

# SCIENTIFIC REPORTS



OPEN

## Stroma-associated master regulators of molecular subtypes predict patient prognosis in ovarian cancer

Received: 17 July 2015

Accepted: 06 October 2015

Published: 04 November 2015

Shengzhe Zhang<sup>1,2,\*</sup>, Ying Jing<sup>1,\*</sup>, Meiyong Zhang<sup>3,4,\*</sup>, Zhenfeng Zhang<sup>1,5</sup>, Pengfei Ma<sup>1</sup>, Huixin Peng<sup>1</sup>, Kaixuan Shi<sup>2</sup>, Wei-Qiang Gao<sup>1,2</sup> & Guanglei Zhuang<sup>1,4</sup>

High-grade serous ovarian carcinoma (HGS-OvCa) has the lowest survival rate among all gynecologic cancers and is hallmarked by a high degree of heterogeneity. The Cancer Genome Atlas network has described a gene expression-based molecular classification of HGS-OvCa into Differentiated, Mesenchymal, Immunoreactive and Proliferative subtypes. However, the biological underpinnings and regulatory mechanisms underlying the distinct molecular subtypes are largely unknown. Here we showed that tumor-infiltrating stromal cells significantly contributed to the assignments of Mesenchymal and Immunoreactive clusters. Using reverse engineering and an unbiased interrogation of subtype regulatory networks, we identified the transcriptional modules containing master regulators that drive gene expression of Mesenchymal and Immunoreactive HGS-OvCa. Mesenchymal master regulators were associated with poor prognosis, while Immunoreactive master regulators positively correlated with overall survival. Meta-analysis of 749 HGS-OvCa expression profiles confirmed that master regulators as a prognostic signature were able to predict patient outcome. Our data unraveled master regulatory programs of HGS-OvCa subtypes with prognostic and potentially therapeutic relevance, and suggested that the unique transcriptional and clinical characteristics of ovarian Mesenchymal and Immunoreactive subtypes could be, at least partially, ascribed to tumor microenvironment.

High-grade serous ovarian carcinoma (HGS-OvCa) is the most lethal gynecological cancer and represents a clinically heterogeneous disease<sup>1–3</sup>. For example, essentially all patients diagnosed with advanced disease undergo very similar standard treatment, which is aggressive surgical debulking followed by multi-cycles of platinum-based combination chemotherapy<sup>4</sup>. However, approximately 30% of cases exhibit intrinsic chemoresistance and gain little or no benefit. Additionally, a large percentage of chemosensitive patients develop acquired resistance and eventually relapse within various time windows<sup>5,6</sup>. Therefore, it is important to leverage novel prognostic tools to stratify seemingly identical patients and redirect them to more precise therapies that may be potentially efficacious.

<sup>1</sup>State Key Laboratory of Oncogenes and Related Genes, Renji-Med X Clinical Stem Cell Research Center, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. <sup>2</sup>School of Biomedical Engineering & Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China. <sup>3</sup>Department of Obstetrics and Gynecology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. <sup>4</sup>Shanghai Key Laboratory of Gynecologic Oncology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. <sup>5</sup>State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to W.-Q.G. (email: gao.weiqiang@sjtu.edu.cn) or G.Z. (email: zhuangguanglei@gmail.com)

To complement conventional histopathology, major efforts have recently been focused on the molecular classifications enabled by large-scale global gene expression profiling studies. Several groups have used microarray-based gene expression datasets to retrospectively classify HGS-OvCa patients into prognostic and/or molecular subtypes<sup>7</sup>. Using k-means clustering, Tothill *et al.* reported six molecular subtypes in 285 serous and endometrioid tumors, and defined a poor prognosis subtype by a reactive stroma gene expression signature<sup>8</sup>. Tan *et al.* presented a meta-analysis of epithelial ovarian cancer and identified five distinct subgroups, which exhibited significantly different patient outcome<sup>9</sup>. Nevertheless, these classification schemes have not yet achieved widespread application, partly due to the lack of imperative understanding of biologic rationale that determines the transcriptional and clinical characteristics of diverse subtypes.

Recently, the Cancer Genome Atlas (TCGA) network identified four HGS-OvCa subtypes<sup>10</sup>, namely Differentiated, Mesenchymal, Immunoreactive and Proliferative, which were subsequently validated in an independent patient cohort (Mayo Clinic cohort)<sup>11</sup>. Surprisingly, however, survival time did not differ significantly for the transcriptional subtypes in the TCGA HGS-OvCa dataset<sup>10</sup>, in contrast to the clinical relevance of molecular classifiers evident in other cancers<sup>12–14</sup>. Counterintuitively, a statistically significant difference in patient survival was observed in the Mayo Clinic cohort, i.e. the Immunoreactive subtype had the longest survival time, while the Mesenchymal subtype had the shortest. These inconsistent findings necessitate further prudent investigations before employing the TCGA subtyping in patient stratification.

