Stemness and NK-like signatures define a subtype of esophageal squamous cell carcinoma with poor prognosis

Guozhong Jiang

the first hospital of zhengzhou university

Zhizhong Wang

Zhengzhou University <https://orcid.org/0000-0002-2967-811X>

Zhenguo Cheng

Zhengzhou University

WeiWei Wang

Zhengzhou University

Zifang Zhang

Zhengzhou University

Shuangshuang Lu

Zhengzhou University

Huisha Xu

Zhengzhou University

Yunshu Dong

Institute of Biophysics, Chinese Academy of Sciences

Peinan Chen

Zhengzhou University

Zhongxian Zhang

Zhengzhou University

Dongling Gao

Zhengzhou University

Zhimin Wang

Zhengzhou University

Jinxin Miao

Zhengzhou University

Xia Xue

Zhengzhou University <https://orcid.org/0000-0002-5783-1281>

Pengju Wang

Zhengzhou University

Lirong Zhang

Zhengzhou University

Rathi Gangeswaran

Queen Mary University of London

Peng Liu

Queen Mary University of London

Louisa Chard Dunmal

Queen Mary University of London

Junkuo Li

Anyang Cancer Hospital

Yongjun Guo

Zhengzhou University <https://orcid.org/0000-0003-1641-2969>

Jianzeng Dong

First Affiliated Hospital of Zhengzhou University <https://orcid.org/0000-0001-7299-7805>

Nicholas Lemoine

Queen Mary University of London

Wencai Li

Zhengzhou University

Jun Wang

Barts Cancer Institute, Queen Mary University of London <https://orcid.org/0000-0003-2509-9599>

Yaohe Wang (✉ yaohe.wang@qmul.ac.uk)

Queen Mary University of London <https://orcid.org/0000-0003-2367-6313>

Article

Keywords: esophageal cancer, ESCC, stemness

Posted Date: August 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-712426/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

To explore the biology of esophageal squamous cell carcinoma (ESCC) and identify new therapeutic opportunities, comprehensive genomic-transcriptomic characterizations of 120 Chinese ESCC patients were performed. Here we show that ESCC can be categorized into differentiated, metabolic, immunogenic and stemness subtypes based on transcriptomics, with each subtype exhibiting unique molecular and histopathological features. The stemness subgroup, with four key signature genes (*WFDC2*, *SFRP1*, *LGR6* and *VWA2*), has the poorest prognosis and was associated with downregulated immune activities, a high frequency of *EP300* mutation/activation, functional mutation enrichment in Wnt signalling and the highest level of intratumoural heterogeneity. The immune profiling further revealed ESCC tumour cells overexpress NK cell markers *XCL1/2* and *CD160* as a new immune evasion strategy. Strikingly, *XCL1* expression also affects the sensitivity of ESCC cells to common chemotherapy drugs. This study opens up new avenues for ESCC treatment and provides a unique public resource to better understand and treat ESCC.

Introduction

Esophageal cancer (EC) is an aggressive and invasive disease, associated with one of the highest mortality rates (509,000 per year) and incidence rates (572,000 new cases per year) worldwide in 2018¹. The global incidence and mortality of EC are predicted to increase in the coming decades^{1,2}. The highest prevalence of EC occurs in Asia and Africa, where the most common subtype is esophageal squamous cell carcinoma (ESCC), whereas adenocarcinoma is more common in North America and Western Europe¹. Despite advances in therapeutic options, including novel targeted therapies and cancer immunotherapies, the prognosis of ESCC remains poor, with a five-year survival of < 15%^{3,4}. The major challenges facing ESCC treatment are the aggressive progression and late diagnosis. Therefore, studies on the molecular features of ESCC to identify biomarkers for early diagnosis and key molecular alterations affecting the prognosis of the disease are crucial for an early intervention and an improved therapeutic strategy. Several major international studies have made important advances in identifying the molecular landscapes and understanding the molecular mechanisms of ESCC⁵⁻⁹. They highlighted common deregulation of RTK/RAS/PI3K and WNT/Notch pathways, cell cycle regulation, frequently mutated genes such as *TP53*, *FAT1*, *NOTCH1*, *KMT2D*, *NFE2L2* and *ZNF750*, and epigenetic alterations of ESCC¹⁰. However, the genetic events associated with the heterogeneous behaviour of ESCC are still poorly understood, resulting in a lack of robust biomarkers for predicting prognosis or designing effective targeted therapeutics. Moreover, the immune landscape and precise immune escape mechanisms of ESCC have not been fully revealed and there is no effective immunotherapeutic regime available for ESCC even though immunotherapeutic agents will be featured in standard systemic treatment for ESCC in the near future as it is significantly associated with inflammation and host immunity against dysplastic cells¹¹. Therefore, the integrated multi-omics investigation of ESCC to decipher the molecular and immune heterogeneity is required for a thorough understanding of disease pathogenesis, especially for patients from the region with the highest incidence of ESCC.

Here we present a comprehensive genomics and transcriptomics analysis of tumours with matched normal tissue in untreated ESCC patients with more than four years follow up after surgical resections. We explored the transcriptomic subtypes and diverse immune microenvironments, as well as their prognostic potential, and uncovered novel tumour-intrinsic immune escape mechanisms. We further identified significant gene and pathway alterations that contributed to the adverse phenotypes. Our study broadens the knowledge of ESCC molecular and histological diversity, and provides novel potential therapeutic targets for treatment of ESCC.

