



Characterization of Neoantigen Load Subgroups in Gynecologic and Breast Cancers

Yue Zhu, Xiaowei Meng, Xinjia Ruan, Xiaofan Lu, Fangrong Yan* and Fei Wang*

State Key Laboratory of Natural Medicines, Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing, China

Objective: Although gynecologic and breast (Pan-Gyn) cancers share a variety of similar characteristics, their response to immunotherapy is different. Immune checkpoint inhibitor therapy is not effective in all patients, while neoantigen load (NAL) may be a predictive biomarker. However, the selection of a NAL cutoff point and its predictive effect remain to be elucidated.

Methods: We divided 812 Pan-Gyn cancer samples from The Cancer Genome Atlas into three groups based on 60 and 80% of their load percentile. We then correlated the identified NAL subgroups with gene expression, somatic mutation, DNA methylation, and clinicopathological information. We also characterized each subgroup by distinct immune cell enrichment, PD-1 signaling, and cytolytic activity. Finally, we predicted the response of each subgroup to chemotherapy and immunotherapy.

Results: Across Pan-Gyn cancers, we identified three distinct NAL subgroups. These subgroups showed differences in biological function, genetic information, clinical variables, and immune infiltration. Eighty percent was identified as a meaningful cutoff point for NAL. In all patients, a higher NAL (top 20%) was associated with better overall survival as well as high immune infiltration and low intra-tumor heterogeneity. Furthermore, an interesting lncRNA named AC092580.4 was found, which was associated with two significantly different immune genes (CXCL9 and CXCL13).

Conclusions: Our novel findings provide further insights into the NAL of Pan-Gyn cancers and may open up novel opportunities for their exploitation toward personalized treatment with immunotherapy.

Keywords: gynecologic and breast cancer, neoantigen load, immune infiltrate, intratumor heterogeneity, immunotherapy

INTRODUCTION

In recent years, cancer immunotherapy, especially immune checkpoint inhibitor (ICI) treatment, has revolutionized the traditional treatments of patients with advanced tumor. Antibodies targeting CTLA-4 or PD-1/PD-L1 are effective in treating a variety of malignant tumors (Callahan et al., 2016). However, durable benefits are limited to a minority of patients. The

OPEN ACCESS

Edited by:

Yun Xiao,
Harbin Medical University, China

Reviewed by:

Gustavo Fioravanti Vieira,
University La Salle Canoas, Brazil
Zhan Zhou,
Zhejiang University, China

*Correspondence:

Fangrong Yan
f.r.yan@163.com
Fei Wang
blue6923@163.com

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 30 January 2020

Accepted: 04 June 2020

Published: 13 July 2020

Citation:

Zhu Y, Meng X, Ruan X, Lu X,
Yan F and Wang F (2020)
Characterization of Neoantigen Load
Subgroups in Gynecologic and Breast
Cancers.
Front. Bioeng. Biotechnol. 8:702.
doi: 10.3389/fbioe.2020.00702

biological mechanisms that drive the individual heterogeneity of these responses are not fully understood, but are important for the design of personalized immunotherapy strategies. Recently, some phase 3 clinical trials reported negative results for both non-selective patients and selected groups, emphasizing the importance of better predictive biomarkers for clinical demand (Carbone et al., 2017; Powles et al., 2018). Early findings have indicated that PD-L1 immunohistochemistry, peripheral blood markers, several relative gene expression signatures, and T cell receptor clonality may be associated with clinical response (Gibney et al., 2016). For example, tumor-associated macrophages (TAM) and regulatory T cells (Tregs) are associated with tumor-promoting function (Nishikawa and Sakaguchi, 2014), while CD8⁺ T cells are associated with improved clinical outcomes and immunotherapeutic responses (Tumeh et al., 2014). The antitumor activity of antigen-specific CD8⁺ T cells may be the basis for the efficacy of immune checkpoint blocking therapy since the amount and activity of CD8⁺ T cells increase with these drugs (Rizvi et al., 2015). Conversely, a low T cell density is associated with poor prognosis (Galon et al., 2007). Additionally, a correlation between a high mutation load and clinical benefit to immune checkpoint blockade was observed in a small cohort of patients with melanoma and lung and colon cancers (Snyder et al., 2014; Le et al., 2015; Rizvi et al., 2015; Hugo et al., 2016). Moreover, a high tumor mutational burden (TMB) usually significantly correlates with higher tumor-infiltrating lymphocyte (TIL) levels (Thomas et al., 2018), and TMB has been proven to be a predictive biomarker for clinical benefit after immunotherapy (Samstein et al., 2019).

Nowadays, with the development of new sequencing technologies, specialized computational methods, and human leukocyte antigen (HLA) binding predictions, neoantigen has been utilized. These neoepitopes with specific amino acid sequence variations generated by cancer somatic mutations can be recognized by the immune system. Generally, patients with high TMB have more neoantigens. However, it is unclear whether a high neoantigen load (NAL) is robustly predictive of clinical benefit across diverse human tumors.

Gynecologic and breast (Pan-Gyn) cancers share several kinds of similar characteristics: the Müllerian duct's development, the influence of female hormones, and the special gynecologic oncology effect (Mullen and Behringer, 2014). Recently, similarities in molecular characteristics have been found in Pan-Gyn cancers in a comprehensive pan-cancer analysis study (Berger et al., 2018; Hoadley et al., 2018). These cancers have been proven to be highly immune-infiltrating tumors in various clinical and genomic studies (Bregar et al., 2017). However, the mechanisms of malignant tumor immunity infiltration and immune response to treatment are still poorly understood.

In this study, we investigated the association between NAL and overall survival across the following five The Cancer Genome Atlas (TCGA) cancer types: breast carcinomas (BRCA), uterine cervical carcinomas (CESC), ovary carcinomas (OV), endometrial carcinomas (UCEC), and uterine carcinosarcomas (UCS). They represent the most frequent and aggressive gynecologic cancers. To better understand the complex impact of neoantigens, we further comprehensively characterized the NAL

subgroups in a multiple omics view including somatic mutation, gene expression, DNA methylation, and long non-coding RNA expressions in these five gynecologic cancers. We characterized the subgroups by immunity infiltration state, intra-tumor heterogeneity (ITH), PD-L1 immune blocking point inhibitor, and other immune signatures. We also predicted the response of each subgroup to immunotherapy and chemotherapy. The present research can provide necessary biological information for NAL, guidance on personalized immunotherapy options, and decision on patients' management.

MATERIALS AND METHODS

Patients Tumor Samples

The mutation annotation file (MAF) files containing Pan-Gyn cancers' somatic mutation information and DNA methylation beta value were obtained from The Cancer Genome Atlas Project (TCGA) pan-cancer analyses data portal¹. Transcriptomic sequencing (RNA-Seq) raw count data of the Pan-Gyn cancers with 2,199 tumor samples, including 1,049 BRCA, 186 CESC, 419 OV, 488 UCEC, and 57 UCS, were downloaded from the GDC data portal². Nine hundred and thirty one immune-related Pan-Gyn samples were selected with $p < 0.05$ by the CIBERSORT algorithm (Newman et al., 2015). The corresponding clinical and pathologic information files were obtained from Firehose³. The 4,165 gynecologic tumor-specific potential neoantigens predicted by NetMHCpan 2.8 were available from TSNAdb⁴ (Hoof et al., 2009; Wu et al., 2018).