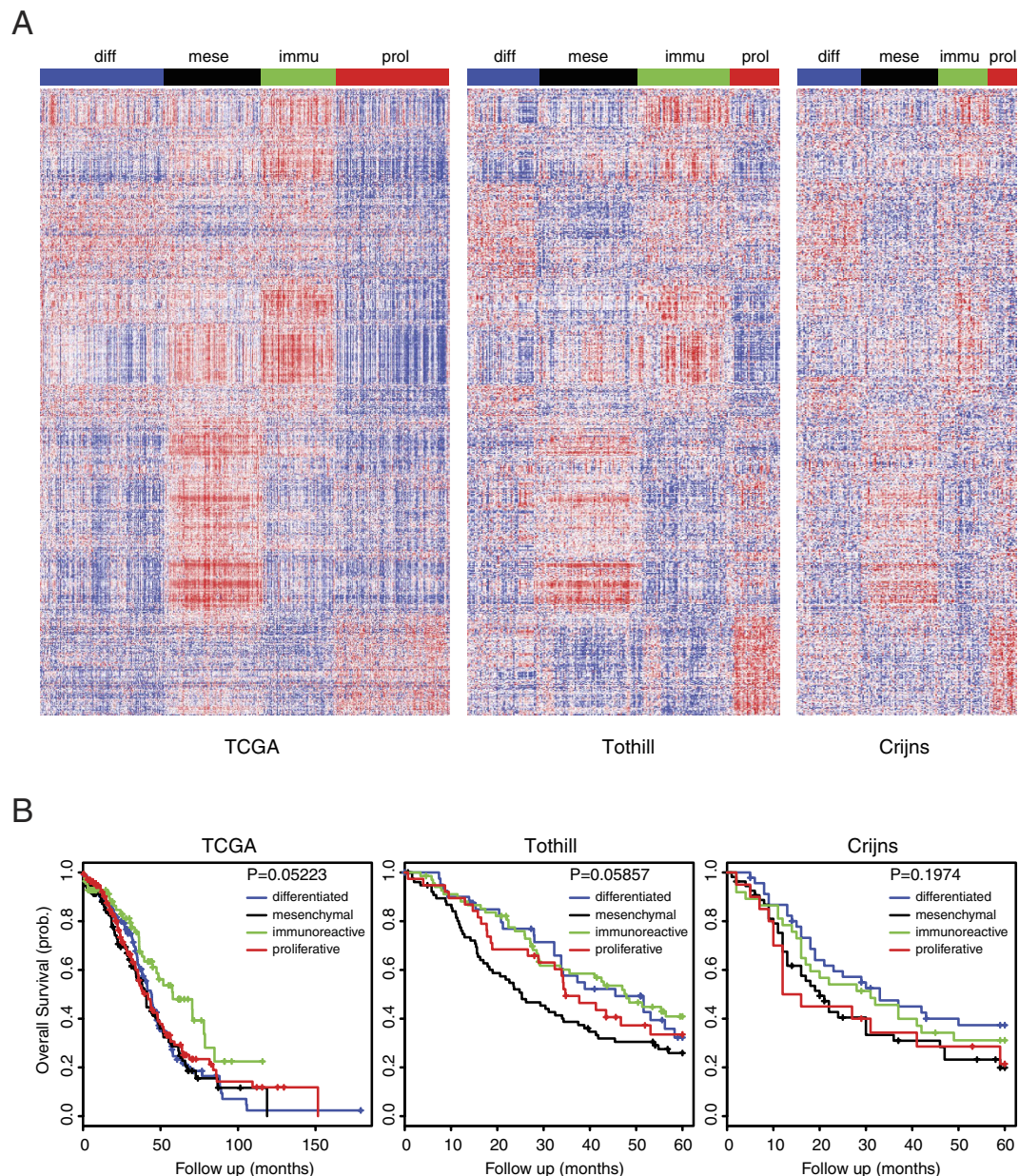
We reasoned that a more thorough understanding of the biological and regulatory mechanisms underlying the distinct subtypes might facilitate the development of novel prognostic signatures and subtype-specific therapeutic strategies in HGS-OvCa. For example, numerous studies have implicated tumor-associated stroma in tumor progression and patient prognosis<sup>15–17</sup>. Interestingly, it has been recently discovered that stromal genes significantly contributed to the stem/serrated/mesenchymal transcriptional subtype in colorectal cancer<sup>18,19</sup>. Although the Mesenchymal and Immunoreactive subtypes of ovarian cancer are known to contain infiltrating stromal cells and lymphocytes, respectively, it remains to be determined whether and to what extent tumor microenvironment influences the assignment of transcriptional subtypes. In this study, we designed an analytical approach to delineate the cellular and molecular underpinnings of HGS-OvCa subtypes, with a specific focus on the involvement of tumor stromal constituents.

## Results

**The TCGA subtypes are not associated with patient prognosis.** Both non-negative matrix factorization (NMF) method (Supplementary Figure S1) and k-means clustering algorithm (Supplementary Figure S2) yielded four robust high-consensus molecular subtypes in the TCGA dataset, thus verifying previous classifications<sup>10</sup>. We calculated silhouette width<sup>20</sup> to identify samples most representative of each cluster and obtained a ‘core’ set of 388 tumors (Supplementary Figure S3). Subsequently, we derived a 749-gene classifier (Supplementary Table S1) with the lowest prediction error using significance analysis of microarrays (SAM)<sup>21</sup>, followed by prediction analysis for microarrays (PAM)<sup>22</sup>. We applied the 749-gene signature and NMF consensus clustering in two independent HGS-OvCa gene expression profiles (Tothill and Crijns)<sup>8,23</sup>, and validated the four molecular subtypes (Fig. 1A; Supplementary Figure S4–5). However, in all three datasets, the HGS-OvCa molecular subtypes were not prognostically relevant (Fig. 1B). These unexpected results prompted us to further investigate the cellular and molecular determinants of HGS-OvCa clusters.

