cancer associated fibroblasts
CD4+ Tcm	CD4+ central memory T-cells
CD4+ Tem	CD4+ effector memory T-cells
CD8+ Tcm	CD8+ central memory T-cells
CD8+ Tem	CD8+ effector memory T-cells
cDC	Conventional dendritic cells
CI	Confidence interval
CLP	Common lymphoid progenitors
CMP	Common myeloid progenitors
CTLA-4	Cytotoxic T-lymphocyte-associated antigen-4
DC	Dendritic cells
EGFR	Epidermal growth factor receptor
ES	Enrichment score
FDA	Food and Drug Administration
FDR	False discovery rate
GMP	Granulocyte-macrophage progenitors
GO	Gene ontology
GSEA	Gene set enrichment analysis
HLA	Human leukocyte antigen
HR	Hazard ratio
HSC	Hematopoietic stem cells
ICI	Immune-checkpoint inhibitor
iDC	Immature dendritic cells
iGSEA	Integrative gene set enrichment analysis
LCE	Lung cancer explorer project
LOH	Loss of heterozigosity
LRT	Likelihood ratio test
ly Endothelial cells	Lymphatic endothelial cells
MAPE	Meta-analysis for pathway enrichment
Mb	Megabase
MEP	Megakaryocyte-erythroid progenitors
MHC	Major histocompatibility complex
MPP	Multipotent progenitors
MSC	Mesenchymal stem cells
MSigDB	Molecular signatures database
MSKCC	Memorial Sloan Kettering Cancer Center
mv Endothelial cells	Microvascular endothelial cells
NES	Normalized enrichment score
NKT	Natural killer T-cells

NSCLC	Non-small cell lung cancer
OR	Odds ratio
OS	Overall survival
PD-1	Programmed cell death protein 1
pDC	Plasmacytoid dendritic cells
PD-L1	Programmed death-ligand 1
PFS	Progression-free survival
PH	Proportional hazard
RCS	Restricted cubic spline
RCT	Randomized controlled trial
SCLC	Small cell lung cancer
ssGSEA	Single sample gene set enrichment analysis
TCGA	The Cancer Genome Atlas
TCR	T-cell receptor
Tgd cells	Gamma delta T-cells
Th1 cells	Type 1 T-helper cells
Th2 cells	Type 2 T-helper cells
TIDE	Tumor immune dysfunction and exclusion
TILS	Tumor infiltrating lymphocytes
TMB	Tumor mutational burden
TME	Tumor microenvironment
Tregs	Regulatory T-cells
WES	Whole exome sequencing

Chapter 1

Background

Cancer arises from the transformation of normal cells into tumor cells in a multistage process that generally progresses from a pre-cancerous lesion to a malignant tumor. If the unregulated cellular proliferation invades other parts of the body, we observe metastases.[1]

Cancer is the second leading cause of death worldwide accounting for an estimated 9.6 million deaths in 2018. The most common types of cancer in men are cancers of the lung, prostate, colorectum, stomach and liver while in women are cancers of the breast, colorectum, lung, cervix and thyroid.[1]

Most cancers, if diagnosed at an early stage, can be cured. The situation is more complex when the disease is already advanced since the chance of a cure is significantly reduced and treatments like radiotherapy, chemotherapy and surgery most of the time can only improve quality of life and prolong survival.[2]

Cancer cells are generally unstable and proliferate evading the immune system surveillance. The following mechanisms may be exploited:

- genetic changes of tumor cells that make them less visible to the immune system;
- proteins on the surface of tumor cells that turn off immune cells;
- changes in the normal cells around the tumor able to interfere with how the immune system responds to the cancer cells.[3]

1.1 Immune-checkpoint inhibitors

Immunotherapy is the new emerging class of treatments for certain types of cancer (also in advanced stages). It can be used as monotherapy or combined with other types of treatment, like chemotherapy or radiotherapy. Also called biotherapy or biologic therapy, it boosts the body's natural defenses to fight against cancer, ideally resulting in the eradication of the disease.[4–6]

Several types of immunotherapy are used to treat cancer, like T-cell transfer therapy, monoclonal antibodies, vaccines and immune system modulators. A revolutionary milestone in the field was the development of immune-checkpoint inhibitors (ICIs).[3]

T-cells are a type of lymphocyte that play a central role in the immune response. On the surface of T-cells proteins (called 'Checkpoints') are expressed. These proteins act as 'Off' switches and naturally prevent T-cells from attacking other healthy cells in the body.[4]

ICIs strengthen the antitumor immune response by interrupting the inhibition of T-cells allowing the body to recognize the tumor cells and triggering the immune system to attack and eliminate the cancerous cells. Antibodies have been developed and tested for most of the inhibitory receptors and only few of them have already been approved for use in humans.[4,7]

Immune checkpoint molecules that suppress the T-cell activity include cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4), programmed cell death protein 1 (PD-1) and programmed death-ligand 1 (PD-L1).

CTLA-4 inhibitor therapy: CTLA-4 is a protein on the surface of T-cells and its main function is to regulate the amplitude of early stage T-cell activation.[6] Initial T-cell activation involves antigen presentation by the major histocompatibility complex (MHC) molecules on the antigen presenting cells (APCs) to the corresponding T-cell receptor (TCR) on naïve T-cells. The interaction of the costimulatory T-cell receptor CD28 with the B7 ligand is required for full activation. T-cells lose the capacity to recognize cancer cells when CTLA-4 attaches to protein B7 on the APC surface (T-cell exhaustion). CTLA-4 inhibitors bind CTLA-4 and allow the T-cells to be activated and to kill cancer cells (Figure 1.1).[6,8]

Ipilimumab and Tremelimumab are CTLA-4 inhibitors.

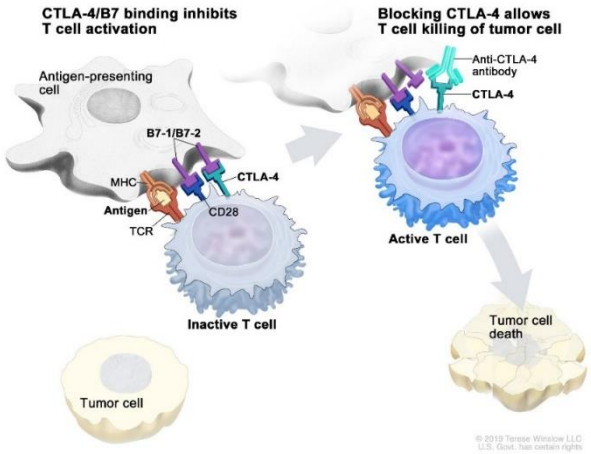

Figure 1.1 CTLA-4 inhibitor. The binding of B7-1/B7-2 to CTLA-4 keeps the T-cells in the inactive state so they are not able to kill tumor cells in the body (left panel). Blocking the binding of B7-1/B7-2 to CTLA-4 with an immune checkpoint inhibitor (anti-CTLA-4 antibody) allows the T-cells to be active and to kill tumor cells (right panel).[5]

PD-1 and PD-L1 inhibitor therapy: PD-1 is a protein on the surface of T-cells while PD-L1 is a protein found on some types of cancer cells.[5] The main role of PD-1 is to down-regulate T-cell activity in peripheral tissues to avoid autoimmunity reactions during inflammatory response to infection.[6] When PD-1 binds PD-L1, signals that prevent phosphorylation of key signaling intermediates in the T-cells are generated; as a consequence, T-cells are down-regulated and are unable to kill cancer cells. PD-1 and PD-L1 inhibitors keep PD-1 and PD-L1 proteins from attaching to each other and this allows the T-cells activation, expansion and effector functions (Figure 1.2).[8]

Pembrolizumab, Nivolumab and Cemiplimab are PD-1 inhibitors while Atezolizumab, Avelumab and Durvalumab are PD-L1 inhibitors.

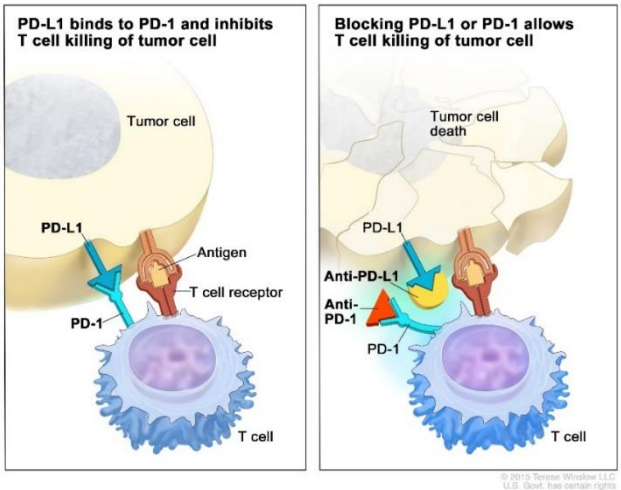

Figure 1.2 PD-1 and PD-L1 inhibitor. The binding of PD-L1 to PD-1 keeps T-cells from killing tumor cells in the body (left panel). Blocking the binding of PD-L1 to PD-1 with an immune checkpoint inhibitor (anti-PD-L1 or anti-PD-1) allows the T-cells to kill tumor cells (right panel).[5]

CTLA-4 therapies are considered 'leaders' of the immune checkpoint inhibitors because CTLA-4/B7 synapse stops potentially autoreactive T-cells at initial stage of naïve T-cells activation, mainly within lymph nodes. Whereas, the PD-1/PD-L1 pathway acts later protecting the tumor cells from T-cells attack through the PD-L1 expression. Thus, the PD1/PD-L1 pathway represents an adaptive immune resistance mechanism exerted by tumor cells in response to endogenous anti-tumor activity.[8]

In 2011 the Food and Drug Administration (FDA) approved an anti-CTLA-4 (Ipilimumab) for the treatment of melanoma. In 2014, Nivolumab and Pembrolizumab were approved for the treatment of melanoma and received indications for use in other type of cancers, like non-small cell lung cancer (NSCLC), renal cell carcinoma, head and neck squamous cell

carcinoma, urothelial carcinoma and microsatellite instability-high colorectal cancer. Recently, Cemiplimab was approved for the treatment of metastatic cutaneous squamous cell carcinoma.[6,9]

The clinical use of anti-CTLA-4 and anti-PD-1/PD-L1 therapies has extraordinarily increased overall survival (OS) in patients with cancer.[10] Despite the potential success and advantages of ICIs, only a fraction of patients benefit from ICIs.[4,7,9]

Failure of the ICI therapy can occur when there is a lack of anti-tumor T-cells generation, when there is an inadequate tumor-specific T-cells function or when a damage in the T-cell memory occurred.[4] A crucial part of this game is played by tumor-infiltrating T-cells and, in many cases, the effectiveness of ICIs is limited by the lack of adequate antitumor immunity in the tumor microenvironment (TME) (it consists of factors extrinsic to cancer cells like immune and stromal cells, vasculature, extracellular matrix and cytokines that influence response to therapy).[9–11] Analyzing the TME, three major types of scenario of infiltration of immune cells can be seen: immune desert, immune excluded and immune inflamed. Immune desert is characterized by the absence of T-cells in the TME and the lack of suitable T-cell priming or activation. The immune excluded phenotype exhibits the presence of multiple chemokines, vascular factors or mediators and stromal-based inhibition; however, accumulated T-cells are unable to infiltrate the TME. Immune inflamed tumors demonstrate infiltration of multiple immune cell subtypes, recruited to help them shield from attack by the immune system.[7,8,11]

Moreover, patients undergoing ICI treatment might develop, at some point in time, an innate and adaptive resistance to immunotherapy restricting its effectiveness.[4,9] They might also develop severe immune-related adverse events because of the immunologic hyper-activation of normal tissues.[7]

Therefore, the development of predictive biomarkers is critical for differentiating responders and non-responders, to avoid any adverse effects, and to identify subgroups of patients that can potentially derive the greatest benefit from ICI treatment.[4,7,12]

PD-L1 immunohistochemistry, T-cell infiltration levels, T-cell receptor clonality, gene expression signatures, and peripheral blood markers have been correlated with clinical response.[13] Additionally, in small cohorts of melanoma patients treated with anti-CTLA-4 and in NSCLC, melanoma, and bladder cancer patients treated with PD-1/PD-L1 inhibitors, an association between high mutational load, measured by tumor mutational burden (TMB), and clinical benefit was observed.[13] However, tumors with similar TMB can have very

different response to ICI so additional mechanisms are under play and should be investigated.[9]

1.2 Sex-based differences in response to immune-checkpoint inhibitors

In medical research, despite growing evidence that sex (i.e., differences between men and women at biological level) might be a disease treatment modifier, sex-based differences in the presentation, progression, adverse events as well as in the prophylactic or therapeutic treatment of diseases were rarely analyzed and reported.[14]

This is an inheritance from decades of inadequate representation of females in biomedical and clinical studies that began in 1977 when the U.S. FDA recommended to exclude females of childbearing age from clinical trials of drugs. Clearly this decision was taken to protect pregnant females and their fetuses from adverse outcomes but, as an obvious consequence, now there is a medical and social need to understand how drugs work in the overall population, which includes females. Only in the early 1990s, the FDA and the National Institute for Health in the USA recommended the inclusion of female subjects in clinical trials. Although females are now included in clinical trials, analyses of whether outcomes differ between females and males and whether there are different mechanisms mediating the difference are still inadequate.[14]

Sex is an important factor in the pathogenesis and prognosis of many diseases including cancers (outside the reproductive tract) and can modulate pharmacokinetics, pharmacodynamics, and toxicities of drugs.[15]

Throughout the life course, males have a higher risk of malignancies and two-times higher risk of mortality from cancers than females (greatest differences observed for larynx, esophagus, bladder, and lung cancers).[14]

Sex affects also innate and adaptive immune response. In fact, on average, women mount stronger innate and adaptive immune responses than men do, clearing more rapidly pathogens on one side, and developing more systemic autoimmune diseases on the other side (roughly 80% of patients with systemic autoimmune diseases are women). At the cellular level, sex-based differences in innate and adaptive immune response were reported in previous studies, and this probably reflects complex interactions between genes, hormones, the environment, and commensal microbiome composition.[14–16]

The X chromosome contains a large number of immune-related genes that code for proteins involved in the regulation of innate and adaptive immunity. Immune-related genes encoded

on the X chromosome may escape X inactivation, resulting in higher expression levels in women than men.[14]

Also, sex hormones modulate the development and function of multiple immune-cell populations and proteins, like PD-1 and PD-L1, and constitute another major determinant of sex differences in immunity.[16–19]

Moreover, sex-related differences have been described also in the composition and amount of intratumoral immune infiltrates as well as in tumor expression levels of PD-L1 across a large spectrum of tumors.[20,21] It has been shown that an intratumoral immune infiltrate, enriched of CD8+ T-cells expressing high levels of CTLA-4 and PD-1, strongly correlates with higher response to anti-PD-1 monotherapy. Additionally, tumors of male patients with melanoma had a statistically significantly larger proportion of these immune infiltrate cells compared with those of women' tumors.[21]

Despite the growing acknowledged sex-related dimorphism in immune system response, little is known about the effect of patients' sex on the efficacy of different regimens of ICI as cancer treatment.

In 2017, Botticelli et al.[22] performed a meta-analysis on phase II and III randomized controlled trials (RCTs) in order to determine sex-differential effects of anti-CTLA-4, anti-PD-1 and anti-PD-L1 treatment in cancer patients. They analyzed nine trials with available information on OS and six trials on progression-free survival (PFS). They observed a benefit in OS for males treated with anti-CTLA-4 compared to females (male OS-HR 0.65, 95% CI 0.55 to 0.77 vs female OS-HR 0.79, 95% CI 0.65 to 0.96, p-value=0.08), while no relevant results were observed with anti-PD-1 treatment, neither for OS nor for PFS.

Later in 2018, Conforti and colleagues[23] hypothesized that male patients could derive a larger relative benefit from ICIs than female patients, and they published a systematic review and meta-analysis to assess the heterogeneity of ICI efficacy on different solid tumors. They found, for the first time, that male patients affected by several types of advanced cancers, and treated with ICIs as monotherapy or as a combination therapy (anti-CTLA-4 plus anti-PD-1 antibodies), had a significantly reduced risk of death compared with men in control groups (pooled OS-hazard ratio (HR) 0.72, 95% confidence interval (CI) 0.65 to 0.79), while in females the benefit obtained with immune checkpoint inhibitors compared with control groups was smaller (pooled OS-HR 0.86, 95% CI 0.79 to 0.93; interaction p-value=0.002).[23]

However in 2019, Wallis et al.[24] reproduced the meta-analysis of Conforti et al.[23] adding seven recently published RCTs and reported similar efficacy of cancer immunotherapy in men and women.

To strengthen the results of Conforti et al.[23], we showed with a preliminary 'exercise' in this field (Study I)[25], that tumors categorized in four groups based on both TMB and T-cell inflamed GEP (a gene expression signature of 18 genes) of patients treated with Pembrolizumab as monotherapy were differently distributed in males and females. The four groups were defined as follows: GEP low and TMB low, GEP low and TMB high, GEP high and TMB low, GEP high and TMB high. In particular, we re-analyzed patient-level data publicly available from Cristescu et al.[26] according with patients' sex and found that, in the entire patient population, the percentage of tumors with low levels of both TMB and GEP score (a condition that strongly predicts the absence of response to Pembrolizumab) was nearly double in women as compared with men (29% in women versus 15% in men, prevalence ratio 1.9, 95% CI 1.22 to 2.96; Figure 1.3). By contrast, the percentage of tumors characterized by high TMB and high GEP score (associated with a high probability of response to Pembrolizumab) was almost halved in women compared with men (26% in women versus 42% in men, prevalence ratio 0.61, 95% CI 0.43 to 0.88). We observed these differences in all the three independent cohorts of patients analyzed in the original publication (i.e., melanoma, head and neck squamous cell carcinoma and the pancancer cohort, that includes 20 different cancer types). Results remained stable after controlling for age and tumor histotype in a logistic multivariable model (p-value=0.03).

