



OPEN ACCESS

EDITED BY

Claudio Sette,
Catholic University of the Sacred
Heart, Rome, Italy

REVIEWED BY

Camelia Quek,
Melanoma Institute Australia, Australia
Wenbin Li,
Chinese Academy of Medical
Sciences, China
Timo Gaiser,
University of Heidelberg, Germany

*CORRESPONDENCE

Yueqiong Lao
laoyq0804@163.com

SPECIALTY SECTION

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

RECEIVED 11 May 2022

ACCEPTED 24 October 2022

PUBLISHED 11 November 2022

CITATION

Lao Y, Wang Y, Yang J, Liu T, Ma Y,
Luo Y, Sun Y, Li K, Zhao X, Niu X, Xi Y
and Zhong C (2022) Characterization
of genomic alterations and
neoantigens and analysis of immune
infiltration identified therapeutic and
prognostic biomarkers in
adenocarcinoma at the
gastroesophageal junction.
Front. Oncol. 12:941868.
doi: 10.3389/fonc.2022.941868

COPYRIGHT

© 2022 Lao, Wang, Yang, Liu, Ma, Luo,
Sun, Li, Zhao, Niu, Xi and Zhong. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Characterization of genomic alterations and neoantigens and analysis of immune infiltration identified therapeutic and prognostic biomarkers in adenocarcinoma at the gastroesophageal junction

Yueqiong Lao*, Yuqian Wang, Jie Yang, Tianyuan Liu,
Yuling Ma, Yingying Luo, Yanxia Sun, Kai Li, Xuan Zhao,
Xiangjie Niu, Yiyi Xi and Ce Zhong

Department of Etiology and Carcinogenesis, National Cancer Center/National Clinical Research Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Objectives: Adenocarcinoma at the gastroesophageal junction (ACGEJ) refers to a malignant tumor that occurs at the esophagogastric junction. Despite some progress in targeted therapies for HER2, FGFR2, EGFR, MET, Claudin 18.2 and immune checkpoints in ACGEJ tumors, the 5-year survival rate of patients remains poor. Thus, it is urgent to explore genomic alterations and neoantigen characteristics of tumors and identify CD8+ T-cell infiltration-associated genes to find potential therapeutic targets and develop a risk model to predict ACGEJ patients' overall survival (OS).

Methods: Whole-exome sequencing (WES) was performed on 55 paired samples from Chinese ACGEJ patients. Somatic mutations and copy number variations were detected by Strelka2 and FACETS, respectively. SigProfiler and SciClone were employed to decipher the mutation signature and clonal structure of each sample, respectively. Neoantigens were predicted using the MuPeXI pipeline. RNA sequencing (RNA-seq) data of ACGEJ samples from our previous studies and The Cancer Genome Atlas (TCGA) were used to identify genes significantly associated with CD8+ T-cell infiltration by weighted gene coexpression network analysis (WGCNA). To construct a risk model, we conducted LASSO and univariate and multivariate Cox regression analyses.

Results: Recurrent *MAP2K7*, *RNF43* and *RHOA* mutations were found in ACGEJ tumors. The COSMIC signature SBS17 was associated with ACGEJ progression. *CCNE1* and *VEGFA* were identified as putative CNV driver genes. *PI3KCA* and *TP53* mutations conferred selective advantages to cancer cells. The Chinese ACGEJ patient neoantigen landscape was revealed for the first time, and 58

potential neoantigens common to TSNAdb and IEDB were identified. Compared with Siewert type II samples, Siewert type III samples had significant enrichment of the SBS17 signature, a lower *TNFRSF14* copy number, a higher proportion of samples with complex clonal architecture and a higher neoantigen load. We identified 10 important CD8+ T-cell infiltration-related Hub genes (*CCL5*, *CD2*, *CST7*, *GVINP1*, *GZMK*, *IL2RB*, *IKZF3*, *PLA2G2D*, *P2RY10* and *ZAP70*) as potential therapeutic targets from the RNA-seq data. Seven CD8+ T-cell infiltration-related genes (*ADAM28*, *ASPH*, *CAMK2N1*, *F2R*, *STAP1*, *TP53INP2*, *ZC3H3*) were selected to construct a prognostic model. Patients classified as high risk based on this model had significantly worse OS than low-risk patients, which was replicated in the TCGA-ACGEJ cohort.

Conclusions: This study provides new neoantigen-based immunotherapeutic targets for ACGEJ treatment and effective disease prognosis biomarkers.

KEYWORDS

adenocarcinoma at the gastroesophageal junction, genome and transcriptome, tumor neoantigens, CD8+ T cells, therapeutic biomarkers, prognostic prediction

Introduction

There were over 1.6 million new cases of esophagogastric (esophageal, gastroesophageal junction or gastric) cancer and 1.3 million related deaths worldwide in 2020, with esophagogastric cancer ranking third in terms of incidence and second in terms of mortality (1). The incidence of adenocarcinoma of the gastroesophageal junction (ACGEJ) has increased rapidly worldwide in the past four decades (2). According to the widely used Siewert classification, ACGEJ is divided into three subtypes based on the location of the tumor epicenter (3). Siewert type I is treated as esophageal cancer, while Siewert types II and III are considered true gastric cardia tumors and gastric tumors. The National Comprehensive Cancer Network (NCCN) guidelines recommend surgery for patients with early-stage ACGEJ, preoperative chemoradiation for patients with locally advanced ACGEJ, and trastuzumab plus chemotherapy for patients with HER2-positive metastatic ACGEJ (4, 5). Immune checkpoint inhibitor (ICI) treatments, such as nivolumab and pembrolizumab (anti-PD-1 antibodies), have shown encouraging efficacy in patients with unresectable advanced or metastatic ACGEJ (6). However, the prognosis of patients with advanced-stage gastric cancer and ACGEJ remains poor, with a median overall survival (OS) of approximately 1 year (6). Multiple factors, including tumor-specific mutant antigens (neoantigens) and the tumor immune microenvironment (TME), may affect the effectiveness of ICIs. Recently, a study from China demonstrated that Siewert type I tumors and Siewert type II/III tumors had different driver genes, mutational signatures and disrupted pathways (7). In our previous study, we performed whole

genome and transcriptome sequencing in 124 Chinese ACGEJ tumors and categorized them as tumor with chromosomal instability dominated by CNVs and identified alterations vulnerable to drugs and that could be used as prognostic biomarkers for ACGEJ patients (8). However, comprehensive molecular characteristics, especially those of tumor neoantigens, and of different Siewert types in ACGEJ remain unclear. Therefore, it is urgent to investigate the molecular characteristics of ACGEJ patients and identify tumor-specific targets for treatments and potential prognostic markers.

Neoantigens are generated by various mutation types, including nonsynonymous single nucleotide variants (SNVs), insertions or deletions (indels) and gene fusions (9, 10). Neoantigens can be loaded into MHC class I molecules and recognized by cytotoxic CD8+ T cells, which increase immune cell infiltration and enhance the efficacy of cancer immunotherapy (11, 12). Importantly, neoantigen load has been related to prolonged survival in patients with melanoma, non-small-cell lung cancer and mismatch repair-deficient colorectal cancer treated with ICIs (13–16). Neoantigens in ACGEJ can serve as excellent targets for immunotherapy. Many previous studies have revealed that neoantigen vaccine treatment is inhibited by the TME, which impedes the function of immune cells and even hinders the immune response (17–26). For instance, regulatory T (Treg) cells can suppress the induction of effective neoantigen-specific T-cell responses in tumors (17). Neoantigen-specific resident memory T (TRM) cells induced by therapeutic vaccination may provide long-term immune surveillance and prevent disease recurrence (18). In addition, neoantigen vaccines spontaneously upregulate the expression of inhibitory receptors

(such as PD-1 and TIM3) in CD4+ T cells and CD8+ T cells, indicating that functional blockade of these receptors contributes to the generation of efficient neoantigen-specific T-cell responses (19–23). Although few studies have reported the association of cancer-associated fibroblasts (CAFs) with neoantigen-targeted therapy, CAFs were identified as potential mediators of the response to immune checkpoint inhibitor (ICI) treatments (24–26). The combination of neoantigen vaccines with other therapies, such as ICIs, chemotherapy, radiotherapy and immunosuppressive factor-targeted therapies, generates a stronger antitumor response. Moreover, preclinical data demonstrated that the combination of the neoantigen vaccine VB10.NEO and NKTR-214, a T-cell-proliferation inducer, can induce clonal expansion of natural killer (NK), CD4+ T and CD8+ T cells and induce strong and durable neoantigen-specific T-cell responses (27). CD8+ T cells are central effector cells in the tumor immune microenvironment and are primarily responsible for recognizing and killing tumor cells. Several studies have investigated the relationship between intratumoral CD8+ T cells and the prognosis of gastric cancer, two of which reported that a high abundance of CD8+ T cells was associated with better OS in patients with gastric cancer (28–30). Unfortunately, we know little about the role of CD8+ T cells in ACGEJ. Exploring CD8+ T-cell infiltration-related genes can provide targets for immunotherapy that can enhance the antitumor effect of tumor neoantigen vaccines and biomarkers to predict the OS of patients in ACGEJ.

In this study, we explored the molecular characteristics of ACGEJ and compared SNVs, copy number variations (CNVs), clonal patterns, and neoantigen load between Siewert types II and III and among differentiation grades. We also searched for CD8+ T-cell infiltration-related Hub genes and conducted weighted gene coexpression network analysis (WGCNA) to identify prognostic markers for the construction of a risk prediction model.

Materials and methods

Biospecimen collection

This study included 55 Chinese ACGEJ patients seen at the Linzhou Cancer Hospital and Linzhou Esophageal Cancer Hospital (Henan Province, China) from 2014 to 2016. All patients received no treatments before surgery and signed written informed consent forms. ACGEJ tumor, tumor-adjacent (> 5 cm from the tumor margins) and normal tissues and peripheral blood samples were collected during surgery. Pathological diagnoses were independently confirmed by at least two pathologists. Clinical information of patients was collected from medical records. Clinical follow-up data were obtained by phone interview, the most recent interview was conducted in September 2021 and the median follow-up period of patients was 50.3 months. This research

was approved by the Institutional Review Board of Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College.

Whole-exome sequencing

Genomic DNA was extracted from 55 matched tumor and blood samples using the Allprep DNA Kit (Qiagen) and QiaAmp Blood Midi Kit (Qiagen). Exon target capture was performed using an Agilent SureSelect Human All ExonV5 kit (Agilent) following the manufacturer's protocol. After quality checking, the libraries were sequenced on an Illumina HiSeq xTen with 2X 150 bp paired-end reads. Average sequencing depths of 380x and 130x were achieved for paired tumor and blood samples, respectively (Supplementary Tables 1, 2). Importantly, WES data in 55 Chinese ACGEJ patients have not been published.