### Tumor-associated stromal content contributes to defining Mesenchymal and Immunoreactive subtypes.

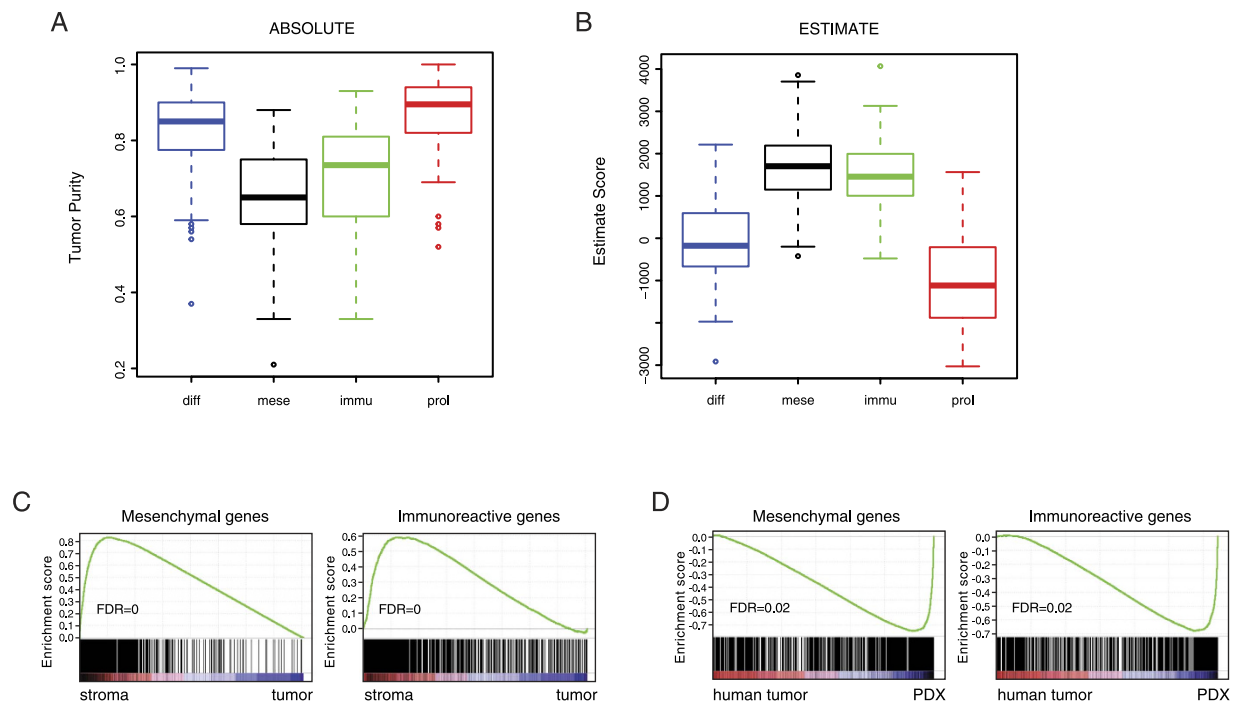
The four TCGA subtypes were initially termed Differentiated, Mesenchymal, Immunoreactive and Proliferative on the basis of expressed genes in the clusters<sup>10</sup>. For example, the Mesenchymal subtype was defined by high expression of FAP, fibronectin and collagens, whereas chemokine ligands (CXCL9, CXCL10, CXCL11) and receptors (CXCR3, CXCR6) characterized the Immunoreactive subtype. Recent studies have revealed a significant contribution of tumor stromal genes to stem/serrated/mesenchymal transcriptional subtype in colorectal cancer<sup>18,19</sup>. We sought to assess whether a similar interplay between stromal components and molecular characteristics existed in HGS-OvCa. To this end, tumor purity was inferred by the ABSOLUTE algorithm<sup>24</sup> and the average purity estimates of Mesenchymal and Immunoreactive samples were significantly lower than those of Differentiated and Proliferative samples (Fig. 2A). An alternative approach using the ESTIMATE method<sup>25</sup> to predict the fraction of stromal and immune cells produced consistent results in the TCGA dataset (Fig. 2B), as well as in Tothill and Crijns cohorts (Supplementary Figure S6). These data suggested that a higher stromal content was associated with the Mesenchymal and Immunoreactive subtypes and might dominate the observed transcriptional traits. To analytically test this hypothesis, we identified signature genes that were upregulated in the Mesenchymal or Immunoreactive subtypes (Supplementary Table S2). Gene Set Enrichment Analysis (GSEA)<sup>26</sup> indicated that Mesenchymal and Immunoreactive gene signatures were significantly enriched in the microdissected stroma components in comparison to paired tumor tissues<sup>8</sup> (Fig. 2C). Additionally, we analyzed gene expression profiles of nine pairs of ovarian tumors and matched patient-derived xenografts (PDXs), in which human stromal cells were substituted by mouse cells<sup>27</sup>. Mesenchymal and Immunoreactive gene transcripts were accordingly depleted in PDXs (Fig. 2D). Collectively, these



**Figure 1. The TCGA subtypes are not associated with patient prognosis.** (A) Tumors from TCGA, Tothill and Crijns datasets were separated into four clusters on the basis of gene expression. (B) Kaplan Meier curves for four molecular subtypes in the TCGA, Tothill and Crijns datasets.

findings demonstrated that tumor-associated stromal content substantially influenced the transcriptional profiles and molecular subtypes of HGS-OvCa.

**Regulatory networks and master regulators of Mesenchymal and Immunoreactive subtypes.** We employed a network-based strategy<sup>28–30</sup> to uncover the molecular mechanism underlying the discrete HGS-OvCa molecular subtypes, particularly Mesenchymal and Immunoreactive gene programs most correlated with stromal rather than epithelial origin. First, a regulatory network was constructed for ovarian cancer based on the TCGA dataset using a genome-wide reverse engineering approach<sup>28</sup>. Next, we applied the Master Regulator Analysis (MRA) algorithm to the network<sup>29</sup>, to identify regulons showing statistically significant overlap with Mesenchymal or Immunoreactive genes. From a list of 1111 transcription factors (TFs) (Supplementary Table S3), MRA inferred 6 Mesenchymal-specific TFs (Fig. 3A; Supplementary Figure S7; Supplementary Table S4) and 10 Immunoreactive-specific TFs (Fig. 3B; Supplementary Figure S8; Supplementary Table S5), as master regulators (MRs) of HGS-OvCa molecular subtypes. Interestingly, most Mesenchymal MRs had been implicated in epithelial-mesenchymal