Given the limited and contrasting literature available, further studies are needed to deeply evaluate the role of sex in cancer immunotherapy efficacy.

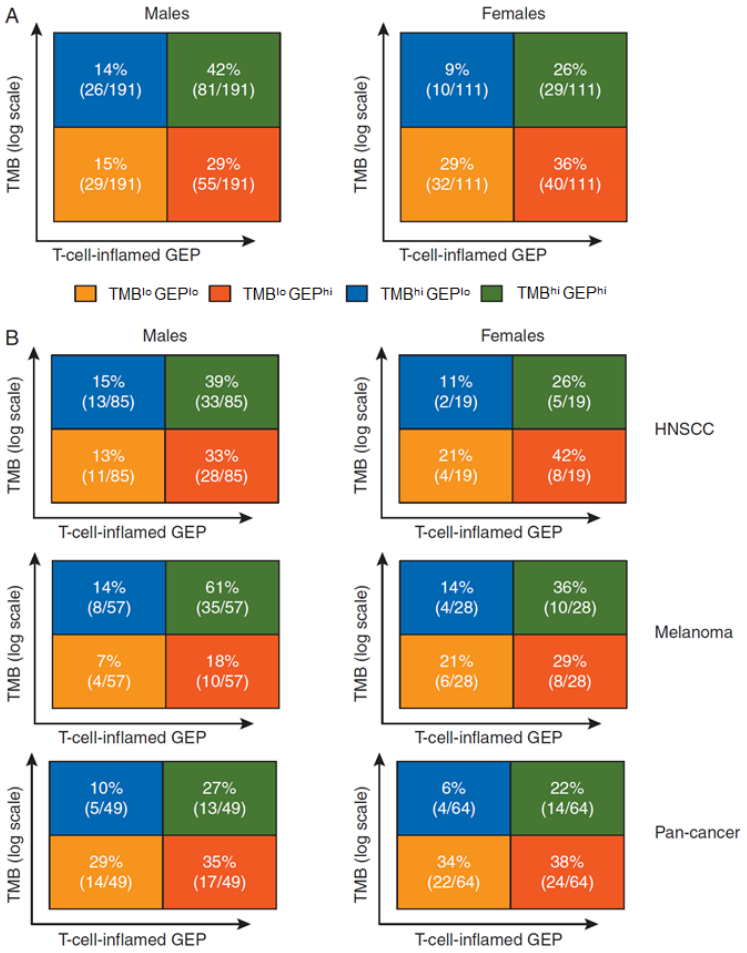

Figure 1.3 Tumors distribution among the four biomarker-defined groups according with patients' sex[25]

Chapter 2

Hypotheses and Aims

2.1 Hypotheses

Sex dimorphism of the immune system is complex. On average, women mount stronger immune response then men.[27] Potential consequences of this are the following:

- the risk of mortality from cancer might be lower in women;[14]
- tumors in women might develop a more intense immune-editing process to become metastatic and this could make advanced female tumors less immunogenic and enriched with stronger mechanisms of immune escape;[23]
- women could become more resistant to immunotherapies and develop more ICI-related adverse events.[12,23]

First, following Conforti et al.[23], we hypothesized that women could derive larger benefit than men from strategies other than therapy with ICI alone. The reason underlying the investigation of sex-based efficacy of ICI plus chemotherapy strategies compared to ICI alone was that chemotherapy can increase the mutational load of tumors and, consequently, the immunogenicity of tumor cells. The already strong immune environment of the female body can then eliminate these tumors with high antigenicity more efficiently than the male body.[27]

On the basis of the results of the previous work, we designed the second study conjecturing that the heterogeneity of response to different immunotherapy strategies might be due to differences in the molecular mechanisms that drive anticancer immune response in men and women. We hypothesized sex-based differences in cell type composition of the immune infiltrate and expression levels of immune checkpoint molecules in the TME, as well as sex-based differences in the mechanisms of immune evasion.

3.2 Aims

The overall aim of this doctoral thesis was to extend previous (limited) knowledge about sex-based differences in cancer immunotherapy efficacy and to further explore mechanisms at molecular level that regulate anticancer immune response in men and women.

More specifically the aims were:

- to investigate, through meta-analysis on published data, the interaction between the efficacy of anti-PD-1/PD-L1 combined with chemotherapy (compared to standard chemotherapy) and patients' sex (Study II);
- to assess, through meta-analysis on published data, the interaction between the efficacy of anti-PD-1/PD-L1 given alone or combined with chemotherapy (compared to standard chemotherapy) and patients' sex (Study II);
- to study, through meta-analysis of public database on genetic patient-level data, sex-based differences in key elements of anticancer immune response (Study III);
- to develop and implement a new statistical method to perform meta-analysis of gene set enrichment analysis (GSEA) data (Study III).

All the analyses were based on patients with lung cancer due to the great availability of RCTs publications and publicly available patient-level data.

Chapter 3

Statistical methods

3.1 Meta-analysis

During the review process, the quantitative analysis of the data extracted from each study is called 'meta-analysis'.

A meta-analysis[28,29] combines results from several studies to produce an estimate of the overall, or average, effect of interest. The overall estimate is a weighted mean of the estimated effects from all the studies, with weights generally equal to the inverse of the variance of each estimate. Artificially increasing the number of patients included, a meta-analysis has a higher study power and precision compared with those of the original studies.

In case of no evidence of statistical heterogeneity, a fixed-effects model can be fitted. This model assumes that the true effect of interest is the same in each study, so that all included studies come from the same population, and variations observed are solely due to sampling errors. This is hardly ever the case since studies are not conducted according to a standard common protocol and there will be always variations in patients' groups, clinical setting, methods of delivery of treatments/interventions, etc. Thus, we cannot assume that all studies come from the same 'population' of studies and a random-effects model becomes a better choice.

The random-effects model not only assumes that the effects of the individual studies deviate from the true effect due to sampling errors (within-study variability), but that there is another source of variability introduced by the fact that the studies do not stem from one single population, but are drawn from a 'universe' of different populations (between-study variability). In a random-effects model the variance of the average effect of interest incorporates both the within- and the between-study variability and therefore the standard error of the estimate is greater, the confidence interval for the true average effect wider and its p-value larger, than the comparable quantities obtained from a fixed-effects model.

In a fixed-effects model, the pooled estimate θ_{IV} is equal to:

$$\theta_{IV} = \frac{\sum_{i=1}^{k} \omega_i \theta_i}{\sum_{i=1}^{k} \omega_i}$$

with θ_i the parameter of interest in each study ($i = 1, \dots, k$).

In case of the most common 'inverse variance method', weights are the following:

$$\omega_i = \frac{1}{SE(\theta_i)^2}$$

The standard error of θ_{IV} is given by:

$$SE(\theta_{IV}) = \frac{1}{\sqrt{\sum_{i=1}^{k} \omega_i}}$$

The Q statistic, which investigates whether the variation in the individual effects is compatible with a hypothesis of chance alone, is used to perform the test for heterogeneity. It follows a Chi-square distribution with $(k-1)$ degrees of freedom and it is equal to:

$$Q = \sum_{i=1}^{k} \omega_i (\theta_i - \theta_{IV})^2$$

Under a random-effects model, θ_i is assumed to follow a Normal distribution with mean θ and variance τ^2. According to DerSimonian and Laird[30], τ^2 has the following equation:

$$\tau^2 = \frac{Q - (k-1)}{\sum_{i=1}^{k} \omega_i - \left(\frac{\sum_{i=1}^{k} \omega_i^2}{\sum_{i=1}^{k} \omega_i} \right)}$$

Once the between-study variability τ^2 is calculated, it is possible to estimate new weights and summary measures:

$$\omega_i' = \frac{1}{SE(\theta_i)^2 + \tau^2}$$

$$\theta_{IV} = \frac{\sum_{i=1}^{k} \omega_i' \theta_i}{\sum_{i=1}^{k} \omega_i'}$$

$$SE(\theta_{IV}) = \frac{1}{\sqrt{\sum_{i=1}^{k} \omega_i'}}$$

Note that when $\tau^2 = 0$, weights reduce to those given in a fixed-effects model ($\omega_i' = \omega_i$).

The $100(1-\alpha)\%$ confidence interval for the overall estimate θ_{IV} is given by the equation:

$$\left\{ \theta_{IV} - \left(z_{1-\frac{\alpha}{2}} \times SE(\theta_{IV}) \right) \; ; \; \theta_{IV} + \left(z_{1-\frac{\alpha}{2}} \times SE(\theta_{IV}) \right) \right\}$$

Finally, another measure of interest is the percentage of the total variation across studies due to heterogeneity:

$$I^2 = 100 \times \frac{[Q - (k-1)]}{Q}$$

It takes values from 0% to 100%, with a value of 0% indicating no observed heterogeneity.

3.2 Meta-analysis of treatment effect in subgroups (analysis of interaction)

Single studies are typically underpowered to explore whether the treatment effect differs between groups of participants and the only way to reliably detect interactions is often through meta-analysis. The analysis of the interaction between the treatment effect and a covariate of interest in a meta-analytical context raises additional complications compared to the ones faced by standard meta-analysis. These complications are often overlooked by reviewers.[31]

Suppose we are interested in the analysis of the interaction between the treatment effect and the participants' sex. The research question is: dose the effect of the experimental treatment compared to the effect of the control treatment differ between females and males?

The easiest way to answer this question might be, first, to estimate the meta-analytical treatment effect separately in female and male patients and then to compare the two meta-analytical estimates with a classical test of heterogeneity. Despite being powerful, this approach is prone to ecological bias since the variations of the treatment effect in the subgroups at the participant level (within-study interaction) could be exaggerated or masked by the across-trial interaction.[31]

To avoid the risk of ecological bias, the best approach should be, first, to assess the interaction of interest within each relevant trial and then to combine the single studies interactions using standard meta-analytical techniques. Despite losing power compared to the previous method (studies that contributes to only one subgroup are not included) this is the solely approach to study interaction avoiding the risk of ecological bias.[31]

3.3 Cox Proportional Hazard Regression Model

The Cox Proportional Hazard (PH) Model[32] is the most common regression method for analyzing survival data. The Cox model is a semi-parametric model and has the form:

$$h(t) = h_0(t)\exp(X_i\beta)$$

where t is the survival time, $h(t)$ is the hazard function, X_i are the covariates and β their estimated coefficients.

The term $h_0(t)$ is the baseline hazard and it corresponds to the value of the hazard when all the X_i are equal to zero. The model is semi-parametric since no assumptions about the shape of the baseline hazard function are made.

The focus of the model is on the hazard $h(t)$, that can be interpreted as the instantaneous risk of experiencing the event of interest at time t. Formally, it is the conditional probability of the event of interest to occur in the interval $[t, t + dt)$, given that it has not yet occurred.

The measure of association between the exposure and the outcome is given by the HR, defined as the hazard in the exposed, divided by the hazard in the unexposed. It simply corresponds to $\exp(\beta)$ for a given covariate. Values of HR>1 indicate covariates positively associated with the event probability, values of HR<1 indicate covariates whose exposure reduce the hazard of event, while HR=1 indicates no association between the exposure and the outcome.

The strongest assumption of the model is the proportionality of the hazards which means that the effects of the predictor variables upon survival are constant over time. The hazards may vary over time, but their ratio is supposed to be the same at all time points. The underlying PH assumption can be tested with both graphical and formal methods (e.g., using Schoenfeld residuals' test) and extensions of the traditional Cox model are available in case of departures from proportionality.

3.4 Restricted Cubic Splines

Including a continuous variable as a predictor in a regression model is not straightforward. The simplest, and often reasonable, relationship we can think of, between a continuous variable and an outcome, is the linear one. The problem is that if the assumption of linearity does not hold, the model will be misspecified.

A naïve solution is to divide the continuous variable into categories. This is very attractive from a decision-making point of view, but the effect of the predictor is forced to be flat within each category and, if the model is used for prediction, the predictive power of the variable is reduced. Possible alternatives are (i) to include the predictor as a continuous variable finding a transformation that produces a linear relationship, (ii) to use quadratic or

cubic polynomials to model the relationship, or (iii) to use splines or fractional polynomials.[33]

A spline function is a smoothly joined piecewise polynomial of degree n. The term 'piecewise' means that the range of values of the predictor is subdivided using a set of knots and separate regression lines or curves ('polynomial') are fitted between the knots. 'Smoothly joined' means that the function has the constraint to be continuous, with continuous derivative, at the join points (knots). The researcher has to choose both the number and position of knots and the degree of polynomials to be used between the knots.[34]

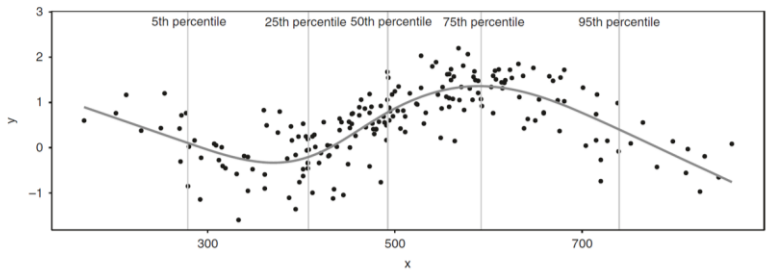

Figure 3.1 Example of a restricted cubic spline[35]

Cubic splines are frequently used since they are really flexible for fitting data, visually smooth, and include fewer constants for fit than higher degree splines do. Restricted cubic splines (RCS) are a particular type of cubic spline with the additional constraint to be linear in the tails (Figure 3.1) and are represented by the following equation[34]:

$$S(x) = \beta_{00} + \beta_{01}x + \sum_{i=1}^{K-2} \beta_{i3}\left[(x - t_i)_+^3 - \frac{(x - t_{K-1})_+^3 \times (t_K - t_i)}{t_K - t_{K-1}} + \frac{(x - t_K)_+^3 \times (t_{K-1} - t_i)}{t_K - t_{K-1}}\right]$$

where $i = 1, \ldots, K$ is the number of nodes, $(x - t_i)_+^3$ is equal to $(x - t_i)^3$ if $(x - t_i) > 0$ and equal to 0 if $(x - t_i) \leq 0$.

Modelling the effect of a predictor using splines simply involves the introduction into the (linear, logistic, Cox, etc.) model of new variables. To evaluate the significance of the non-linearity of the relationship between the predictor of interest and the outcome, we can simply compare, using a likelihood ratio test (LRT), the log-likelihood for a model with the spline variable to the log-likelihood for a model with only a linear effect of the covariate.[34]

3.5 Gene Set Enrichment Analysis

In 2005, Subramanian et al.[36] developed a method called Gene Set Enrichment Analysis (GSEA) to evaluate microarray data at the level of gene set and to determine whether the gene set(s) analyzed is(are) correlated with the phenotypic class of interest (e.g., normal samples vs tumor samples, females vs males, etc.). The identification of the enriched gene set(s) can provide crucial information of molecular functions and mechanisms underlying different diseases. It directly answers the question whether a certain pathway is activated in a given tissue, under some phenotypic condition of interest, and whether the pathway is more active than other pathways.[37]

A gene set is an a priori collection of genes, defined on the basis of biological knowledge, encoding products in a metabolic pathway or located in the same cytogenetic band or sharing the same gene ontology (GO) category.[36]

The original methodology proposed by Subramanian et al.[36] requires the following input information:

- an expression dataset *D* with *N* genes (rows) and *k* samples (columns);
- a correlation/ranking metric between the expression level of *N* genes and the phenotype of interest (e.g., T-test statistic, signal-to-noise ratio, etc.);
- an exponent value *p* to control the weight of the step;
- a selected gene set *S* composed of N_H genes.

The Enrichment Score (ES) reflects the degree to which the set of genes in *S* is overexpressed at the extremes (top or bottom) of the entire ranked list *L*, built on the basis of the chosen correlation metric. The score is calculated by walking down the list *L*, increasing a running-sum statistic when it is encountered a gene in *S* and decreasing it when genes not in *S* are encountered. The magnitude of the increment depends on the correlation of the gene with the phenotype. So, genes with very low correlations do not contribute so much to the enrichment score of the set.[36]

More specifically, as described in Subramanian et al.[36], the ES is calculated as follows:

1) Order the *N* genes in *D* according to the correlation/ranking metric [$r_i = r(g_i)$] of the individual gene g_i with the phenotype, to form the ranked list $L = \{g_1, \dots, g_N\}$, with $i = 1, \dots, N$ number of genes.
2) Evaluate the fraction of genes in *S* ('hit') weighted by their correlation and the fraction of genes not in *S* ('miss') present up to a given position i in *L*:

$$P_{hit}(S,i) = \sum_{\substack{g_j \in S \\ j \le i}} \frac{|r_j|^p}{N_R}, \qquad \text{where } N_R = \sum_{g_j \in S} |r_j|^p$$

$$P_{miss}(S,i) = \sum_{\substack{g_j \notin S \\ j \le i}} \frac{1}{(N - N_H)}$$

The ES for the gene set *S* [ES(S)] will be the maximum deviation from zero of $P_{hit} - P_{miss}$.

The test is a Kolmogorov-Smirnov-like statistic. For randomly distributed genes of *S*, ES(S) will be relatively small, but if genes of *S* are concentrated at the top or bottom of the list L, or otherwise are non-randomly distributed, then ES(S) will be correspondingly high.
The choice of the exponent value *p* determines the influence of the single genes on the score. When $p = 0$ an unweighted statistic is obtained, when $p = 1$ genes in *S* are only weighted by their absolute value of correlation with the phenotype. The suggestion is to use $p < 1$ in order to penalize a lack of coherence in a set and $p > 1$ if the correlation of a small subset of *S* with the phenotype is sufficient to call a set enriched. Subramanian et al. recommended to always use $p = 1$ since other settings should be used only by deeply experienced researchers.