Neoantigen Load Assessment

The MAF file with 812 Pan-Gyn cancer samples was filtered by tumor-specific neoantigens. The total number of neoantigens identified was normalized to the exonic coverage sequenced. The R package "maftools" was used to compute the Pan-Gyn NAL with the MAF file (Mayakonda et al., 2018). Neoantigen load cutoffs of 60 and 80% were selected based on the different immune states, obtaining 163 samples as the neoantigen load-high (NAL-H) group, 161 samples as the neoantigen load-middle (NAL-M) group, and 488 samples as the neoantigen load-low (NAL-L) group.

RNA Analysis

The Ensembl ID for genes was annotated in GENCODE27 to obtain gene symbol names. The protein coding genes [messenger RNA (mRNA)] and long non-coding RNA (lncRNA) were selected. Raw count data were then converted into FPKM (the fragments per kilobase of exon per million fragments mapped) values for analysis. To reduce noise, we filtered out low-expression genes with FPKM values below 1 in at least 90% of the samples. Batch effect removal was performed by the R package "combat." Differential expression

¹<https://portal.gdc.cancer.gov/about-data/publications/pancanatlas>

²<https://portal.gdc.cancer.gov/>

³<http://www.firehose.org/>

⁴<http://biopharm.zju.edu.cn/tsnadb/>

analysis among the NAL subgroups was performed by the R package “limma” with the standard comparison mode. The significantly differentially expressed genes were obtained with a false discovery rate (FDR) < 0.05 and fold change greater than 2 for overexpression or less than 0.5 for down-expression. Gene Ontology (GO) annotation was then performed using the R package “clusterProfiler” to characterize the subgroups according to the differentially expressed mRNAs. The correlation between the lncRNAs and mRNAs was computed, and differentially expressed lncRNAs were filtered with a correlation higher than 0.6. lncRNA functions were predicted with their highly correlated genes using gene set enrichment analysis (GSEA) in the R package “clusterProfiler” (Yu et al., 2012).

Integrative Analysis of DNA Methylation and mRNA Expression

We performed integrative analysis among the DNA methylation and mRNA expression to explore epigenetically silenced or activated genes (epi-silenced/activated genes). Specifically, we chose methylation probes that were differentially expressed in the NAL subgroups and excluded the correlations between probes and mRNAs. Differentially methylated analysis was performed by the R package “limma” on defined subgroups. The significantly different methylation probes were obtained with FDR < 0.05. We combined the differential expression information with the differential methylation results by correlation analysis. The methylation probes of interest were filtered with correlation higher than 0.6.

Mutation Analysis

MutSigCV was used to infer significant tumor-mutated genes ($q < 0.05$) with default parameters (Lawrence et al., 2013). Significantly different mutations among neoantigen load subgroups were obtained using the R package “limma” (FDR < 0.05). According to the hg19 human reference genome⁵, we also analyzed 30 mutation signatures with the MAF file and compared the mutation signatures among the identified subgroups.

Chemotherapeutic Response Prediction

We predicted the chemotherapy response for each sample based on the largest public pharmacogenomics database, the Genomics of Drug Sensitivity in Cancer (GDSC). Three commonly used and three other useful chemicals were selected, namely, cisplatin, docetaxel, paclitaxel, etoposide, vinorelbine, and gemcitabine. The prediction process was performed by the R package “pRRophetic,” in which the IC₅₀ of the sample was estimated by ridge regression and the prediction accuracy was evaluated by 10-fold cross-validation based on the GDSC training set. All parameters were set by the fault value, with “combat” for batch effect removal and the expression of repeated genes summarized as the average (Geeleher et al., 2014).

⁵<http://hgdownload.cse.ucsc.edu/>

Statistical Analysis

All statistical tests were performed by R/3.6.1, using χ^2 or Fisher’s exact test for the categorical data, two-sample Wilcoxon test (Mann–Whitney test) for continuous data, and Kaplan–Meier curve of log-rank test and Cox regression for the hazard ratio (HR). Survival analysis was executed by using the R package “survival.” Log-rank test was used to estimate the p value. Fisher’s exact independence test was used to statistically test the association between the categorical clinical information and the identified neoantigen subgroups. For all statistical analyses, a p value less than 0.05 was considered statistically significant. All statistical modeling and visualization were performed using the R language.

RESULTS

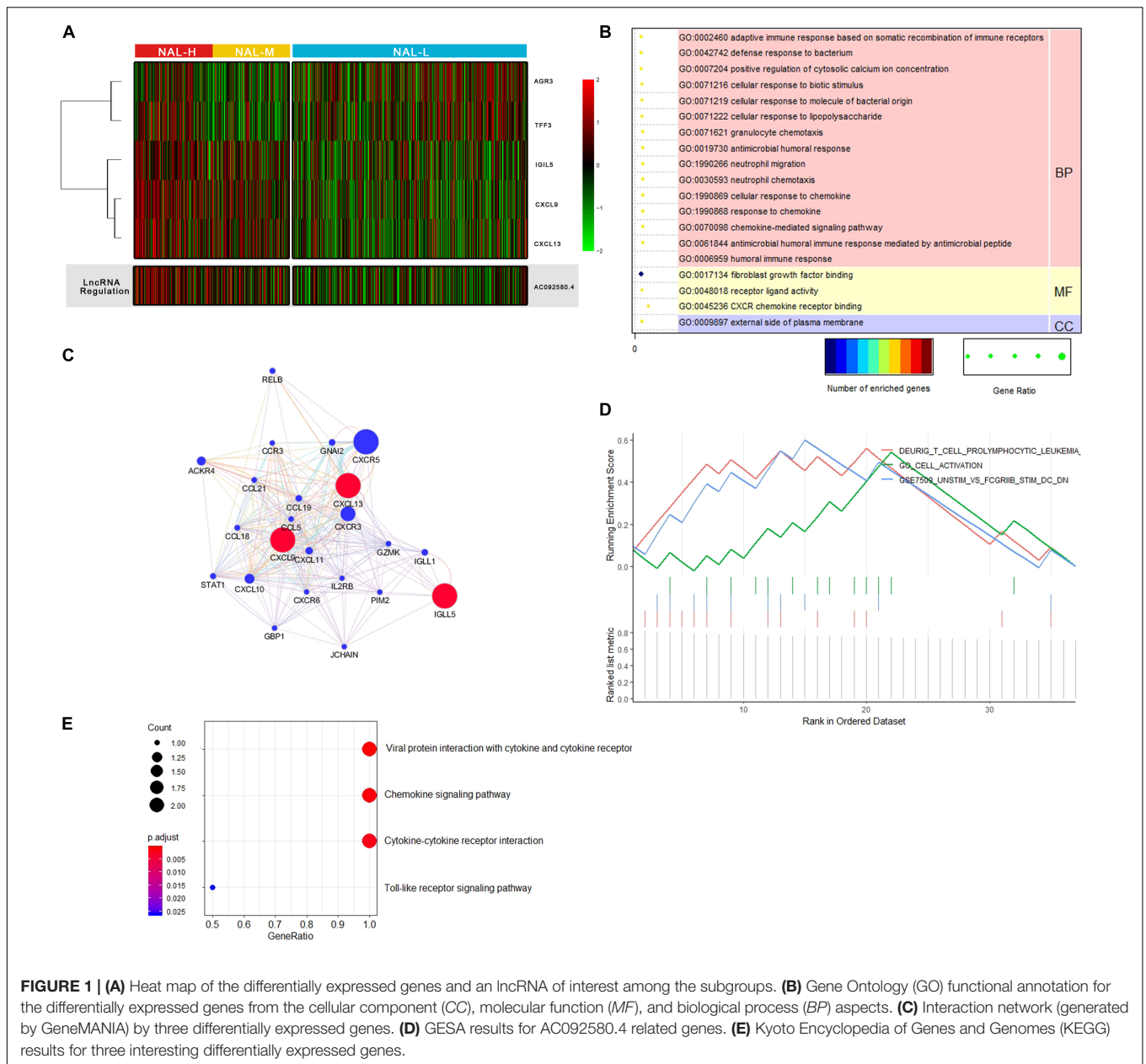
Overview of Sample Selection and Subgroup Identification