Data collection from database

We obtained somatic mutation data of 105 The Cancer Genome Atlas (TCGA) ACGEJ samples from the Broad Institute GDAC Firehose website (<https://gdac.broadinstitute.org/>) (31) and an additional 40 ACGEJ samples from the Tumor Portal (<http://www.tumorportal.org>) (32) (Supplementary Table 3). Transcriptomic data (FASTQ format) of the 55 Chinese ACGEJ patients described above were derived from our earlier publication (33). Combining two published studies of our group on ACGEJ tumors (8, 33), we obtained bulk RNA sequencing data of 178 ACGEJ samples, which consisted of 123 patients from one cohort (8) and 55 patients from the other (33). We conducted follow-up telephone interviews until September 2021, with 47.7% (85/178) of patients being lost. Eventually, the gene expression (transcripts per million, TPM) data of 93 Chinese ACGEJ samples in a period of 50.3 months (median) were included and also publicly available from NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo>, GSE159721). The gene expression (TPM) data of 86 TCGA ACGEJ samples (tumors located in the cardia/GEJ rather than other sites of the stomach (body of stomach, fundus of stomach, gastric antrum, pylorus, etc) and the corresponding clinicopathological information were obtained from the TCGA Pan-cancer Atlas (<https://gdc.cancer.gov/about-data/publications/pancanatlas>) (34). Moreover, we obtained the tumor-specific neoantigen data of 441 TCGA stomach adenocarcinoma (STAD) samples from TSNADB (<http://biopharm.zju.edu.cn/tsnadb>) (35). We then downloaded a list of gastric adenocarcinoma neoepitope that bound to MHC class I molecules from the Immune Epitope Database (IEDB) (36).

Somatic mutation identification

DNA sequence reads were mapped to the human reference genome GRCh37 (Ensembl) using BWA-MEM (v0.1.22) (37) with default parameters. Somatic SNVs and indels in each tumor sample were called by Strelka2 (v2.8.3) (38) and annotated by Ensembl Variant Effect Predictor (VEP, release 90) (39). Low-quality and potential germline variants were filtered by gatk-tools (v0.2.2).

Identification of mutational signatures

Based on nonnegative matrix factorization (NMF), SigProfiler was applied to identify single-base-substitution (SBS) mutational signatures (40). Six base substitutions (including C>A, C>T, C>G, T>A, T>C, T>G) and their trinucleotide sequence context were considered in this analysis. We used SigProfilerMatrixGenerator and SigProfilerExtractor to categorize the SNV mutations and then extract *de novo* mutational signatures. After matching to COSMIC signatures, we obtained 5 mutational signatures and determined the contribution of each signature in each individual sample.

Copy number alteration analysis

FACETS (v0.5.14) (41), an allele-specific copy number analysis tool, was used to estimate integer copy number calls corrected for tumor purity, ploidy and clonal heterogeneity. VCF files from FACETS were used as input for GISTIC2.0 (v.2.0.23) (42) to identify significantly amplified or deleted regions of the genome for each patient. The Seg.CN value required by GISTIC2.0 was calculated with TCN_EM values from FACETS as $\text{Seg.CN} = \log_2(\text{TCN_EM}) - 1$. Arm- and focal-level CNV regions with $\text{FDR } q < 0.25$ were considered significantly aberrant regions.

Clonality analysis of somatic mutations

We used the R package SciClone (43) to infer clonal and subclonal architectures by clustering similar variant allele frequencies of somatic mutations in a single sample using a variational Bayesian binomial mixture model. The regions of CNVs and loss of heterozygosity (LOH) inferred by FACETS and somatic mutations of sufficient depth (20x or greater coverage) called by Strelka2 from WES were used as inputs. Due to lack of enough points to cluster or lack of copy number neutral regions, two ACGEJ specimens (GEJ183 and GEJ191) were excluded; subsequently, a total of 53 samples were analysed.

Neoantigen prediction

We used WES data of blood samples or tumor-adjacent normal tissues in FASTQ format to perform human leukocyte antigen (HLA) typing. OptiType (v1.3.5) (44) was applied to predict class-I HLA typing with default settings. The raw RNA sequencing (RNA-seq) data of tumor tissues in FASTQ format were processed by Kallisto (v0.46.0) (45) to obtain the expression values (TPM) with the GRCh38 v78 coordinates. Tumor-specific somatic variant calls, RNA-seq expression values and class-I HLA genotyping were given as inputs to the MuPeXI pipeline (v1.2) (46) to predict 9 amino acid peptides. Notably, somatic mutation calls from the GRCh37 alignment were converted to GRCh38 by running liftover in MuPeXI. The binding affinities of a mutant peptide with the major histocompatibility complex class I (MHC-I) molecules of patients were predicted by the NetMHCpan (v4.0) (47) algorithm in MuPeXI. Mutant peptides with an eluted ligand percentile rank (EL% rank) score $\leq 2\%$ and RNA expression level (TPM) > 0.1 were predicted to be neoantigens. Neoantigens with a percentile rank score $< 0.5\%$ were considered high-affinity neoantigens.

Immune cell infiltration estimation

MCP-counter (48), which uses a deconvolution approach, was applied to produce the absolute abundance scores for endothelial cells and fibroblasts as well as 8 immune cell types, including CD8+ T cells. xCell, another cell type enrichment analysis method, was used to quantify the relative abundance scores for immune and stromal cells (49). The Tumor Immune Estimation Resource 2.0 (50) (TIMER2.0, <http://timer.cistrome.org/>) web server is a comprehensive resource integrating multiple immune infiltration estimation algorithms, including MCP-counter and xCell. In this study, we imported the TPM-normalized gene expression matrix of ACGEJ samples into TIMER2.0 and obtained the MCP-counter-based and xCell-based deconvolution profiles of CD8+ T cells.

Construction of the coexpression network and identification of significant modules

The WGCNA (51) package was used to construct a weighted gene coexpression network of the 6,000 most variable genes based on RNA-seq data (based on median absolute deviation) among ACGEJ samples. A total of 93 tumor samples from our previously published data were included after removing one outlier specimen with sample hierarchical clustering. The gene expression matrix was converted into a similarity matrix, and the appropriate weighting coefficient β was chosen. Subsequently, an

adjacency matrix was generated and transformed into a topological overlap matrix (TOM). Next, we performed hierarchical clustering with the dynamic tree cut method and identified modules with the following parameters: $\text{minModuleSize} = 30$ and $\text{mergeCutHeight} = 0.25$. Afterwards, we calculated the correlation between module eigengenes (MEs) and clinical traits to identify clinically significant related modules ($P < 0.05$). In general, the turquoise module was considered the key module.

Functional enrichment analysis

To investigate the biological functions of turquoise module genes, we conducted gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses by using the clusterProfiler R package (v3.14.3) (52). The thresholds were set as $P < 0.01$ and $\text{FDR} < 0.05$.

Hub gene identification

Hub genes, highly interconnected with the nodes in a given module, are generally determined based on gene significance (GS) and module membership (MM). GS is defined as the correlation between the gene and a clinical trait, while MM is described as the correlation between the gene and a module. In this study, genes with $\text{GS} > 0.6$ and $\text{MM} > 0.8$ were selected as Hub genes. We further used the CytoHubba and MCODE functions of Cytoscape software (v 3.8.2, <https://cytoscape.org/>) to select and visualize the 10 most important Hub genes in the turquoise module.

Construction and validation of a risk prediction model

A total of 179 patients with ACGEJ from two independent cohorts were included in our study. Bulk RNA-seq data and clinical information were collected. Based on the turquoise module genes, we developed a risk prediction model in the training cohort (containing 93 ACGEJ samples from our previous studies) and validated it in an external cohort (containing 86 ACGEJ samples from the TCGA database). In the training set, we conducted univariate Cox regression analysis to identify the genes that were closely associated with the OS of patients ($P < 0.05$). Subsequently, least absolute shrinkage and selection operator (LASSO) Cox regression analysis was performed to narrow down the list of candidate genes. Finally, multivariate Cox proportional hazards regression analysis was used to select independent prognostic genes

and construct the risk score model. The risk score was calculated using the following formula:

$$\text{risk score} = \sum (\text{expression mRNA}_i \times \text{coefficient}_i)$$

According to the median value of the risk score, we stratified patients into high-risk and low-risk groups and performed Kaplan–Meier analysis (with log-rank test) to compare the survival differences. Additionally, the performance of the model was evaluated by area under the curve (AUC) analysis using the “ROCR” package in R.

Statistical analyses

All statistical analyses were performed using R3.6.3. The Wilcoxon rank-sum test was utilized to compare nonnormally distributed continuous variables between two groups. Categorical data were compared using Fisher’s exact test. Spearman’s rank correlation analysis was applied to measure the correlation between two continuous variables. Based on the correlation between gene expression and patient survival, the “surv-cutpoint” function in the survminer R package was used to find the optimal cut-off point. The survival curves for the prognostic analysis were created by the Kaplan–Meier method, and log-rank tests were conducted to evaluate the significance of differences. Statistical significance was described as follows: ns, not significant; * $P < 0.05$; ** $P < 0.01$; **** $P < 0.0001$.

Results

Driver genes of ACGEJ

We performed WES of tumor and matched blood samples from 55 Chinese ACGEJ patients (Supplementary Table 4) with a median age of 64 years (range 42–80 years old); the cohort included 41 males (74.5%). The majority of the patients had TNM stage III disease (67.3%). Moreover, 23.6% of patients had poorly differentiated disease, and 45.5% had Siewert III subtype disease. All tumors were microsatellite stable (96.4%) or had microsatellite instability (3.6%). To better understand the genomic alterations in ACGEJ, we downloaded and analysed the coding-region mutation data for 145 patients with ACGEJ from TCGA and the Tumor Portal database. The two cohorts were similar to our cohort in age, sex and microsatellite status though differed in tumor stage. There was a significantly higher proportion of stage II and IV patients in the TCGA/Tumor Portal cohort than in our ACGEJ cohort (45.5% vs. 21.8%, 10.3% vs. 0%, respectively; $P < 0.05$) and a lower proportion of stage III patients in the TCGA/Tumor Portal cohort (24.1% vs. 67.3%, $P < 0.05$).