The statistical significance of the ES(S) is calculated by using an empirical phenotype-based permutation test procedure. Specifically, the phenotype labels are permuted for a selected number of times (e.g., 1000 times) and the ES of the gene set is recomputed on the permuted data, generating a null distribution of the ES. The nominal p-value for *S* is then calculated using the positive or negative portion of the empirical null distribution on the basis of the sign of the observed ES(S).

When an entire dataset of gene sets is evaluated, the adjustment for multiple hypothesis testing should be performed with the following procedure:

1) Determine ES(S) for each gene set in the database;
2) For each *S* and 1000 fixed permutation (π) of the phenotype labels, recalculate the correlation metric, reorder the genes in *L* and determine ES(S, π);
3) Normalize the ES(S, π) and the observed ES(S), separately rescaling the positive and the negative scores by dividing by the mean of the positive ES(S, π) and by the mean of the negative ES(S, π), respectively, to yield the normalized enrichment scores NES(S, π) and NES(S);
4) Estimate a global nominal p-value for each NES(S) by the percentage of all (S, π) with NES(S, π)≥NES(S);
5) Compute the false discovery rate (FDR) to control the ratio of false positives to the total number of gene sets analyzed.

3.6 Meta-analysis of Gene Set Enrichment Analysis data

GSEA has become a standard practice and there are many websites and downloadable programs that will provide public datasets and through which it is possible to run the analysis (e.g., GSEA software proposed by the Broad Institute in cooperation with the Molecular Signatures Database (MSigDB) https://www.gsea-msigdb.org/gsea/index.jsp).

In the last decades, we have witnessed an explosion of gene expression profiling studies with publicly available data. However, these studies present with a lot of noisy data, and with small sample sizes (relative to an overwhelming number of genes) that often implies results with inconsistent biological conclusions, even among independent studies targeting the same disease or biological problem. Focusing on a single dataset, as considered in Subramanian et al.[36], we cannot make full utilization of the rich amount of public expression data available. Thus, the idea of meta-analysis was applied also in this context to synthesize multiple microarray studies increasing the robustness and reliability of results.[38–40]

Multiple methodological approaches for meta-analysis of GSEA data have been proposed. In 2010, Shen and Tseng[38] proposed two approaches of meta-analysis for pathway enrichment (MAPE) by combining statistical significance across studies at the gene level (MAPE_G) or at the pathway level (MAPE_P). In MAPE_G, first the association scores with phenotype are calculated in each study, then a meta-analysis on the association scores at gene level is performed, and finally the gene set enrichment analysis is applied and the corresponding Q-values for FDR correction are generated. The framework of MAPE_P is

identical to MAPE_G in the first step. Then, the gene set enrichment analysis is applied in each individual study and a meta-analysis at the gene set level is performed to derive a combined ES with the corresponding Q-value. Both for MAPE_G and MAPE_P, the authors have foreseen to perform the permutation tests using gene permutation or phenotype permutation: the first takes the background information into consideration while destroys the gene correlation structure in the data, the second preserves the gene correlation structures while ignores the background information.[38]

Later in 2013, Chen et al.[39] proposed a Bayesian model to improve the detection of enriched gene sets, simultaneously modelling gene set information and original gene expression data from multiple studies. The computational characteristics of the method becomes excellent when the number of genes or gene sets gets large. However, detecting the convergence of Markov chains and selecting the starting points for parameters may require a great human effort.

In 2018 Lu et al.[40] focused on the development of a new method for meta-analysis of GSEA data, called the integrative GSEA (iGSEA). This method extended meta-analysis approaches developed for genome-wide association studies, based on fixed-effects and random-effects models, to integrate gene set analyses from multiple studies. They proposed a hybrid strategy based on adaptive testing for choosing either random-effects or fixed-effects model to achieve efficiency and stability in performance.

All the above methods can be implemented using packages in R software. We made some attempts to use them but, unfortunately, we found results difficult to interpret and/or unreliable. In addition, in case of problems with the R code, we found difficulties in modifying it. Thus, we decided to develop and implement an 'in-house' approach using the statistical software SAS over which we have full control.

In case of a binary phenotype label, we followed the steps listed below:

1) For each dataset k ($k = 1, \dots, K$), calculate the correlation/ranking metric [$r_{ik} = r(g_{ik})$] of the individual gene g_{ik} of dataset k with the phenotype of interest ($i = 1, \dots, N$ genes). Consider T-test statistic for a binary phenotype label.
2) Apply a random-effects model to estimate the overall strength of the association between gene g_i expression and the phenotype:

$$\hat{r_i} = \frac{\sum_{k=1}^{K} \omega'_{ik} r_{ik}}{\sum_{k=1}^{K} \omega'_{ik}}$$

with ω'_{ik} calculated as described in the paragraph 3.1

3) Order the N genes according to the meta-analyzed correlation/ranking metric $\hat{r}_i$ to form the ranked list $L = \{g_1, \dots, g_N\}$.
4) Evaluate the fraction of genes in S ('hit') weighted by their correlation and the fraction of genes not in S ('miss') present up to a given position i in L:

$$P_{hit}(S,i) = \sum_{\substack{g_j \in S \\ j \le i}} \frac{|\hat{r}_j|^p}{N_R}, \qquad \text{where } N_R = \sum_{g_j \in S} |\hat{r}_j|^p$$

$$P_{miss}(S,i) = \sum_{\substack{g_j \notin S \\ j \le i}} \frac{1}{(N - N_H)}$$

The ES for the gene set S [ES(S)] will be the maximum deviation from zero of $P_{hit} - P_{miss}$. When an entire dataset of gene sets is evaluated, adjustment for multiple hypothesis testing should be performed.

We applied the following procedure:

1) Determine ES(S) for each gene set in the database;
2) For each S and 1000 fixed permutation (π) of the gene labels, determine ES(S, π);
3) Normalize the ES(S, π) and the observed ES(S), separately rescaling the positive and the negative scores by dividing by the mean of the positive ES(S, π) and by the mean of the negative ES(S, π), respectively, to yield the normalized enrichment scores NES(S, π) and NES(S);
4) Estimate a global nominal p-value for each NES(S) by the percentage of all (S, π) with NES(S, π)≥NES(S);
5) Compute the FDR to control the ratio of false positives to the total number of gene sets analyzed.

In comparison to the original methodology proposed by Subramanian et al.[36] (i) we modified the initial steps of GSEA ordering the N genes on the basis of the meta-analyzed T-test statistics instead of simple T-test statistics and (ii) we permuted gene labels instead of phenotype labels, both following Shen and Tseng's[38] MAPE_G method.

Different studies may have different numbers of genes and genes cannot be matched across studies. In case of genes present in only one dataset, we proposed not to eliminate the genes,

but to use for those genes a simple (not meta-analyzed) correlation metric like simple T-test instead.

Chapter 4

Sex-based heterogeneity in response to lung cancer immunotherapy

This chapter is related to study II:

Conforti F, Pala L, Bagnardi V, Viale G, De Pas T, Pagan E, Pennacchioli E, Cocorocchio E, Ferrucci PF, De Marinis F, Gelber RD, Goldhirsch A. **Sex-Based Heterogeneity in Response to Lung Cancer Immunotherapy: A Systematic Review and Meta-Analysis.** *J Natl Cancer Inst.* 2019;111(8):772-781. doi:10.1093/jnci/djz094

4.1 Materials

Data sources and searches

Two systematic reviews on all phase II and III RCTs testing the combination of anti-PD-1/PD-L1 inhibitors alone or with chemotherapy in patients with advanced lung cancer were conducted.

We searched articles in PubMed, MEDLINE, Embase from the inception of each database to October 22, 2018. We also reviewed abstracts and presentations from all major conference proceedings, including the American Society of Clinical Oncology, the International Association for the Study of Lung Cancer, and the European Society for Medical Oncology, from January 1, 2010, to October 22, 2018. We also reviewed the references of articles included in the final selection.

The search terms were 'PD-1', 'programmed death receptor 1', 'PD-L1', 'programmed death ligand 1', 'nivolumab', 'pembrolizumab', 'avelumab', 'durvalumab', 'atezolizumab.'

A first systematic review (and meta-analysis) was conducted on all RCTs testing the combination of anti-PD-1/PD-L1 inhibitors plus chemotherapy. The following inclusion criteria were used: 1) RCTs testing the combination of an anti-PD-1/PD-L1 with chemotherapy against chemotherapy, and 2) data available on HR for PFS and/or OS, according to patients' sex subgroup.

A second systematic review (and meta-analysis) considered all RCTs testing anti-PD-1/PD-L1 given either alone or combined with chemotherapy, as first-line systemic treatment for patients with advanced NSCLC. The following inclusion criteria were used: 1) RCT testing an anti-PD-1 or anti-PD-L1 given either alone or combined with chemotherapy, against

chemotherapy alone, as first-line systemic treatment for patients with advanced NSCLC; and 2) data available on HR for OS, according to patients' sex subgroup.

Study selection and data extraction

The list of retrieved articles was reviewed to choose potentially relevant articles, and disagreements were discussed and resolved with the consensus of all investigators.

From each study, the following data were extracted: name of study, first author and year of publication, study design and blinding, study phase, number of patients, age distribution, sex distribution, patients' smoking status distribution, patients' performance status distribution, patients' histotype, type of ICI used, HRs for PFS and/or OS in the overall population, and HRs according to patients' sex. We included only the most recent and complete report of controlled trials when duplicate publications were identified.

Risk of bias assessment

The methodological quality of studies was assessed by using the six-point Jadad ranking system, which evaluates quality of randomization, double-blinding, and the flow of patients (withdrawals and dropouts).[41] A clinical trial could receive a Jadad score ranging between zero (poor methodological quality) and five (optimal methodological quality).

4.2 Methods

We assessed both the interaction between the efficacy of anti-PD-1/PD-L1 combined with chemotherapy and patients' sex (first meta-analysis) and the interaction between the efficacy of anti-PD-1/PD-L1 given alone or combined with chemotherapy and patients' sex (second meta-analysis).

As far as the first meta-analysis is concerned, the primary endpoint was the difference in efficacy of the combination of an anti-PD-1 or anti-PD-L1 with chemotherapy between men and women, measured in terms of the ratio of the HR for progression or death in the intervention arm compared with those in the control arm reported in men, to the same HR reported in women.

The HRs for progression or death in the intervention arm compared with those in the control arm, along with their 95% CI, were derived from each included study, separately for male and female patients. HRs and CIs were translated into log-hazard ratios and the corresponding variances. The pooled HR of progression or death was calculated in men and women, using a random-effects model. Each study log(HR) was weighted by the inverse of

its within-study variance plus the between-study variance component τ^2. The moment estimator of the between-study variance proposed by DerSimonian and Laird[30] was used. The Q test was performed to assess between-study heterogeneity, and the I^2 statistics were also calculated.[42] To avoid the risk of ecological bias, the null hypothesis that the difference of treatment effect between women and men is zero was tested using the following approach: first, a trial specific ratio of HRs was calculated from the ratio of the reported HRs in men and in women; second, these trial-specific ratios of HRs were combined across trials using a random-effects model.[31] A pooled HR ratio estimate lower than 1 indicates a greater treatment effect in men, and higher than 1 a greater effect in women.

In the second meta-analysis, analysis was performed using the same approach described above for the first meta-analysis. Moreover, a z-test was used to test the heterogeneity of the pooled ratios of HRs, measuring the difference in efficacy between men and women, for the two evaluated immunotherapeutic strategies (i.e., anti-PD-1/PD-L1 given alone or combined with chemotherapy).

All analyses were performed with R software (version 3.4.0).

4.3 Results

First systematic review and meta-analysis

We found eight eligible RCTs[43–50] reporting results of the combination of an anti-PD-1 or anti-PD-L1 plus chemotherapy vs chemotherapy in patients with advanced lung cancer. Table 4.1 reports details of trials included.

Risk of bias assessment through Jadad score for each trial was analyzed. Randomized treatment allocation sequences were generated in all trials; four trials were double-blinded. The Jadad mean score was 4 (range 3-5); no trial received a low-quality score (i.e., Jadad score of 0-1-2). All the included studies had a low risk of reporting bias, attrition bias, and other biases.

Table 4.1 Main features and results of studies included in the first meta-analysis

Trial	Men	Women	Histotype	Experimental arm	Control arm	PFS-HR (95% CI)			OS-HR (95% CI)		
						ITT population	Men	Women	ITT population	Men	Women
KEYNOTE-407[43]	455	104	Only squamous NSCLC	Pembrolizumab + CT	CBDCA + Txl/nTxl	0.56 (0.45 to 0.70)	0.58 (0.46 to 0.73)	0.49 (0.30 to 0.81)	0.64 (0.49 to 0.85)	0.69 (0.51 to 0.94)	0.42 (0.22 to 0.81)
KEYNOTE-189[44]	363	253	Only non-squamous NSCLC	Pembrolizumab + CT	CDDP / CBDCA + PEM	0.52 (0.43 to 0.64)	0.66 (0.50 to 0.87)	0.40 (0.29 to 0.54)	0.49 (0.38 to 0.64)	0.70 (0.50 to 0.99)	0.29 (0.19 to 0.44)
IMpower130[45]	400	279	Only non-squamous NSCLC	Atezolizumab + CT	CBDCA + nTxl	0.64 (0.54 to 0.77)	0.67 (0.54 to 0.85)	0.59 (0.45 to 0.78)	0.79 (0.64 to 0.98)	0.87 (0.66 to 1.15)	0.66 (0.46 to 0.93)
IMpower131[46]	557	126	Only squamous NSCLC	Atezolizumab + CT	CBDCA + nTxl	0.71 (0.60 to 0.85)	0.71 (0.59 to 0.85)	0.66 (0.45 to 0.97)	NA	NA	NA
IMpower132[47]	384	194	Only non-squamous NSCLC	Atezolizumab + CT	CDDP / CBDCA + PEM	0.60 (0.49 to 0.72)	0.64 (0.51 to 0.79)	0.51 (0.36 to 0.71)	NA	NA	NA
IMpower133[48]	261	142	SCLC	Atezolizumab + CT	CBDCA + VP16	0.77 (0.62 to 0.96)	0.87 (0.67 to 1.13)	0.59 (0.41 to 0.85)	0.70 (0.54 to 0.91)	0.74 (0.54 to 1.02)	0.65 (0.42 to 1.00)
PACIFIC[49]	500	213	Squamous (326) and non-squamous (387) NSCLC	Durvalumab + CT + RT	CT + RT	0.55 (0.45 to 0.68)	0.56 (0.44 to 0.71)	0.54 (0.37 to 0.79)	0.68 (0.47 to 0.99)	0.78 (0.59 to 1.03)	0.46 (0.30 to 0.73)
IMpower150[50]	425	267	Only non-squamous NSCLC	Atezolizumab + CT + Bev	CBDCA + Txl + Bev	0.62 (0.52 to 0.74)	0.55 (0.44 to 0.67)	0.73 (0.54 to 0.96)	0.78 (0.64 to 0.96)	NA	NA

Bev=Bevacizumab; CBDCA=Carboplatin; CDDP=Cisplatin; CT=Chemotherapy; ITT=Intention to treat; NA=Not available; nTxl=Nab-paclitaxel; Pem=Pemetrexed; RT=Radiotherapy; Txl=Paclitaxel; V16=Etoposide

All eight trials had available data on PFS according to patients' sex subgroups and were included in the analysis for such endpoint.

Two trials, KEYNOTE 407 and 189, tested the combination of chemotherapy with the anti-PD-1 Pembrolizumab in patients with advanced non-squamous and squamous NSCLC, respectively. Two trials, IMpower 130 and 132, tested the combination of chemotherapy plus the anti-PD-L1 Atezolizumab in advanced non-squamous NSCLC, and one trial, IMpower 131, tested the combination of chemotherapy plus Atezolizumab in squamous NSCLC. One trial, IMpower 150, tested the combination of Atezolizumab plus chemotherapy plus Bevacizumab vs chemotherapy plus Bevacizumab in advanced non-squamous NSCLC. One trial, IMpower 133, tested the combination of chemotherapy plus Atezolizumab in advanced SCLC. Finally, in the PACIFIC trial patients with locally advanced NSCLC were randomly assigned to chemoradiotherapy plus the anti-PD-L1 Durvalumab vs chemoradiotherapy alone (Table 4.1).

The analysis for PFS included 4923 patients, of whom 3345 (67.9%) were men and 1578 (32.1%) were women; 2952 (60.0%) patients had non-squamous NSCLC, 1568 (31.9%) squamous NSCLC, and 403 (8.2%) SCLC (Table 4.1); 4437 (90.1%) patients were former or current smokers, and only 486 (9.9%) were never smokers. All these trials enrolled only EGFR and ALK wild-type tumors, except for the PACIFIC trial, in which 43 EGFR mutated NSCLC were enrolled.