Among all 2,199 Pan-Gyn tumor samples, 931 significant immune-related samples were diagnosed ($p < 0.05$) by the CIBERSORT algorithm. Based on the mutation information of the neoantigens in the MAF file, 812 samples were selected, including 441 BRCA, 71 CESC, 96 OV, 180 UCEC, and 24 UCS (Supplementary Figure S2). The 812 Pan-Gyn samples with full survival and clinic pathological information were retained for downstream analysis.

Four thousand one hundred and sixty-five potential neoantigens from the TSNADB gene sites were selected by integrating the information obtained above. According to the MAF file, the mutation information on the 4,118 neoantigen gene sites were kept finally. We used the R package “maftools” to describe the landscape of the filtered MAF files and compute the NAL (Supplementary Figures S1, S2). Selecting 60 and 80% as the NAL cutoffs, 163 samples as the NAL-H group, 161 samples as the NAL-M group, and 488 samples as the NAL-L group were identified.

Different Functional Pathways Among the Neoantigen Load Subgroups

We firstly used “combat” to remove the batch effect in the expression profiles (Supplementary Figure S2). Then differential expression analysis identified five significantly dysregulated genes (CXCL9, CXCL13, IGLL5, AGR3, and TFF3) with a threshold of FDR < 0.05 and absolute log₂(fold change) > 1 (Supplementary Table S1). Among them, three genes (CXCL9, CXCL13, and IGLL5) were all upregulated for the NAL-H and NAL-M subgroups compared with the NAL-L subgroup (Figure 1) and were involved in immunological processes such as T cell trafficking, B lymphocyte migration, and antigen binding. Gene Ontology annotation for these genes indicated enrichment of the immune-related terms (Figure 1 and Supplementary Table S2), such as humoral immune response (FDR = 0.001), chemokine-mediated signaling pathway (FDR = 0.002), and adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains (FDR = 0.008). The Kyoto



Encyclopedia of Genes and Genomes (KEGG) pathway analysis denoted that these genes were involved in cytokine–cytokine receptor interaction, Toll-like receptor signaling pathway, and leukocyte transendothelial migration (Figure 1). The genetic interaction network of the three differentially expressed genes was generated by GeneMANIA. Different line and node colors in Figure 1 represent different types of interactions and different immune-related functions.

We compared the correlation between the lncRNA genes and the three genes above. Interestingly, we found a lncRNA named AC092580.4, which had strong correlations with CXCL9 and CXCL13 and was significantly upregulated in the NAL-H and NAL-M subgroups (Figure 1). To better understand its function, we selected 37 genes that were highly related with AC092580.4

as a gene set (Supplementary Table S3) and used GSEA to analyze this pre-ranked gene list (Figure 1). Nineteen terms were enriched under this lncRNA (Supplementary Table S4), such as immune-related cells [e.g., CD8 T cells, CD4 T, and cells and natural killer (NK) cells] and pathways (e.g., lymphocyte activation, leukocyte activation, cell death, and positive regulation of immune system process).

Somatic Mutation Landscape of the Neoantigen Load Subgroups

Under a stringent threshold of $q < 0.05$, MutSigCV identified 30 significantly mutated genes (SMGs) among all Pan-Gyn tumor samples, including 18 neoantigen gene sites of greater interest

(Figure 2 and Supplementary Table S5). All these mutated genes differed in the frequency of somatic mutations among the NAL subgroups (FDR < 0.05). For the top 5 most frequently mutated genes [TP53 (48% of the samples mutated), PIK3CA (33%), PTEN (22%), ARID1A (15%), and PIK3R1 (12%)], we described their distributions in the NAL subgroups (Figure 2). These five genes have been reported in previous papers (Berger et al., 2018). However, among all the TCGA gynecologic landmark papers, there were no previous reports of ACVR2A (Supplementary Table S5), one of the 18 significantly different neoantigen mutations (Berger et al., 2018). ACVR2A is a member of the transforming growth factor beta superfamily that plays a role in pathways associated with tumor progression and suppression (Ikushima and Miyazono, 2010).

Here, we also evaluated 30 mutation signatures to better understand the complex mutational processes. Mutational signatures provide insight into the mechanisms of tumor development and contribute to patient treatment decisions (Helleday et al., 2014). Among them, five significantly different signatures were obtained, including signature 1, signature 3, signature 6, signature 13, and signature 30 (FDR < 0.05) (Figure 2). NAL-H was enriched in signature 6, which is related to a DNA mismatch repair defect that suggests sensitivity to checkpoint inhibitors (Berger et al., 2018). NAL-M was enriched in signatures 3, 13, and 30. Signature 3 is closely associated with germline and somatic BRCA1 and BRCA2 mutations in breast, pancreatic, and ovarian cancers. Signature 13 indicates AID/APOBEC family activity of cytidine deaminase. Signature