We then identified 4,018 somatic mutations (3,761 SNVs and 257 indels) in our ACGEJ specimens. The median tumor mutation burden (TMB) of our and TCGA/Tumor Portal cohorts were 1.04 (range 0.02–3.32) and 2.61 (range 0–67.94) per Mb, respectively, and the difference was significant (Wilcoxon rank-sum test, $P < 0.0001$; [Supplementary Figure 1A](#)). The TMB of Siewert type II cases was similar to that of Siewert type III cases, but there was no significant difference in TMB among the three differentiation grades ([Supplementary Figures 1B, C](#)). By combining two methods (dNdScv and MutSigCV), *TP53* (mutated in 52.7% of samples) was identified as the only significant driver gene in our ACGEJ cohort (FDR $q < 0.1$). To increase the detection power and accuracy, we combined our samples and the TCGA/Tumor Portal cohort samples (a total of 200 ACGEJ samples) and found 11 significant driver genes (FDR $q < 0.1$): *TP53*, *ARID1A*, *MUC6*, *SMAD4*, *PIK3CA*, *KRAS*, *PTEN*, *CDKN2A*, *MAP2K7*, *RNF43* and *RHOA*. ([Figure 1A](#)). Recurrent hotspot mutation analysis of our data and TCGA/Tumor Portal cohort data indicated that *TP53* had 3 common mutations: p.Arg248Gln, p.Arg248Trp and p.Arg273His ([Supplementary Figure 1D](#)). In TCGA/Tumor Portal tumors, the most frequent mutation of *TP53* was p.Arg175His, accounting for 13.2% (7/53) of the missense mutations. These results suggested that low driver mutation abundance is a dominant feature of Chinese ACGEJ samples due to their lower TMB.

Mutational signature of ACGEJ

The predominant somatic mutation type was 5'-C [T > G] T-3' in 96 substitution mutations in tumor samples from our cohort ([Supplementary Figure 1E](#)). To further explore the underlying mutational processes in ACGEJ tumors, we performed *de novo* single base substitution (SBS) mutational signature extraction and identified five mutational signatures, namely, SBS1, SBS5, SBS17a, SBS17b and SBS40 ([Figure 1B](#)). SBS1 and SBS5, which are common in tumors, are two clock-like and ageing-related endogenous mutational signatures, and they accounted for 48.4% of somatic mutations in ACGEJ ([Figure 1C](#)). SBS40, also correlated with ageing, accounted for 36.2% of somatic mutations in ACGEJ samples ([Figure 1C](#)). SBS17 (accounting for 15.4% of somatic mutations) can be split into SBS17a, defined by T>C substitution in the CTT trinucleotide context, and SBS17b, characterized by T>G conversion in any of the NTT trinucleotide contexts. Interestingly, we found that SBS17a and SBS17b were stage-related in ACGEJ patients, mainly appearing in TNM stage II and III patients, but were rare in stage I patients ([Figure 1B](#)). Compared with Siewert type II cases, Siewert type III cases showed significantly increased SBS17a and SBS17b signatures

(both $P < 0.05$; [Figures 1D, E](#)). These results indicate that SBS17 may be related with the progression of ACGEJ.

Somatic CNVs and functionally aberrant pathways in ACGEJ

In our ACGEJ tumors, we identified several chromosome arm level CNVs, including gains of 1q, 2q, 3q, 5p, 7p, 7q, 8q, 10p, 13q, 16p, 19q, 20p and 20q and losses of 4p, 4q, 9p, 14q, 15q, 17p, 18q, 19p, 21p and 21q (all $q < 0.25$ and frequency $> 30\%$; [Figure 2A](#) and [Supplementary Table 5](#)). We also identified 29 recurrent amplification peaks and 43 recurrent deletion peaks ($q < 0.25$), for which 1,153 and 1,300 genes overlapped, respectively ([Figure 2B](#) and [Supplementary Tables 6, 7](#)). Among specimens with different Siewert types or differentiation grades ([Supplementary Figures 2A, B](#)), we found similar gene-level CNVs. Among the 2,453 genes located in these CNV regions, 268 (10.9%) with copy number gain and 233 (9.5%) with copy number loss had mRNA levels significantly correlated with copy number (Spearman's $r > 0.3$ and FDR $q < 0.05$). Intersection of the above-described genes and the driver gene *TP53* with the cancer-related genes in COSMIC yielded a list of 29 genes, which were found to be functionally involved in the p53/cell cycle and PI3K/AKT signalling pathways by KEGG analysis. As shown in [Supplementary Table 8](#), p53/cell cycle pathway aberration was related to *TP53* mutation (52.7%) and amplifications of *CCNE1* (63.6%), *CDK6* (49.1%), *CASP8* (25.5%) and *MDM2* (21.8%) as well as deletion of *SMAD4* (40%). PI3K/AKT pathway aberration was related to amplifications of *FGFR2* (36.4%), *KRAS* (27.3%) and *CREB1* (23.6%). These results further confirmed the suggestion of our previous studies that *CCNE1* and *VEGFA*, which showed recurrent gene-level CNV, are CNV driver genes (FDR $q \leq 0.1$, in $\geq 10\%$ samples, [Figures 2C, D](#)). In addition to *ERBB2* amplification, a well-recognized target, other CNVs, such as *CCNE1* amplification and *VEGFA* amplification, could be targeted therapeutically. Moreover, we found that the copy number of *TNFRSF14* was significantly lower in Siewert type III than in Siewert II cases ([Figure 2E](#), $P < 0.05$).

Clonal and subclonal architectures of ACGEJ

Clonal and subclonal architectures reflect intratumor heterogeneity that may drive cancer progression and drug resistance. We used SciClone to reconstruct the clonal and subclonal architectures in each patient and obtained the number and genetic composition of clones. We identified four

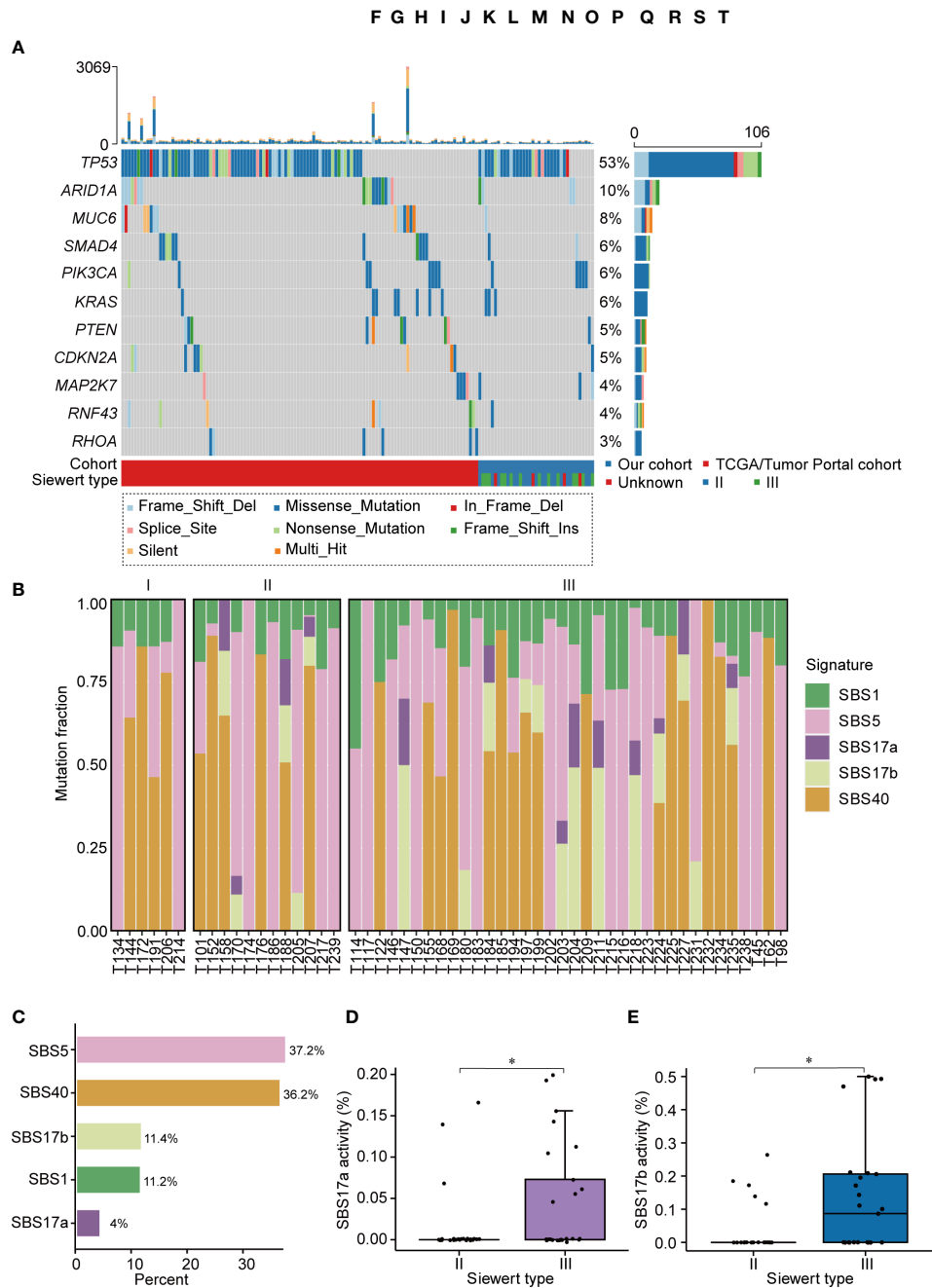


FIGURE 1

The landscape of somatic mutations and mutational signatures in ACGEJ. **(A)** The mutational landscape of driver genes in the 200 combined ACGEJ samples. Each column represents the tumor genome of one patient, and each row represents a gene. The top histogram shows the number of SNVs and indels in each patient. The histogram on the right displays the number of SNVs and indels in each gene. Driver genes (dNdScv, FDR $q < 0.1$) are ordered by the fraction of mutated tumor samples. **(B)** Bar plots showing the defined mutational signatures in our ACGEJ samples. Samples are ordered by TNM stage. **(C)** Bar plots comparing the defined mutational signatures and their relative contributions to somatic mutations. **(D, E)** SBS17a activity **(D)** and SBS17b activity **(E)** in different ACGEJ Siewert types. *P* values were derived from Wilcoxon rank-sum tests. Statistical significance was described as follows: **P* < 0.05.

clonal patterns in 55 ACGEJ patients (Figure 3A), including monoclonal (a single dominant clone), minor subclone (a single dominant clone with a minor subclone), biclonal (two major clones) and complex (more than two clones) clonal patterns. The

complex clonal pattern was the most common pattern in our samples (Figure 3B). We then analysed the distribution of the clonal patterns across samples with different histopathological features and observed that the complex clonal status was