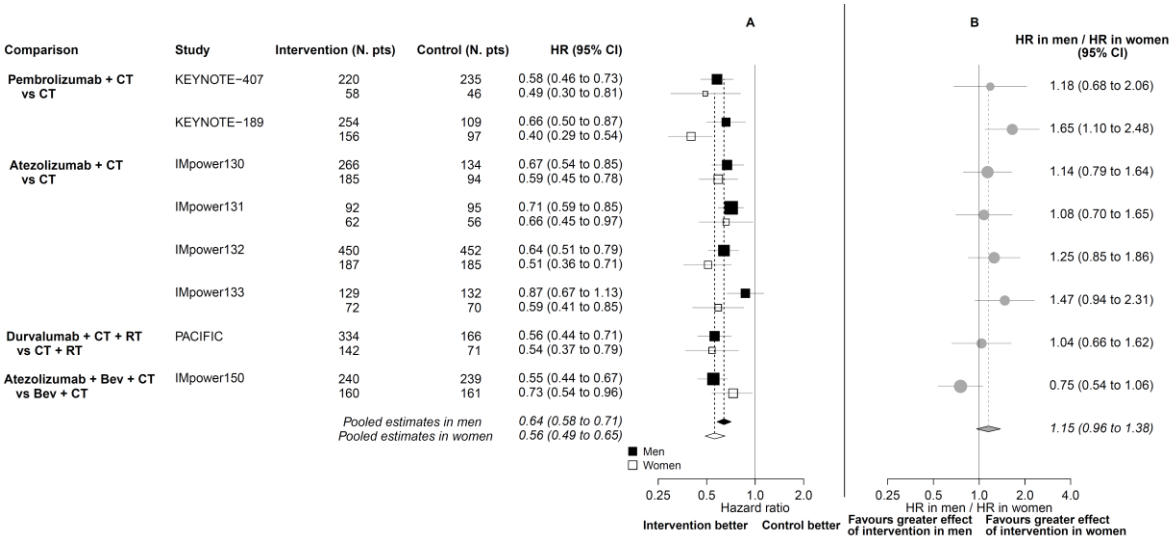

Figure 4.1 HRs of progression or death according to patients' sex

PFS results, reported in Figure 4.1, showed that men treated with anti-PD-1 or anti-PD-L1 plus chemotherapy had a reduced risk of progression or death compared with men treated in control arm (pooled PFS-HR 0.64, 95% CI 0.58 to 0.71; Figure 4.1A). In women, the benefit obtained with anti-PD-1 or anti-PD-L1 plus chemotherapy compared with the control arm was larger (pooled PFS-HR 0.56, 95% CI 0.49 to 0.65; Figure 4.1A). No heterogeneity among single-study estimates was observed in either male patients (Q=10.9, p-value=0.14, I^2=36.0%) or in female patients (Q=9.9, p-value=0.19, I^2=29.6%).

The pooled ratio of PFS hazard ratios reported in women versus those reported in men in each trial was 1.15 (95% CI 0.96 to 1.38; Figure 4.1B) indicating a larger benefit in women compared with men.

Five of the eight RCTs (i.e., KEYNOTE-407, KEYNOTE-189, IMpower130, IMpower133, and PACIFIC trial) had available data on OS according with patients' sex subgroups and were included in the analysis for such endpoint (Table 4.1).

The analysis for OS included 2970 patients, of whom 1979 (66.6%) were men and 991 (33.4%) were women; 1682 (56.6%) patients had non-squamous NSCLC, 885 (29.8%) squamous NSCLC, and 403 (13.6%) SCLC (Table 4.1); 2715 (91.4%) patients were former or current smokers, and 255 (8.6%) were never smokers.

Men treated with anti-PD-1 or anti-PD-L1 plus chemotherapy had a reduced risk of death compared with men treated in the control arm (pooled OS-HR 0.76, 95% CI 0.66 to 0.87; Figure 4.2A). In women, the OS benefit obtained with anti-PD-1 or anti-PD-L1 plus chemotherapy compared with the control arm was larger (pooled OS-HR 0.48, 95% CI 0.35

to 0.67; Figure 4.2A). Heterogeneity among single-study estimates was observed in female patients (Q=10.89, p-value=0.03, I^2=63.3%), but not in male patients (Q=1.55, p-value=0.81, I^2=0.0%). The pooled ratio of OS-HRs reported in women versus those reported in men in each trial was 1.56 (95% CI 1.21 to 2.01; Figure 4.2B), indicating a larger benefit in women compared with men.

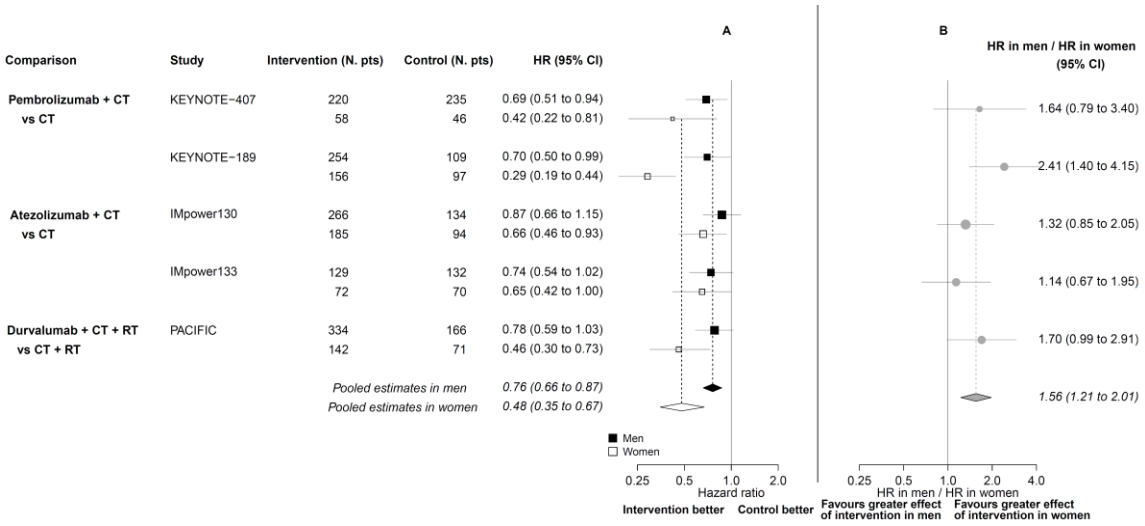

Figure 4.2 HRs of death according to patients' sex

We conducted two sensitivity analyses using as endpoint the OS. In the first scenario, we assessed sex-based heterogeneity of efficacy of the combination of chemotherapy plus anti-PD-1/PD-L1 excluding the PACIFIC trial because patients included in this trial also received radiotherapy. In the second sensitivity analysis to assess sex-based heterogeneity of efficacy in trials enrolling only patients with NSCLC we excluded the IMpower133 trial that enrolled patients with SCLC. The pooled ratio of OS-HRs reported in women versus those reported in men was respectively 1.53 (95% CI 1.10 to 2.13) for the first sensitivity analysis and 1.68 (95% CI 1.28 to 2.19) for the second, confirming a larger benefit in women compared with men (data not shown).

Second systematic review and meta-analysis

We found six eligible RCTs[43–45,51–53] reporting results on OS according with patients' sex for anti-PD-1/PD-L1 given either alone (KEYNOTE-024, KEYNOTE-042 and CheckMate 026 trials) or combined with chemotherapy (KEYNOTE-407, KEYNOTE-189 and IMpower130 trials) as first-line systemic treatment for patients with advanced NSCLC (Table 4.2).

Table 4.2 Main features and results of studies included in the second meta-analysis

Trial	Men	Women	NSCLC histotype	Experimental arm	Control arm	OS-HR (95% CI)		
						ITT population	Men	Women
KEYNOTE-407[43]	455	104	Only squamous	Pembrolizumab + CT	CBDCA + Txl/nTxl	0.64 (0.49 to 0.85)	0.69 (0.51 to 0.94)	0.42 (0.22 to 0.81)
KEYNOTE-189[44]	363	253	Only non-squamous	Pembrolizumab + CT	CDDP/CBDCA + PEM	0.49 (0.38 to 0.64)	0.70 (0.50 to 0.99)	0.29 (0.19 to 0.44)
IMpower130[45]	400	279	Only non-squamous	Atezolizumab + CT	CBDCA + nTxl	0.79 (0.64 to 0.98)	0.87 (0.66 to 1.15)	0.66 (0.46 to 0.93)
KEYNOTE-024[51]	187	118	Squamous (56) and non-squamous (249)	Pembrolizumab	CBDCA + TXL/GEM/PEM or CDDP + GEM/PEM	0.60 (0.41 to 0.89)	0.54 (0.36 to 0.80)	0.96 (0.56 to 1.64)
KEYNOTE-042[52]	902	372	Squamous (492) and non-squamous (782)	Pembrolizumab	CBDCA +TXL or CBDCA+PEM	0.81 (0.71 to 0.93)	0.80 (0.68 to 0.94)	0.89 (0.68 to 1.17)
CheckMate 026[53]	332	209	Squamous (130) and non-squamous (411)	Nivolumab	CBDCA + TXL/GEM/PEM or CDDP + GEM/PEM	1.07 (0.86 to 1.33)	0.97 (0.74 to 1.26)	1.15 (0.79 to 1.66)

CBDCA=Carboplatin; CDDP=Cisplatin; CT=Chemotherapy; GEM=Gemcitabine; ITT=Intention to treat; nTxl=Nab-paclitaxel; PEM=Pemetrexed; Txl=Paclitaxel

These six trials had similar designs and enrolled comparable patient populations, in the same setting of disease. All trials enrolled only EGFR and ALK wild-type tumors and had a standard first-line, platinum-based chemotherapy as a control arm.

Analysis included 3974 patients, of whom 2639 (66.4%) were men and 1335 (33.6%) were women; 2737 (68.9%) patients had non-squamous tumors, and 1237 (31.1%) squamous tumors; 2120 patients were treated with anti-PD-1 alone (1579 with Pembrolizumab and 541 with Nivolumab), and 1854 with anti-PD-1/PD-L1 plus chemotherapy (1175 with Pembrolizumab plus chemotherapy and 679 with Atezolizumab plus chemotherapy) (Table 4.2); 3423 (86.1%) patients were former or current smokers, and 544 (13.7%) were never smokers, and for 7 patients, data were unknown.

Male patients treated with anti-PD-1 alone had a reduced risk of death as compared with men treated with standard chemotherapy (pooled OS-HR 0.78, 95% CI 0.60 to 1.01; Figure 4.3A). In women, anti-PD-1 alone was not better than standard chemotherapy (pooled OS-HR 0.97, 95% CI 0.79 to 1.19). By contrast, anti-PD-1/PD-L1 administered in combination

with chemotherapy was associated with a very large OS advantage compared with chemotherapy alone in women, but a smaller benefit was seen in men (female-pooled OS-HR 0.44, 95% CI 0.25 to 0.76; male-pooled OS-HR 0.76, 95% CI 0.64 to 0.91).

The pooled ratio of OS-HRs reported in women versus those reported in men in each trial was 0.83 (95% CI 0.65 to 1.06) for anti-PD-1 alone, indicating a greater effect of anti-PD-1 alone in men (Figure 4.3B), and 1.70 (95% CI 1.16 to 2.49) for anti-PD-1/PD-L1 plus chemotherapy, indicating a greater effect for such a therapeutic strategy in women. Thus, the difference in efficacy for men versus women was statistically significantly different for the two immunotherapeutic strategies (i.e., anti-PD-1/PD-L1 alone or combined with chemotherapy; p-value=0.002).

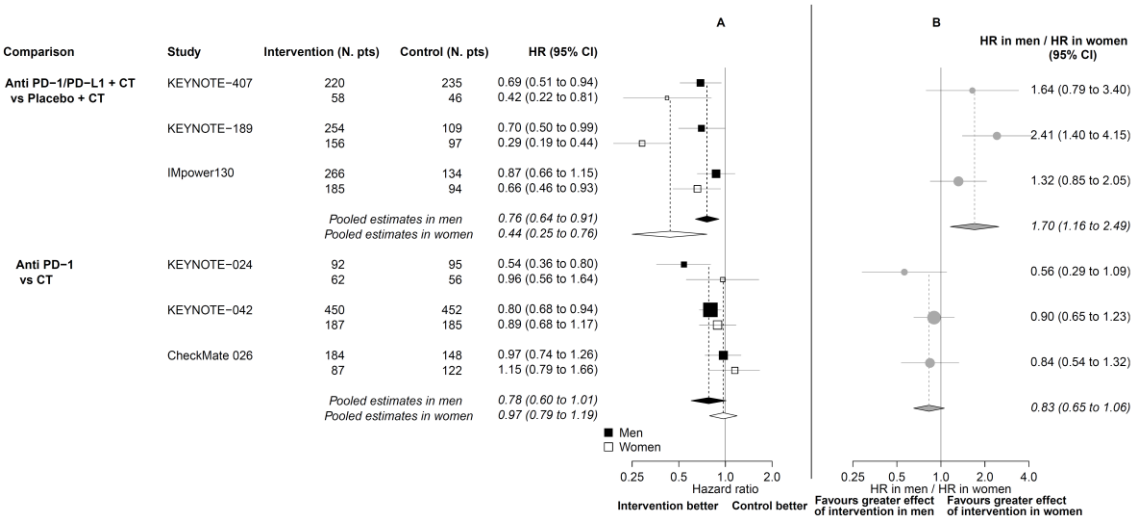

Figure 4.3 HRs of death according to patients' sex and type of immunotherapeutic strategy

4.4 Discussion

In a previous work, Conforti et al.[23] showed that male patients with several solid tumors, treated with anti-CTLA-4 or anti-PD-1 drugs given in monotherapy or as a combination therapy, obtain larger benefit than women. In the discussion of that previous work it was anticipated that, given the complexity of the sex-dimorphism of the immune system function and responses, it would be possible that women derive larger benefit than men from different immunotherapeutic strategies.[23]

Here, we showed that women with advanced lung cancer experience a larger benefit from the addition of chemotherapy to an anti-PD-1 or anti-PD-L1 than men.

It could be hypothesized that such heterogeneity of response is due to the ability of chemotherapy to increase the mutational burden and neoantigenic load of female lung

cancer tumors and this could be also a potential biological rationale to explain the lower efficacy of anti-PD-1 alone in women.[25,54,55]

Furthermore, different efficacy of chemotherapy in modulating the anticancer immune responses of men and women could be speculated, given the sex-related differences in the amount and composition of intratumoral immune infiltrates reported.[20,21,56]

We also provided data suggesting that the interaction between patients' sex and the efficacy of different immunotherapeutic strategies could be important when choosing therapeutic options for female and male patients with NSCLC. We meta-analyzed six RCTs testing two different immunotherapeutic approaches (i.e., anti-PD-1/PD-L1 alone or in combination with chemotherapy) in comparable populations, and we found an interaction between patients' sex and the efficacy of both these therapeutic strategies, with opposite direction/interaction in men and women.

Historically, results from immunotherapy trials for advanced breast cancer have been disappointing.[57,58] However, these prior studies have primarily tested immunotherapy alone. Recently, the results of the IMpassion130 study showed that the combination of Atezolizumab plus chemotherapy improved results compared with chemotherapy alone for women with advanced triple-negative breast cancer, especially if their tumors expressed PD-L1.[59] This provides additional support for the need to use combination of immunotherapy plus chemotherapy to improve outcomes for female patients.[58]

The results reported here highlight a relevant methodological issue. Given the complexity of the sex-based dimorphism of the immune system, involving multiple elements of innate and adaptive immune responses, the proper way to assess the interaction between patients' sex and the efficacy of different anticancer immunotherapies is to assess such interaction separately for each type of immunotherapeutic strategy.[16,17] Indeed, the direction of the interaction could be the opposite for different types of treatment, as we showed for anti-PD-1/PD-L1 given alone or combined with chemotherapy. Analyzing together different immunotherapeutic approaches, with interaction between their efficacy and patients' sex, but with the opposite direction of such interaction, would lead to diluting results and to misleading conclusions, as it happened in the meta-analysis of Wallis et al.[24]

A limitation of our meta-analysis is that it relies on published results rather than on individual patients' data. This precludes the possibility of exploring relevant issues, such as the menopausal status of female patients on the efficacy of immunotherapeutic treatments, which deserves to be investigated given the key role exerted by sex hormones in the regulation of the immune system and for the potential therapeutic implications.[16,17]

Our findings are hypothesis generating because they are based on meta-analyses of aggregated published results derived from RCTs and, as such, require further validation in future prospective trials before being considered sufficient to support any change in clinical practice.

Chapter 5

Sex-based dimorphism of anticancer immune response

This chapter is related to study III:

Conforti F, Pala L, Pagan E, Bagnardi V, De Pas T, Tortora G, Bria E, Minucci S, Joffe H, Gelber RD, Viale G, Giaccone G, Goldhirsch A. **Molecular characterization of sex-based dimorphism of anticancer immune response and mechanisms of immune evasion, in patients with early non-small cell lung cancer.**

Work in progress.

5.1 Materials

LCE-project

The Lung Cancer Explorer (LCE) project provides the largest publicly available collection of expression and clinical data from 56 studies for over 6700 patients with lung cancer (http://lce.biohpc.swmed.edu/). Data were reprocessed to be suitable for computational meta-analysis.[60]

Description of the procedures adopted for reprocessing and normalizing expression data, quality control assessment and standardization of the datasets has been previously reported.[60]

We focused on the largest datasets on adenocarcinoma and squamous cell carcinoma available in the LCE-project. We included in the analysis five datasets with data on more than 250 tumor samples, and at least 25 samples from female patients, with adenocarcinoma (The Cancer Genome Atlas (TCGA) LUAD[61], Rousseaux et al.[62], Sato et al.[63], Schabath et al.[64], Shedden et al.[65]) and two datasets with more than 100 tumor samples, and at least 10 samples from female patients, with squamous NSCLC (TCGA LUSC[66], Noro et al.[67]).

For each dataset, patients' epidemiological and clinical data were available, including tumor histology, patient demographics, stage at diagnosis and smoking status.

TRACERx cohort

TRACERx (ClinicalTrial.gov identifier NCT01888601) is a prospective study of patients with primary NSCLC which aims to define the evolutionary trajectories of lung cancer in

both space and time through multiregion and longitudinal tumour sampling and genetic analysis from diagnosis to relapse. More details on the patient cohort enrolled in the TRACERx lung study have been previously reported.[68–70]

We analyzed Whole Exome Sequencing (WES) data of 327 tumor regions from 100 patients with NSCLC and RNAseq data of a subset of 164 tumor regions from 64 tumors.

MSKCC, POPLAR and OAK databases

Samstein et al.[13] examined a cohort of 1661 patients with a variety of cancer types treated with at least one dose of ICI therapy at Memorial Sloan Kettering Cancer Center (MSKCC). All tumors were profiled by next generation sequencing. Individual patient data on TMB and patients' outcome are publicly available at https://www.cbioportal.org/study/summary?id=tmb_mskcc_2018. We considered only data of the 350 patients with NSCLC.