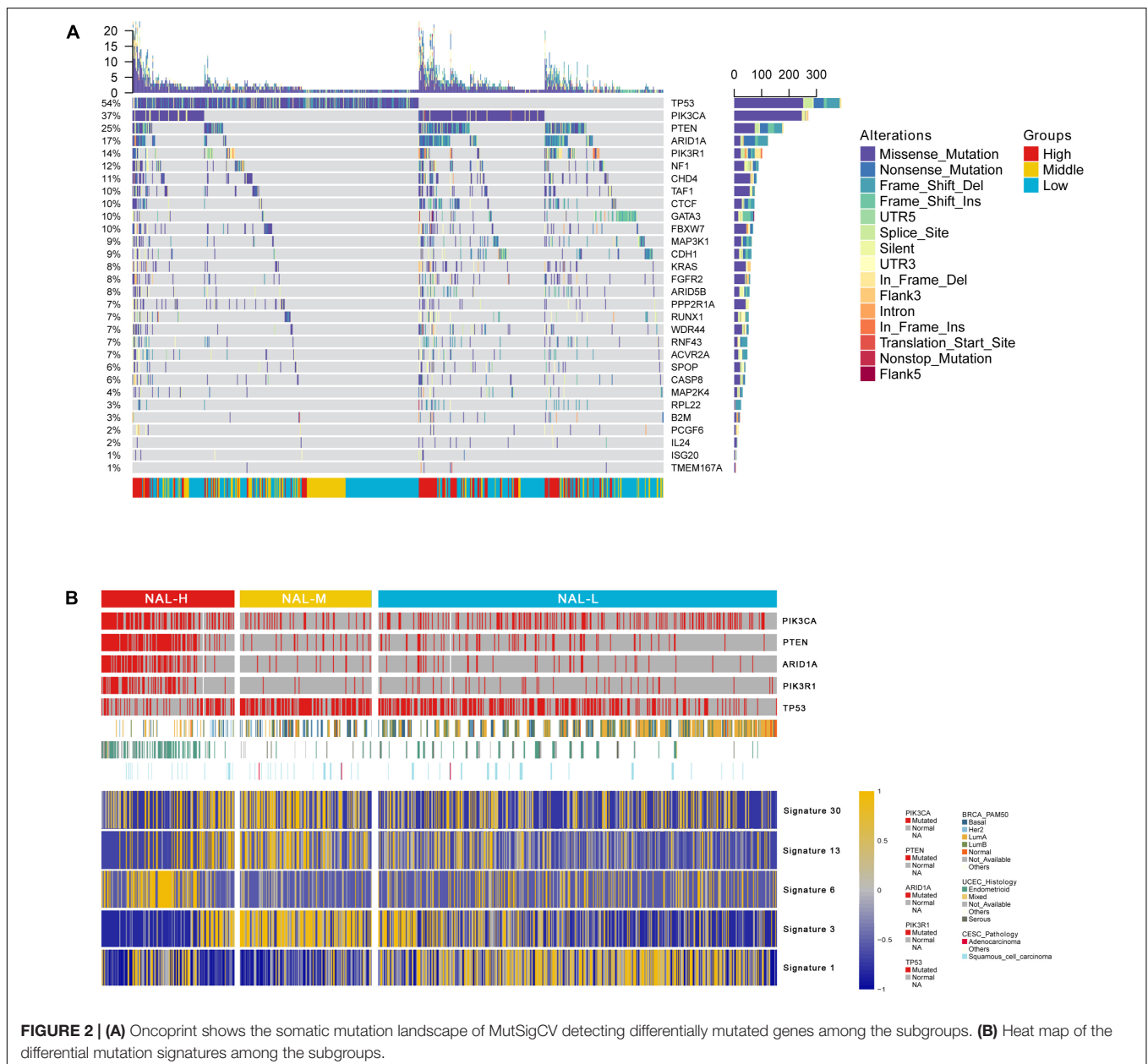


FIGURE 2 | (A) Oncoprint shows the somatic mutation landscape of MutSigCV detecting differentially mutated genes among the subgroups. **(B)** Heat map of the differential mutation signatures among the subgroups.

30 has been observed in a small subset of breast cancers. NAL-L was enriched in signature 1, indicating that the endogenous process was initiated by the spontaneous demethylation of 5-methylcytosine.

Association Between Neoantigen Load Subgroups and the Clinical Outcomes

As expected, NAL-H showed the best survival compared with NAL-M and NAL-L ($p = 0.048$, HR = 0.578, 95% CI = 0.361–0.924) (Figure 3). We then compared the differences in the clinical covariates among the identified subgroups (Table 1).

Based on the clinical information of patients, we examined multiple variables including age, gender, tumor stage, tumor type, clinical stage, histological grade, and menopause status. For the age of patients belonging to the continuous variable, all samples were divided into two groups with a cutoff point equal to 55. For other categorical variables, we used their classification information for group comparison. Four significantly different variables were obtained, including age, tumor type, clinical stage, and histological grade. We then converted them into latent variables and performed univariate Cox regression to determine whether these parameters affected patient outcome (Figure 3). Older patients had an increased risk of poor prognosis