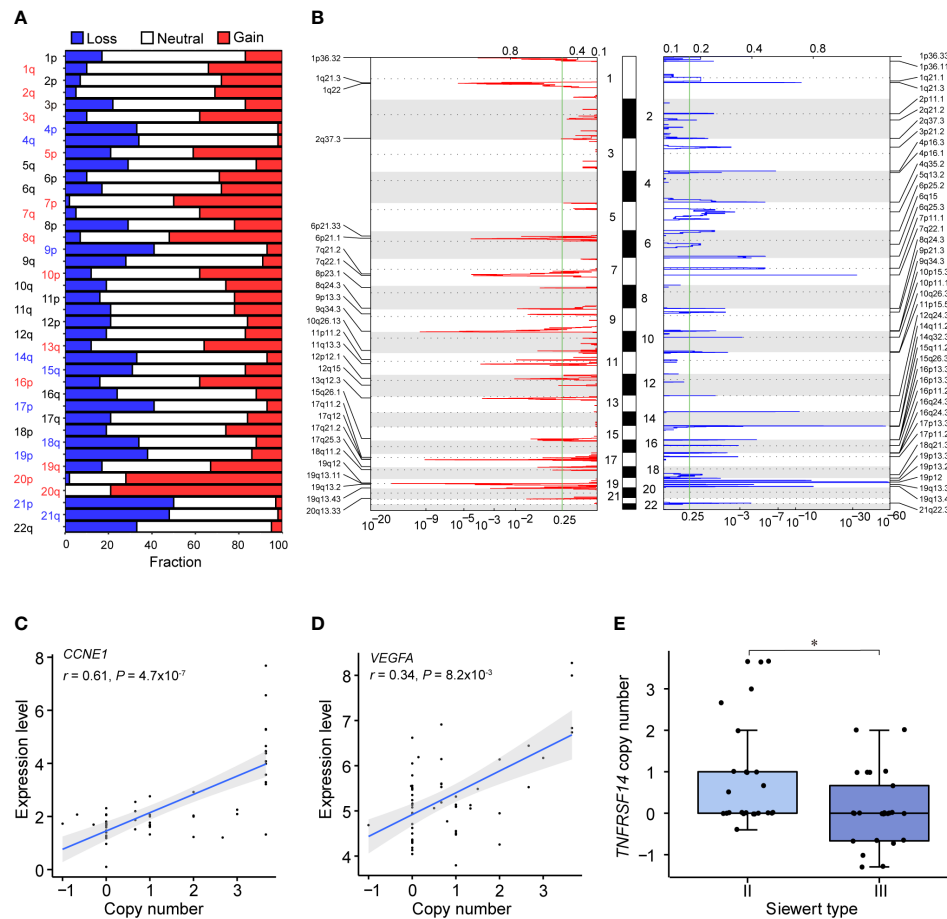


FIGURE 2

Somatic CNVs identified in our ACGEJ samples. **(A)** CNVs at the chromosome arm level. The bar graphs show the frequency of arm-level copy-number alterations, and the horizontal axis represents chromosome arms. Significant chromosome arm gains and losses ($q < 0.25$ and frequency $> 30\%$) are indicated in red and blue, respectively. **(B)** Focal peaks of amplifications (left) and deletions (right) detected by GISTIC 2.0. The x axis represents the false discovery rate, and the y axis represents the chromosome. The green line represents the significance threshold ($q = 0.25$). **(C, D)** Correlations between the copy number and expression level of two putative CNV driver genes, *CCNE1* (left) and *VEGFA* (right). The copy number data on the horizontal axis were derived from the GISTIC output (all_data_by_genes.txt file). *P* values were derived from Spearman's correlation tests. **(E)** Comparison of *TNFRSF14* copy number between Siewert II and III samples. *P* values were derived from Wilcoxon rank-sum tests. Statistical significance was described as follows: * $P < 0.05$.

significantly higher in Siewert type III than in Siewert type II samples (64.0% vs. 33.3%, $P < 0.05$; Figure 3C). However, ACGEJ samples with good, moderate or poor differentiation grades exhibited similar clonal architectures. We then defined tumors with monoclonal, minor subclone and biclonal patterns as having low intratumor heterogeneity and those with complex clonal patterns as having high intratumor heterogeneity. As shown in Figure 3D, higher TMB was significantly associated with higher intratumor heterogeneity ($P < 0.05$). We further explored the clonality of recurrently mutated genes identified in ACGEJ or gastric cancer samples (7, 53–56) and found that mutations of *PIK3CA* (E545K), *TP53* (R248Q), *SOX9* (Q369P) and *NEFH* (V267M) were located in the dominant founding clone, indicating that mutations in these genes were acquired at

an earlier time point (Figure 3E). These results emphasized the importance of *PIK3CA* and *TP53* mutations in the initiation of ACGEJ.

Characteristics of neoantigens in ACGEJ samples

Although genomic and transcriptomic alterations in ACGEJ have been characterized in some studies, the characteristics of neoantigens remain unclear. To the best of our knowledge, here, we describe the landscape of neoantigens in Chinese ACGEJ samples for the first time. We combined WES and RNA-seq data from our ACGEJ samples to predict neoantigens. First, we

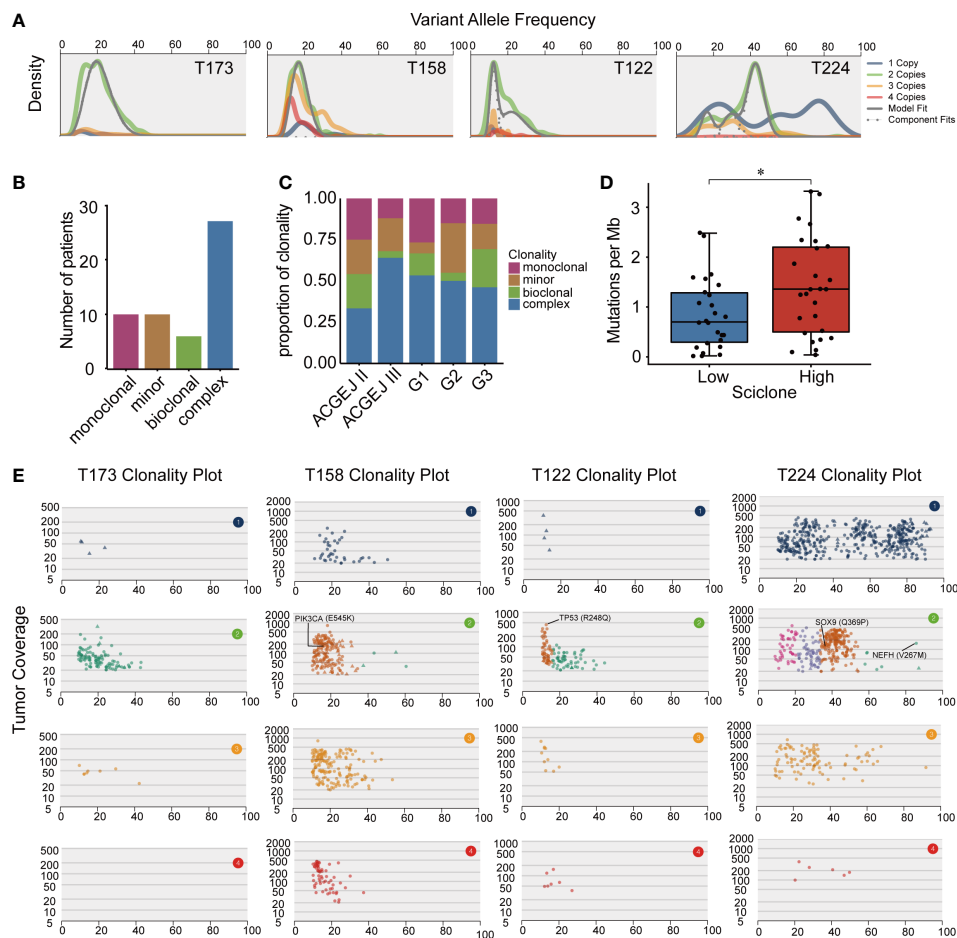


FIGURE 3

The clonal and subclonal architectures inferred in our ACGEJ samples. **(A)** Kernel density plots of variant allele frequency (VAF) across regions with one, two, three or four copies, posterior predictive densities summed over all clusters for copy-number-neutral variants and posterior predictive densities for each cluster/component. **(B)** Clonal patterns of 53 ACGEJ samples. **(C)** The distribution of clonal patterns across histopathological subtypes. ACGEJ II: Siewert type II, ACGEJ III: Siewert type III, G1: well differentiated, G2: moderately differentiated, G3: poorly differentiated or undifferentiated. **(D)** Comparison of TMB between patients with low and high intratumor heterogeneity defined by clonal patterns. **(E)** VAFs versus tumor coverage for each of the four copy number regions for four representative samples: T173, T158, T122 and T224. Mutations in previously reported driver genes are labelled. Statistical significance was described as follows: * $P < 0.05$.

employed OptiType to identify the 4-digit HLA class I alleles in 55 ACGEJ samples and found that they were quite different from those in stomach adenocarcinoma (STAD) samples. HLA-C*01:02, HLA-C*06:02 and HLA-C*07:02 were the three most common neoantigen-binding sites in ACGEJ samples, while HLA-A*02:01, HLA-C*24:02 and HLA-C*03:01 were the top 3 in STAD samples (Figures 4A, B). Moreover, we identified 12,285 neoantigens originating from missense mutations and 1,723 neoantigens originating from frame-shift and in-frame indels, and 27.2% (3,808/14,008) of neoantigens were high-affinity neoantigens (binding affinity < 0.5% rank) (Supplementary Table 9). The neoantigen load in cancers usually shows a positive correlation with TMB (14, 16, 57). In ACGEJ, we found that the number of neoantigens was also

positively correlated with TMB ($r = 0.65$, $P = 1 \times 10^{-7}$; Figure 4C). In STAD samples from TCGA, 2.09 neoantigens were generated per missense mutation and 2.53 neoantigens were generated per indel (10, 58). Similarly, we found that missense mutations generated fewer neoantigens (mean, 2.11 per mutation) than frame-shift indels (mean, 6.44 per mutation) in our specimens. In the ACGEJ samples, tumor neoantigens were mainly derived from missense mutations (Figure 4D), which accounted for 87.7% of neoantigens. Unsurprisingly, only 63.1% (45.6% - 76.4%) of nonsilent mutations generated neoantigens (Figure 4E), which indicated that amino acid sequence changes might not necessarily generate neoantigens. We further investigated whether frequently altered genes have the advantage of generating neoantigens and found that *TP53*,