The POPLAR phase II study (ClinicalTrial.gov identifier NCT01903993) and the OAK phase III study (ClinicalTrial.gov identifier NCT02008227) enrolled 287 and 850 patients, respectively, and evaluated the efficacy and safety of second/third-line Atezolizumab compared with Docetaxel in patients with advanced or metastatic NSCLC after failure of platinum-based chemotherapy. Individual patient data on blood-based TMB and patients' outcome were publicly available as supplementary materials of Gandara et al. (https://www.nature.com/articles/s41591-018-0134-3#Sec19).[71]

5.2 Methods

Assessment of sex-based differences in the tumor immune infiltrate and in the expression levels of immune checkpoint molecules in the tumor microenvironment

Gene-expression data were analyzed through the previously validated xCell algorithm, to estimate the abundance of different cell types in the microenvironment of each tumor sample included in the datasets of the LCE-project as well as of the TRACERx lung study. The entire pipeline of xCell has been previously described.[36,72]

Briefly, xCell is a computational algorithm that integrates single sample gene set enrichment analysis (ssGSEA) with deconvolution approach, and permitting estimation of the cell-type composition of the TME, evaluating the expression levels of a compendium of 489 gene signatures that identify 64 different cell types, including multiple adaptive and innate immunity cells, hematopoietic progenitors, epithelial cells, and extracellular matrix cells.[72]

For each single dataset, mean values of ES for the 64 different cell-types were calculated in tumors of men and women and then compared using a multivariable linear regression model adjusted for patient age, stage at diagnosis, tumor-histotype, smoking status, and EGFR and ALK mutational status.

We then performed a meta-analysis of the adjusted sex-related differences obtained in each single dataset using a random-effects model. The FDR was used to correct for multiple comparisons. Pooled estimate higher than 0 indicated a greater ES in females, and lower than 0 a greater ES in males.

Moreover, a curated list of 78 genes with a key role in anticancer immune response was derived from Thorson et al.[20], and the expression levels of each gene were compared to assess differences between tumors of male and female patients following the methodology described above.

Assessment of sex-based differences in mechanisms of immune-evasion

Gene-expression data from the LCE-project datasets were analyzed through the validated Tumor Immune Dysfunction and Exclusion (TIDE) tool, that permits quantification of the activation status of two major mechanisms of immune-evasion exploited by tumors: the induction of T-cell dysfunction (T-cell dysfunction mechanism) and the inhibition of T-cell infiltration into TME (T-cell exclusion mechanism).[73]

For each tumor sample we calculated two scores, the 'T-cell dysfunction score' and the 'T-cell exclusion score': both scores range from -4 to +4, with higher score levels being associated with greater activation status of the corresponding mechanism of immune-evasion.

For each single dataset, mean values of the 'T-cell dysfunction score' and 'T-cell exclusion score' were calculated in tumors of men and women and then compared using a multivariable linear regression model adjusted for patient age, stage at diagnosis, tumor-histotype, smoking status, and EGFR and ALK mutational status.

We then performed a meta-analysis of the adjusted sex-related differences obtained in each single dataset using a random-effects model. The Q test was performed to assess between-study heterogeneity, and the I^2 statistic, which expresses the percentage of the total observed variability due to heterogeneity, was also calculated.

A pooled-estimate score higher than 0 indicated a greater activation status in females of the corresponding mechanism of immune-evasion, and lower than 0 a greater activation in males.

Assessment of sex-based differences in TCR repertoire diversity, tumor neoantigens load and alterations in neoantigens presentation machinery

Multiregion Bulk RNAseq and WES data from TRACERx lung study were employed to assess for each tumor and/or patient the following elements:

- TCR abundance and Shannon diversity score;[74,75]
- amount of ubiquitous expanded TCRs;[69,70]
- number of predicted tumor neoantigens and their clonal and subclonal distribution;[68,76,77]
- occurrence of loss of heterozygosity (LOH) events at the human leukocyte antigen (HLA) class I locus[69], as well as genetic disruptive events (i.e., non-silent mutations or copy-number loss defined relative to ploidy) in other antigen presentation pathway genes, including: CIITA, IRF1, PSME1, PSME2, PSME3, ERAP1, ERAP2, HSPA (also known as PSMA7), HSPC (also known as HSPBP1), TAP1, TAP2, TAPBP, CALR, CNX (alias CANX), PDIA3 and B2M.

More details on analyses performed are reported in the Appendix.

Assessment of sex-based differences in biological processes, cellular component and molecular functions

Gene-expression data from the LCE-project datasets were analyzed through the GSEA method proposed by Subramanian et al.[36] using the C5 collection of the MSigDB. The C5 collection includes hundreds of gene sets curated and validated by the GO project, and captures all the relevant biological process, cellular component, and molecular function of human cells and tissues.

GSEA was originally proposed to evaluate microarray data at the level of gene set to determine whether the gene set(s) analyzed is(are) correlated with the phenotypic class of interest.[36] We modified the original methodology calculating first sex-related differences for each individual gene in each dataset and then applying a random-effects model to estimate the overall strength of the association between each gene and sex. The GSEA method using the C5 collection of the MSigDB was then applied on the meta-analytic T-test statistics. When we encountered a gene present in only one dataset, in the GSEA method we used the simple T-test statistic for that gene. Gene labels were permuted to estimate the FDR-corrected statistical significance of the NES. ES (and NES) higher than 0 indicated a greater enrichment of the gene set in females, and lower than 0 a greater enrichment in males. The dataset from Noro et al.[67] on squamous NSCLC was not included in the present analysis.

Both meta-analysis and GSEA were implemented and performed in SAS software v. 9.4 (SAS Institute, Cary, NC).

Assessment of sex-based differences in the association between tumor mutational burden and patients' outcome

We analyzed data from MSKCC dataset on patients treated with anti-PD-1 or anti-PD-L1, to assess the association between tissue-TMB (tTMB) and OS according with patients' sex. Data of patients treated with the anti-PD-L1 Atezolizumab in the OAK and POPLAR RCTs were evaluated to assess the association between blood-TMB (bTMB) and PFS according with patients' sex. We did not explore sex-based differences in the association between bTMB and OS, since in the original analysis performed on the whole population a significant predictive value of bTMB was reported for PFS but not for OS.[78]

For all the analyses reported, the tissue and blood TMB were analyzed as continuous variables. Wilcoxon rank-sum test was used to compare the distribution of tissue and blood TMB between female and male patients. Cox proportional hazard regression model was used to evaluate the association between tTMB and bTMB and patient OS and PFS, respectively. Male and female subgroups were analyzed separately.

Departure from linearity in the relationship between tissue or blood TMB and the hazard of death was investigated with RCS models with four knots located at the 20^{th}, 40^{th}, 60^{th} and 80^{th} percentiles of the TMB distribution of female and male patients, respectively. The LRT was used to determine whether the RCS model significantly increased the likelihood function compared with a simpler model that assumed a linear relationship.

Multivariable analyses were performed excluding patients with tumors harboring EGFR gene mutation or ALK gene translocation and considering only those patients with available data on the following adjustment factors: age, smoking history, tumor histotype and type of specimen analyzed, number of metastatic sites at enrollment, sum of longest diameter of target lesions at baseline, and PD-L1 expression levels.

All statistical analyses were performed with SAS software v. 9.4 (SAS Institute, Cary, NC) and R software (version 3.4.1).

5.3 Results

Totally, 2575 tumor samples from the seven LCE-project databases selected were analyzed: 1528 tumors (59.3%) were from men and 1047 (40.7%) from women (Table 5.1).

Seven hundred thirty-two tumors (28.4%) were from patients younger than 60 years, 934 (36.3%) from patients aged between 61 and 70 years and 853 (33.1%) from patients older than 70 years; 1880 patients (73.0%) were current or former smokers, 206 (8.0%) non-smokers, and for 489 patients (19.0%) the smoking history was unknown. The majority of patients (i.e., 95.9%) had stage I-III tumors. Five datasets reported data of 1967 (76.4%) adenocarcinoma and 2 datasets of 608 (23.6%) squamous-NSCLCs (Table 5.1).

Tumor samples from 100 patients from the TRACERx lung study were also studied.[68–70] The cohort consisted of 68 men and 32 women, with a median age of 68 years. Eighty-eight patients were current or former smokers and only 12 patients were non-smokers. Sixty-one tumors were adenocarcinomas, 32 squamous-cell carcinomas, and 7 were classified as other histology. The cohort was predominantly early-stage: 62 stage I patients, 24 stage II, 13 stage IIIa and 1 stage IIIb. Totally, 327 tumor regions (323 primary tumor regions and 4 lymph-node metastases) were analyzed.

Table 5.1 Baseline demographic and clinical characteristics of patients in each LCE-project dataset included

	TCGA LUAD[61] N=515	TCGA LUSC[66] N=501	ROUSSEAUX[62] N=293	SATO[63] N=275	SCHABATH[64] N=442	SHEDDEN[65] N=442	NORO[67] N=107	All datasets N=2575
Gender, N (%)								
Male	238 (46.2)	371 (74.1)	250 (85.3)	148 (53.8)	202 (45.7)	223 (50.5)	96 (89.7)	1528 (59.3)
Female	277 (53.8)	130 (25.9)	43 (14.7)	127 (46.2)	240 (54.3)	219 (49.5)	11 (10.3)	1047 (40.7)
Age (years), N (%)								
≤60	157 (30.5)	108 (21.6)	130 (44.4)	91 (33.1)	70 (15.8)	147 (33.3)	29 (27.1)	732 (28.4)
61-70	174 (33.8)	195 (38.9)	96 (32.8)	91 (33.1)	152 (34.4)	164 (37.1)	62 (57.9)	934 (36.3)
>70	165 (32.0)	189 (37.7)	66 (22.5)	87 (31.6)	199 (45.0)	131 (29.6)	16 (15.0)	853 (33.1)
Unknown	19 (3.7)	9 (1.8)	1 (0.3)	6 (2.2)	21 (4.8)	0	0	56 (2.2)
Smoking status, N (%)								
Smoker	0	0	0	244 (88.7)	335 (75.8)	0	0	579 (22.5)
Current Smoker	119 (23.1)	133 (26.5)	0	0	0	32 (7.2)	54 (50.5)	338 (13.1)
Former Smoker	307 (59.6)	338 (67.5)	0	0	0	268 (60.6)	50 (46.7)	963 (37.4)
Non-Smoker	75 (14.6)	18 (3.6)	0	28 (10.2)	33 (7.5)	49 (11.1)	3 (2.8)	206 (8.0)
Unknown	14 (2.7)	12 (2.4)	293 (100.0)	3 (1.1)	74 (16.7)	93 (21.0)	0	489 (19.0)
Pack-years, N (%)								
≤20	162 (31.5)	64 (12.8)	0	28 (10.2)	33 (7.5)	0	5 (4.7)	292 (11.3)
21-50	177 (34.4)	209 (41.7)	0	0	0	0	73 (68.2)	459 (17.8)
>50	86 (16.7)	169 (33.7)	0	0	0	0	29 (27.1)	284 (11.0)
Unknown	90 (17.5)	59 (11.8)	293 (100.0)	247 (89.8)	409 (92.5)	442 (100.0)	0	1540 (59.8)
Race, N (%)								
White	388 (75.3)	349 (69.7)	0	244 (88.7)	399 (90.3)	294 (66.5)	0	1674 (65.0)
Black	52 (10.1)	30 (6.0)	0	16 (5.8)	13 (2.9)	12 (2.7)	0	123 (4.8)
Asian	8 (1.6)	9 (1.8)	0	7 (2.5)	3 (0.7)	6 (1.4)	0	33 (1.3)
Other	1 (0.2)	0	0	8 (2.9)	2 (0.5)	1 (0.2)	0	12 (0.5)
Unknown	66 (12.8)	113 (22.6)	293 (100.0)	0	25 (5.7)	129 (29.2)	107 (100.0)	733 (28.5)
Stage, N (%)								
I/II	402 (78.1)	409 (81.6)	226 (77.1)	182 (66.2)	334 (75.6)	371 (83.9)	107 (100.0)	2031 (78.9)
III	84 (16.3)	84 (16.8)	53 (18.1)	86 (31.3)	63 (14.3)	68 (15.4)	0	438 (17.0)
IV	27 (5.2)	7 (1.4)	5 (1.7)	6 (2.2)	17 (3.8)	0	0	62 (2.4)
Unknown	2 (0.4)	1 (0.2)	9 (3.1)	1 (0.4)	28 (6.3)	3 (0.7)	0	44 (1.7)
EGFR mutation, N (%)								
No	195 (37.9)	275 (54.9)	0	0	395 (89.4)	0	0	865 (33.6)
Yes	80 (15.5)	21 (4.2)	0	0	47 (10.6)	0	0	148 (5.7)
Unknown	240 (46.6)	205 (40.9)	293 (100.0)	275 (100.0)	0	442 (100.0)	107 (100.0)	1562 (60.7)
KRAS mutation, N (%)								
No	39 (7.6)	14 (2.8)	0	0	288 (65.2)	0	0	341 (13.2)
Yes	23 (4.5)	1 (0.2)	0	0	154 (34.8)	0	0	178 (6.9)
Unknown	453 (88.0)	486 (97.0)	293 (100.0)	275 (100.0)	0	442 (100.0)	107 (100.0)	2056 (79.8)
ALK mutation, N (%)								
No	211 (41.0)	275 (54.9)	0	0	0	0	0	486 (18.9)
Yes	34 (6.6)	9 (1.8)	0	0	0	0	0	43 (1.7)
Unknown	270 (52.4)	217 (43.3)	293 (100.0)	275 (100.0)	442 (100.0)	442 (100.0)	107 (100.0)	2046 (79.5)
Subtype, N (%)								
Adenocarcinoma	515 (100.0)	0	293 (100.0)	275 (100.0)	442 (100.0)	442 (100.0)	0	1967 (76.4)
Squamous-cell	0	501 (100.0)	0	0	0	0	107 (100.0)	608 (23.6)

Sex-based differences in the tumor immune infiltrate, in the expression levels of immune checkpoint molecules in the tumor microenvironment, TCR repertoire diversity, tumor neoantigens load and alterations in neoantigens presentation machinery

We first assessed differences in the cell-type composition of the immune infiltrate between tumors of male and female patients.

Figure 5.1 shows sex-based differences in the abundance of immune cells found in each of the seven LCE-project datasets as well as the pooled meta-analytic results and the TRACERx results.

In the pooled analysis, the innate and adaptive immune cells found enriched in the TME of women (red dots) as compared with men, at FDR cut-off <0.05 were (Figure 5.1):

- Dendritic cells (including plasmocytoid dendritic cells, conventional dendritic cells, and activated dendritic cells);
- CD4+ T-cells (including CD4+ naïve T-cells and CD4+ central memory T-cells);
- B-cells (including Memory B-cells and Class-switched memory B-cells);
- Mast-cells.

Innate and adaptive immune cells found enriched in the TME of female patients (red dots), at FDR cut-off <0.25 were (Figure 5.1):

- Eosinophils;
- Regulatory T-cells;
- Natural killer T-cells;
- Macrophages M1-type;
- CD8+ T-cells.

The TME of female patients was also enriched at FDR <0.05 in cancer-associated fibroblast (CAF), HSC and GMP, that are respectively mesenchymal and myeloid-derived cells known to exert immunosuppressive activity in the TME and to play a role in cancer immune evasion (Figure 5.1).[9,79]

The only immune cell type found enriched in the TME of men (blue dots) at FDR <0.05, was type 2 T-helper cell.

Analysis of tumors from the TRACERx lung study cohort, confirmed the enrichment of aDC, HSC and CAF at FDR <0.05 in the TME of women in univariable analysis (Figure 5.1). Furthermore, a trend for enrichment in the TME of women was also observed for CD4+ naïve T-cells (p-value=0.24; Figure 5.1).