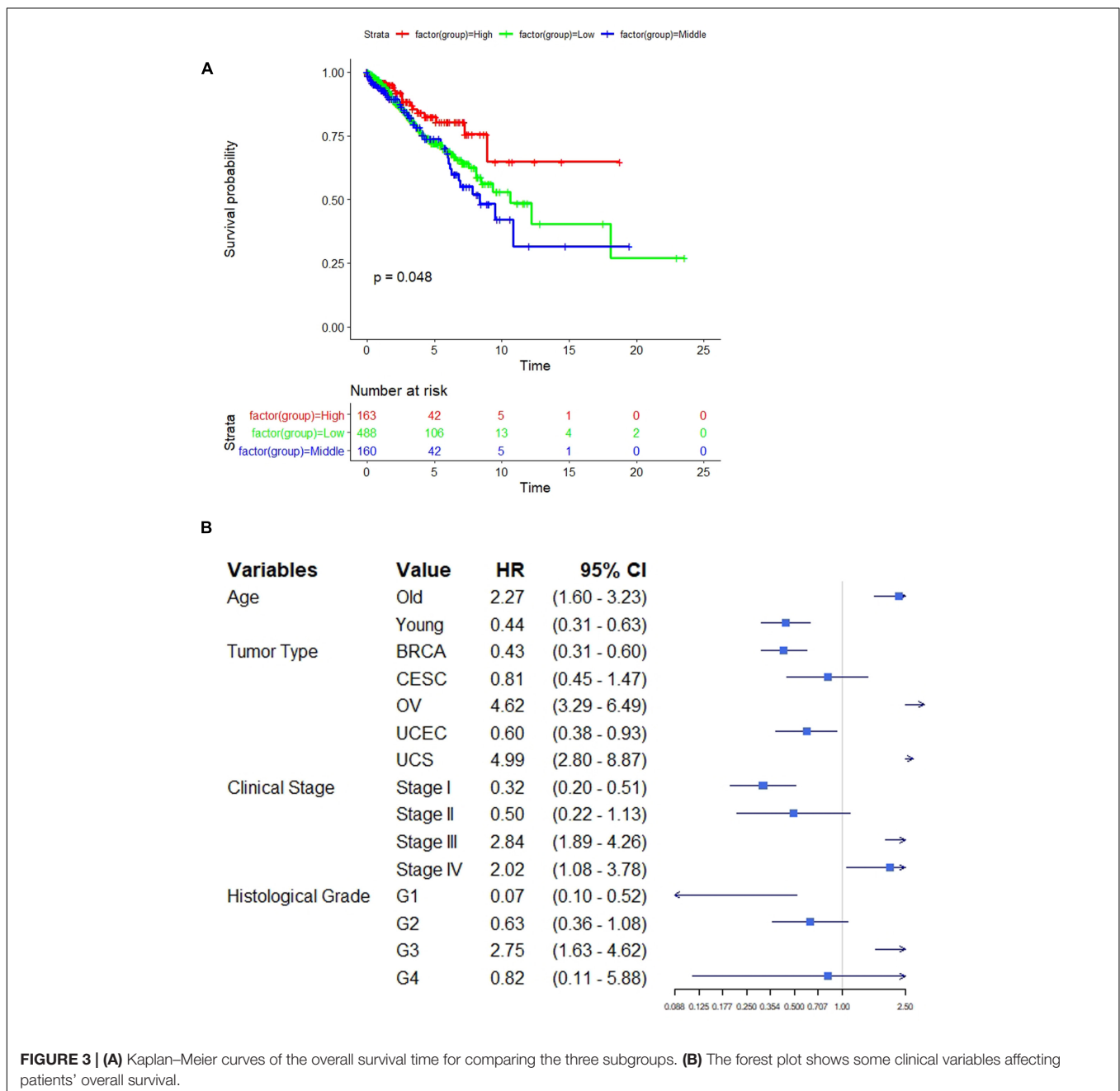


FIGURE 3 | (A) Kaplan–Meier curves of the overall survival time for comparing the three subgroups. **(B)** The forest plot shows some clinical variables affecting patients' overall survival.

TABLE 1 | Comparison of cohorts according to clinical parameters.

Index	Group	Subgroups			p value
		NAL-H	NAL-M	NAL-L	
Age (years)	≤55	103	106	271	0.0276*
	>55	58	55	217	
Gender	Female	163	161	484	0.4797
	Male	0	0	4	
Tumor stage	Stage I	5	10	56	0.7686
	Stage II	28	52	191	
	Stage III	7	14	61	
	Stage IV	0	1	11	
Tumor type	BRCA	40	78	323	0.0005***
	CESC	17	20	34	
	OV	13	45	38	
	UCEC	91	16	73	
	UCS	2	2	20	
Clinical stage	Stage I	74	20	79	1.42e-105***
	Stage II	11	11	12	
	Stage III	32	47	57	
	Stage IV	5	4	15	
Histological grade	G1	18	3	18	0.0387*
	G2	28	15	37	
	G3	71	62	88	
	G4	4	1	1	
Menopause status	Premenopausal	24	24	87	0.5402
	Postmenopausal	105	79	295	
	Indeterminate	10	4	32	

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

(HR = 2.27). BRCA, CESC, and UCEC patients had better prognosis (HR = 0.43, 0.81, and 0.60, respectively). Early tumor stage and a lower histological grade could favor the prognosis of patients (stage I: HR = 0.32; stage II: HR = 0.50; G1: HR = 0.07; G2: HR = 0.63).

Immune Infiltration Differences Associated With the Neoantigen Load Subgroups

We evaluated several immune-related signatures to gain further immunologic insights on these subgroups by single-sample GSEA. Interestingly, there were significant differences in immune enrichment among the three subgroups. As shown, NAL-H and NAL-M had significantly higher degrees of adaptive immune infiltration in T cells, B cells, and cytotoxic lymphocytes, while NAL-L enriched with innate immune infiltration in eosinophils, NK cells, mast cells, and interdigitating cells (iDC), and the differences within these subgroups need to be further studied (Figure 4). Further research is needed on the causes of the different immune infiltrations among subgroups.

Specifically, we noticed that NAL-H had significantly lower ITH, which corresponded with early findings that a low ITH was an important determinant of good response to checkpoint therapy (Wolf et al., 2019). Intra-tumor heterogeneity is independent of tumor mutation burden and can affect

tumor invasiveness and immunity. A high-ITH tumor has a strong immunosuppressive tumor microenvironment. Moreover, NAL-H was significantly enriched in T helper cells. A recent paper reported that the activity of tumor antigen-specific CD8⁺ and CD4⁺ T cells could induce antitumor response in immunotherapy. The immune response of T helper cells is that CD4⁺ T cells can recognize MHC II antigen, which plays an important role in anticancer activity (Alspach et al., 2019).

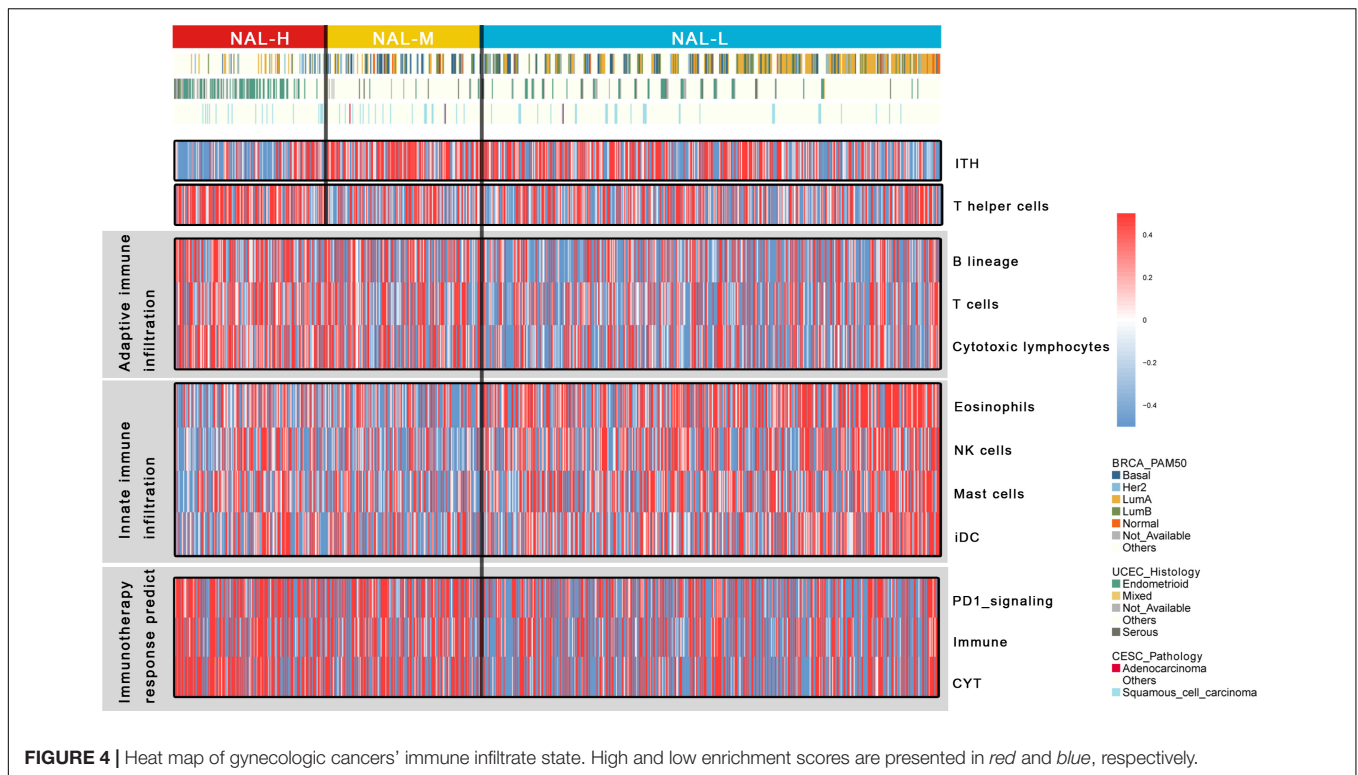
Differential Response With Chemotherapy and Immunotherapy in Subgroups