Results

Transcriptomics subtypes of ESCC with distinct prognosis

To fully understand the transcriptomic heterogeneity and molecular signatures of ESCC associated with prognosis, we investigated the transcriptomic landscape of 120 treatment-naïve ESCC tumours prospectively collected under strict protocols (Supplementary Table 1). Unsupervised clustering of RNA-seq profiling using non-negative matrix factorization revealed four stable clusters (Fig. 1a and Extended Data Fig. 1a, Supplementary Table 2). Gene set enrichment analysis (GSEA) of representative genes in each cluster annotated these subtypes as 'differentiated', 'immunogenic', 'metabolic' and 'stemness' (Fig. 1b, Supplementary Table 2 and 3). Gene signatures of these four subtypes were validated in three independent patient cohorts¹²⁻¹⁴ (Extended Data Fig. 1b). Keratinocyte differentiation and epidermis development genes such as *LCE3D*, *CDSN* and *KLK5* defined the differentiated subtype. B-cell surface markers *MS4A1*, *CD79A* and *MZB1* and T-cell chemokine ligands *CXCL9*, *CXCL13* and *CXCL10* characterized the immunogenic subtype. The metabolic subtype is associated with the upregulation of genes involved in drug metabolism by cytochrome P450 and retinol metabolism such as *GSTA1*, *ADH7*, *UGT1A10* and *UGT1A3*. High expression of *WFDC2*, *PEG10*, Wnt signalling modulator *SFRP1* and squamous cell carcinoma (SCC) stem cell marker *LGR6*¹⁵ defined the stemness subtype (Supplementary Table 2). All immune-associated pathways, such as the interferon gamma pathway, TCR pathway and chemokine-signalling pathway, were significantly downregulated in the stemness subtype (Fig. 1b). The transcription factor (TF) profiling further highlighted subtype-specific TFs, including the upregulated activity of *MYB*, *SOX10*, *SP5* and *ARNT2* in the stemness group (Extended Data Fig. 2). Interestingly, these four molecular subtypes of ESCC presented distinguishing histopathological features (Fig. 1c), reflecting the molecular characteristics. For example, keratin pearls within tumours were readily seen in the differentiated subgroup; immunogenic subgroup showed extensive immune cell infiltration within tumours; some tumour cells in the metabolic subgroup demonstrated eosinophilic cytoplasm with less immune cell infiltration, whereas the stemness subgroup presented poorly differentiated tumour cells and had few immune cell infiltrating within the tumour. Importantly, the stemness subtype was associated with the worst overall survival of all subtypes (Fig. 1d, log rank test $P = 0.028$). The significance as a prognostic biomarker of the top four genes (*WFDC1*, *SFRP1*, *LGR6* and *VWA2*) related to the stemness signature was further proven in an independent cohort of 65 ESCC patients using RT-PCR (Fig. 1e, $P = 0.031$). Strikingly, overexpression of *SFRP1* in ESCC cell line KYSE-70 could dramatically increase cell

proliferation *in vitro* and xenograft tumour growth *in vivo*, while SFRP1 knockdown in KYSE-520 cells had the opposite effects (Fig. 1f and Extended Data Fig. 3).

Tumour immune cell infiltration and NK-like phenotype

We next examined the tumour immune environment of patient tissues based on RNA-seq and determined how this correlated with the four transcriptomic subtypes and prognosis. Several *in silico* immune deconvolution published methods¹⁶⁻²¹ were benchmarked against molecular protein staining of CD4⁺ and CD8⁺ cells, tumour cellularity and copy number profiles (Methods, Extended Data Fig. 4), and the best performing method, the Danaher et al.¹⁸ signature, was used to investigate the tumour-infiltrating immune cell populations within the 120 ESCC tumours. Consensus clustering based on the immune cell estimates revealed three distinct clusters: C1 is associated with a generally low level of immune cell infiltration, but with relatively high levels of neutrophils, dendritic and mast cells; C2 is marked by a high level of immune infiltration of B cells, T cells and macrophages; C3 is correlated with a low level of infiltration of all immune cell populations, except for NK cell markers (Fig. 2a). The C3 cluster is significantly associated with high expression of NK cell markers *XCL1*, *XCL2* (Fig. 2b) and *CD160* (Extended Data Fig. 5a).

The prognostic values for all profiled immune cells were then assessed in our cohort using a multivariate Cox regression analysis, accounting for patient age, drinking and smoking history, tumour stage and grade, and the higher NK cell abundance was the strongest poor prognostic factor (Fig. 2c and 2d (China), log rank test $P = 0.019$). This negative correlation with overall survival for NK cell markers was also seen in the TCGA cohort of 90 ESCC samples (Fig. 2d (TCGA), $P = 0.015$). The comparison between the four transcriptomic subtypes and three immune profile clusters demonstrated a non-random correlation (Fisher's exact test, $p = 9.37\text{e-}11$). Fifteen of 31 cases (48.4%) from the C3 cluster were stemness tumours, while the differentiated and immunogenic subtypes were overrepresented in the C1 and C2 clusters, respectively (Fig. 2e). Of note, we also observed significantly positive correlations of mRNA expression between *LGR6* and *XCL1* ($r = 0.40$, $p < 0.0001$), *XCL2* ($r = 0.39$, $p < 0.0001$) and *CD160* ($r = 0.32$, $p = 0.0003$) (Fig. 2f).

Given the strong association between stemness and NK cell estimates in contrast to the significant absence of immune cell infiltration within tumours of the stemness subgroup (Fig. 1c), immunohistochemistry staining for XCL1, CD160 and LGR6 was performed in the matched serial sections of tumour tissues in order to determine the spatial composition and cell types that expressed these markers. Surprisingly XCL1 and CD160 were predominantly expressed in tumour cells (Fig. 2g), with a few positive immune cells infiltrated in the stroma. This was also supported by the strong positive correlation between *XCL1/2* gene expression and tumour cellularity (Extended Data Fig. 5b). XCL1 or CD160 expression colocalized with LGR6 expression (Fig. 2g, Extended Data Fig. 5c), and co-expression within tumours was seen in 16% (17/107) of our cohort for XCL1 and LGR6, and 25% (25/107) for CD160 and LGR6. Of note, XCL1 was exclusively expressed in cancer cells showing adenocarcinoma morphology and dysplastic cells in the submucosa glands, while CD160 could be expressed in both

squamous carcinoma and adenocarcinoma cells (Fig. 2g, Extended Data Fig. 5c) as well as in the proliferative and dysplastic cells of the submucosa glands (Extended Data Fig. 5d). Interestingly, we observed that *XCL1*-expressing dysplastic cells in the submucosal gland in most of the cases are completely separated from the squamous cell carcinoma, suggesting that this subgroup of patients might concurrently have both squamous cell carcinoma and adenocarcinoma or this adeno-squamous carcinoma might be derived from submucosa gland epithelial cells. Our finding of tumour cells overexpressing NK cell markers *XCL1/2* and *CD160* correlating with the lowest level of immune cell infiltration suggests a new tumour-intrinsic immune evasion mechanism of ESCC.