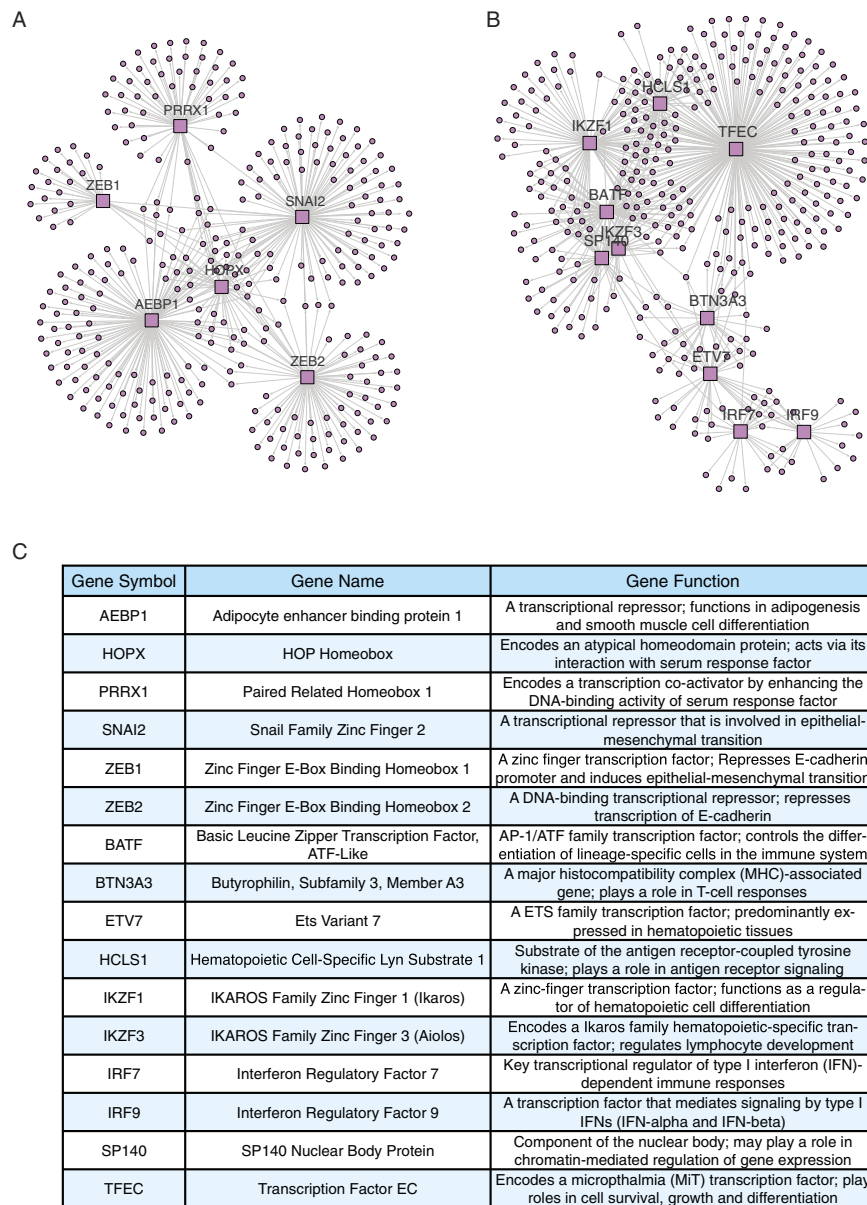
**Figure 2. Tumor-associated stromal content contributes to defining Mesenchymal and Immunoreactive subtypes.** (A) Tumor purity estimated by ABSOLUTE analysis for four molecular subtypes of TCGA samples. (B) ESTIMATE scores for four molecular subtypes of TCGA samples. (C) GSEA for upregulation of Mesenchymal and Immunoreactive genes in microdissected tumor stroma versus epithelial tissues. (D) GSEA for downregulation of Mesenchymal and Immunoreactive genes in patient-derived xenografts (PDX) versus matched primary tumors.

transition and Immunoreactive MRs were predominantly transcriptional regulators of immune function (Fig. 3C), thus validating the robustness of our approach.

### Mesenchymal and Immunoreactive MRs correlate with tumor stroma and patient survival.

Using MRs as refined physiologically relevant gene signatures, we performed single-sample Gene Set Enrichment Analysis (ssGSEA)<sup>31</sup> to assess MRs compound scores for the TCGA HGS-OvCa samples. The ssGSEA scores correlated well with those produced by the Gene Set Variation Analysis (GSVA)<sup>32</sup> as an independent method (Supplementary Figure S9). As expected, the Mesenchymal and Immunoreactive subtypes showed relatively higher levels of Mesenchymal and Immunoreactive MRs expression as well as compound scores, respectively (Fig. 4A). The binary scores indicated that 78% (380/489) of expression profiles showed activity of Mesenchymal or Immunoreactive MRs, and that 28% of tumor samples could be assigned to both Mesenchymal and Immunoreactive subtypes (Fig. 4A), confirming previous observations that HGS-OvCa is highly heterogeneous<sup>11,33</sup>. To investigate the relative contributions of epithelial tumor cells or stromal cells to the expression of MRs, we analyzed two gene sets of ovarian tumor samples<sup>8,34</sup>, in which epithelial and stromal components had been microdissected and profiled separately. Both Mesenchymal and Immunoreactive MRs were significantly upregulated in tumor stroma in comparison to epithelial tumor areas (Fig. 4B). Consistently, Mesenchymal and Immunoreactive MRs were significantly downregulated in patient-derived xenografts relative to paired primary tumors (Fig. 4C). Therefore, as with HGS-OvCa molecular subtypes, Mesenchymal and Immunoreactive MRs were associated with noncancerous tumor stroma. Further detailed analysis indicated that levels of Mesenchymal and Immunoreactive MRs increased upon metastasis (Fig. 4D) and chemotherapy (Fig. 4E)<sup>35,36</sup>, supporting their involvement in tumor progression and response to treatment. Interestingly, the Mesenchymal MRs correlated with gene expression of multiple IGF-related molecules including IGF, IGFBP4, IGFBP6 and IGFBP7, indicative of IGF pathway activation in the Mesenchymal tumors (Supplementary Figure S10).