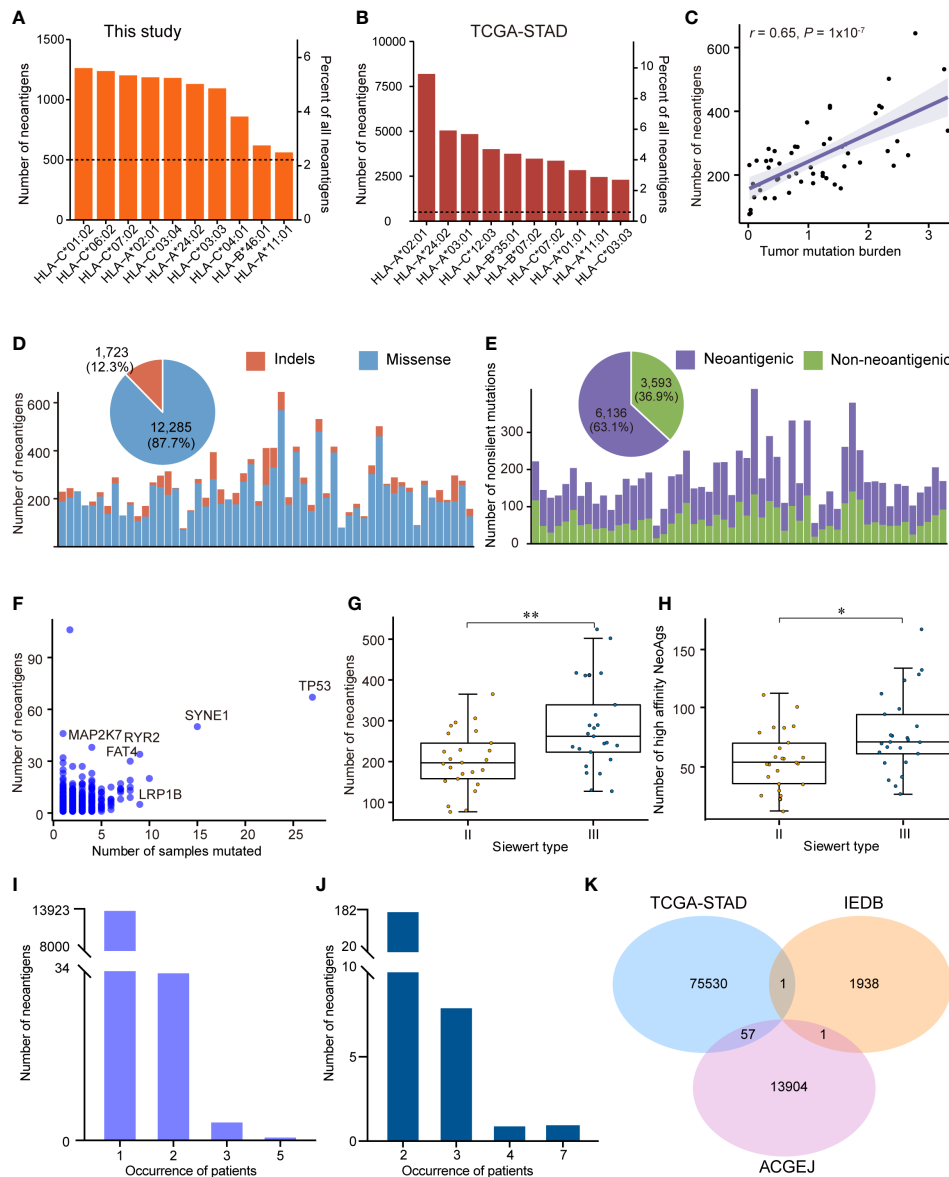


FIGURE 4

The neoantigen landscape in ACGEJ. (A, B) Ten predicted neoantigen sites of our ACGEJ samples (A) and TCGA-STAD samples (B). (C) Correlations between neoantigen load and tumor mutation burden (Spearman's correlation test). (D) Bar plots showing the number of neoantigens for each individual and their variant types. The pie plot on the top shows the fraction of variant types that produced neoantigens in 55 ACGEJ samples. (E) Bar plot and pie plot displaying the composition of nonsilent mutations that produced neoantigens. (F) Frequently altered genes with a high ability to generate neoantigens. The graph displays the number of samples containing specific genetic mutations and the relevant number of neoantigens. (G, H) Number of neoantigens (G) and number of high-affinity neoantigens (H) in different ACGEJ Siewert types. *P* values were derived from Wilcoxon rank-sum tests. NeoAgs: neoantigens. (I, J) Neoantigen recurrence in our patients (I) and TCGA-STAD patients (J). (K) Venn diagram displaying the intersections of neoantigens from our ACGEJ patients, TCGA-STAD patients and IEDB. Statistical significance was described as follows: **P* < 0.05; ***P* < 0.01.

SYNE1, *MAP2K7*, *RYR2*, *FAT4* and *LRP1B* generate more neoantigens than other genes across all samples (Figure 4F). We also found that Siewert type III tumors possessed more neoantigens or high-affinity neoantigens than Siewert type II tumors (Figures 4G, H; *P* < 0.05). However, there were no significant differences in neoantigen load or the number of

high-affinity neoantigens among tumors of different TNM stages (I, II and III) (Supplementary Figures 3A, B). Similar to the results across TNM stage, neoantigen load and the number of high-affinity neoantigens were not significantly different among well, moderate and poorly differentiated tumors (Supplementary Figures 3C, D).

As the vast majority of the potential neoantigens that have been reported are patient-specific (58, 59), we next explored the heterogeneity of predicted neoantigens in ACGEJ. The results suggested that there were 13,962 unique neoantigens in 55 ACGEJ patients; 99.72% (13923/13962) of the neoantigens were found in only one patient, while only 0.28% (39/13962) of the neoantigens were found in at least two patients (Figure 4I). The most common shared neoantigen was “GRQKRSDSL”, which was generated by *VSX1* (V138L). We also calculated the frequencies of neoantigens in TCGA-STAD samples from TSNAdb. Similarly, we found that there were 85,283 neoantigens in a total of 411 samples; only 0.21% (182/85,283) of the neoantigens were found in two samples, 8 neoantigens were found in three samples, one neoantigen was found in four samples, and one neoantigen was found in seven samples (Figure 4J). Furthermore, we compared the predicted neoantigens from our ACGEJ samples and TCGA-STAD samples to those found using IEDB data to identify common neoantigens (Figure 4K and Supplementary Table 10). We found 57 common neoantigens between the TCGA-STAD and ACGEJ samples, and only one neoantigen was common between the IEDB and ACGEJ samples. These results collectively point to the difficulty in ubiquitous neoantigen identification for ACGEJ.

Construction of the weighted gene coexpression network and identification of Hub genes in ACGEJ

We examined all components in the TME, including immune and non-immune cells. Specifically, we applied xCell to estimate the abundance scores of immune and stromal cell populations in ACGEJ samples, and the results showed that the ratio of neoantigen to mutation was positively correlated with immune score, stromal score and microenvironment score (Supplementary Figures 4A–C). The ratio of neoantigen to mutation was significantly associated with TME cells, including B cell, myeloid dendritic cell, endothelial cell, cancer-associated fibroblasts, and T cell (Supplementary Figures 4D–O). Considering the importance of the CD8+ T-cell infiltration level for neoantigen-targeted therapy and the prognosis of patients with ACGEJ, we conducted WGCNA to obtain CD8+ T-cell-related Hub genes. In addition, to gain a more comprehensive understanding of CD8+ T-cell infiltration, we assembled a relatively large-scale cohort of 93 Chinese ACGEJ patients with RNA-seq data and complete follow-up information by integrating our previously published data. In this dataset, clinical information included age, gender, TNM stage and the absolute abundance of CD8+ T cells (quantified by MCP-counter method) as well as survival status and OS. A soft-thresholding power of $\beta = 4$ was chosen to construct a scale-free network (scale-free $r^2 = 0.88$, slope = -1.62) (Supplementary Figures 5A, D). A total of 12 gene coexpression modules were

identified from the hierarchical clustering tree (Figures 5A, B). As shown in the module–trait relationship heatmap (Figure 5B), the correlation coefficients and p values between module eigengenes and clinical traits were determined to identify which modules were associated with clinical information. The turquoise module was identified to be significantly positively related to the infiltration of CD8+ T cells in the ACGEJ samples ($r = 0.61$, $P = 8 \times 10^{-11}$). Furthermore, we applied GO and KEGG analyses to explore the biological functions of genes in the turquoise module. The GO analysis showed that the turquoise module genes were mainly related to the terms cytokine receptor activity, external side of the plasma membrane and T-cell activation (Figure 5C). The KEGG pathway analysis revealed that the genes were mainly enriched in the terms haematopoietic cell lineage, Th1/Th2 cell differentiation, Th17-cell differentiation, chemokine signalling pathway and cytokine–cytokine receptor interaction (Figure 5D). These results suggest that the turquoise module genes may be involved in the response to immunotherapy. To further investigate new potential targets for immunotherapies (immunosuppressive therapies or T-cell proliferation-inducing therapies), we used cut-off thresholds ($GS > 0.6$ and $MM > 0.8$; Figure 5E) to identify key genes in the turquoise module. We identified the ten most important Hub genes in the module that were related to CD8+ T-cell infiltration: *CCL5*, *CD2*, *CST7*, *GVINP1*, *GZMK*, *IL2RB*, *IKZF3*, *PLA2G2D*, *P2RY10* and *ZAP70* (Figure 5F).