Molecular characteristics of *XCL1*-high NK-like cells

Characterization of the NK-like tumour cells is essential to uncover novel molecular mechanisms and pathways which may offer alternative therapeutic targets and candidate drugs for this subgroup of ESCC. We first utilized the Cancer Cell Line Encyclopedia (CCLE) RNA-seq data to analyze the association of *XCL1/2* with ESCC. Interestingly, EC cell lines had the highest level of *XCL1* mRNA expression out of all cancer types profiled (Fig. 3a). Hierarchical clustering based on differentially expressed genes ($n = 413$, $p < 0.01$) between *XCL1* high ($n = 11$) and low ($n = 11$) ESCC cell lines clearly separated the two groups (Fig. 3b, Supplementary Table 4). The GSEA results further showed that *XCL1*-high cells exhibited upregulation of drug metabolism cytochrome P450, retinol metabolism and biological oxidations (false discovery rate, FDR < 0.0001), while asparagine N-linked glycosylation, N-Glycan biosynthesis, cytosolic tRNA aminoacylation and signalling by NTRK1 were highly over-represented in *XCL1*-low cells (Fig. 3c). We also observed a significantly positive correlation in mRNA expression between *XCL1* and *LGR6* for ESCC lines (Fig. 3d, Pearson's correlation $r = 0.59$, $p = 0.004$), further highlighting the strong association between NK-like and stemness phenotypes. Interestingly, *XCL1*-high cells also showed significantly downregulated cell cycle activities compared to *XCL1*-low cells (Fig. 3e), suggesting that *XCL1*-overexpressing cells are more quiescent, slow-cycling cells.

Given the upregulation of drug metabolism cytochrome P450 in *XCL1*-high ESCC cells, we next explored the drug sensitivity to fluorouracil (5-FU), the first-line chemotherapy drug for ESCC, between *XCL1*-high and -low cell lines. All *XCL1*-high cells lines tested exhibited a much higher IC₅₀ than *XCL1*-low cells (Fig. 3f, Mann-Whitney rank test, $p = 0.007$). Impressively, over-expression of *XCL1* in an *XCL1*-low cell line KYSE-150 resulted in a two-fold increase in IC₅₀ of 5-FU compared to the control cells (Fig. 3g, Extended Data Fig. 6a and 6b) while overexpression of *XCL1* did not affect the cell proliferation (Extended Data Fig. 6d). We then explored the cancer drug sensitivity data of 367 compounds from the Genomics of Drug Sensitivity in Cancer (GDSC) resource²². Compared to *XCL1*-low cells, *XCL1*-high cell lines displayed a higher level of resistance to almost all drugs screened (Fig. 3h and Supplementary Table 5, IC₅₀ z -score t -test $p < 0.05$). *XCL1*-low cell lines exhibited a significantly higher sensitivity to Wnt-C59, LGK974 (both targeting PORCN and WNT signalling) and CP724714 (targeting ERBB2 and RTK

signalling). Deregulated pathways associated with *XCL1*-high cells may represent novel therapeutic targets for this subtype of ESCC with poor prognosis.

The genomic landscape among molecule subtypes of ESCC

Given the heterogeneous transcriptomic and immune signatures observed in our cohort, we next examined whether the genomic landscapes differ among ESCC subtypes. Whole-exome sequencing of 103 samples with matched RNA-seq were performed and analysed (Supplementary Table 6). We identified on average 358 somatic mutations (range 21 – 3,583) and 152 non-silent mutations (range 9 – 1,242) per sample respectively. Overall and subclonal mutation burden and tumour purity were not correlated (Extended Data Fig. 7a). The overall mutational load, the number of non-silent mutations, and somatic copy number profiling were similar among the four ESCC subtypes (Extended Data Fig. 7b). Three driver gene detection methods, MutsigCV^{23,24}, dNdScv²⁵, OncodriveFM²⁶, were applied, and 10 ESCC driver genes were identified by at least two methods (Extended Data Fig. 7c), which had all been detected in ESCC previously^{5,7-9,27,28}. Incorporating representative genomic alterations of the three molecular subtypes ESCC1/2/3 previously identified from 90 TCGA ESCCs¹³ and driver genes that have been implicated in large series of Chinese ESCCs^{5,27}, we found that although the alteration frequencies (i.e., mutations, amplifications and deletions) of most significant genes in ESCC were similar among our subtypes, the stemness type had a higher frequency of *NOTCH1* and *EP300* alterations, affecting 39% and 26% of stemness tumours, respectively (Fig. 4a, Fisher's exact test, $p = 0.018$ and 0.008 for *NOTCH1* and *EP300*, respectively). Interestingly, no alterations in *NOTCH1* and *EP300* were detected in the metabolic subtype, and no *ZNF750* alterations were detected in the differentiated subtype. Stemness tumours seemed to have the highest overall frequencies of ESCC1 and ESCC2 alterations (74% and 65% of cases, respectively). Across our cohort, although the highest overall frequency was observed for ESCC1 alterations (61%), other subtype alterations also occurred substantially, 50% for ESCC2 and 52% for ESCC3 (Extended Data Fig. 7d).

Next, we investigated the genomic alterations of ESCC functional genes identified above across ESCC cell lines. Interestingly, in *XCL1*-high cells, alterations in *EP300* occurred in 4/11 (36%) cases, compared to zero cases in *XCL1*-low cells (Fig. 4b), corresponding to our result of the highest *EP300* alteration frequency in the stemness group. Whereas alterations in *NOTCH1*, *FAT1* and *KDM6A* seemed to be more prevalent, 64%, 55% and 45%, in *XCL1*-low cells, respectively, than in *XCL*-high cells, 18%, 9% and 9%, respectively (Fig. 4b). *EP300* is a histone acetyltransferase, and its mutation has been shown to be associated with poor outcome in ESCC and HNSCC^{29,30}. We also observed significantly poorer overall survival for patients with *EP300* mutations in our cohort (Fig. 4c, HR = 5.248, $p = 8e-06$). When combining with *CREBBP* mutations, the *EP300/CREBBP* mutation status was still predictive of worse survival (Fig. 4d, HR = 2.229, $p = 0.02$). Although *CREBBP* mutations were not enriched in the stemness subtype, alterations in *EP300* and *CREBBP* tended to co-occur ($n=4$) only in stemness samples (Fig. 4a). Furthermore, *EP300*-mutated patient samples ($n=8$, 7.8%) had significantly higher *EP300* expression

levels than *EP300*-wildtype samples (Fig. 4e, Wilcoxon rank test, $p < 0.01$), and this was largely due to the unbalanced overexpression of the mutated alleles (Fig. 4f), suggesting the gain-of-function mode for these mutations.