To assess Mesenchymal and Immunoreactive MRs as prognostic biomarkers, we analyzed their expression in all HGS-OvCa samples included in curatedOvarianData for which overall survival information was available<sup>37</sup>. All 6 Mesenchymal MRs significantly correlated with poor patient outcome (Supplementary Figure S11). Conversely, 7 of 10 Immunoreactive MRs showed significant association with improved overall survival (Supplementary Figure S12). Based on these findings, we subdivided the TCGA samples into three clusters, i.e. ‘immu + mese-’, ‘immu - mese+’ and ‘mixed’, according to Mesenchymal



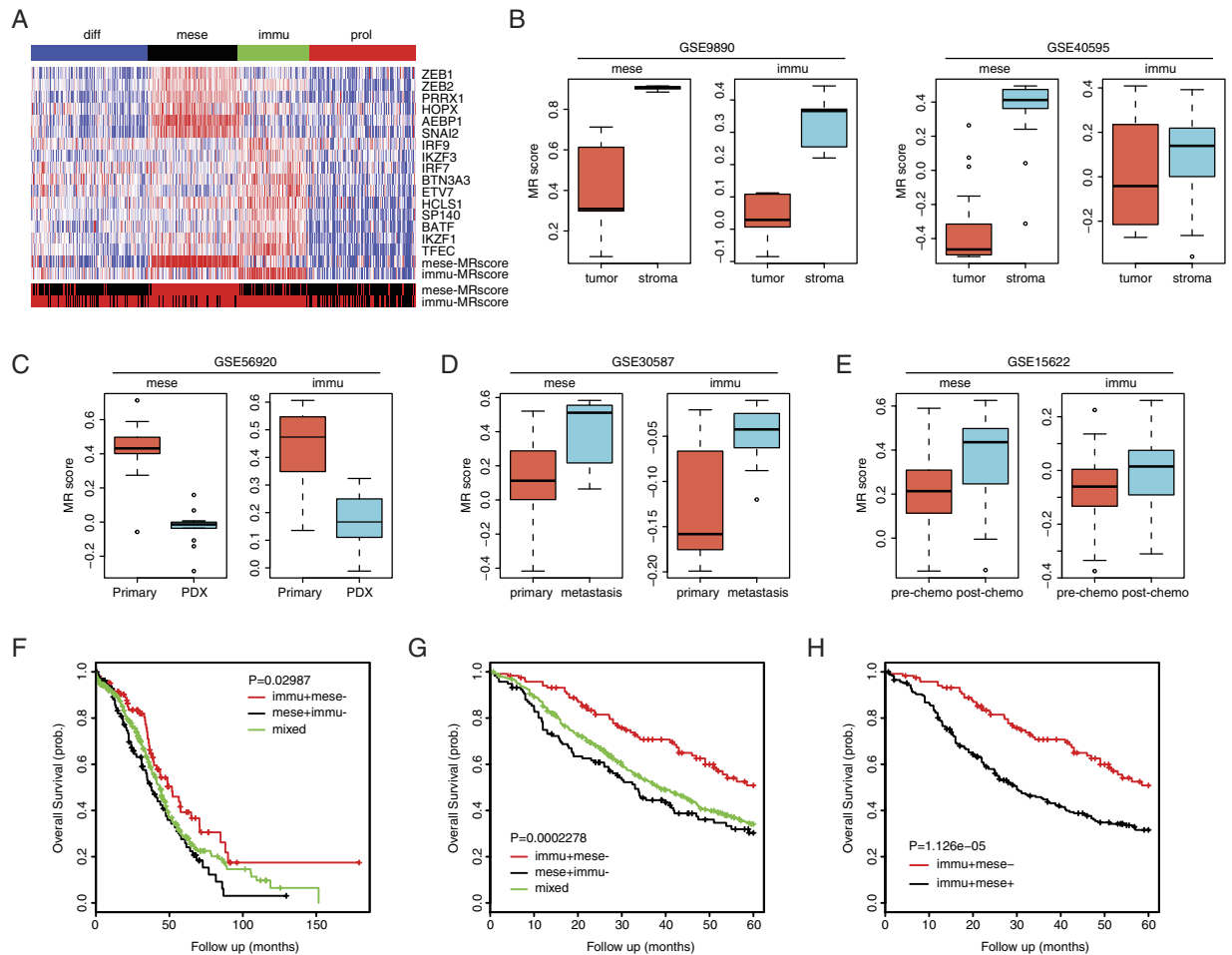
**Figure 3. Regulatory networks and master regulators of Mesenchymal and Immunoreactive subtypes .**

(A) The Mesenchymal regulatory network showing the six MRs (square nodes) and all inferred targets (round nodes). (B) The Immunoreactive regulatory network showing the ten MRs (square nodes) and all inferred targets (round nodes). (C) The list of Mesenchymal and Immunoreactive MRs.

and Immunoreactive compound scores. We found that patients classified as ‘immu – mese+’ had significantly shorter survival than ‘immu + mese–’ patients, and the ‘mixed’ group showed intermediate outcome (Fig. 4F). To independently corroborate the prognostic value of MRs signatures, we performed meta-analysis of 749 HGS-OvCa expression profiles across five datasets<sup>8,23,38–40</sup>. The difference in survival between the three groups was highly statistically significant (Fig. 4G). Interestingly, compared with ‘immu + mese–’ subtype, patients whose tumor samples express both Mesenchymal and Immunoreactive MRs (‘immu + mese+’) had statistically significantly worse survival (Fig. 4H). On the contrary, we did not observe survival difference between ‘immu – mese+’ and ‘immu + mese+’ groups (Supplementary Figure S13). These data highlighted the dominant role of Mesenchymal MRs in predicting patient outcome.

## Discussion

In this study, we presented a detailed analysis of the four molecular subtypes based on TCGA HGS-OvCa expression data. We showed that Mesenchymal and Immunoreactive subtypes were characterized by transcriptional traits dominated by tumor-infiltrating stromal cells. By systematically interrogating



**Figure 4. Mesenchymal and Immunoreactive MRs correlate with tumor stroma and patient survival .**

(A) Heatmap of Mesenchymal and Immunoreactive MRs expression and ssGSEA scores for four molecular subtypes of TCGA samples. Binary scores were shown to indicate whether a tumor sample activated Mesenchymal or Immunoreactive MRs. Red, activated; black, not activated. (B) Mesenchymal and Immunoreactive MRs scores in microdissected tumor stroma (5 samples in GSE9890 and 31 samples in GSE40595) versus epithelial tissues (5 samples in GSE9890 and 32 samples in GSE40595). (C) Mesenchymal and Immunoreactive MRs scores in PDX versus matched primary tumors (9 samples). (D) Mesenchymal and Immunoreactive MRs scores in tumor metastasis versus primary tumors (9 paired samples). (E) Mesenchymal and Immunoreactive MRs scores in tumors treated with chemotherapy (34 samples) versus non-treated tumors (35 samples). (F) Kaplan Meier curves for three prognostic groups of TCGA samples classified by Mesenchymal and Immunoreactive MRs signatures. G. Kaplan Meier curves for meta-analysis of 749 HGS-OvCa expression profiles across five cohorts. H. Kaplan Meier curves for ‘immu + mese-’ and ‘immu + mese+’ patients.

subtype-specific regulatory networks, we further identified the transcriptional module and master regulators that drove the expression of Mesenchymal and Immunoreactive signatures. This approach led to the identification of novel transcription factors as potentially critical regulators of tumor-associated microenvironment, which also served as robust prognostic biomarkers of aggressive ovarian cancer.