Construction and validation of the risk prediction model

Previous studies have reported that high neoantigen load was associated with good survival in many cancers (13–16). To determine whether the potential neoantigens could predict OS in patients with ACGEJ, we divided patients into high and low groups based on the median of neoantigen load and the ratio of neoantigen to mutation. Regrettably, we found that neither neoantigen load nor the ratio of neoantigen to mutation was associated with patient OS (Supplementary Figures 3E, F). However, both methods, MCP-counter and xCell (Supplementary Figure 3G and Figure 4N), showed that the abundance of CD8+ T cells was positively correlated with the ratio of neoantigen to mutation (MCP-counter: $r = 0.28$, $P = 0.044$; xCell: $r = 0.36$, $P = 0.007$). To further explore the prognostic value of the CD8+ T-cell infiltration-related module genes and identify potential prognostic factors of ACGEJ, we determined the prognostic value of the turquoise module genes. We obtained 48 genes significantly related to ACGEJ OS through univariate Cox regression analysis. After LASSO Cox and multivariate Cox regression analyses (Figures 6A, B and Supplementary Figure 6A), we identified seven genes (*ADAM28*, *ASPH*, *CAMK2N1*, *F2R*, *STAP1*, *TP53INP2*, *ZC3H3*) and used these genes to construct a

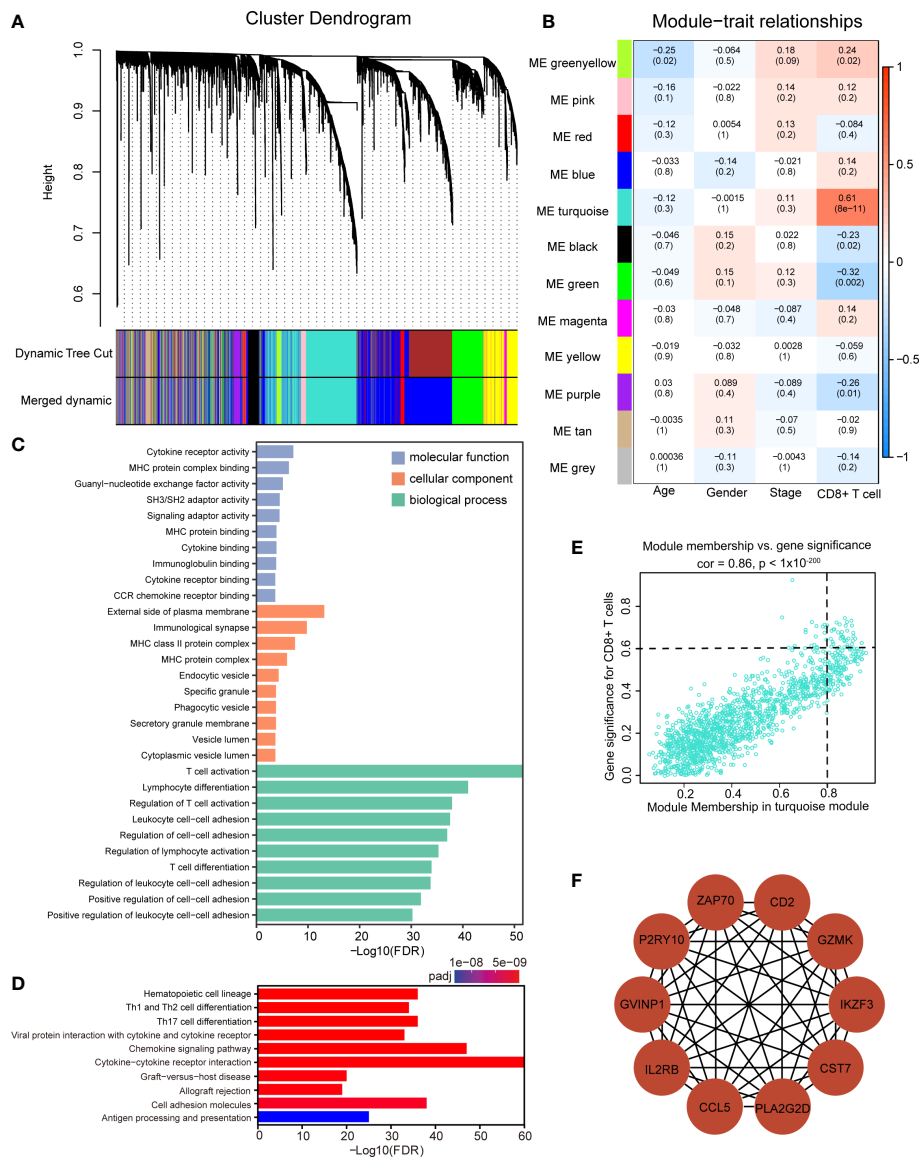


FIGURE 5

Identification of significant modules and Hub genes via WGCNA. (A) Gene dendrogram obtained by average linkage hierarchical clustering. The color row underneath the dendrogram indicates 12 different coexpression gene modules (the grey module represents unassigned genes). (B) The module-trait relationship heatmap between module eigengenes (row) and clinical traits (column). The correlation coefficients and p values are annotated in each box. (C, D) GO analysis (C) and KEGG pathway enrichment analysis (D) of all genes in the turquoise module. (E) Scatter plot of gene significance for the infiltration of CD8+ T cells (y axis) versus module membership (x axis) in the turquoise module. (F) Protein-protein interaction network of the ten Hub genes.

prognosis prediction model with a training cohort containing our ACGEJ samples ($n=93$) and an external validation cohort containing TCGA ACGEJ samples ($n=86$). The risk score for all ACGEJ samples was calculated by the following formula: risk score = expression level of *ZC3H3**1.83 + expression level of *STAP1**1.25 + expression level of *ADAM28**0.77+ expression level of *TP53INP2**(-1.03) + expression level of *CAMK2N1**1.05 + expression level of *F2R**0.84 + expression level of *ASPH**1.10. Taking the median risk score as the cut-off

value, ACGEJ samples were divided into high- and low-risk score groups. The results showed that patients in the high-risk group had a significantly worse prognosis than those in the low-risk group in both the training cohort ($P < 0.001$, Figure 6C) and validation cohort ($P < 0.01$, Figure 6D). As shown in Supplementary Figures 6B, C, the area under the curve (AUC) was 0.729 in the training cohort and 0.712 in the validation cohort, which indicated that the risk model had good prediction performance. Next, multivariate Cox regression analysis

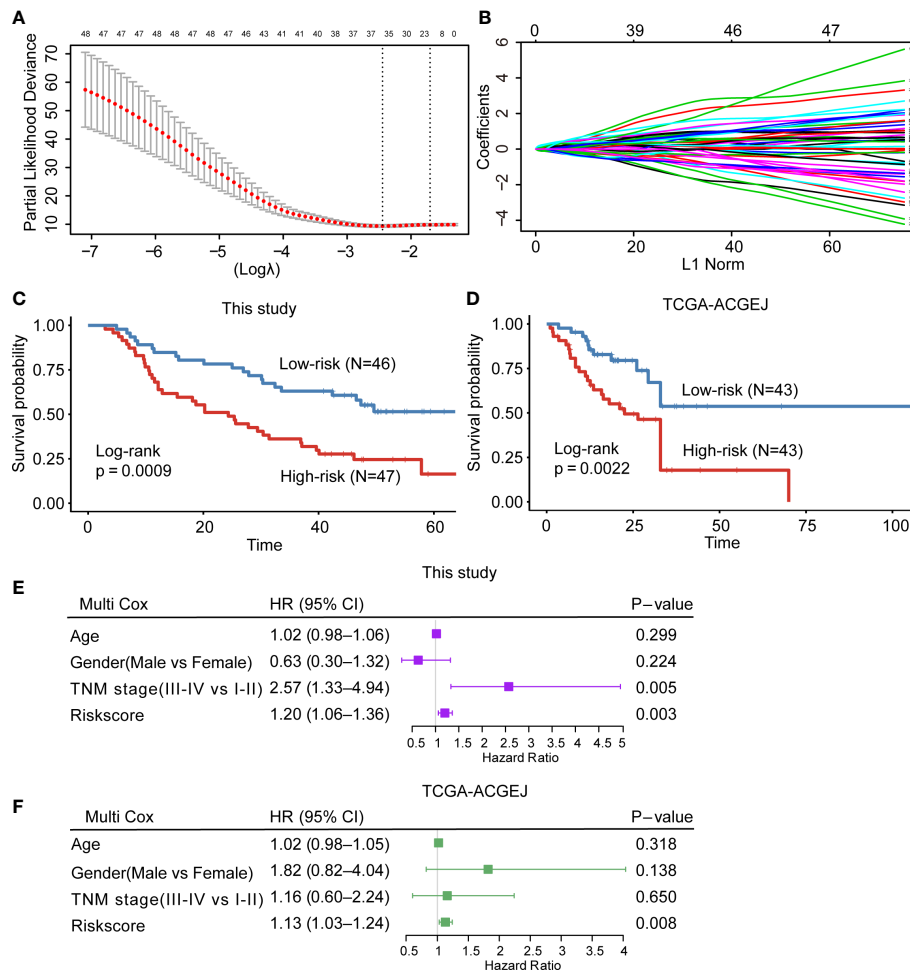


FIGURE 6

Construction and validation of the CD8+ T-cell infiltration-related prognostic prediction model. (A, B) LASSO Cox regression with 10-fold cross-validation to determine the prognostic value of 48 prognosis-related genes in the turquoise module. (C, D) Kaplan–Meier curve for the training cohort (C) and external validation cohort (D). (E, F) Multivariate Cox regression analysis of clinicopathological factors and the risk score for our ACGEJ patients (E) and TCGA-ACGEJ patients (F).

confirmed that the risk score was an independent prognostic factor after adjusting for age, gender and TNM stage in both our ACGEJ cohort (HR =1.20, 95% CI = 1.06–1.36, $P < 0.01$; Figure 6E) and the TCGA-ACGEJ cohort (HR =1.13, 95% CI = 1.03–1.24, $P < 0.01$; Figure 6F). In summary, our results showed that the risk score can serve as an effective predictor for the risk classification of ACGEJ patients.

Discussion

In this study, we determined the genomic alteration and neoantigen landscapes in samples collected from Chinese patients with ACGEJ. Comparing the genomic alterations in our 55 WES samples with those in 124 whole-genome sequencing samples revealed similar mutational signature

patterns and chromosome arm level CNVs and showed that *TP53* was the only significant driver gene, *CCNE1* and *VEGFA* amplifications occurred frequently and were correlated with gene expression upregulation. However, our WES data seem to have fewer putative CNV driver genes, probably because of the smaller sample size. We found some differences between Siewert II and III samples, including differences in the mutational signatures SBS17a and SBS17b, copy number of *TNFRSF14*, the prevalence of the complex clonal pattern and neoantigen load. We identified of 58 shared tumor neoantigens, which could provide candidate targets for neoantigen-based targeted therapy. In addition, we identified 10 CD8+ T-cell infiltration-related Hub genes that may serve as immunotherapeutic biomarkers. Moreover, we constructed a risk prediction model based on CD8+ T-cell infiltration-related genes. Our findings in the present study provide a deeper understanding of ACGEJ that

may lead to improved candidate target selection, and we proposed a novel prognostic prediction model.

We identified 11 recurrently mutated genes in two cohorts, 3 of which (*RNF43*, *MAP2K7* and *RHOA*) were newly discovered driver genes in ACGEJ. The E3 ubiquitin ligase RING finger protein 43 (*RNF43*) inhibits WNT signalling by ubiquitinating Frizzled receptors and targeting them for degradation (60). Loss of *RNF43* function weakens the DNA damage response, leading to resistance to radiotherapy and chemotherapy in gastric cancer (61). *RNF43* was found to be frequently mutated in tumors with microsatellite instability and was identified as a significantly mutated driver gene in gastric adenocarcinoma (62). In our cohort, the vast majority of samples were microsatellite stable, and only 2% of samples had *RNF43* mutation. Mitogen-activated protein kinase 7 (*MAP2K7*) is an important tumor suppressor in gastric cancer, and frequent loss-of-function mutations activate the JNK pathway (63). Ras homologue family member A (*RHOA*) belongs to the Rho family of GTPases and is a driver gene in diffuse-type gastric carcinoma (64, 65). *RHOA* mutation was not found in the Chinese patients in this study, and its mutation might be related to differences in genetic or environmental risk factors, but this conclusion needs to be further verified in larger cohorts.