***EP300* mutations and overexpression and their relationship with stemness and NK-like phenotype in ESCC**

We next examined whether *EP300* mutations and expression had any associations with stemness and NK-like gene signatures. We first performed differential expression analysis between *EP300*-mutated and wildtype samples, followed by GSEA against representative (i.e., upregulated) gene sets of four transcriptomic subtypes and *XCL1*-high genes derived from cell lines. We found that the gene set representative of stemness subtype was upregulated in *EP300*-mutated compared to wildtype samples (NES = 1.36, FDR = 0.05), while all gene sets of other three subtypes were massively downregulated in *EP300*-mutated samples (Fig. 5a and 5b, FDR < 0.001, Extended Data Fig. 8a). Although there was no upregulation of *XCL1*-high genes, we observed significantly increased gene expression of *CD160* in *EP300*-mutated samples (Fig. 5c, $p = 0.003$). Interestingly, the *EP300* gene expression was also the highest in the stemness subtype, and there was a positive correlation in expression between *EP300*, and *XCL1/2* and *CD160* (Extended Data Fig. 8b and 8c).

Next, focusing on *EP300*-wildtype samples, we selected top 20 *EP300*-high and 20 *EP300*-low expression samples, and found that the stemness gene set and *XCL1*-high genes were significantly upregulated in the *EP300*-high group (Fig. 5a and 5b, 'stemness up' gene set, NES = 2.49, FDR < 0.0001; *XCL1*-high gene set, NES = 1.47, FDR = 0.014), whereas the differentiated gene set was greatly downregulated (NES = -3.47, FDR < 0.0001). Interestingly, we found the metabolic gene set was significantly upregulated in *EP300*-high samples (NES = 2.71, FDR < 0.0001), while the immunogenic gene set was not notably affected by *EP300* expression levels (Fig. 5a). Taking all these together, our results demonstrate a strong positive association between *EP300* mutations, overexpression and stemness / NK-like signatures, suggesting that *EP300* may have a role promoting tumours to become a more aggressive subtype in ESCC.

Functional mutation enrichment in pathways among ESCC subtypes

To further investigate whether specific pathway-related mutations were associated with transcriptomic subtypes, we performed functional mutation enrichment analysis adjusted for sample mutation load using the 50 cancer hallmark gene sets (see Methods). We identified functionally relevant mutations in Wnt/ β -catenin signalling which were significantly more enriched in the stemness samples, while functional mutations in inflammatory response, angiogenesis and hypoxia were highly enriched in the immunogenic tumours (Fig. 5d, Kruskal–Wallis test, $p < 0.05$, Extended Data Fig. 9a). Of note, the

hallmark gene set expression profile using GSVA also confirmed the functional differences identified among the four subtypes (Extended Data Fig. 9b). We then evaluated whether the pathway functional mutation enrichment affected the expression of pathway genes, and found that mutation enrichment in Wnt/ β -catenin signalling, inflammatory response and hypoxia were positively correlated with pathway gene expression activities for the corresponding ESCC subtypes (Fig. 5e), suggesting that high level of Wnt signalling expression in the stemness group could be a consequence of functional mutations in regulators of the Wnt pathway, and these functional mutations were the most enriched in stemness samples. This is also the case for the high pathway activity of inflammatory response and hypoxia for the immunogenic subtype. The mutations in the Wnt signalling pathway were highly deleterious as predicted by SIFT³¹ and PolyPhen-2³², and the stemness group seemed to have the highest proportion of deleterious Wnt mutations (93%), significantly higher than that of the differentiated subtype (60%, Fisher's exact test, $p = 0.035$, Extended Data Fig. 9c).

Clonality analysis among molecule subtypes of ESCC

Given the well-documented interplay between immune infiltration and tumour clonal evolution^{33,34}, we finally investigated the level of intratumoural heterogeneity (ITH, measured by Shannon diversity index) based on variant allele frequencies of all mutations in each tumour among the immune and transcriptomic subtypes. Our results showed the greatest ITH in the C3 cluster, whereas in C2 where the immune infiltration was the highest, the diversity was greatly reduced (Fig. 5f). In addition, the highest ITH presented in the stemness subtype while the immunogenic subtype had the lowest ITH (Fig. 5f). Our results suggest that the immune microenvironment strongly influences the tumour evolution and (sub)clonal architecture. The stemness / C3 group with the highest level of immune exclusion also had the highest tumour (sub)clonal diversity. Importantly, high ITH was also a significant marker of poor prognosis in ESCC (Fig. 5g, multivariate analysis, hazard ratio HR = 2.214, log rank $P = 0.007$).

Discussion

To our knowledge, this study is the first comprehensive genomic-transcriptomic characterization of the largest cohort of treatment-naïve ESCC patients to date with long-term follow up data from high incidence areas. Here we performed a thorough integrated analysis of genetic alterations, gene expression and immune cell infiltration, and examined their correlations with clinical and pathological data. We validated our results using independent datasets, and *in vitro* and *in vivo* biological experiments. Our study offers important insights into genetic events that drive ESCC diversity and progression, and reveals a new mechanism of ESCC immune evasion.