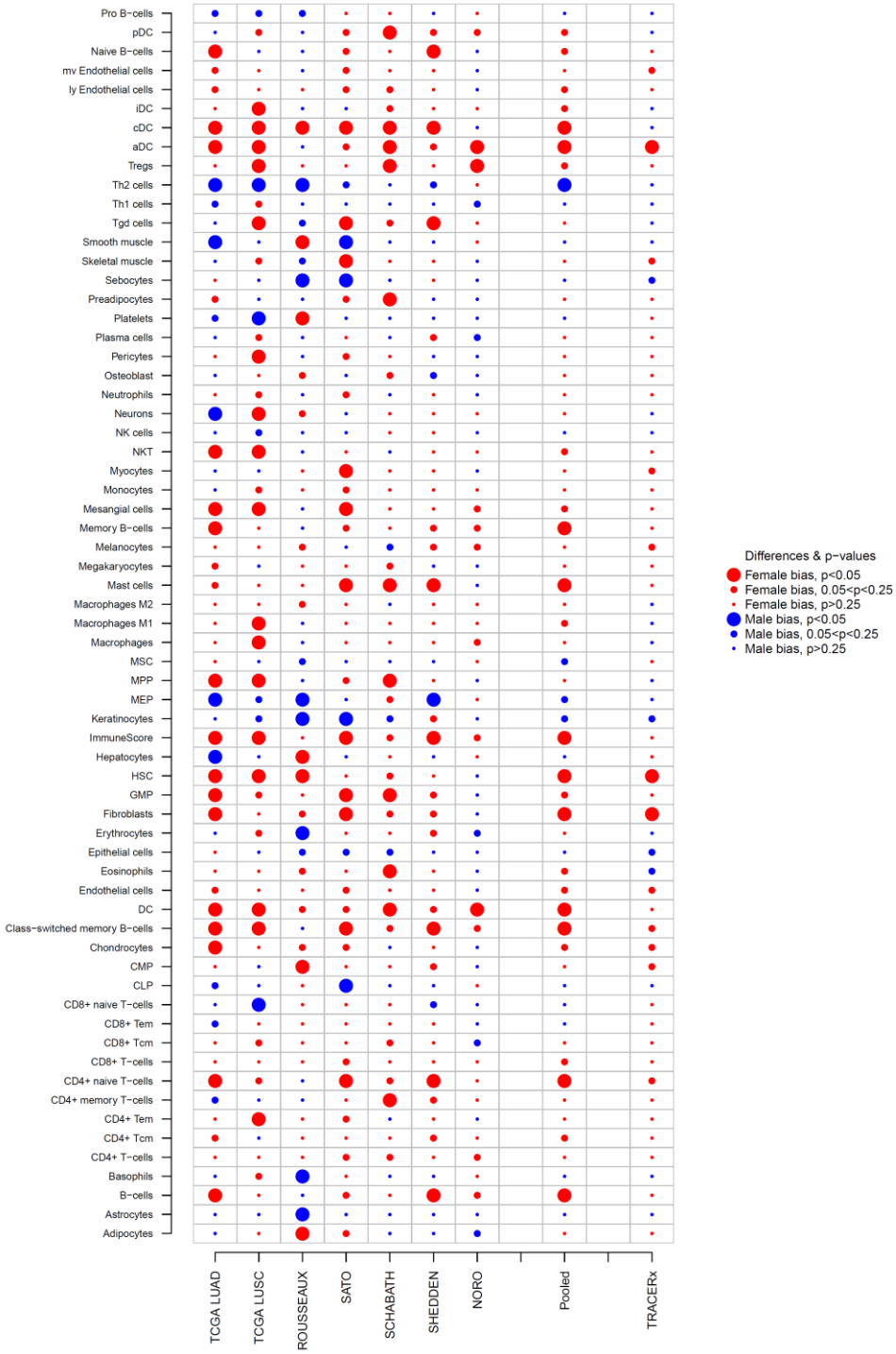

Figure 5.1 Summary plot of sex-based differences in the composition of the immune infiltrate

Abbreviations:
aDC: Activated dendritic cells; **CD4+ Tcm**: CD4+ central memory T-cells; **CD4+ Tem**: CD4+ effector memory T-cells; **CD8+ Tcm**: CD8+ central memory T-cells; **CD8+ Tem**: CD8+ effector memory T-cells; **cDC**: Conventional dendritic cells; **CLP**: Common lymphoid progenitors; **CMP**: Common myeloid progenitors; **DC**: Dendritic cells; **GMP**: Granulocyte-macrophage progenitors; **HSC**: Hematopoietic stem cells; **iDC**: Immature dendritic cells; **ly Endothelial cells**: Lymphatic endothelial cells; **MEP**: Megakaryocyte-erythroid progenitors; **MPP**: Multipotent progenitors; **MSC**: Mesenchymal stem cells; **mv Endothelial cells**: Microvascular endothelial cells; **NKT**: Natural killer T-cells; **pDC**: Plasmacytoid dendritic cells; **Tgd cells**: Gamma delta T-cells; **Th1 cells**: Type 1 T-helper cells; **Th2 cells**: Type 2 T-helper cells; **Tregs**: Regulatory T-cells.

In line with these data, showing a more abundant immune infiltrate in tumors of women as compared with men, we found that TME of women was characterized by higher expression levels of a number of chemokines, receptors and integrins involved in leukocytes extravasation and tumor infiltration. In the pooled analysis of the seven LCE-project datasets, molecules found highly expressed in the TME of women at FDR <0.05 were: CCL5, CX3CL1, CXCL9, BTN3A2, ICAM1 and LFA1 (Figure 5.2).

To further confirm a greater immune infiltration of TME of women as compared with men, we characterized the TCR repertoire of tumor infiltrating lymphocytes (TILs) of tumor samples from the TRACERx lung study cohort.

As expected, we found a greater TCR repertoire diversity in TILs of women (median TCR entropy score 0.84 in females versus 0.67 in males, p-value=0.03; Figure 5.3A), that confirmed a higher TME infiltration by T-cells in women.

Previous works demonstrated that the amount of expanded TCRs present in all tumor regions assessed in multiregional tumor analysis (the so called expanded ubiquitous TCRs), is a proxy for immune response against tumor antigens.[61,69,70]

Coherently with all other findings reported above, we found a numerically higher amount of expanded ubiquitous TCRs in tumors of women as compared with men (the median number of ubiquitous expanded TCRs was 33 in females and 25 in males, p-value=0.24; Figure 5.3B).

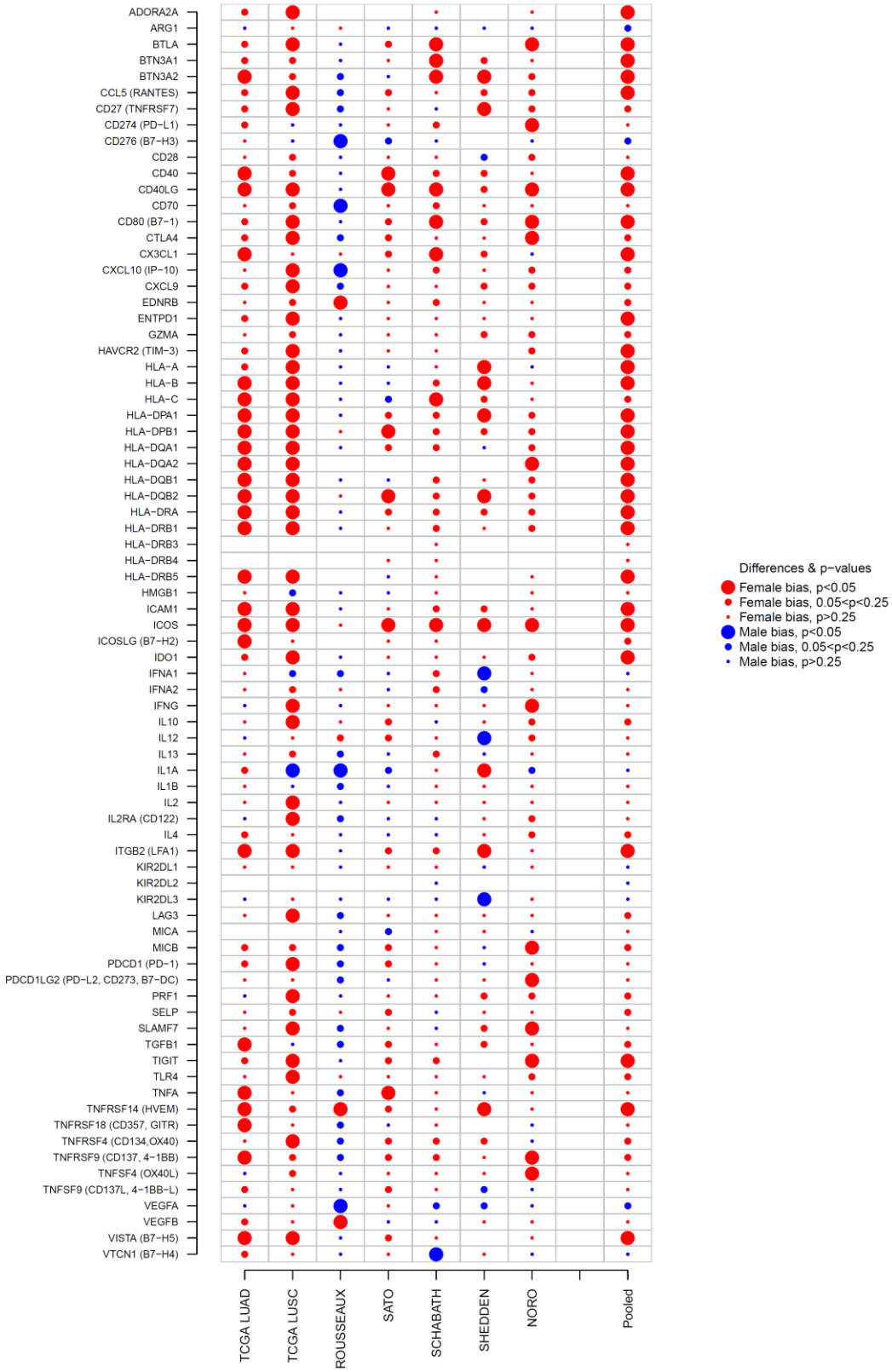

Figure 5.2 Summary plot of sex-based differences in expression levels of immune-related genes

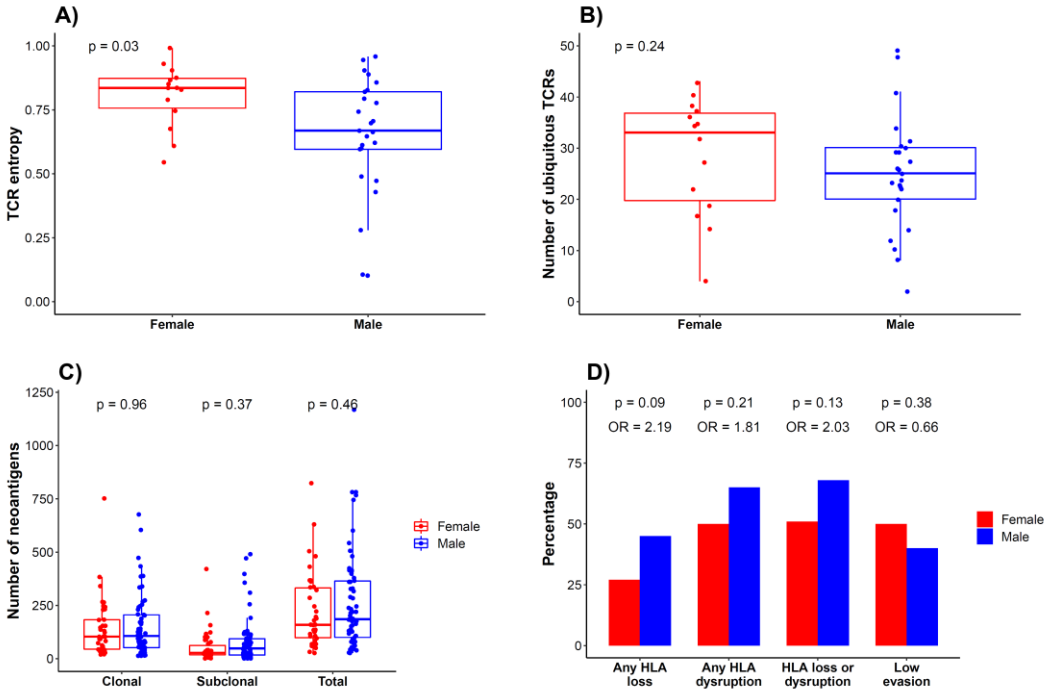

Figure 5.3 Sex-based differences in TCR repertoire diversity, tumor antigenicity, and alterations in antigen presentation machinery

Sex-based differences in mechanisms of immune-evasion

To assess sex-based differences in mechanisms exploited by tumors to evade immune response, we analyzed the seven datasets of LCE-project through the TIDE tool.

In all the 7 datasets, the mean value of the 'T-cell dysfunction' score was higher in tumors of women as compared with men. The pooled adjusted difference estimate was 0.09 (95% CI 0.05 to 0.13, p-value<0.001), confirming greater T-cell dysfunction status in tumors of women (Figure 5.4).

On the contrary, the mean value of the 'T-cell exclusion' score was always higher in tumors of men. The pooled adjusted difference estimate was -0.08 (95% CI -0.12 to -0.03; p-value=0.001), confirming greater activation status of such mechanism of immune-evasion in tumors of men (Figure 5.5).

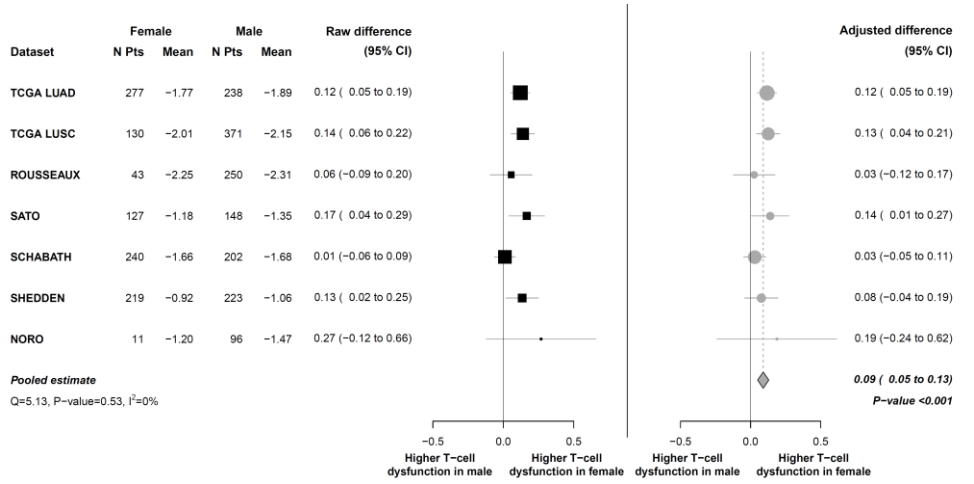

Figure 5.4 Meta-analysis of differences in the 'T-cell dysfunction score' assessed in tumors of female and male patients

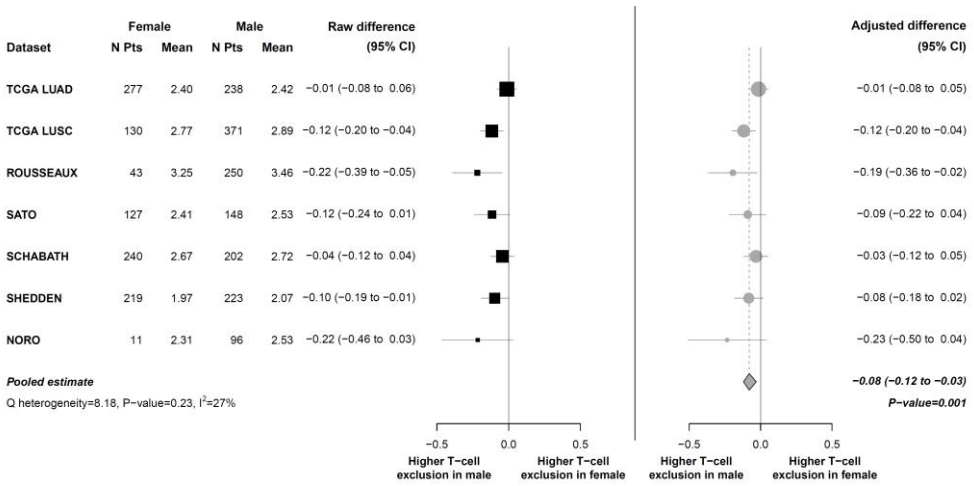

Figure 5.5 Meta-analysis of differences in the 'T-cell exclusion score' assessed in tumors of female and male patients

Coherently with these findings, the pooled analysis of the seven LCE-project datasets presented in Figure 5.2 have revealed a higher expression levels of a number of inhibitory immune-checkpoints in TME of females (including TIM3, TIGIT, BTLA, IDO1, ADORA2A, ENTPD1, BTN3A1, TNFRSF14, and VISTA at FDR <0.05 and LAG3 at FDR=0.09), which are known to play a key role in T-cells exhaustion and dysfunction mechanisms.

In summary, the lower abundance of a number of immune cell types in the TME, the significantly higher T-cell exclusion score, the smaller TCR repertoire diversity and the lower amount of ubiquitous expanded TCRs observed in tumors of men are all elements suggesting a less efficient tumor recognition and infiltration by immune-system.

Since several mechanisms have been shown to impair infiltration of the TME by immune system, the TRACERx cohort tumors were analyzed to estimate sex-based differences in the predicted tumor neoantigens load and/or in the neoantigens clonal distribution, and no sex-based difference in the total number of predicted tumor neoantigens, nor in the number of clonal or subclonal neoantigens were found (Figure 5.3C).

TRACERx lung study cohort and the LCE-project datasets were also used to explore the hypothesis that tumors from men had impaired antigen presentation mechanisms as compared with women.

As hypothesized, we found lower expression levels of both HLA Class I and Class II molecules in the TME of men as compared with women (FDR <0.05; Figure 5.2), and also a trend for higher frequency of LOH events at the HLA class I locus in tumors of men (Odds ratio [OR] for HLA-LOH in men versus women: 2.19, p-value=0.09; Figure 5.3D) as well as a numerically higher frequency for all other disruptive events (i.e., non-synonymous mutations or copy number loss) in genes involved in tumor antigen presentation mechanisms (OR for disruptive events in men versus women: 1.81, p-value=0.21; Figure 5.3D).

Assessment of sex-based differences in biological processes, cellular component and molecular functions

To assess sex-based differences in biological processes, cellular component and molecular functions we analyzed six datasets of the LCE-project through GSEA using the C5 collection of the MSigDB.

We analyzed 5917 gene sets and among the top 1% gene sets significantly enriched in the TME of women there were 39 (65%) gene sets directly related to anticancer immune response (Figure 5.6).

In particular, we found the following gene sets significantly enriched in the TME of women:

- *interferon gamma mediated signaling pathway, response to interferon gamma, response to type I interferon*, that are pathways and functions with a known key role in spontaneous anticancer immune response as well as in patients' response to treatment with ICIs;
- *MHC protein complex, antigen binding, lymphocyte activation*, that are very early processes necessary for tumor recognition by immune system;
- *leukocyte migration, leukocyte cell-cell adhesion*, that are involved in the TME infiltration by immune system.

None of the top 1% gene sets found significantly enriched in tumors arising in men as compared with women were directly related to anticancer immune responses (Figure 5.6).