SBS17a and SBS17b have been reported to appear at an early stage of esophageal and gastric adenocarcinoma and to be related to gastric-oesophageal reflux (66, 67). SBS17a and SBS17b may be markers of oxidative damage and risk factors for chromosomal instability in ACGEJ (8, 68). SBS17b accounted for more somatic mutations than SBS17a in our cohort, which is in line with the phenomenon that 5'-C [T > G] T-3' was the most common somatic single base substitution in ACGEJ samples. Due to differences in anatomical position, we believe that Siewert type III tumors are more likely to feature oxidative stress, which was supported by the result that the SBS17a and SBS17b signatures were significantly more prominent in Siewert type III samples than in Siewert type II samples. With high-depth (mean coverage 380X) WES, we confirmed that *CCNE1* and *VEGFA* are robust CNV driver genes in ACGEJ. We also identified *TP53* mutations and CNVs related to the p53/cell cycle and PI3K/AKT signalling pathways. By analysing clonal and subclonal profiles, we demonstrated that the complex clonal pattern was most common in ACGEJ samples, reflecting the intratumor heterogeneity in this cancer. A limitation of our study is that we could not show clonal evolution by drawing fishplots and Bayesian trees in ACGEJ because of the difficulty in collecting matched pre- and post-treatment samples, paired primary and metastatic samples, or samples taken from multiple sites in an individual.

We also performed a comprehensive analysis of genomic alterations and transcriptomic changes to predict neoantigens.

We found that the number of neoantigens produced by *SYNE1* mutation was second to that produced by *TP53* mutation. Although *SYNE1* was not identified as a driver gene in our ACGEJ samples, it has been identified as a significantly mutated gene in Western patients with ACGEJ and Chinese patients with gastric cancer (53, 69). As such, we plan to study the role of *SYNE1* mutation-derived neoantigens in future research. Based on analysis with two neoantigen-related databases, TSNAdb and IEDB, we obtained 58 neoantigens involving 50 genes (Supplementary Table 10) that may serve as candidate targets for neoantigen vaccines.

Several clinical trials have demonstrated that neoantigen vaccines based on dendritic cells, peptides and mRNAs can induce CD8+ T-cell-specific responses, emphasizing the considerable potential of neoantigens in immunotherapy (21, 22, 70). However, neoantigen vaccines spontaneously upregulate the expression of surface molecules (such as PD-1, TIM3 and CTLA-4) in neoantigen-specific T cells and/or PD-L1 in tumor cells, which in turn impedes the function of the neoantigen vaccines (71). Thus, a combination of neoantigen vaccines and immunosuppressive therapies to improve vaccine-induced T-cell responses is recommended. In addition, the combination of neoantigen vaccines and T-cell proliferation-inducing therapy, such as VB10.NEO plus NKTR-214, was confirmed to induce strong immunogenic CD8+ T-cell responses in preclinical models (27). Subsequently, based on WGCNA, we identified important Hub genes in ACGEJ positively associated with the infiltration of CD8+ T cells, providing new insights for research on the mechanisms of action of immunotherapy and offering potential targets for combination therapies in ACGEJ. We also constructed a risk prediction model based on the CD8+ T-cell infiltration -related genes, and the risk score showed good performance in predicting the OS of ACGEJ patients with in both the training cohort and the external validation cohort.

Conclusion

Our study reveals the genomic characteristics and neoantigen features of ACGEJ. Through WGCNA, we established a CD8+ T-cell infiltration-related prediction model. Furthermore, many newly identified potential therapeutic targets can be pursued.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Y.Lao performed bioinformatics analysis. Y.W., J.Y., T.L., Y.M., Y.Luo, Y.S., K.L. and X.Z. contributed to statistical analysis. X.N., Y.X. and C.Z. responded for sample collection and performed experiments. All authors contributed to the article and approved the submitted version.

Funding

This project was supported by National Key Research and Development Program of China (2016YFC1302700 to CW), National Science Fund for Distinguished Young Scholars (81725015 to CW), Medical and Health Technology Innovation Project of Chinese Academy of Medical Sciences (2016-I2M-4-002 to CW, 2019-I2M-2-001 to DL and CW), Beijing Outstanding Young Scientist Program (BJJWZYJH01201910023027 to CW).

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.

The reviewer WL declared a shared affiliation with the authors to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or

claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.941868/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

The landscape of somatic mutations and mutational signatures of ACGEJ. (A) Violin plots comparing tumor mutation burden differences across the two cohorts. (B, C) Box plots comparing tumor mutation burden in our ACGEJ samples with different Siewert types (B) or differentiation grades (C). (D) Comparison of distribution of non-synonymous TP53 somatic mutations between our tumor samples and TCGA/Tumor Portal samples. (E) Mutational spectra of our ACGEJ samples. *P* values were derived from Wilcoxon rank-sum tests; *****P* < 0.0001; ns, not significant; G1: well differentiated; G2: moderately differentiated; G3: poorly differentiated or undifferentiated.

SUPPLEMENTARY FIGURE 2

(A, B) Box plots comparing gene level CNVs in our ACGEJ samples with different Siewert types (A) or differentiation grades (B). *P* values were derived from Wilcoxon rank-sum tests. Somatic CNVs are limited to homozygous deletion and high-level amplifications (GISTIC score of -2 and 2 respectively).

SUPPLEMENTARY FIGURE 3

(A, C) Box plots comparing number of neoantigens in our ACGEJ samples with different TNM stages (A) and differentiation grades (C). (B, D) Box plots comparing number of high-affinity neoantigens in ACGEJ samples with different TNM stages (B) and differentiation grades (D). NeoAgs: neoantigens. *P* values were derived from Wilcoxon rank-sum tests. (E, F) Kaplan–Meier curve for neoantigen load (E) and the ratio of neoantigen to mutation (F). (G) Correlations between CD8+ T cell infiltration level (quantified by MCP-counter method) and the ratio of neoantigen to mutation (Spearman's correlation test).

SUPPLEMENTARY FIGURE 4

(A–O) Association of the ratio of neoantigen to mutation and immune score, stromal score and microenvironment score and TME cells (quantified by xCell method) (Spearman's correlation test).

SUPPLEMENTARY FIGURE 5

(A) Network topology was plotted by different soft thresholding powers (*x*-axis) and correlation coefficient between log (*k*) and log [*P*(*k*)] (*y*-axis). The red line represents a correlation coefficient of 0.9. An approximate soft-thresholding powers, $\beta=4$, was obtained. (B) Mean connectivity under different soft-thresholding powers. (C, D) Assessing the scale-free topology when β was set to 4.

SUPPLEMENTARY FIGURE 6

(A) Multivariate Cox regression analysis of 17 OS-related factors. (B, C) ROC curve of the training cohort (B) and external validation cohort (C).

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
- Bartel M, Brahmabhatt B, Bhurwal A. Incidence of gastroesophageal junction cancer continues to rise: Analysis of surveillance, epidemiology, and end results (SEER) database. *J Clin Oncol* (2019) 37(4_suppl):40. doi: 10.1200/JCO.2019.37.4_suppl.40