Firstly, four distinct subtypes of ESCC (differentiated, immunogenic, metabolic and stemness) have been identified and validated in our study, and each subtype shows a unique molecular signature and histopathological changes. Previously, the integrated genomic characterization of EC had divided ESCC

into three molecular subtypes^{13,35}. These ESCC subtypes showed trends for geographic associations while their correlations with clinical outcome were not demonstrated. The limitation also lies in the fact that the cases originated from the regions with moderate or lower ESCC incidence. Upon investigating their subtype-specific alterations in our cohort and subtypes, we found that the overall frequencies were high and comparable for all three previous subtypes, although stemness tumours seemed to have the highest frequencies of ESCC1 and ESCC2 alterations. The ESCC samples of our study were from the high incidence populations of China where ESCC accounts for the vast majority of EC (> 90%). Identification of robust biomarkers in these ESCC has tremendous clinical implications, given that 70% of global EC cases occur in China. Indeed, we found that the stemness subgroup has the poorest prognosis compared to other subgroups. We functionally validated the role of one of the top stemness genes, *SFRP1* (a WNT signaling modulator) in the progression of ESCC. We proved that *SFRP1* significantly enhances the malignant phenotypes of ESCC cells *in vitro* and progression *in vivo*, suggesting that targeting the WNT pathway via *SFRP1* may be a promising strategy for treatment of ESCC as ESCC has a high frequency (up to 86%) of alterations in the WNT pathway⁷.

Immune evasion is a hallmark of cancer³⁶. Cancer immunotherapy has revolutionized the profile of cancer therapies over the past decade. While long-term survival is observed for a fraction of cancer patients, the majority of patients, including EC patients currently do not benefit from immunotherapy treatments such as immune checkpoint blockade therapy¹¹, emphasizing the need to identify the genomic and molecular determinants underpinning immune evasion of ESCC. We, for the first time, identified ESCC cells overexpressing natural killer (NK) cell markers such as XCL1/2 and CD160. The overexpression of these markers is significantly associated with a shorter overall survival of ESCC patients. XCL1 is a C class chemokine and is produced predominantly by NK and activated CD8⁺ T cells^{37,38}, and the XCL1-XCR1 axis normally plays a crucial role in the induction of effective cytotoxic immunity³⁹. A recent study demonstrated that XCL1 expression correlates with CD8-positive T cell infiltration and PD-L1 expression in mature ovarian cystic teratomas, but there is no correlation between XCL1 expression and prognosis or clinical stage⁴⁰. However, in our samples, the *XCL1*-overexpressing tumours showed the lowest immune cell infiltration. Interestingly, the Human Protein Atlas data also provided the evidence of *XCL1*-overexpressing tumour cells in a number of cancer types, and this overexpression is significantly correlated with a shorter overall survival rate in patients with colorectal and renal cancer (Extended Data Fig. 6c). The immunogenic role of XCL1 and its underlying functional mechanisms in the development of ESCC and other tumours need further investigation. CD160, a glycosylphosphatidylinositol-anchored Ig domain protein that is expressed on NK cells, $\gamma\delta$ T cells and a minor subset of CD4⁺ and CD8⁺ T cells, is also overexpressed in some ESCC tumours in our study. Previous studies demonstrated that CD160 is dramatically increased in B cell malignancies^{41,42}. CD160 can function as both a co-activating and co-inhibitory receptor, depending on which receptor/ligand is operating in the context of neighboring interactions⁴³. Given the negative correlation of CD160 expression with clinical prognosis in ESCC and other tumours such as renal cancer (Extended Data Fig. 5e), we postulate that CD160-expressing cancer cells could use the HVEM and BTLA inhibitory

pathways to inhibit T and NK cells activities^{43,44}. These interesting preliminary data warrant further investigation into the detailed pathways of how CD160 expression in cancer cells modulates host immunity. The interactions of CD160 with its ligands may be important in the pathophysiology of ESCC, offering new targets for therapeutic manipulation.

Recent studies have showed that cancer stem cells (CSCs) have the ability to hide from the immune system *ab initio*, evading the immunosurveillance phase^{45,46}. Interestingly, we found that in the C3 immune cluster of ESCC characterized by high level of *XCL1* and *CD160*, ~ 50% of tumours are from the stemness subtype which presents a significant WNT alteration. A recent study demonstrated that latency competent cancer (LCC) cells from early stage human lung and breast carcinoma can enter a quiescence state through expression of the WNT inhibitor DKK1 and a broad downregulation of ULBP ligands for NK cells. These cells evade NK-cell-mediated clearance to remain latent for extended periods⁴⁶. In this study, we also observed the significant association between WNT dysregulation and a stemness slow-cycling state, and identified an unreported immune evasion of ESCC stem-like cells: masking the tumour cells with expression of NK cell markers. These findings, together with those previously discussed, highlight the critical interaction between key signaling pathways crucial for stem-cell propagation and the mechanisms that guide immune evasion. A deeper understanding of the unique interactions between cancer stem-like cells and the immune system could provide ground for developing novel therapeutic strategies that can harness the immune system against the “hardest immune-evaders”.

When investigating the differences in genomic changes among ESCC subtypes, we noted a strong positive association between *EP300* mutations / overexpression and stemness / NK-like signatures. Histone acetyltransferase p300 is a crucial transcriptional coactivator. Along with CBP, they regulate transcription of thousands of genes in a cell via chromatin remodeling and histone modification, playing an important role in several fundamental biological processes including proliferation, cell cycle, cell differentiation, and DNA damage response⁴⁷. Of note, how cancer cells utilize p300 seems to be context dependent; both loss- and gain-of-function genetic alterations in p300 have been reported across solid tumours and hematological malignancies^{47–50}. Our data suggest that *EP300* alternations potentially have an activating role in ESCC, promoting tumours to become stem-like phenotype, which leads to immune exclusion, drug resistance and worse clinical outcome. It has been shown that overexpression of EP300 led to up-regulation of mesenchymal and stemness markers, increases in migration, invasion, anchorage-independent growth and drug resistance in a breast cancer cell model⁵¹. Similar observation was also obtained in non-small cell lung cancer and nasopharyngeal carcinoma cells^{52,53}. EP300 knockdown, on the other hand, reduces cancer stem cell phenotype, EMT, tumour growth and metastasis in these cancers^{52–54}, further supporting its oncogenic role and potential involvement in stem-like phenotype. Indeed, EP300 mutations and overexpression correlate with adverse prognosis in many solid cancers, including ESCC^{29,30}. Importantly, it has been shown that pharmacological inhibition of CEP/p300 KAT activities sensitizes cells to DNA-damaging chemotherapeutic agents and radiation^{30,55}. Thus, more work lies ahead to test similar combinational therapies to treat these resistant stemness / NK-like ESCC cells *in vitro* and *in vivo*.