We provided several complementary lines of evidence supporting that Mesenchymal and Immunoreactive signature genes were mostly expressed by tumor stromal components. First, tumor purity inferred by the ABSOLUTE or ESTIMATE algorithm was significantly lower in Mesenchymal and Immunoreactive samples than in Differentiated and Proliferative samples, suggesting that a higher stromal content was associated with the Mesenchymal and Immunoreactive subtypes. Second, analysis of the expression profiles of microdissected ovarian cancer demonstrated that Mesenchymal and Immunoreactive gene signatures were significantly enriched in the stroma components in comparison to paired tumor tissues. Third, in nine pairs of ovarian tumors and matched patient-derived xenografts (PDXs), where human stroma was substituted by mouse stroma, Mesenchymal and Immunoreactive gene transcripts were depleted in PDXs. Notably, these findings were confirmed in multiple independent

datasets, indicating that abundant stroma was an intrinsic feature of some HGS-OvCa samples rather than a technical artifact.

The four TCGA subtypes, namely Differentiated, Mesenchymal, Immunoreactive and Proliferative, were originally defined by specifically expressed genes in the clusters<sup>10</sup>. However, regulatory mechanisms at the molecular level remain to be discovered within each subtype, in order to elucidate causal drivers and to identify relevant targets for effective cancer treatment. We took a systems biology approach to gain insight into transcriptional networks associated with molecular subtypes and identified central transcription factors as master regulators of Mesenchymal and Immunoreactive phenotypes. Our analysis identified ZEB1, ZEB2, SNAI2, PRRX1, AEBP1 and HOPX as Mesenchymal MRs, most of which had been implicated in epithelial-mesenchymal transition<sup>41</sup>, and IRF7, IRF9, IKZF1, IKZF3, BATF, ETV7, BTN3A3, SPI40, HCLS1 and TFEC as Immunoreactive MRs, which were predominantly transcriptional regulators of immune function<sup>42–44</sup>. In line with stromal enrichment of Mesenchymal and Immunoreactive signature genes, Mesenchymal and Immunoreactive MRs were associated with noncancerous tumor stroma. It would be interesting to investigate the biological functions of these MRs in future studies.

Importantly, a statistically significant correlation was observed between the MRs and patient survival. In contrast, the simple classification into four molecular subtypes did not seem to have prognostic relevance. It was previously shown that subtype assignment of HGS-OvCa expression profiles were not mutually exclusive<sup>20</sup>. Therefore, MRs might have increased power to detect Mesenchymal and Immunoreactive traits that were otherwise concealed by sample heterogeneity. Indeed, we found that Mesenchymal or Immunoreactive MRs were detectable in the majority of TCGA samples (78%) at different levels of activation. Furthermore, there is an emerging consensus that network-based biomarker candidates, especially with sound biologic rationale, often exhibit high reproducibility and sensitivity in ovarian cancer<sup>45,46</sup>. The facts that stroma-associated MRs predicted clinical outcome and that Mesenchymal and Immunoreactive MRs differentially linked to opposite prognostic categories prompted us to deduce that the stromal composition and interplay might have a significant, if not major, contribution to the HGS-OvCa prognosis. These findings may reconcile controversies on the prognostic importance of HGS-OvCa molecular subtypes and propose a simple way to subgroup patients for accurate management. Further validation of stromal MRs in forecasting disease outcome will require prospective clinical studies, preferably using feasible assays that can process many samples on a routine basis, e.g. quantitative PCR or immunohistochemistry.

In addition to patient prognosis, we also found that Mesenchymal and Immunoreactive MRs were upregulated in tumor metastasis or in response to chemotherapy. Moreover, the Mesenchymal MRs compound scores correlated with expression levels of multiple IGF-related genes, suggesting that the IGF pathway might be a druggable target in ovarian tumors with Mesenchymal MRs signature. Future studies should further elucidate the intricate relationship between tumor subtypes, MRs expression, tumor progression, response to treatment and patient outcome. For example, individuals with Immunoreactive MRs signature may benefit from treatments targeting tumor cells via the immune response, such as recently approved immune checkpoint inhibitors or vaccine therapies in the adjuvant setting. On the other hand, patients expressing Mesenchymal MRs exhibit poor outcome and probably need to be treated more aggressively with chemotherapy. We speculate that beyond disease prognosis, insights into stroma-associated MRs may contribute to the selection and development of new therapeutic strategies.

In summary, integrative analyses presented in this work determined the cellular and molecular underpinnings of HGS-OvCa subtypes. We provided evidence that infiltrating stromal cells had a profound effect on the expression patterns of HGS-OvCa, particularly Mesenchymal and Immunoreactive clusters. A stroma-associated gene signature composed of transcriptional master regulators was inferred by unbiased reverse engineering algorithm, and proved to effectively stratify patients into different prognostic groups. Therefore, systematic interrogation of genome-wide context-specific networks may not only advance our understanding of the regulatory programs underlying cancer phenotypes, but also enable accurate prediction of patient prognosis. Importantly, the MRs signature only consists of 16 genes, making clinical implementation using various gene/protein profiling platforms feasible. We envision that our findings should provide a basis for improved stratification of patients with HGS-OvCa that may ultimately lead to more precise therapies.