3. Siewert JR, Stein HJ. Classification of adenocarcinoma of the oesophagogastric junction. *Br J Surg* (1998) 85(11):1457–9. doi: 10.1046/j.1365-2168.1998.00940.x
4. Ajani JA, D'Amico TA, Bentrem DJ, Chao J, Corvera C, Das P, et al. Esophageal and esophagogastric junction cancers, version 2.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* (2019) 17(7):855–83. doi: 10.6004/jnccn.2019.0033
5. Boku N, Ryu MH, Kato K, Chung HC, Minashi K, Lee KW, et al. Safety and efficacy of nivolumab in combination with s-1/Capecitabine plus oxaliplatin in patients with previously untreated, unresectable, advanced, or recurrent Gastric/Gastroesophageal junction cancer: Interim results of a randomized, phase II trial (Attraction-4). *Ann Oncol* (2019) 30(2):250–8. doi: 10.1093/annonc/mdy540
6. Nakamura Y, Kawazoe A, Lordick F, Janjigian YY, Shitara K. Biomarker-targeted therapies for advanced-stage gastric and gastro-oesophageal junction cancers: An emerging paradigm. *Nat Rev Clin Oncol* (2021) 18(8):473–87. doi: 10.1038/s41571-021-00492-2
7. Geng Q, Lao J, Zuo X, Chen S, Bei JX, Xu D. Identification of the distinct genomic features in gastroesophageal junction adenocarcinoma and its siwert subtypes. *J Pathol* (2020) 252(3):263–73. doi: 10.1002/path.5516
8. Lin Y, Luo Y, Sun Y, Guo W, Zhao X, Xi Y, et al. Genomic and transcriptomic alterations associated with drug vulnerabilities and prognosis in adenocarcinoma at the gastroesophageal junction. *Nat Commun* (2020) 11(1):6091. doi: 10.1038/s41467-020-19949-6
9. Turajlic S, Litchfield K, Xu H, Rosenthal R, McGranahan N, Reading JL, et al. Insertion-and-Deletion-Derived tumor-specific neoantigens and the immunogenic phenotype: A pan-cancer analysis. *Lancet Oncol* (2017) 18(8):1009–21. doi: 10.1016/s1470-2045(17)30516-8
10. Yang W, Lee KW, Srivastava RM, Kuo F, Krishna C, Chowell D, et al. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat Med* (2019) 25(5):767–75. doi: 10.1038/s41591-019-0434-2
11. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* (2015) 348(6230):69–74. doi: 10.1126/science.aaa4971
12. Blass E, Ott PA. Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nat Rev Clin Oncol* (2021) 18(4):215–29. doi: 10.1038/s41571-020-00460-2
13. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* (2015) 350(6257):207–11. doi: 10.1126/science.aad0095
14. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* (2016) 351(6280):1463–9. doi: 10.1126/science.aaf1490
15. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* (2015) 348(6230):124–8. doi: 10.1126/science.aaa1348
16. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* (2015) 372(26):2509–20. doi: 10.1056/NEJMoa1500596
17. Plitas G, Rudensky AY. Regulatory T cells in cancer. *Annu Rev Cancer Biol* (2020) 4(1):459–77. doi: 10.1146/annurev-cancerbio-030419-033428
18. Steinert EM, Schenkel JM, Fraser KA, Beura LK, Manlove LS, Igyarto BZ, et al. Quantifying memory CD8 T cells reveals regionalization of immunosurveillance. *Cell* (2015) 161(4):737–49. doi: 10.1016/j.cell.2015.03.031
19. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanović S, Gouttefangeas C, et al. Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature* (2019) 565(7738):240–5. doi: 10.1038/s41586-018-0810-y
20. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, et al. Neoantigen vaccine generates intratumoral T cell responses in phase I glioblastoma trial. *Nature* (2019) 565(7738):234–9. doi: 10.1038/s41586-018-0792-9
21. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* (2017) 547(7662):217–21. doi: 10.1038/nature22991
22. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* (2017) 547(7662):222–6. doi: 10.1038/nature23003
23. Guo Z, Yuan Y, Chen C, Lin J, Ma Q, Liu G, et al. Durable complete response to neoantigen-loaded dendritic-cell vaccine following anti-PD-1 therapy in metastatic gastric cancer. *NPJ Precis Oncol* (2022) 6(1):34. doi: 10.1038/s41698-022-00279-3
24. Feig C, Jones JO, Kraman M, Wells RJ, Deonaraine A, Chan DS, et al. Targeting CXCL12 from fap-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proc Natl Acad Sci U.S.A.* (2013) 110(50):20212–7. doi: 10.1073/pnas.1320318110
25. Dominguez CX, Müller S, Keerthivasan S, Koepfen H, Hung J, Gierke S, et al. Single-cell RNA sequencing reveals stromal evolution into LRRCL15(+) myofibroblasts as a determinant of patient response to cancer immunotherapy. *Cancer Discovery* (2020) 10(2):232–53. doi: 10.1158/2159-8290.Cd-19-0644
26. Obradovic A, Graves D, Korner M, Wang Y, Roy S, Naveed A, et al. Immunostimulatory cancer-associated fibroblast subpopulations can predict immunotherapy response in head and neck cancer. *Clin Cancer Res* (2022) 28(10):2094–109. doi: 10.1158/1078-0432.Ccr-21-3570
27. Granum S, Zell-Flagstad H, Bersaas A, Skullerud L, Müller E, Axelsen M, et al. Abstract 2256: Combination of neoantigen DNA plasmid vaccine VB10.NEO and NKTR-214, a CD122-biased immunostimulatory cytokine, induces strong neoantigen-specific T cell responses and sustained tumor regression in pre-clinical models. *Cancer Res* (2019) 79(13_Supplement):2256. doi: 10.1158/1538-7445.Am2019-2256
28. Thompson ED, Zahurak M, Murphy A, Cornish T, Cuka N, Abdelfatah E, et al. Patterns of PD-L1 expression and CD8 T cell infiltration in gastric adenocarcinomas and associated immune stroma. *Gut* (2017) 66(5):794–801. doi: 10.1136/gutjnl-2015-310839
29. Wang Y, Zhu C, Song W, Li J, Zhao G, Cao H. PD-L1 expression and CD8⁺ T cell infiltration predict a favorable prognosis in advanced gastric cancer. *J Immunol Res* (2018) 2018:4180517. doi: 10.1155/2018/4180517
30. Wang J, Li R, Cao Y, Gu Y, Fang H, Fei Y, et al. Intratumoral CXCR5⁺CD8⁺T associates with favorable clinical outcomes and immunogenic contexture in gastric cancer. *Nat Commun* (2021) 12(1):3080. doi: 10.1038/s41467-021-23356-w
31. Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, et al. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* (2014) 513(7517):202–9. doi: 10.1038/nature13480
32. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* (2013) 45(5):478–86. doi: 10.1038/ng.2591
33. Wang Y, Chen Y, Yang J, Lin Y, Luo Y, Zhang S, et al. Identification and analysis of key genes in association with development and progression of adenocarcinoma at the gastroesophageal junction. *Basic Clin Med* (2021) 41(6):831–6. doi: 10.16352/j.issn.1001-6325.2021.06.010
34. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell* (2018) 33(4):721–35.e8. doi: 10.1016/j.ccell.2018.03.010
35. Wu J, Zhao W, Zhou B, Su Z, Gu X, Zhou Z, et al. TSNAdb: A database for tumor-specific neoantigens from immunogenomics data analysis. *Genomics Proteomics Bioinf* (2018) 16(4):276–82. doi: 10.1016/j.gpb.2018.06.003
36. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res* (2019) 47(D1):D339–d43. doi: 10.1093/nar/gky1006
37. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics* (2013). doi: 10.48550/arXiv.1303.3997
38. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: Fast and accurate calling of germline and somatic variants. *Nat Methods* (2018) 15(8):591–4. doi: 10.1038/s41592-018-0051-x
39. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensemble variant effect predictor. *Genome Biol* (2016) 17(1):122. doi: 10.1186/s13059-016-0974-4
40. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature* (2020) 578(7793):94–101. doi: 10.1038/s41586-020-1943-3
41. Shen R, Seshan VE. FACETS: Allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* (2016) 44(16):e131. doi: 10.1093/nar/gkw520
42. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhir R, Getz G. Gistic2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* (2011) 12(4):R41. doi: 10.1186/gb-2011-12-4-r41
43. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: Inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol* (2014) 10(8):e1003665. doi: 10.1371/journal.pcbi.1003665
44. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. Optitype: Precision HLA typing from next-generation sequencing data. *Bioinformatics* (2014) 30(23):3310–6. doi: 10.1093/bioinformatics/btu548
45. Bray NI, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* (2016) 34(5):525–7. doi: 10.1038/nbt.3519

46. Bjerregaard AM, Nielsen M, Hadrup SR, Szallasi Z, Eklund AC. Mupexi: Prediction of neo-epitopes from tumor sequencing data. *Cancer Immunol Immunother* (2017) 66(9):1123–30. doi: 10.1007/s00262-017-2001-3
47. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* (2017) 199(9):3360–8. doi: 10.4049/jimmunol.1700893
48. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17(1):218. doi: 10.1186/s13059-016-1070-5
49. 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
50. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* (2020) 48(W1):W509–W14. doi: 10.1093/nar/gkaa407
51. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559
52. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
53. Wang H, Ding Y, Chen Y, Jiang J, Chen Y, Lu J, et al. A novel genomic classification system of gastric cancer *Via* integrating multidimensional genomic characteristics. *Gastric Cancer* (2021) 24(6):1227–41. doi: 10.1007/s10120-021-01201-9
54. Wang R, Song S, Harada K, Ghazanfari Amlashi F, Badgwell B, Pizzi MP, et al. Multiplex profiling of peritoneal metastases from gastric adenocarcinoma identified novel targets and molecular subtypes that predict treatment response. *Gut* (2020) 69(1):18–31. doi: 10.1136/gutjnl-2018-318070
55. Hao D, He S, Harada K, Pizzi MP, Lu Y, Guan P, et al. Integrated genomic profiling and modelling for risk stratification in patients with advanced oesophagogastric adenocarcinoma. *Gut* (2021) 70(11):2055–65. doi: 10.1136/gutjnl-2020-322707
56. Xing X, Jia S, Wu J, Feng Q, Dong B, Li B, et al. Clonality analysis of synchronous gastro-oesophageal junction carcinoma and distal gastric cancer by whole-exome sequencing. *J Pathol* (2017) 243(2):165–75. doi: 10.1002/path.4932
57. Jardim DL, Goodman A, de Melo Gagliato D, Kurzrock R. The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell* (2021) 39(2):154–73. doi: 10.1016/j.ccell.2020.10.001
58. Lin M, Zhang XL, You R, Yang Q, Zou X, Yu K, et al. Neoantigen landscape in metastatic nasopharyngeal carcinoma. *Theranostics* (2021) 11(13):6427–44. doi: 10.7150/thno.53229
59. Ren Y, Cherukuri Y, Wickland DP, Sarangi V, Tian S, Carter JM, et al. HLA class-I and class-II restricted neoantigen loads predict overall survival in breast cancer. *Oncoimmunology* (2020) 9(1):1744947. doi: 10.1080/2162402x.2020.1744947
60. Koo BK, Spit M, Jordens I, Low TY, Stange DE, van de Wetering M, et al. Tumor suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of wnt receptors. *Nature* (2012) 488(7413):665–9. doi: 10.1038/nature11308
61. Neumeier V, Brutau-Abia A, Allgäuer M, Pfarr N, Weichert W, Falkeis-Weits C, et al. Loss of RNF43 function contributes to gastric carcinogenesis by impairing DNA damage response. *Cell Mol Gastroenterol Hepatol* (2021) 11(4):1071–94. doi: 10.1016/j.jcmgh.2020.11.005
62. Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* (2014) 46(6):573–82. doi: 10.1038/ng.2983
63. Hudson AM, Stephenson NL, Li C, Trotter E, Fletcher AJ, Katona G, et al. Truncation- and Motif-based pan-cancer analysis reveals tumor-suppressing kinases. *Sci Signal* (2018) 11(526):eaan6776. doi: 10.1126/scisignal.aan6776
64. Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, et al. Recurrent gain-of-Function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet* (2014) 46(6):583–7. doi: 10.1038/ng.2984
65. Zhang H, Schaefer A, Wang Y, Hodge RG, Blake DR, Diehl JN, et al. Gain-of-Function rhoa mutations promote focal adhesion kinase activation and dependency in diffuse gastric cancer. *Cancer Discovery* (2020) 10(2):288–305. doi: 10.1158/2159-8290.Cd-19-0811
66. Murugaesu N, Wilson GA, Birkbak NJ, Watkins T, McGranahan N, Kumar S, et al. Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discovery* (2015) 5(8):821–31. doi: 10.1158/2159-8290.Cd-15-0412
67. Tomkova M, Tomek J, Kriaucionis S, Schuster-Böckler B. Mutational signature distribution varies with DNA replication timing and strand asymmetry. *Genome Biol* (2018) 19(1):129. doi: 10.1186/s13059-018-1509-y
68. Kauppi J, Räsänen J, Sihvo E, Nieminen U, Arkkila P, Ahotupa M, et al. Increased oxidative stress in the proximal stomach of patients with barrett's esophagus and adenocarcinoma of the esophagus and esophagogastric junction. *Transl Oncol* (2016) 9(4):336–9. doi: 10.1016/j.tranon.2016.06.004
69. Chong IY, Cunningham D, Barber LJ, Campbell J, Chen L, Kozarewa I, et al. The genomic landscape of oesophagogastric junctional adenocarcinoma. *J Pathol* (2013) 231(3):301–10. doi: 10.1002/path.4247
70. Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, et al. Cancer immunotherapy: a dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* (2015) 348(6236):803–8. doi: 10.1126/science.aaa3828
71. Peng M, Mo Y, Wang Y, Wu P, Zhang Y, Xiong F, et al. Neoantigen vaccine: An emerging tumor immunotherapy. *Mol Cancer* (2019) 18(1):128. doi: 10.1186/s12943-019-1055-6