In summary, this in-depth analysis of genomics-transcriptomics of Chinese ESCC patients provides insights into the nature of ESCC tumours. These findings pave the way for us to develop more effective diagnostic and therapeutic approaches for ESCC. In addition, the data created from this study, especially from treatment-naïve ESCC patients with follow-up more than four years after surgery, provides a unique public resource to better understand and treat ESCC.

Declarations

Acknowledgments

This project is supported by the National Key R&D program of China (2016YFE0200800), the Nature Sciences Foundation of China (U1704282 and 81771776) and the core funding for development of the Cell and Gene Therapy Programme by Zhengzhou University. JW, NRL and YW also acknowledge support from the Cancer Research UK Centre of Excellence Award to Barts Cancer Centre (C355/A25137). JW acknowledges support from the Academy of Medical Sciences Springboard Award (SBF003\1025), and Cancer Research UK (C355/A26819) and FC AECC and AIRC under the Accelerator Award Program.

Author Contributions

YW conceived and supervised this study. YW, JW and WL were responsible for all aspects of study design and managed the project with the support by DG; GJ and WW were in charge of the collection of all samples and clinical data, DNA and RNA extraction, with the help from PC, ZZ(Zhongxian Zhang) and JL. ZW (Zhizhong Wang) led the bioinformatic analysis with support by YD and XX under the supervision by JW; ZC led and performed the *in vitro* and *in vivo* biological function validation studies with HX, ZZ (Zifang Zhang), SL, JM, ZW(Zhimin Wang) and PL; DG and RG did histopathology and IHC staining, YW, WL, ZC, RG and YW reviewed H&E and IHC slides and prepared the representative pictures; LZ, LSCD, YG, J.D and NL participated in interpretation of some experiments and critically reviewed the manuscript. JW, ZW (Zhizhong Wang), ZC and YW interpreted all results and wrote the manuscript.

Declaration of Interests

All the authors declare no competing financial interests.

All animal care and experiments were approved by the Ethical Committee of the Zhengzhou University and were in accordance with the Provision and General Recommendation of Chinese Experimental Animals Administration Legislation.

References

1. Bray, F., *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394–424 (2018).
2. Malhotra, G.K., *et al.* Global trends in esophageal cancer. *J Surg Oncol* **115**, 564–579 (2017).
3. Abnet, C.C., Arnold, M. & Wei, W.Q. Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology* **154**, 360–373 (2018).
4. Smyth, E.C., *et al.* Oesophageal cancer. *Nat Rev Dis Primers* **3**, 17048 (2017).
5. Cui, Y., *et al.* Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma. *Cell Res* **30**, 902–913 (2020).
6. Sawada, G., *et al.* Genomic Landscape of Esophageal Squamous Cell Carcinoma in a Japanese Population. *Gastroenterology* **150**, 1171–1182 (2016).
7. Song, Y., *et al.* Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* **509**, 91–95 (2014).
8. Wu, C., *et al.* Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. *Nat Genet* **46**, 1001–1006 (2014).
9. Yan, T., *et al.* Multi-region sequencing unveils novel actionable targets and spatial heterogeneity in esophageal squamous cell carcinoma. *Nat Commun* **10**, 1670 (2019).
10. Cao, W., *et al.* Multi-faceted epigenetic dysregulation of gene expression promotes esophageal squamous cell carcinoma. *Nat Commun* **11**, 3675 (2020).
11. Kojima, T., *et al.* Randomized Phase III KEYNOTE-181 Study of Pembrolizumab Versus Chemotherapy in Advanced Esophageal Cancer. *J Clin Oncol*, JCO2001888 (2020).
12. Li, J., *et al.* LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut* **63**, 1700–1710 (2014).
13. Cancer Genome Atlas Research, N., *et al.* Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).
14. Sawada, G., *et al.* An Integrative Analysis to Identify Driver Genes in Esophageal Squamous Cell Carcinoma. *PLoS One* **10**, e0139808 (2015).
15. Huang, P.Y., *et al.* Lgr6 is a stem cell marker in mouse skin squamous cell carcinoma. *Nat Genet* **49**, 1624–1632 (2017).
16. Chen, B., Khodadoust, M.S., Liu, C.L., Newman, A.M. & Alizadeh, A.A. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* **1711**, 243–259 (2018).
17. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
18. Danaher, P., *et al.* Gene expression markers of Tumor Infiltrating Leukocytes. *J Immunother Cancer* **5**, 18 (2017).
19. Becht, E., *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* **17**, 218 (2016).

20. Davoli, T., Uno, H., Wooten, E.C. & Elledge, S.J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**(2017).
21. Li, T., *et al.* TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res* **77**, e108-e110 (2017).
22. Yang, W., *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* **41**, D955-961 (2013).
23. Tokheim, C.J., Papadopoulos, N., Kinzler, K.W., Vogelstein, B. & Karchin, R. Evaluating the evaluation of cancer driver genes. **113**, 14330–14335 (2016).
24. Lawrence, M.S., *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
25. Martincorena, I., *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029–1041 e1021 (2017).
26. Gonzalez-Perez, A. & Lopez-Bigas, N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res* **40**, e169 (2012).
27. Chang, J., *et al.* Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. *Nat Commun* **8**, 15290 (2017).
28. Li, X.C., *et al.* A mutational signature associated with alcohol consumption and prognostically significantly mutated driver genes in esophageal squamous cell carcinoma. *Ann Oncol* **29**, 938–944 (2018).
29. Bi, Y., *et al.* EP300 as an oncogene correlates with poor prognosis in esophageal squamous carcinoma. *J Cancer* **10**, 5413–5426 (2019).
30. Kumar, M., *et al.* CREBBP/EP300 mutation is associated with poor outcome in HNSCC and targetable with synthetic cytotoxicity. *bioRxiv* (2020).
31. Ng, P.C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812–3814 (2003).
32. Adzhubei, I.A., *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248–249 (2010).
33. Losic, B., *et al.* Intratumoral heterogeneity and clonal evolution in liver cancer. *Nat Commun* **11**, 291 (2020).
34. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).
35. Frankell, A.M., *et al.* The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet* **51**, 506–516 (2019).
36. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
37. Kelner, G.S., *et al.* Lymphotactin: a cytokine that represents a new class of chemokine. *Science* **266**, 1395–1399 (1994).
38. Lei, Y. & Takahama, Y. XCL1 and XCR1 in the immune system. *Microbes Infect* **14**, 262–267 (2012).