## Methods

**Microarray datasets.** We used various microarray datasets of HGS-OvCa in the public domain. Combined and filter TCGA gene expression data were downloaded from [https://tcga-data.nci.nih.gov/docs/publications/ov\\_2011/](https://tcga-data.nci.nih.gov/docs/publications/ov_2011/). The five patient cohorts (Tohill, Crijns, Bonome, Yoshihara and Denkert) for meta-analysis have been described previously<sup>8,23,38–40</sup>, and processed data were downloaded from a recent report<sup>20</sup>. Other microarray datasets, including GSE9890, GSE15622, GSE30587 and GSE56920, are publicly available in NCBI GEO database.

**Identification of molecular subtypes and signature genes.** We classified TCGA HGS-OvCa based on non-negative matrix factorization (NMF) consensus clustering originally used to define the four molecular subtypes<sup>10</sup>. NMF is an unsupervised technique to reduce the dimensionality of gene expression data. A small number of metagenes are defined as a positive linear combination of many genes. The metagene expression patterns provide a robust clustering of samples<sup>47</sup>. The NMF classification

was confirmed using k-means clustering algorithm implemented in R package ‘ConsensusClusterPlus’<sup>48</sup>. The k-means clustering algorithm aims to classify a given dataset into k clusters, which have been specified a priori. The samples are assigned to the nearest k cluster centroids, and each cluster center is recalculated as the mean value of cluster members, followed by sample reassignment. This process is repeated until the distance between consecutive cluster centers converges. Both the NMF and k-means clustering methods yielded four robust transcriptomic clusters. To minimize the impact of outlier samples on the identification of subtype markers, the silhouette width was computed to filter out expression profiles with negative values, which excluded 101 samples that were not likely a robust representative of the subclass. To identify subtype-specific signature genes, we used significance analysis of microarrays (SAM) to identify genes significantly differentially expressed across the four subtypes. These genes were trained by prediction analysis for microarrays (PAM) to achieve the lowest prediction error, which resulted in the 749-gene signature (Supplementary Table S1). To validate the presence of four molecular subtypes in additional datasets, the 749-gene signature was applied to Tothill and Crijns cohorts, followed by consensus-based NMF analysis. Heatmaps were generated using GenePattern<sup>49</sup>.

**GSEA and ssGSEA.** Gene Set Enrichment Analysis (GSEA) was performed as described<sup>26</sup>, using Mesenchymal and Immunoreactive signatures as gene sets. We downloaded the GSEA software from the Broad Institute GSEA portal. Single sample GSEA (ssGSEA) was applied to generate compound scores for Mesenchymal and Immunoreactive master regulators as gene signatures<sup>31</sup>. The procedure was similar to GSEA but gene expression values were ranked for a given sample, and an enrichment score was calculated based on the normalized rank difference in Empirical Cumulative Distribution Functions (ECDF) of the genes in the signature and the remaining genes. We normalized the scores by the absolute difference between the minimum and the maximum for all samples within a dataset before combining ssGSEA scores across different datasets. The Gene Set Variation Analysis (GSVA) was performed to compare the enrichment scores produced by independent methods<sup>32</sup>.

**Tumor purity analysis.** For tumor purity analysis, we used two different previously validated approaches, ABSOLUTE and ESTIMATE<sup>24,25</sup>. The ABSOLUTE method predicts tumor purity based on the allelic copy-ratio profiles derived from SNP arrays. The ESTIMATE analyses quantify non-tumor constituents by identifying specific gene signatures related to the infiltration of normal cells in tumor tissues. Tumor purity inferred by the ABSOLUTE algorithm was obtained from the TCGA working group. ESTIMATE scores, which predict the level of infiltrating non-tumor cells, were calculated by performing ssGSEA as reported. We first defined stromal and immune scores based on the genes related to stromal tissue and immune cell infiltration, and then combined the stromal and immune scores as the ESTIMATE scores.

**Transcriptional network inference and master regulator analysis.** Combined and filter TCGA HGS-OvCa gene expression data were used to build a gene regulatory network. The analysis was conducted with Bioconductor package ‘RTN’<sup>30</sup>, which re-implemented ARACNe in R for reconstruction and analysis of transcriptional networks using mutual information (MI). Transcriptional regulatory units, termed regulons, were assembled by computing the MI between transcription factors (TF) and all potential targets using gene expression data, followed by multiple hypothesis testing corrections (Benjamini-Hochberg). The TF list was derived from a previous publication<sup>50</sup>. Unstable TF-gene interactions were removed by bootstrap analysis, and Data Processing Inequality (DPI) algorithm was used to remove redundant interactions and preserve the dominant TF-gene pairs. In order to identify subtype-specific transcription factors, the Master Regulator Analysis (MRA) pipeline was applied to estimate the statistical significance of the overlap between the regulons and Mesenchymal or Immunoreactive signature genes using the hypergeometric distribution. GSEA was performed to validate the MRA results. For network visualization we used the Bioconductor package ‘RedeR’<sup>51</sup>.

**Survival analysis.** To test the individual Mesenchymal and Immunoreactive master regulators as prognostic markers, we analyzed the hazard ratio of MRs expression and generated forest plots using the ‘curatedOvarianData’ Bioconductor package<sup>37</sup>. For meta-analysis of 749 HGS-OvCa expression profiles, ssGSEA compound scores of Mesenchymal and Immunoreactive master regulators were computed for each sample. The patients were dichotomized into a high-score and a low-score group, using the median ssGSEA score as the threshold value. Based on both Mesenchymal and Immunoreactive compound scores, we stratified samples into three clusters, i.e. ‘immu + mese-’, ‘immu - mese+’ and ‘mixed’. Overall survival curves were calculated using the Kaplan–Meier method, and statistical significance was assessed using the log-rank test. The analyses were conducted with the R Bioconductor ‘survival’ package.

## References

1. Bowtell, D. D. The genesis and evolution of high-grade serous ovarian cancer. *Nat Rev Cancer* **10**, 803–808 (2010).
2. Liu, J. & Matulonis, U. A. New strategies in ovarian cancer: translating the molecular complexity of ovarian cancer into treatment advances. *Clin Cancer Res* **20**, 5150–5156 (2014).
3. Bookman, M. A. *et al.* Better therapeutic trials in ovarian cancer. *J Natl Cancer Inst* **106**, dju029 (2014).
4. Coleman, R. L., Monk, B. J., Sood, A. K. & Herzog, T. J. Latest research and treatment of advanced-stage epithelial ovarian cancer. *Nat Rev Clin Oncol* **10**, 211–224 (2013).