To evaluate the traditional chemotherapeutic response of the NAL subgroups, we trained a predictive model using the GDSC cell line dataset by ridge regression and assessed its satisfactory prediction accuracy by 10-fold cross-validation. We estimated the IC₅₀ for each sample in the Pan-Gyn cancers based on a predictive model of six chemical drugs (Supplementary Figure S3). We identified significant differences in the estimated IC₅₀ for these subgroups of all these chemotherapeutic drugs (cisplatin: $p = 0.0005$; paclitaxel: $p = 0.0057$; docetaxel: $p = 0.0001$; etoposide: $p = 0.0076$; vinorelbine: $p = 0.0032$; and gemcitabine: $p = 0.0010$). NAL-H and NAL-M could be more sensitive to all six drugs than NAL-L. Specifically, NAL-M had a significantly sensitive response to docetaxel, etoposide, and paclitaxel. NAL-M was composed of most basal-like breast carcinomas (BLBC). This finding was consistent with the report that BLBC has a relative sensitivity to chemotherapy, which may provide opportunities for optimizing treatment (Bertucci et al., 2012).

Considering that immunotherapy has revolutionized the treatment of patients with cancers, we used some immune signatures to evaluate the ICI treatment response. Early studies have shown that a high immune cytolytic activity (CYT) is significantly associated with significant pan-cancer survival benefits (Rooney et al., 2015) and effectively corresponds to anti-CTLA-4 and anti-PD-L1 immunotherapy (Ji et al., 2012; Herbst et al., 2014). We found significant high CYT in NAL-H and NAL-M, corresponding to the upregulation of PD-1 signaling signature genes (Figure 4). In conclusion, the presence of CYT, high immune infiltration, and PD-1 signaling may suggest that NAL-H and NAL-M have good response to immunotherapy, especially ICI treatment.

Further Exploration of the Difference Between NAL-H and NAL-M

Based on the above analysis, NAL-H and NAL-M were similar in many aspects, including differentially expressed genes, levels of immune infiltration, response to treatment, etc. In order to further understand the differences between these two subgroups, differential expression analysis identified 74 significantly dysregulated genes with a threshold of FDR < 0.05 and absolute log₂(fold change) > 0.5, including 34 overexpressed and 40 down-expressed genes (Figure 5). The GSEA for these genes indicated enrichment of 15 terms, and three of them caught our interest, including ESR1 targets down in NAL-H (FDR = 0.0078), ESR1 up in NAL-H (FDR = 0.0078),



and martens tretinoin response up in NAL-M (FDR = 0.014) (**Supplementary Table S6**). Gene Ontology annotations of ESR1 included DNA-binding transcription factor activity and identical protein binding. A volcano plot was used to display the fold difference of these genes between the subgroups (**Figure 5**).

We also compared the correlation between the significantly different DNA methylation probes and the 74 genes above. Specifically, we found an interesting methylation probe named cg17240454, which had strong negative correlation with SPDEF (**Figure 5**).

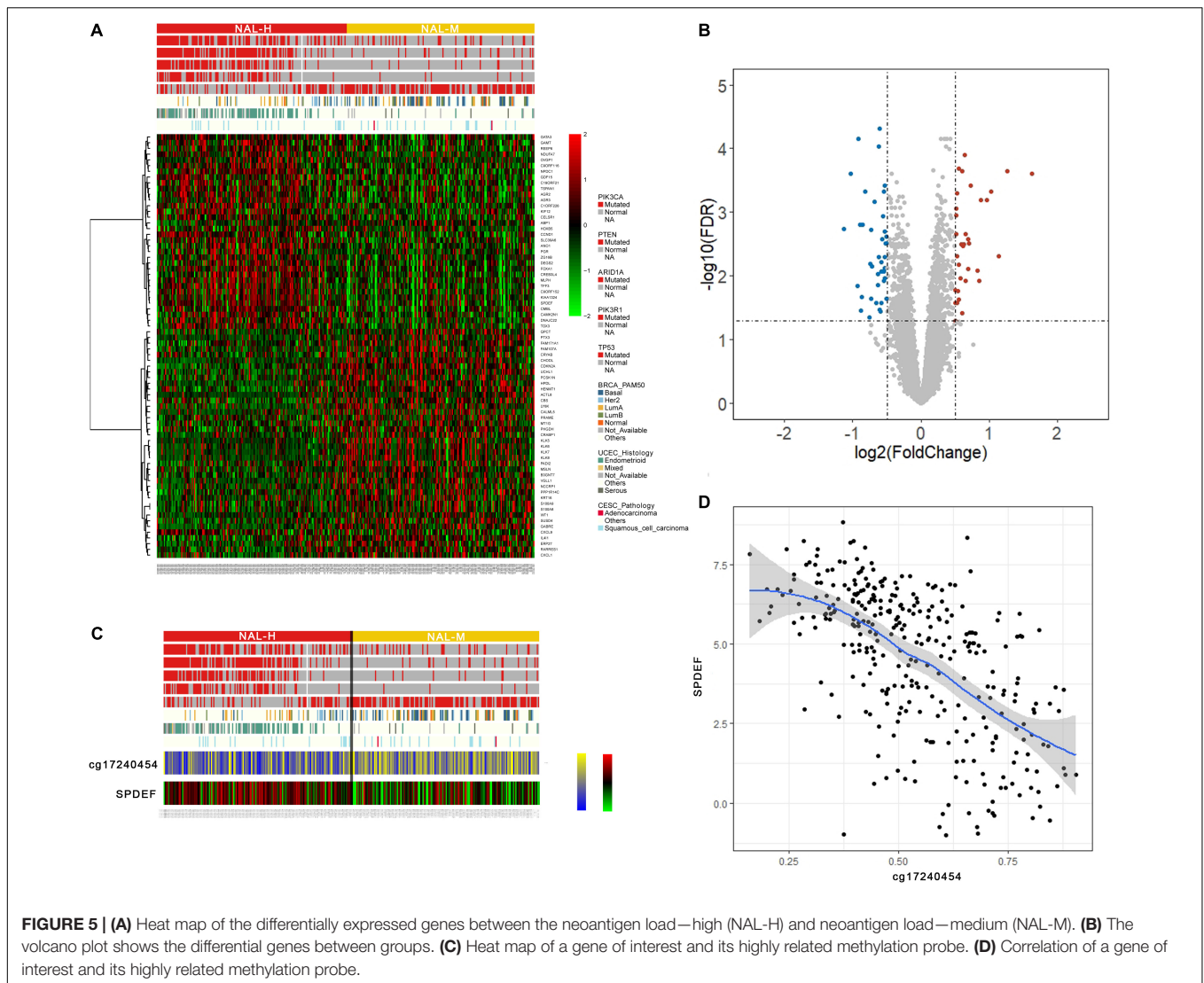
DISCUSSION

Good predictive biomarkers are needed to predict survival after immunotherapy in Pan-Gyn cancers. In this study, to investigate the ability of NAL as a predictive biomarker, we performed an integrative, multi-omics data analysis of the TCGA Pan-Gyn cancers based on 812 samples. We studied the relationship between the NAL subgroups and overall survival, the characterization of the NAL subgroups with multiple omics data, the estimation of the immune infiltrate status from several immune signatures, and the ability to predict the outcome of immunotherapy and chemotherapy. As NAL provides important information for clinical immunotherapy selection, we attempted to explain the biological mechanism behind it.