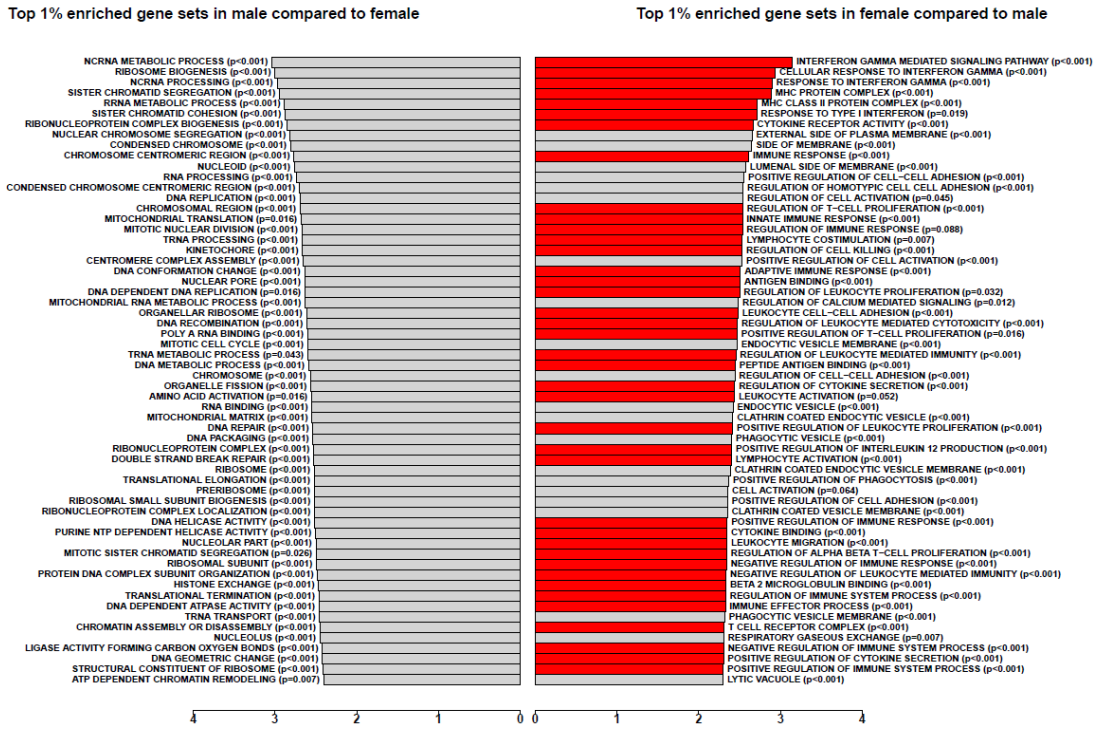

Figure 5.6 Top 1% enriched gene sets in male compared to female (left) and top 1% enriched gene sets in female compared to male (right). Red bars highlight gene sets related to anticancer immune response.

<u>Sex-based differences in the association between tumor mutational burden and patients' outcome</u>

Since we found that tumor of men had impaired neoantigens presentation mechanisms as compared with women, we analyzed data of the cohort of 329 patients from MSKCC dataset, with advanced NSCLC and treated with anti-PD-1 or anti-PD-L1 as monotherapy, to test the hypothesis of sex-based differences in the association between tTMB and patients outcome.

We analyzed data from 167 women and 162 men. During a median follow-up of 9 months (IQR 3-18 months), 99 and 111 deaths occurred in women and men, respectively. The median tTMB value was 7.9 mutations/Megabase (Mb) in female (min-max range 0-55, IQR 3.9-12.3) and 6.9 mutations/Mb in males (min-max range 0-100, IQR 4.4-12.8; Wilcoxon p-value=0.98).

A higher tTMB was associated with improved OS in both women (OS-HR for increase of 10 unit of tTMB/Mb: 0.72, 95% CI 0.54 to 0.95; p-value=0.02) and men (OS-HR for increase of 10 unit of tTMB/Mb: 0.76, 95% CI 0.62 to 0.94; p-value=0.01).

However, there was a sex-based difference in the linearity of the association between tTMB and patients OS. A linear trend towards decreasing HR of death for progressively increasing tTMB values was observed in women, along the entire range of tTMB values (test for linearity p-value=0.26; Figure 5.7A), whereas the association between tTMB and OS was not linear in men (test for linearity p-value=0.006, Figure 5.7B).

Results from spline regression analyses suggested an OS advantage starting from tTMB values >10 mutations/Mb in women (Figure 5.7A), while in males it appeared only for higher tTMB values (i.e., tTMB >20 mutations/Mb, Figure 5.7B).

Individual patient data on age, smoking history, tumor histotype and type of specimen analyzed were available for 262 patients. 224 out of 262 patients - respectively 112 women and 112 men - had an EGFR and ALK wild type tumor and were further analyzed through multivariable analyses.

Adjusting for age, smoking history, tumor histotype and type of specimen analyzed, we confirmed that tTMB retained a significant linear association with better OS in women (adjusted OS-HR for increase of 10 unit of tTMB/Mb: 0.58, 95% CI 0.37 to 0.90; p-value=0.02) but not in men, where the association became much less strong and not significant (adjusted OS-HR for increase of 10 unit of tTMB/Mb: 0.81, 95% CI 0.60 to 1.08; p-value=0.15).

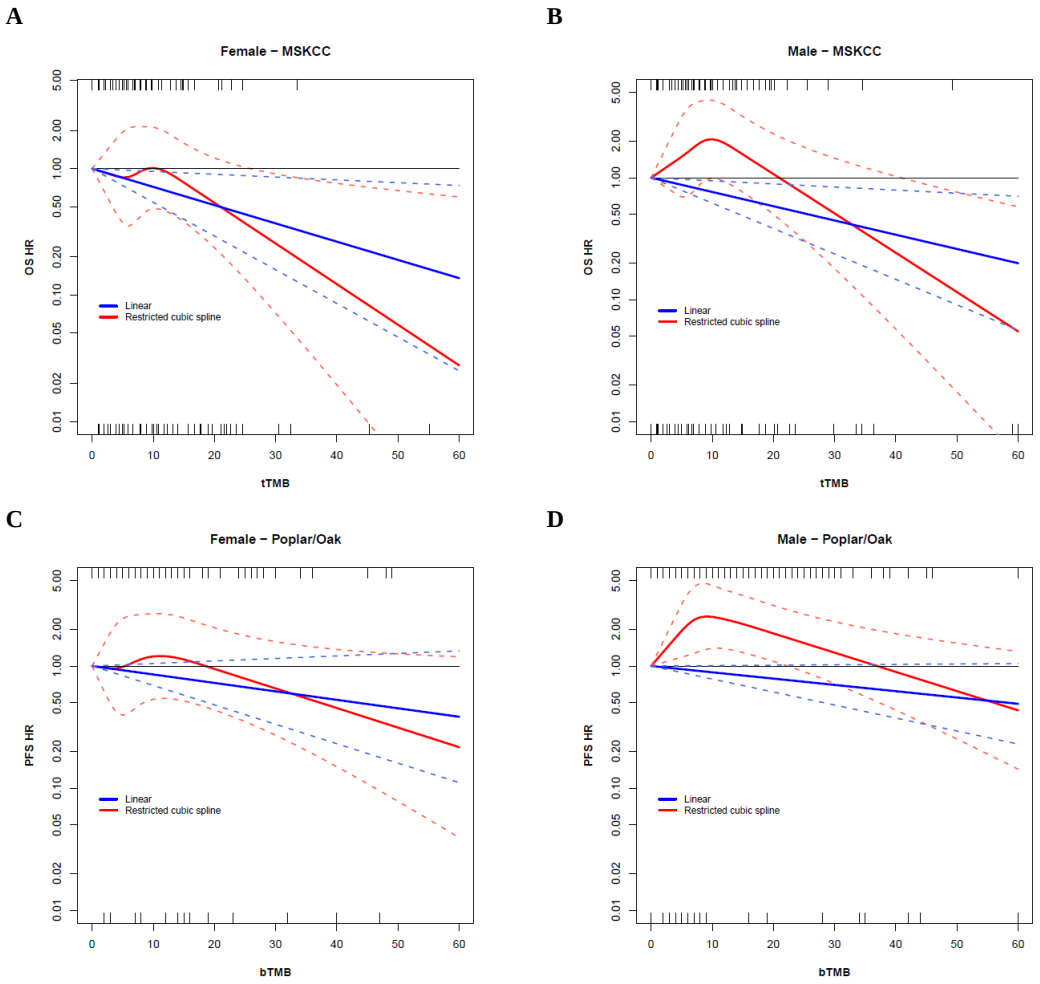

Figure 5.7 OS and PFS by TMB and gender in MSKCC (panel A and B) and Poplar/Oak (panel C and D) database

To confirm significant sex-based difference in the linearity of the association between TMB and patients outcome, we further analyzed data from 391 patients (128 women and 263 men), with advanced EGFR and ALK wild-type NSCLC treated with the anti-PD-L1 Atezolizumab in the OAK and POPLAR RCTs.

During a median follow-up of 11 months (IQR 5-20 months), 110 and 236 PFS events occurred in women and men, respectively. The median bTMB value was 6.0 mutations in female (min-max range 0-49, IQR 3.0-14.5) and 9.0 mutations in males (min-max range 0-67, IQR 5.0-17.0; Wilcoxon p-value=0.002).

Adjusting for age, tumor histology, number of metastatic sites at enrollment, sum of longest diameter of target lesions at baseline, smoking history and PD-L1 expression levels, we confirmed a significant sex-based difference in the linearity of the association between bTMB and PFS. A linear trend towards decreasing HR of PFS for progressively increasing bTMB values was observed in women, along the entire range of bTMB values (test for

linearity p-value=0.34; Figure 5.7C), whereas the association between bTMB and PFS was not linear in men (test for linearity p-value<0.001, Figure 5.7D).

Results from spline regression analyses suggested a PFS advantage starting from TMB values ≥20 mutations in women (Figure 5.7C), while in males appeared only for higher TMB values (i.e., TMB >37 mutations, Figure 5.7D).

5.4 Discussion

Taken together, data reported here show meaningful sex-based differences in the cell-type composition of the immune infiltrate of patients with NSCLC as well as in mechanisms exploited by tumors to evade immune response. Importantly, such differences are not related to other variables potentially associated with sex such as age, stage of disease, tumor histotype and smoking status.

Differences observed suggest that, on average, women mount stronger and more structured immune-response against NSCLC, as highlighted by the higher intratumor abundance of activated dendritic cells, memory CD4+ T-cells and B-cells including Class-switched memory cells.[9,79]

Our data also suggest that to evade such more efficient initial immune recognition and response, NSCLC arising in women develop more complex and redundant mechanisms of resistance, as revealed by the higher expression of multiple immune-checkpoint molecules with inhibitory functions, as well as by the higher abundance of immune-suppressive cells in the TME, such as CAFs and MDSCs, and also activated Tregs in squamous-NSCLC.[9,79]

These findings could explain the observed greater dysfunction status of T-cells infiltrating the TME of NSCLC of females revealed by TIDE.

Notably, it has been shown that the TIDE dysfunction signature specifically reflects the profile of dysfunctional T-cells strongly resistant to ICIs reprogramming.[73]

On the contrary, the TME of NSCLC arising in males was ‘colder’ as compared with females.

This was revealed by the lower abundance of a number of innate and adaptive immune cell types and confirmed by the significantly higher T-cells exclusion score as well as by the smaller TCR repertoire diversity.

We found that such poorer immune infiltration of tumors of men could depend on a less efficient tumor neoantigens presentation to the immune-system, due to lower expression

levels of HLA class I and II molecules, higher frequency of HLA type I LOH and/or alterations in other component of the antigen presentation machinery.

Notably, we did not find differences in the tumor neoantigens load between men and women. This further shows that the different prevalence of smoking habits between men and women cannot be the main reason explaining the sex-based dimorphism of anticancer immune response observed, since smoking mainly affects the anticancer immune response by increasing the TMB.[9,79]

Results obtained through GSEA using the C5 collection of the MsigDB further confirmed and reinforced our hypothesis of significant sex-based differences in anticancer immune response, being strictly in line with results found with the other analyses performed. Our results clearly showed that, among the molecular pathways and biological processes most significantly enriched in the TME of women as compared with men, there were many directly related to the anticancer immune response, while none of the gene sets found significantly enriched in tumors arising in men as compared with women were directly related to anticancer immune response. All our findings are in line and in part explain the significant enrichment of a number of innate and adaptive immune cells found in TME of women as well as the significant different association between the TMB and the outcome of patients treated with ICIs.

We previously demonstrated significant sex-based heterogeneity of response to different type of immunotherapy strategies in patients with advanced NSCLC.[23,80]

The sex-based dimorphism in key elements of anticancer immune response and in mechanisms of tumor immune-evasion showed here could explain our previous observations.

For example, the higher abundance of MDSCs, CAFs and Tregs found in TME of females, could explain both, the smaller survival benefit experienced by women when treated with anti-PD-1 as monotherapy - since it has been recurrently reported that these immune suppressive cells play a major role in ICIs resistance[9,79] - and also the impressively larger survival benefit observed in women in RCTs treated with the combination of anti-PD-1/PD-L1 with chemotherapy - considering the ability of chemotherapy to target these suppressive cell types.[81,82]

We provided a clear example of the potential clinical implications of our findings, showing significant differences in the association between TMB and survival benefit observed in men and women treated with anti-PD-1 or anti-PD-L1 antibodies.

The value of TMB as biomarker to select patients who benefit the most from anticancer immunotherapy is debated, since discordant results on its predictive value for OS benefit have been reported in trials testing ICIs.

Our analysis suggested that considering different TMB-thresholds in men and women may improve its predictive value for both, and future research should test such hypothesis.

This work has several strengths. The first is the quality of data analyzed, since we evaluated several independent large cohorts of hundreds of patients with NSCLC, including the TRACERx lung study cohort, that is one of the best and most deeply characterized at the molecular level among all reported.

Furthermore, all the statistical and computational approaches used are straightforward and widely validated.

Finally, this work is supported by a strong rationale, represented by the known meaningful differences of both innate and adaptive immune system between men and women, explaining different prevalence and mortality from autoimmune and infectious diseases and from several types of cancers. The results reported here are in line with and were anticipated in our previous works on this issue.

Yet, this paper has limitations and further studies are needed to better explore such complex issue.

The main limitation is that we only analyzed cohorts of patients with early-stage disease since we did not have data on advanced tumors. However, we expect that the sex-based dimorphism of anticancer immune response and immune evasion mechanisms could be even deeper in advanced tumors, as a consequence of continuous immune-editing process and tumor evolution during tumor progression.

We also did not study other tumor histotypes, while our previous data showed large sex-based differences of ICIs effectiveness also in solid tumors other than NSCLC, including melanoma.

In conclusion, data reported here and in our previous works provided a proof of the importance of the features of the host immune system in shaping the immune response against cancer. This could have several straightforward implications in the context of both translational and clinical research. These include the need to explore differential therapeutic approaches in men and women with cancers to improve results for both.

Chapter 6

Conclusions

The overall aim of this doctoral thesis was to extend previous (limited) knowledge about sex-based differences in cancer immunotherapy efficacy and to further explore mechanisms at molecular level that regulate anticancer immune response in men and women.

First we showed that women with advanced lung cancer derive larger benefit than men from the combination of chemotherapy to an anti-PD-1/PD-L1 immune checkpoint inhibitor. We showed also that there is an interaction between patients' sex and the efficacy of the immunotherapeutic strategy chosen: men derive larger benefit than women with an anti-PD-1 treatment alone, while women have better survival with anti-PD-1/PD-L1 plus chemotherapy.

Then we showed that on average women mount stronger and more structured immune-response against NSCLC compared to men, proved also by the results of the GSEA which highlighted that most of the molecular pathways and biological processes significantly enriched in the TME of women are directly related to the anticancer immune response.

Our data support the need to test different therapeutic strategies in male and female populations to further improve the use of immunotherapy. For example, accrual and design of trials of immunotherapy might best be performed separately for men and women, with proper sample size planning for both.

Clearly our works have some limitations. Part of our works were based on published results rather than on individual patients' data and this precluded the possibility of exploring relevant issues, such as the menopausal status of female patients on the efficacy of immunotherapeutic treatments. The analysis of molecular mechanisms that drive anticancer immune response in men and women regarded only patients with early-stage lung cancer instead of including also advanced tumors and/or other type of tumors. The methodology that we proposed and implemented to apply GSEA to meta-analysis was not yet validated in its operating characteristics.

On the contrary, our works have several strengths. We meta-analyzed results of large RTCs, we used several independent large cohorts publicly available of hundreds of patients with NSCLC, including also the TRACERx lung study cohort. Our rationale was strong and was represented by the known meaningful differences of both innate and adaptive immune system between men and women.

Undoubtedly, all our findings require further validation in future prospective trials before being considered sufficient to support any change in clinical practice.

Our future efforts will be directed to:

- the validation of our proposed method for GSEA meta-analysis through simulations;
- the analysis of sex-based differences in molecular mechanisms of anticancer immune response in advanced-stage tumors as well as in histotypes other than NSCLC.