5. Pignata, S. *et al.* Chemotherapy in epithelial ovarian cancer. *Cancer Lett* **303**, 73–83 (2011).
6. Riester, A. *et al.* Age below 40 or a recently proposed clinical prediction score cannot bypass adrenal venous sampling in primary aldosteronism. *J Clin Endocrinol Metab* **99**, E1035–1039 (2014).
7. Waldron, L. *et al.* Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer. *J Natl Cancer Inst* **106** (2014).
8. Tothill, R. W. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* **14**, 5198–5208 (2008).
9. Tan, T. Z. *et al.* Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. *EMBO Mol Med* **5**, 983–998 (2013).
10. Cancer Genome Atlas Research, N. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
11. Konecny, G. E. *et al.* Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *J Natl Cancer Inst* **106** (2014).
12. Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* **19**, 619–625 (2013).
13. De Sousa, E. M. F. *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med* **19**, 614–618 (2013).
14. Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* **17**, 500–503 (2011).
15. Tlsty, T. D. & Coussens, L. M. Tumor stroma and regulation of cancer development. *Annu Rev Pathol* **1**, 119–150 (2006).
16. Finak, G. *et al.* Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med* **14**, 518–527 (2008).
17. Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309–322 (2012).
18. Isella, C. *et al.* Stromal contribution to the colorectal cancer transcriptome. *Nat Genet* **47**, 312–319 (2015).
19. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet* **47**, 320–329 (2015).
20. Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
21. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98**, 5116–5121 (2001).
22. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA* **99**, 6567–6572 (2002).
23. Crijns, A. P. *et al.* Survival-related profile, pathways, and transcription factors in ovarian cancer. *PLoS Med* **6**, e24 (2009).
24. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* **30**, 413–421 (2012).
25. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* **4**, 2612 (2013).
26. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550 (2005).
27. Ricci, F. *et al.* Patient-derived ovarian tumor xenografts recapitulate human clinicopathology and genetic alterations. *Cancer Res* **74**, 6980–6990 (2014).
28. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat Genet* **37**, 382–390 (2005).
29. Carro, M. S. *et al.* The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318–325 (2010).
30. Fletcher, M. N. *et al.* Master regulators of FGFR2 signalling and breast cancer risk. *Nat Commun* **4**, 2464 (2013).
31. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
32. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
33. Verhaak, R. G. *et al.* Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J Clin Invest* **123**, 517–525 (2013).
34. Yeung, T. L. *et al.* TGF-beta modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res* **73**, 5016–5028 (2013).
35. Brodsky, A. S. *et al.* Expression profiling of primary and metastatic ovarian tumors reveals differences indicative of aggressive disease. *PLoS One* **9**, e94476 (2014).
36. Ahmed, A. A. *et al.* The extracellular matrix protein TGFBI induces microtubule stabilization and sensitizes ovarian cancers to paclitaxel. *Cancer Cell* **12**, 514–527 (2007).
37. Ganzfried, B. F. *et al.* curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome. *Database (Oxford)* **2013**, bat013 (2013).
38. Bonome, T. *et al.* A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res* **68**, 5478–5486 (2008).
39. Yoshihara, K. *et al.* Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. *PLoS One* **5**, e9615 (2010).
40. Denkert, C. *et al.* A prognostic gene expression index in ovarian cancer - validation across different independent data sets. *J Pathol* **218**, 273–280 (2009).
41. Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol* **16**, 488–494 (2014).
42. Ikushima, H., Negishi, H. & Taniguchi, T. The IRF family transcription factors at the interface of innate and adaptive immune responses. *Cold Spring Harb Symp Quant Biol* **78**, 105–116 (2013).
43. Yoshida, T. & Georgopoulos, K. Ikaros fingers on lymphocyte differentiation. *Int J Hematol* **100**, 220–229 (2014).
44. Murphy, T. L., Tussiwand, R. & Murphy, K. M. Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. *Nat Rev Immunol* **13**, 499–509 (2013).
45. Kang, J., D'Andrea, A. D. & Kozono, D. A DNA repair pathway-focused score for prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy. *J Natl Cancer Inst* **104**, 670–681 (2012).
46. Jin, N. *et al.* Network-based survival-associated module biomarker and its crosstalk with cell death genes in ovarian cancer. *Sci Rep* **5**, 11566 (2015).
47. Brunet, J. P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci USA* **101**, 4164–4169 (2004).
48. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
49. Reich, M. *et al.* GenePattern 2.0. *Nat Genet* **38**, 500–501 (2006).
50. Bhat, K. P. *et al.* The transcriptional coactivator TAZ regulates mesenchymal differentiation in malignant glioma. *Genes Dev* **25**, 2594–2609 (2011).

51. Castro, M. A., Wang, X., Fletcher, M. N., Meyer, K. B. & Markowitz, F. RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. *Genome Biol* **13**, R29 (2012).

### Acknowledgements

This work was supported by the National Natural Science Foundation of China (81472537 to G Zhuang, 81130038 and 81372189 to WQ Gao), the Grants from the State Key Laboratory of Oncogenes and Related Genes (No. 91-14- 18 and 91-15- 12 to G Zhuang), the Shanghai Institutions of Higher Learning (Eastern Scholar to G Zhuang), the Chinese Ministry of Science and Technology (2012CB966800 to WQ Gao), the Shanghai Health Bureau Key Discipline and Specialty Foundation and the KC Wong foundation to WQ Gao.

### Author Contributions

G.Z. and W.Q.G. designed and supervised the project. S.Z., Y.J., M.Z., Z.Z., P.M., H.P. and K.S. performed all the analyses. G.Z. wrote the manuscript. All authors contributed to editing the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zhang, S. *et al.* Stroma-associated master regulators of molecular subtypes predict patient prognosis in ovarian cancer. *Sci. Rep.* **5**, 16066; doi: 10.1038/srep16066 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>