We selected 80 and 60% as the NAL cutoff points, which divided the 812 samples into three groups: NAL-H, NAL-M, and NAL-L. Three significantly overexpressed genes (CXCL9, CXCL13, and IGLL5) in NAL-H and NAL-M were found, and

the GO annotation and KEGG analysis indicated that these genes were enriched in immune-related terms and pathways. The TNF- α network activates macrophages and B cells, leading to strong upregulation of the gene expressions of several chemokines, especially CXCL13 (Shu et al., 2012). The IFN- γ network causes CXCL9 to be upregulated in different cells, including macrophages (Shu et al., 2012). Lymphocyte chemoattractants (CXCL9 and CXCL13) as immune-activating factors revealed a high immune state of NAL-H and NAL-M, and they were highly significant predictors of therapy response. Interestingly, we found an lncRNA named AC092580.4, which was significantly correlated with CXCL9 and CXCL13 genes and was significantly upregulated in NAL-H and NAL-M. This lncRNA may promote the expressions of CXCL9 and CXCL13. We further studied the function of AC092580.4 and found that it had an upward regulation effect on some immune-related cells and pathways.

We used several immune signatures to compare the differences of the immune infiltrate status among subgroups. Similarly, we noticed that NAL-H and NAL-M were significantly enriched in higher degree of adaptive immune infiltration, including T cells, B cells, and cytotoxic lymphocytes, while NAL-L was enriched in innate immune infiltration, including eosinophils, NK cells, mast cells, and iDC. To better understand the response of treatment, we evaluated subgroup responses to six traditional chemotherapy drugs. NAL-H and NAL-M could be more sensitive to all six drugs than NAL-L, and the higher enrichment of CYT, immune infiltration, and PD-1 signaling might suggest that NAL-H and NAL-M had good response to immunotherapy. Both NAL-H and NAL-M appeared to respond well to chemotherapy and immunotherapy. Based on



the above analysis, it seems that the 60% cutoff point of NAL can effectively distinguish the different immune status among Pan-Gyn cancers.

However, NAL-H had a significant better prognosis than the other two subgroups, and we found that NAL-H was related with significantly lower ITH and enriched in T helper cells. A recent paper reported that a high-ITH tumor had a strong immunosuppressive tumor microenvironment (Wolf et al., 2019). The relative expressions of the neoantigens in the high-ITH tumor were reduced, reducing the homing ability of TILs to their target cells and sufficient cytotoxic reactions. In our NAL-H subgroups, more neoantigens are exposed to the tumor microenvironment, which enhances the immune system's antitumor ability. Meanwhile, with the higher T helper cells, it can further enhance the antitumor ability in NAL-H because CD4⁺ T cells can recognize MHC II antigen and work together with CD8⁺ T cells. These may be the reasons why NAL-H patients had better survival than did NAL-M patients.

We identified 30 SMGs among the identified subgroups, including 18 neoantigen gene sites of greater interest. The top five most common SMGs were TP53, PIK3CA, PTEN, ARID1A, and PIK3R1. Driver mutant genes (e.g., TP53, KRAS, and PIK3CA) may interfere with genomic stability and may affect the immune status by generating neoantigens (Markowitz and Bertagnoli, 2009). We also used 30 mutation signatures to better understand the complex mutational processes. Interestingly, we found that NAL-H was enriched in signature 6, which was more common in uterine tumors and related to DNA mismatch repair defects. Some studies have proven that mutant signature 6 may indicate that samples are sensitive to the immune inhibitors (Berger et al., 2018). NAL-H may be better responded to by ICI therapy.

To further explore the differences between NAL-H and NAL-M, we integrated gene expression and methylation data. Martens tretinoin response was up in NAL-M. Under physiological concentrations of tretinoin, PML-RAR α has been reported to bind with RXR, which may be crucial to its carcinogenic potential (Martens et al., 2010). NAL-M may have more

complex carcinogenic mechanisms. The ESR1 gene encodes an estrogen receptor, whose expression status distinguishes ER-positive from ER-negative tumors. A recent work showed differential methylation between these two distinct diseases (Dedeurwaerder et al., 2011). We found that a key CpG site, whose Illumina ID is cg17240454, was negatively related with the gene SPDEF. SPDEF is a member of the ETS family, whose high expression can promote the migration and invasion of various cells (Turner et al., 2007). The downregulation of SPDEF in NAL-M may stimulate the migration of cancer cells.

Our analyses were limited by the data from TCGA because pathologists excluded tumor samples with less than 60% tumor cell nucleus from the study and some important immune infiltrate samples might be lost (Berger et al., 2018). We also lacked experimental verification on the targeted classical cellular immunoassays for confirming cell phenotypic distribution. We hope that we could further understand the features of various NAL subgroups across Pan-Gyn cancers.

In summary, this comprehensive analysis of Pan-Gyn cancers revealed that 80% may be a good cutoff point for NAL across multiple tumor types. Among all samples, a higher NAL (highest 20% in each tumor) was associated with better overall survival. The obtained NAL subgroups were characterized by molecular features, immune signatures, and clinical outcome. As we increasingly recognize the impact of NAL on disease progression and treatment response, this biomarker may play an important role in predicting disease outcomes.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found in The Cancer Genome Atlas Project (TCGA) pan-cancer analyses data portal (<https://portal.gdc.cancer.gov/about-data/publications/pancanatlas>).