References

1. WHO. Accessed July 1, 2020. https://www.who.int/health-topics/cancer#tab=tab_1
2. Gale RP. MSD Manual - Overview of cancer. Accessed July 1, 2020. https://www.msdmanuals.com/professional/hematology-and-oncology/overview-of-cancer/overview-of-cancer
3. National Cancer Institute. Accessed July 1, 2020. https://www.cancer.gov/about-cancer/treatment/types/immunotherapy
4. Memon H, Patel BM. Immune checkpoint inhibitors in non-small cell lung cancer: a bird's eye view. *Life Sci*. 2019;233(April):116713. doi:10.1016/j.lfs.2019.116713
5. National Cancer Institute. Accessed July 1, 2020. https://www.cancer.gov/types/lung
6. Ventola CL. Cancer Immunotherapy, Part 1: Current Strategies and Agents. *P&T*. 2017;42(6):375-383.
7. Darvin P, Toor SM, Sasidharan Nair V, et al. Immune checkpoint inhibitors: recent progress and potential biomarkers. *Exp Mol Med*. 2018;50(12):1-11. doi:10.1038/s12276-018-0191-1
8. Oiseth SJ, Aziz MS. Cancer immunotherapy: a brief review of the history, possibilities, and challenges ahead. *J Cancer Metastasis Treat*. 2017;3:250-261. doi:10.20517/2394-4722.2017.41
9. Fares CM, Van Allen EM, Drake CG, Allison JP, Hu-Lieskovan S. Mechanisms of Resistance to Immune Checkpoint Blockade: Why Does Checkpoint Inhibitor Immunotherapy Not Work for All Patients? *Am Soc Clin Oncol Educ B*. 2019;(39):147-164. doi:10.1200/EDBK_240837
10. Capone I, Marchetti P, Ascierto PA, et al. Sexual Dimorphism of Immune Responses: A New Perspective in Cancer Immunotherapy. *Front Immunol*. 2018;9:552. doi:10.3389/fimmu.2018.00552
11. Spranger S, Gajewski TF. Mechanisms of Tumor Cell-Intrinsic Immune Evasion. *Annu Rev Cancer Biol*. 2018;2:213-228. doi:10.1146/annurev-cancerbio-030617-050606
12. Duma N, Abdel-Ghani A, Yadav S, et al. Sex Differences in Tolerability to Anti-Programmed Cell Death Protein 1 Therapy in Patients with Metastatic Melanoma and Non-Small Cell Lung Cancer: Are We All Equal? *Oncologist*. 2019;24:e1148-e1155. doi:10.1634/theoncologist.2019-0094
13. Samstein RM, Lee C-H, Shoushtari AN, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. 2019;51(2):202-206. doi:10.1038/s41588-018-0312-8
14. Klein SL, Morgan R. The impact of sex and gender on immunotherapy outcomes. *Biol Sex Differ*. 2020;11(1):24. doi:10.1186/s13293-020-00301-y
15. Özdemir BC, Csajka C, Dotto G-P, Wagner AD. Sex Differences in Efficacy and Toxicity of Systemic Treatments: An Undervalued Issue in the Era of Precision Oncology. *J Clin Oncol*. 2018;36(26):2680-2683. doi:10.1200/JCO.2018.78.3290

16. Markle JG, Fish EN. SeXX matters in immunity. *Trends Immunol*. 2014;35(3):97-104. doi:10.1016/j.it.2013.10.006

17. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. 2016;16(10):626-638. doi:10.1038/nri.2016.90

18. Kovats S. Estrogen receptors regulate innate immune cells and signaling pathways. *Cell Immunol*. 2015;294(2):63-69. doi:10.1016/j.cellimm.2015.01.018

19. Polanczyk MJ, Hopke C, Vandenbark AA, Offner H. Treg suppressive activity involves estrogen-dependent expression of programmed death-1 (PD-1). *Int Immunol*. 2007;19(3):337-343. doi:10.1093/intimm/dxl151

20. Thorsson V, Gibbs DL, Brown SD, et al. The Immune Landscape of Cancer. *Immunity*. 2018;48(4):812-830.e14. doi:10.1016/j.immuni.2018.03.023

21. Loo K, Tsai KK, Mahuron K, et al. Partially exhausted tumor-infiltrating lymphocytes predict response to combination immunotherapy. *JCI insight*. 2017;2(14). doi:10.1172/jci.insight.93433

22. Botticelli A, Onesti CE, Zizzari I, et al. The sexist behaviour of immune checkpoint inhibitors in cancer therapy? *Oncotarget*. 2017;8(59):99336-99346.

23. Conforti F, Pala L, Bagnardi V, et al. Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. *Lancet Oncol*. 2018;19(6):737-746. doi:10.1016/S1470-2045(18)30261-4

24. Wallis CJD, Butaney M, Satkunasivam R, et al. Association of Patient Sex with Efficacy of Immune Checkpoint Inhibitors and Overall Survival in Advanced Cancers: A Systematic Review and Meta-analysis. *JAMA Oncol*. 2019;5(4):529-536. doi:10.1001/jamaoncol.2018.5904

25. Conforti F, Pala L, Bagnardi V, et al. Sex-based differences of the tumor mutational burden and T-cell inflammation of the tumor microenvironment. *Ann Oncol*. 2019;30(40):653-655. doi:10.1093/annonc/mdz034

26. Cristescu R, Mogg R, Ayers M, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science*. 2018;362(6411):eaar3593. doi:10.1126/science.aar3593

27. Wang S, Cowley LA, Liu X-S. Sex Differences in Cancer Immunotherapy Efficacy, Biomarkers, and Therapeutic Strategy. *Molecules*. 2019;24(18):3214.

28. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: John Wiley & Sons, ed. *Systematic Reviews in Health Care: Meta-Analysis in Context*. ; 2008:285-312. doi:10.1002/9780470693926

29. Petrie A, Sabin C. Systematic reviews and meta-analysis. In: John Wiley & Sons, ed. *Medical Statistics at a Glance*. Eight. ; 2013:130-132.

30. Dersimonian R, Laird N. Meta-Analysis in Clinical Trials. *Control Clin Trials*. 1986;7:177-188.

31. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? *BMJ*. 2017;356:j573. doi:10.1136/bmj.j573

32. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B*. 1972;34(2):187-220. doi:10.2307/2985181

33. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. (Springer-Verlag, ed.).; 2010.

34. Durrleman S, Simon R. Flexible regression models with cubic splines. *Stat Med*. 1989;8(5):551-561. doi:10.1002/sim.4780080504

35. Gauthier J, Wu Q V, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone Marrow Transplant*. 2020;55(4):675-680. doi:10.1038/s41409-019-0679-x

36. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide. *PNAS*. 2005;102(43):15545-15550.

37. Ackermann M, Strimmer K. A general modular framework for gene set enrichment analysis. *BMC Bioinformatics*. 2009;10:1-20. doi:10.1186/1471-2105-10-47

38. Shen K, Tseng GC. Meta-analysis for pathway enrichment analysis when combining multiple genomic studies. *Bioinformatics*. 2010;26(10):1316-1323. doi:10.1093/bioinformatics/btq148

39. Chen M, Zang M, Wang X, Xiao G. A powerful Bayesian meta-analysis method to integrate multiple gene set enrichment studies. *Bioinformatics*. 2013;29(7):862-869. doi:10.1093/bioinformatics/btt068

40. Lu W, Wang X, Zhan X, Gazdar A. Meta-analysis approaches to combine multiple gene set enrichment studies. *Stat Med*. 2018;37:659-672. doi:10.1002/sim.7540

41. Jadad AR, Moore RA, Carroll D, et al. Assessing the Quality of Reports of Randomized Clinical Trials: Is Blinding Necessary? *Control Clin Trials*. 1996;17(1):1-12. doi:10.1016/0197-2456(95)00134-4

42. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539-1558. doi:10.1002/sim.1186

43. Paz-Ares L, Luft A, Vicente D, et al. Pembrolizumab plus Chemotherapy for Squamous Non–Small-Cell Lung Cancer. *N Engl J Med*. 2018;379:2040-2051. doi:10.1056/NEJMoa1810865

44. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *N Engl J Med*. 2018;378:2078-2092. doi:10.1056/NEJMoa1801005

45. Cappuzzo F, McCleod M, Hussein M, et al. IMpower130: Progression-free survival (PFS) and safety analysis from a randomized phase 3 study of carboplatin + nab-paclitaxel (CnP) with or without atezolizumab (atezo) as first-line (1L) therapy in advanced nonsquamous NSCLC. In: *ESMO 2018. Abstract LBA53*.

46. Jotte R, Cappuzzo F, Vynnychenko I, et al. IMpower131: primary PFS and safety analysis of a randomized phase III study of atezolizumab + carboplatin + paclitaxel or nab-paclitaxel vs carboplatin + nab-paclitaxel as 1L therapy in advanced squamous NSCLC. In: *ASCO 2018. Abstract LBA9000*.

47. Papadimitrakopoulou VA, Cobo M, Bordoni R, et al. IMpower132: PFS and Safety Results with 1L Atezolizumab + Carboplatin / Cisplatin + Pemetrexed in Stage IV

Non-Squamous NSCLC. In: *International Association for the Study of Lung Cancer 19th World Conference on Lung Cancer*. ; 2018.

48. Horn L, Mansfield AS, Szczęsna A, et al. First-Line Atezolizumab plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer. *N Engl J Med*. 2018;379:2220-2229. doi:10.1056/NEJMoa1809064

49. Antonia SJ, Villegas A, Daniel D, et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. *N Engl J Med*. 2018;379:2342-2350. doi:10.1056/NEJMoa1809697

50. Socinski MA, Jotte RM, Capuzzo F, et al. Atezolizumab for First-Line Treatment of Metastatic Nonsquamous NSCLC. *N Engl J Med*. 2018;378:2288-2301. doi:10.1056/NEJMoa1716948

51. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med*. 2016;375(19):1823-1833. doi:10.1056/NEJMoa1606774

52. Lopes G, Wu Y-L, Kudaba I, et al. Pembrolizumab (pembro) versus platinum-based chemotherapy (chemo) as first-line therapy for advanced/metastatic NSCLC with a PD-L1 tumor proportion score (TPS) ≥ 1%: Open-label, phase 3 KEYNOTE-042 study. *J Clin Oncol*. 2018;36(18_suppl):LBA4. doi:10.1200/JCO.2018.36.18_suppl.LBA4

53. Carbone DP, Reck M, Paz-Ares L, et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N Engl J Med*. 2017;376(25):2415-2426. doi:10.1056/NEJMoa1613493

54. Salem ME, Xiu J, Lenz H-J, et al. Characterization of tumor mutation load (TML) in solid tumors. *J Clin Oncol*. 2017;35(15_suppl):11517. doi:10.1200/JCO.2017.35.15_suppl.11517

55. Xiao D, Pan H, Li F, Wu K, Zhang X, He J. Analysis of ultra-deep targeted sequencing reveals mutation burden is associated with gender and clinical outcome in lung adenocarcinoma. *Oncotarget*. 2016;7(16):22857-22864. doi:10.18632/oncotarget.8213

56. Wanderley CW, Colón DF, Luiz JPM, et al. Paclitaxel Reduces Tumor Growth by Reprogramming Tumor-Associated Macrophages to an M1 Profile in a TLR4-Dependent Manner. *Cancer Res*. 2018;78(20):5891 LP - 5900. doi:10.1158/0008-5472.CAN-17-3480

57. Santa-Maria CA, Nanda R. Immune Checkpoint Inhibitor Therapy in Breast Cancer. *J Natl Compr Canc Netw*. 2018;16(10):1259-1268. doi:10.6004/jnccn.2018.7046

58. Kok M, Winer EP, Loi S. Passion for immune checkpoint blockade in triple negative breast cancer: Comment on the IMpassion130 study. *Ann Oncol*. 2019;30(1):13-16. doi:10.1093/annonc/mdy473

59. Schmid P, Adams S, Rugo HS, et al. Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N Engl J Med*. 2018;379(22):2108-2121. doi:10.1056/NEJMoa1809615

60. Cai L, Lin S, Girard L, et al. LCE: an open web portal to explore gene expression and clinical associations in lung cancer. *Oncogene*. 2019;38:2551-2564. doi:10.1038/s41388-018-0588-2

61. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511):543-550. doi:10.1038/nature13385

62. Rousseaux S, Debernardi A, Jacquiau B, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med*. 2013;5(186):186ra66. doi:10.1126/scitranslmed.3005723

63. Sato M, Larsen JE, Lee W, et al. Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. *Mol Cancer Res*. 2013;11(6):638-650. doi:10.1158/1541-7786.MCR-12-0634-T

64. Schabath MB, Welsh EA, Fulp WJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene*. 2016;35(24):3209-3216. doi:10.1038/onc.2015.375

65. Shedden K, Taylor JMG, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med*. 2008;14(8):822-827. doi:10.1038/nm.1790

66. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489(7417):519-525. doi:10.1038/nature11404

67. Noro R, Ishigame T, Walsh N, et al. A Two-Gene Prognostic Classifier for Early-Stage Lung Squamous Cell Carcinoma in Multiple Large-Scale and Geographically Diverse Cohorts. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2017;12(1):65-76. doi:10.1016/j.jtho.2016.08.141

68. Jamal-Hanjani M, Wilson GA, McGranahan N, et al. Tracking the Evolution of Non–Small-Cell Lung Cancer. *N Engl J Med*. 2017;376(22):2109-2121. doi:10.1056/NEJMoa1616288

69. McGranahan N, Rosenthal R, Hiley CT, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell*. 2017;171(6):1259-1271.e11. doi:10.1016/j.cell.2017.10.001

70. Rosenthal R, Cadieux EL, Salgado R, et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature*. 2019;567(7749):479-485. doi:10.1038/s41586-019-1032-7

71. Gandara DR, Paul SM, Kowanetz M, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med*. 2018;24(9):1441-1448. doi:10.1038/s41591-018-0134-3

72. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18(1):220. doi:10.1186/s13059-017-1349-1

73. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med*. 2018;24(10):1550-1558. doi:10.1038/s41591-018-0136-1

74. Oakes T, Heather JM, Best K, et al. Quantitative Characterization of the T Cell Receptor Repertoire of Naïve and Memory Subsets Using an Integrated Experimental and Computational Pipeline Which Is Robust, Economical, and Versatile . *Front Immunol* . 2017;8:1267. https://www.frontiersin.org/article/10.3389/fimmu.2017.01267

75. Best K, Oakes T, Heather JM, Shawe-Taylor J, Chain B. Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. *Sci Rep*. 2015;5:14629. doi:10.1038/srep14629

76. Hoof I, Peters B, Sidney J, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*. 2009;61(1):1-13. doi:10.1007/s00251-008-0341-z

77. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. 2016;32(4):511-517. doi:10.1093/bioinformatics/btv639

78. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23(6):703-713. doi:10.1038/nm.4333

79. Hegde PS, Chen DS. Top 10 Challenges in Cancer Immunotherapy. *Immunity*. 2020;52(1):17-35. doi:10.1016/j.immuni.2019.12.011

80. Conforti F, Pala L, Bagnardi V, et al. Sex-Based Heterogeneity in Response to Lung Cancer Immunotherapy: A Systematic Review and Meta-Analysis. *J Natl Cancer Ist*. 2019;111(8):djz094. doi:10.1093/jnci/djz094

81. Vincent J, Mignot G, Chalmin F, et al. 5-Fluorouracil selectively kills tumor-associated myeloid-derived suppressor cells resulting in enhanced T cell-dependent antitumor immunity. *Cancer Res*. 2010;70(8):3052-3061. doi:10.1158/0008-5472.CAN-09-3690

82. Ghiringhelli F, Larmonier N, Schmitt E, et al. CD4+CD25+ regulatory T cells suppress tumor immunity but are sensitive to cyclophosphamide which allows immunotherapy of established tumors to be curative. *Eur J Immunol*. 2004;34(2):336-344. doi:10.1002/eji.200324181

83. Yang L, Forker L, Irlam JJ, Pillay N, Choudhury A, West CML. Validation of a hypoxia related gene signature in multiple soft tissue sarcoma cohorts. *Oncotarget*. 2018;9(3):3946-3955. doi:10.18632/oncotarget.23280

84. Arrieta VA, Cacho-Díaz B, Zhao J, Rabadan R, Chen L, Sonabend AM. The possibility of cancer immune editing in gliomas. A critical review. *Oncoimmunology*. 2018;7(7):e1445458. doi:10.1080/2162402X.2018.1445458

Appendix

Assessment of T-cell receptor abundance and Shannon diversity score

A previously developed quantitative experimental and computational TCR sequencing pipeline was used for the high-throughput sequencing of α and β TCR chains. TCR sequencing was performed on whole RNA extracted from multi-region tumor specimens.[83] A distinct feature of this TCR sequencing protocol is the use of a unique molecular identifier that enables correction for PCR and sequencing errors, thereby providing a quantitative and reproducible method of library preparation.[74,75]

For each tumor region, the Shannon diversity was estimated using the command 'entropy.empirical' from the 'entropy' R package. This was calculated on the basis of the number and prevalence of different TCR subclones found in that region, such that a tumor region that contained only one subclone was assigned a value of 0.

Assessment of ubiquitous expanded TCRs in tumor

We counted the number of TCRs detected with frequencies above a range of frequency thresholds in the tumor samples. To focus on the most expanded TCRs, we examined those present above a threshold frequency of 2/1000 (corresponding to the top 1% of the empirical TCR frequency distribution) in at least one region of the tumor.[69,70]

We calculated the relative abundance of the TCR in the tumor (averaged over all tumor regions) from the same patient. Expanded intratumoral TCRs were subsequently classified as ubiquitous or regional, as previously described.[69,70]

In particular, expanded TCRs were classified as regional if they were absent from at least one region of the tumor and as ubiquitous if otherwise. Ubiquitous TCRs can be absent from the data for specific regions, but this should be attributed to sampling rather than true spatial heterogeneity.[69,70]

For this analysis, we analyzed tumors with TCR data for at least three different regions.

Assessment of predicted neoantigen load

Novel 9-11mer peptides that could arise from identified non-silent mutations present in the sample were determined.[68] The predicted half-maximal inhibitory concentration binding affinities and rank percentage scores - representing the rank of the predicted affinity compared to a set of 400,000 random natural peptides - were calculated for all peptides that bound to each of the patient's HLA alleles using netMHCpan-2.8 and netMHC-4.0.[76,77]

Using established thresholds, predicted binders were considered to be those peptides that had a predicted binding affinity <500nM or a rank percentage score <2% by either tool.[74,75] When RNA-seq data were available, a neoantigen was considered to be expressed if at least five RNA-seq reads mapped to the mutation position, and at least three contained the mutated base.[74,75]

Assessment of HLA LOH and/or alteration in other antigen-presentation-pathway genes

Tumor samples that contain an HLA LOH event were identified using a previously validated method.[69]

Antigen-presentation-pathway genes were compiled from a previous report, and affected the HLA enhanceosome, peptide generation, chaperones or the MHC complex itself.[84]

For each tumor sample, we assessed disruptive events (i.e., non-silent mutations or copy-number loss defined relative to ploidy) of the following genes: CIITA, IRF1, PSME1, PSME2, PSME3, ERAP1, ERAP2, HSPA (also known as PSMA7), HSPC (also known as HSPBP1), TAP1, TAP2, TAPBP, CALR, CNX (alias CANX), PDIA3 and B2M.