39. Dorner, B.G., *et al.* Selective expression of the chemokine receptor XCR1 on cross-presenting dendritic cells determines cooperation with CD8 + T cells. *Immunity* **31**, 823–833 (2009).
40. Tamura, R., *et al.* XCL1 expression correlates with CD8-positive T cells infiltration and PD-L1 expression in squamous cell carcinoma arising from mature cystic teratoma of the ovary. *Oncogene* **39**, 3541–3554 (2020).
41. Farren, T.W., *et al.* Differential and tumor-specific expression of CD160 in B-cell malignancies. *Blood* **118**, 2174–2183 (2011).
42. Liu, F.T., *et al.* CD160 signaling mediates PI3K-dependent survival and growth signals in chronic lymphocytic leukemia. *Blood* **115**, 3079–3088 (2010).
43. Kim, T.J., *et al.* CD160 serves as a negative regulator of NKT cells in acute hepatic injury. *Nat Commun* **10**, 3258 (2019).
44. Cai, G., *et al.* CD160 inhibits activation of human CD4 + T cells through interaction with herpesvirus entry mediator. *Nat Immunol* **9**, 176–185 (2008).
45. Bruschini, S., Ciliberto, G. & Mancini, R. The emerging role of cancer cell plasticity and cell-cycle quiescence in immune escape. *Cell Death Dis* **11**, 471 (2020).
46. Malladi, S., *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* **165**, 45–60 (2016).
47. Attar, N. & Kurdistani, S.K. Exploitation of EP300 and CREBBP Lysine Acetyltransferases by Cancer. *Cold Spring Harb Perspect Med* **7**(2017).
48. Gayther, S.A., *et al.* Mutations truncating the EP300 acetylase in human cancers. *Nat Genet* **24**, 300–303 (2000).
49. Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E. & Panne, D. Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nat Struct Mol Biol* **20**, 1040–1046 (2013).
50. Pasqualucci, L., *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189–195 (2011).
51. Mahmud, Z., *et al.* Oncogenic EP300 can be targeted with inhibitors of aldo-keto reductases. *Biochem Pharmacol* **163**, 391–403 (2019).
52. Hou, X., *et al.* p300 promotes proliferation, migration, and invasion via inducing epithelial-mesenchymal transition in non-small cell lung cancer cells. *BMC Cancer* **18**, 641 (2018).
53. Liao, Z.W., *et al.* P300 promotes migration, invasion and epithelial-mesenchymal transition in a nasopharyngeal carcinoma cell line. *Oncol Lett* **13**, 763–769 (2017).
54. Ring, A., Kaur, P. & Lang, J.E. EP300 knockdown reduces cancer stem cell phenotype, tumor growth and metastasis in triple negative breast cancer. *BMC Cancer* **20**, 1076 (2020).
55. Yan, G., *et al.* Selective inhibition of p300 HAT blocks cell cycle progression, induces cellular senescence, and inhibits the DNA damage response in melanoma cells. *J Invest Dermatol* **133**, 2444–2452 (2013).

56. Andrews, S.a.K., Felix and Segonds-Pichon, Anne and Biggins, Laura and Krueger, Christel and Wingett, Steven. FastQC. (2012).
57. Chen, Y., *et al.* SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1–6 (2018).
58. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
59. Soneson, C., Love, M.I. & Robinson, M.D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
60. Hansen, K.D., Irizarry, R.A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
61. Ritchie, M.E., *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
62. Subramanian, A., *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).
63. Brunet, J.P., Tamayo, P., Golub, T.R. & Mesirov, J.P. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 4164–4169 (2004).
64. Wilkerson, M.D. & Hayes, D.N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
65. Dennis, G., Jr., *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
66. Aran, D., Hu, Z. & Butte, A.J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* **18**, 220 (2017).
67. Rosenthal, R., *et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
68. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
69. McKenna, A., *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
70. Cibulskis, K., *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213–219 (2013).
71. Kim, S., *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* **15**, 591–594 (2018).
72. McLaren, W., *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
73. Favero, F., *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* **26**, 64–70 (2015).

- 74. Murugaesu, N., *et al.* Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov* **5**, 821–831 (2015).
- 75. Roth, A., *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* **11**, 396–398 (2014).
- 76. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
- 77. Oh, B.Y., *et al.* Intratumor heterogeneity inferred from targeted deep sequencing as a prognostic indicator. *Scientific Reports* **9**(2019).

Figures

Figure 1

Transcriptomic subtypes of Chinese ESCC. a, Four distinct transcriptomic subtypes were identified using non-negative matrix factorization (NMF). Expression heatmap of all representative genes from the four clusters is displayed, and the top five representative genes are shown next to each cluster. Each row represents a representative gene, and each column represents a patient. b, Heatmap of top enriched pathways for each subtype is shown. Each row represents a significant pathway curated from the

mSigDB database (v.6.2). Four subtypes are shown in column. The ‘-log₁₀’ transformed p-values from the hypergeometric test were used to generate this heatmap. Red indicates that the pathway is highly enriched for the gene set. Blue indicates that no enrichment was observed for the gene set. For the stemness subtype, results of three mostly downregulated pathways, ‘interferon gamma signaling’, ‘TCR pathway’ and ‘chemokine signaling pathway’ from GSEA are shown. Normalised enrichment scores (NES) and FDR values are also displayed. c, Representative histopathology images for the four subtypes are shown. d, A Kaplan–Meier curve is shown comparing patients from the four subtypes. e, A Kaplan–Meier curve is shown for patient samples with high and low stemness signature in an independent cohort (n = 63). The stemness signature was measured as the average expression readout of four genes, LGR6, VWA2, WFDC1 and SFRP1, by RT-PCR. The patients were split into high and low groups based on an optimal cutoff with R survminer package (see Methods). For all survival curves, significance was determined using a two-sided log-rank test. f, The effect of SFRP1 overexpression (in KYSE-70, n = 6) or knock-down (in KYSE-520, n = 6) on the tumour growth of ESCC was evaluated by the tumour growth of SFRP1 modified ESCC cells in immune-deficient mice. All mice in the overexpression group developed tumours, while two mice in the knockdown group had no tumour formation. The tumour size is presented at the end time point of the study (30 days after transplantation of the ESCC cells).