REFERENCES

- Alspach, E., Lussier, D. M., Miceli, A. P., Kizhvatov, I., DuPage, M., Luoma, A. M., et al. (2019). MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* 12, 1–6.
- Berger, A. C., Korkut, A., Kanchi, R. S., Hegde, A. M., Lenoir, W., Liu, W., et al. (2018). A comprehensive pan-cancer molecular study of gynecologic and breast cancers. *Cancer Cell* 33, 690–705.
- Bertucci, F., Finetti, P., and Birnbaum, D. (2012). Basal breast cancer: a complex and deadly molecular subtype. *Curr. Mol. Med.* 12, 1–10.
- Bregar, A., Deshpande, A., Grange, C., Zi, T., Stall, J., Hirsch, H., et al. (2017). Characterization of immune regulatory molecules B7-H4 and PD-L1 in low and high grade endometrial tumors. *Gynecol. Oncol.* 145, 446–452. doi: 10.1016/j.ygyno.2017.03.006
- Callahan, M., Postow, M., and Wolchok, J. (2016). Targeting T cell co-receptors for cancer therapy. *Immunity* 44, 1069–1078. doi: 10.1016/j.immuni.2016.04.023
- Carbone, D. P., Reck, M., Paz-Ares, L., Creelan, B., Horn, L., Steins, M., et al. (2017). First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *New Engl. J. Med.* 376, 2415–2426.
- Dedeurwaerder, S., Desmedt, C., Calonne, E., Singhal, S. K., Haihe-Kains, B., Defrance, M., et al. (2011). DNA methylation profiling reveals a predominant

AUTHOR CONTRIBUTIONS

FY, FW, and YZ contributed to the conception and design. YZ, XL, and XR developed the methodology. YZ and XL analyzed and interpreted the data. YZ and XR wrote, reviewed, and/or revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the “Double First-Class” University project (CPU2018GY09) and the National Natural Science Foundation of China (81973145).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2020.00702/full#supplementary-material>

FIGURE S1 | Overview of the MAF file.

FIGURE S2 | Distribution of Pan-Gyn cancers and neoantigen load.

FIGURE S3 | Response to chemotherapy among identified subgroups.

TABLE S1 | Five statistically significantly dysregulate genes among identified subgroups.

TABLE S2 | Gene Ontology annotation results of significantly dysregulate genes among identified subgroups.

TABLE S3 | Thirty-seven genes highly related with AC092580.4.

TABLE S4 | GSEA results for 37 genes highly related with AC092580.4.

TABLE S5 | Thirty statistically significantly mutated genes among identified subgroups.

TABLE S6 | GSEA results for differently expression genes between NAL-H and NAL-M.

- immune component in breast cancers. *EMBO Mol. Med.* 3, 726–741. doi: 10.1002/emmm.201100801
- Galon, J., Fridman, W.-H., and Pagès, F. (2007). The adaptive immunologic microenvironment in colorectal cancer: a novel perspective. *Cancer Res.* 67, 1883–1886. doi: 10.1158/0008-5472.can-06-4806
- Geeleher, P., Cox, N. J., and Huang, R. S. (2014). Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* 15, R47–R53.
- Gibney, G. T., Weiner, L. M., and Atkins, M. B. (2016). Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol.* 17, e542–e551. doi: 10.1016/s1470-2045(16)30406-5
- Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* 15, 585–590.
- Herbst, R. S., Soria, J.-C., Kowanetz, M., Fine, G. D., Hamid, O., Gordon, M. S., et al. (2014). Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* 515, 563–568.
- Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., and Drill, E. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 173, 291–304.

- Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., et al. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61, 1–13. doi: 10.1007/s00251-008-0341-z
- Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Hu-Lieskovan, S., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 165, 35–44. doi: 10.1016/j.cell.2016.02.065
- Ikushima, H., and Miyazono, K. (2010). TGF β signalling: a complex web in cancer progression. *Nat. Rev. Cancer* 10, 415–419.
- Ji, R.-R., Chasalow, S. D., Wang, L., Hamid, O., Schmidt, H., Cogswell, J., et al. (2012). An immune-active tumor microenvironment favors clinical response to ipilimumab. *Cancer Immunol. Immunother.* 61, 1019–1031. doi: 10.1007/s00262-011-1172-6
- Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Carter, S. L., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–219.
- Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., et al. (2015). PD-1 blockade in tumors with mismatch-repair deficiency. *New Engl. J. Med.* 372, 2509–2520.
- Markowitz, S. D., and Bertagnolli, M. M. (2009). Molecular basis of colorectal cancer. *New Engl. J. Med.* 361, 2449–2460.
- Martens, J. H., Brinkman, A. B., Simmer, F., Francoijs, K. J., Nebbioso, A., Ferrara, F., et al. (2010). PML-RAR α /RXR alters the epigenetic landscape in acute promyelocytic leukemia. *Cancer Cell* 17, 173–185. doi: 10.1016/j.ccr.2009.12.042
- Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28, 1747–1756. doi: 10.1101/gr.239244.118
- Mullen, R. D., and Behringer, R. R. (2014). Molecular genetics of Müllerian duct formation, regression and differentiation. *Sex. Dev.* 8, 281–296. doi: 10.1159/000364935
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–456.
- Nishikawa, H., and Sakaguchi, S. (2014). Regulatory T cells in cancer immunotherapy. *Curr. Opin. Immunol.* 27, 1–7.
- Powles, T., Durán, I., Van Der Heijden, M. S., Loriot, Y., Vogelzang, N. J., De Giorgi, U., et al. (2018). Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *Lancet* 391, 748–757. doi: 10.1016/s0140-6736(17)33297-x
- Rizvi, N. A., Hellmann, M. D., Snyder, A., Snyder, A., Kvistborg, P., Makarov, V., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124–128.
- Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61. doi: 10.1016/j.cell.2014.12.033
- Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* 51, 202–206.
- Shu, D., Kurtis, J. D., Sunthorn, P. T., Edward, K., Duffy, P. E., and Michal, F. (2012). CXC ligand 9 response to malaria during pregnancy is associated with low-birth-weight deliveries. *Infect. Immun.* 80, 3034–3038. doi: 10.1128/iai.00220-12
- Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. *New Engl. J. Med.* 371, 2189–2199. doi: 10.1056/nejmoa1406498
- Thomas, A., Routh, E. D., Pullikuth, A., Jin, G., Su, J., Chou, J. W., et al. (2018). Tumor mutational burden is a determinant of immune-mediated survival in breast cancer. *Oncoimmunology* 7, e1490854–e1490857.
- Tumeh, P. C., Harview, C. L., Yearley, J. H., Shintaku, I. P., Taylor, E. J. M., Robert, L., et al. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* 515, 568–570.
- Turner, D. P., Omar, M., Moira, S., Fisher, P. B., and Watson, D. K. (2007). Prostate-derived ETS factor is a mediator of metastatic potential through the inhibition of migration and invasion in breast cancer. *Cancer Res.* 67, 1618–1625. doi: 10.1158/0008-5472.can-06-2913
- Wolf, Y., Bartok, O., Patkar, S., Eli, G. B., Cohen, S., Litchfield, K., et al. (2019). UVB-induced tumor heterogeneity diminishes immune response in melanoma. *Cell* 179, 219–235.
- Wu, J., Zhao, W., Zhou, B., Su, Z., Gu, X., Zhou, Z., et al. (2018). TSNAdb: a database for tumor-specific neoantigens from immunogenomics data analysis. *Genomics Proteomics Bioinformatics* 16, 276–282. doi: 10.1016/j.gpb.2018.06.003
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic* 16, 284–287. doi: 10.1089/omi.2011.0118

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhu, Meng, Ruan, Lu, Yan and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.