Figure 2

Heterogeneity of immune cell infiltration in ESCC. a, The immune cell infiltration profiling across our cohort is shown, clustered by the level of estimated immune cell infiltration. Each row represents an immune cell type as estimated by the method used by DanaHER et al. Immune cells are natural killer (NK) cells, neutrophils, B cells, macrophages, CD4+ mature cells, regulatory T cells (Treg), CD56dim NK cells, total T cells (T cells), CD8+ T cells, cytotoxic cells, exhausted CD8+ T cells (exhausted CD8), dendritic

cells (DC) and mast cells. Consensus clustering was performed. Each column represents a patient sample. Three immune infiltration clusters were identified; C1, C2 and C3. b, Levels of gene expression of XCL1 and XCL2 are shown among the three immune subtypes as a box and whisker plot. Significance in each pairwise comparison using the Wilcoxon rank-sum test is shown as, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$. c, The survival analysis of all profiled immune cell types against overall survival is shown. The hazard ratio (HR) derived from the multivariate Cox regression model is shown as a whisker plot. The squared dot indicates the HR value and whiskers shows the 95% confidence interval of HR estimate. Significance is determined using a two-sided log-rank test (■ $p < 0.1$; * $p < 0.05$). d, A Kaplan–Meier curve is shown for NK cell estimates against overall survival for our cohort (China) and TCGA. Multivariate survival analysis was performed for China cohort. HR and p-value derived from log-rank test are shown. e, The number of cases of the four transcriptomic subtypes is shown among the three immune subtypes C1, C2 and C3. f, The scatter plot of expression levels between LGR6 and three NK cell markers, XCL1, XCL2 and CD160, is shown. Pearson's correlation coefficient and associated p-value are displayed. g, IHC (Immunohistochemistry) staining of XCL1 and LGR6 from one patient Sample 427, and IHC of CD160 and LGR6 from a different patient Sample 401 are shown. The IHC results show that XCL1 and LGR6, CD160 and LGR6 are co-expressed in tumour cells.

normalised enrichment score (NES). d, The scatter plot in gene expression $\log_2(\text{RPKM}+1)$ between XCL1 and LGR6 for 22 ESCC lines is shown. e, The cell cycle activities were downregulated in XCL1-high cells compared to XCL1-low cells. f, The cytotoxicity of 5-FU in a panel of human ESCC cell lines is shown. The cells are divided into XCL1 high (red) and low (blue) groups based on the CCLE separation. $\log_2$ transformed IC50 scores are shown. g. The IC50 of 5-FU between control and XCL1 overexpressing cells of KYSE-150 is shown. h, The drug screening profiling between ESCC XCL1-high and low cells is shown, based on data generated by the Genomics of Drug Sensitivity in Cancer (GDSC) resource. The drugs that show significant difference in IC50 (Student's T-test, $p < 0.05$) between the two groups are selected. High and low levels of resistance are indicated in red and blue, respectively.

Figure 4

The landscape of genomic alterations in ESCC. a, The genomic landscape of ESCC driver genes and previously reported ESCC subtype specific genes is shown among four transcriptomic subtypes in our cohort. ESCC1/2/3 denotes the ESCC subtypes identified by TCGA 2017 EC study. Additional significant genes reported from previous large Chinese ESCC cohort studies were also included. For copy number aberrations, only amplifications and deletions were included. Cases of copy gain or loss were not

counted. b, The genomic landscape of significant ESCC genes across XCL1-high and low ESCC cell lines. c, The overall survival of EP300-mutated (Mut) versus wildtype (Wt) cases. d, The overall survival of EP300 and/or CREBBP mutated (Mut) versus wildtype (Wt) samples. e, EP300 gene expression levels between EP300-mutated and wildtype samples, ** $p < 0.01$. Two cases with amplification or deletion only were excluded. f, Reference and alternative allele counts and percentages between DNA WES and RNA-seq for three EP300-mutated cases, 369, 390 and 463. Fisher's exact test was performed to determine the allelic imbalance between DNA and RNA, * $p < 0.05$, ***, $p < 0.001$.

Figure 5

EP300 mutations, overexpression in ESCC and pathway functional mutation enrichment. a, GSEA normalised enrichment score (NES) of EP300 Mut (mutated) versus Wt (wildtype) samples (green bars) and EP300 high versus low expression samples (red bars), against representative (upregulated) gene sets of four transcriptomic subtypes and upregulated genes in XCL-high ESCC cell lines (XCL_up). b, GSEA plots of 'stemness_up' and 'differentiated_up' gene sets for the EP300 Mut (mutated) versus Wt (wildtype) comparison, and the 'stemness_up' and 'XCL1_up' gene sets for the EP300 high versus low-expression comparison. FDR q-values were shown. c, Gene expression of CD160 between EP300 mutated and wildtype samples. d, Pathway functional mutation enrichment adjusted for tumour mutation burden and the correlation between functional mutation enrichment ratio and tumour cellularity are plotted together for each hallmark gene set. P-values derived from the Kruskal–Wallis test comparing the enrichment ratio among the four subtypes were used for y-axis values. For significant hallmark gene sets, they were coloured to represent the subtype which had the highest enrichment for this gene set. e, Significance p-values ($-\log_{10}$ transformed) comparing the enrichment scores among the four subtypes against p-values comparing GSVA values among the four subgroups are plotted for each hallmark gene set. For significant hallmark gene sets that passed both significance threshold ($p < 0.05$), they were coloured to represent the subtype which had the highest enrichment for that gene set in both functional mutation and pathway expression levels. f, The intra-tumour heterogeneity, measured as the Shannon density, is shown across the four transcriptomic and immune subtypes, using the Wilcoxon rank-sum test. g, The survival analysis of Shannon density index against overall survival is shown using a multivariate analysis. All significance shown in the figure, • $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, and **** $p < 0.0001$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [t.pdf](#)
- [Supplementarydatafigurestosubmit.pdf](#)
- [SupplementarytablesESCCpapertosubmit.xlsx](#)