Dysregulation of splicing-related proteins in prostate cancer is controlled by FOXA1

John G. Foster[1], Rebecca Arkell[1], Marco Del Giudice[2,3], Chinedu Anene[1], Andrea Lauria[2,3], John D. Kelly[5,6], Nicholas R. Lemoine[1], Salvatore Oliviero[2,3], Matteo Cereda[2,†] and Prabhakar Rajan[1,4,5,6,7,†]

[1]Centre for Molecular Oncology, Barts Cancer Institute, Cancer Research UK Barts Centre, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK

[2]Italian Institute for Genomic Medicine, Via Nizza 52, 10126, Turin, Italy

[3]Department of Life Science and System Biology, Università degli Studi di Torino, via Accademia Albertina 13, 10123 Turin, Italy

[4]The Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, UK

[5]Division of Surgery and Interventional Science, University College London, 43-45 Foley Street, London, W1W 7TS, UK

[6]Department of Uro-oncology, University College London NHS Foundation Trust, 47 Wimpole Street, London, W1G 8SE, UK

[7]Department of Urology, Barts Health NHS Trust, The Royal London Hospital, Whitechapel Road, London, E1 1BB, UK

† These authors contributed equally to this work

Correspondence to: j.foster@qmul.ac.uk; matteo.cereda@iigm.it; p.rajan@qmul.ac.uk

Abstract

Prostate cancer (PCa) is genomically driven by dysregulation of transcriptional networks involving the transcriptional factors (TFs) FOXA1, ERG, AR, and HOXB13. However, the role of these specific TFs in the regulation of alternative pre-mRNA splicing (AS), which is a proven therapeutic vulnerability for cancers driven by the TF MYC, is not described. Using transcriptomic datasets from PCa patients, we tested for an association between expression of *FOXA1*, *ERG*, *AR*, *HOXB13*, and *MYC*, and genes involved in AS - termed splicing-related proteins (SRPs), which have pleiotropic roles in RNA metabolism. We identified FOXA1 as the strongest predictor of dysregulated SRP gene expression, which was associated with PCa disease relapse after treatment. Subsequently, we selected a subset of FOXA1-binding and actively-transcribed SRP genes that phenocopy the FOXA1 dependency of PCa cells, and confirmed *in vitro* via knockdown and over-expression that FOXA1 regulates SRP gene expression. Finally, we demonstrated the persistence of a FOXA1-SRP gene association in treatment-relapsed castration-resistant PCa (CRPCa) patients. Our data demonstrate, for the first time, that FOXA1 controls dysregulated SRP gene expression, which is associated with poor PCa patient outcomes. Analogous to MYC-driven cancers, our findings implicate the therapeutic targeting of SRPs and AS in FOXA1-overexpressing PCa.

Running title

FOXA1 and splicing-related proteins in prostate cancer

Keywords

FOXA1, alternative pre-mRNA splicing, prostate cancer, gene regulation

Introduction

Prostate cancer (PCa) is the commonest male gender-specific cancer [1]. Genomic characterisation of primary PCa has uncovered several molecular subtypes [2-4], characterised by alterations in genes encoding the transcription factors (TFs) FOXA1, ERG, and AR, which is an existing therapeutic target [5]. The gene encoding the AR-interacting TF HOXB13 has been identified as a candidate PCa susceptibility gene [6]. Co-operatively, these TFs reprogram the AR-associated cistrome in prostate tumourigenesis [7-9]. Additionally, PCa susceptibility loci identified from genome-wide association studies (GWAS) fall within the cistromes of these TFs themselves [10], thereby demonstrating widespread dysregulation of transcriptional networks in PCa.

Recently, widespread genomic and transcriptional dysregulation of genes encoding RNA-binding proteins (RBPs) have been reported across several cancers [11-13]. RBPs are a family of proteins with pleiotropic roles in RNA metabolism including alternative pre-mRNA splicing (AS) [14]. Through the regulation of their target mRNAs, RBPs are associated with different oncogenic processes and patient outcomes [15]. In PCa and other malignancies, mutations in genes coding for RBPs and changes in their expression levels have been observed [11,16]. Moreover, AS appears to represent a cancer therapeutic vulnerability in leukaemia and breast cancer driven by the oncogenic TF MYC [17-19]. MYC has also been shown to transcriptionally regulate expression of AS-associated RBPs in lymphoma, lung cancer and glioma pre-clinical models [20-22]. However, little is known of the mechanisms of transcriptional dysregulation of RBP expression in PCa.

Here, we show that genes encoding RBPs and other proteins involved in AS (referred to, hereafter, as splicing-related proteins; SRPs) are globally dysregulated in PCa, and identify the TF FOXA1 as a key regulator of SRP gene expression.

Results

Dysregulated SRP gene expression is associated with FOXA1 in primary human PCa

We hypothesised that dysregulation of SRP genes and others involved in gene expression processes in PCa is transcriptionally controlled by one or more of the TFs FOXA1, ERG, AR or HOXB13. To test this hypothesis, we utilised published RNA-Seq gene expression data of primary untreated prostate tumours (n=409) included in The Cancer Genome Atlas (TCGA) [2]. Transcriptomes were analysed based on expression levels (*i.e.* transcript per million reads, TPMs) of *FOXA1*, *ERG*, *AR* and *HOXB13* genes (Fig. 1A). We included *MYC* as a positive control as it is implicated in the regulation of SRP expression [20-22]. Samples were stratified for expression of genes encoding these five TFs with a cut-off of the top 25% of gene expression by TPM defining high expression (HE) and the remainder as Rest (Fig. 1A and B, and Supplementary Data 1) [23].

To determine the biological processes that are altered upon HE of the TFs, we performed a gene set analysis (GSA) (see Methods). For each TF, we compared the cumulative TPM values of genes in 16 gene sets representing Genetic Information Processing pathways accordingly to the Kyoto Encyclopedia of Genes and Genomes (KEGG) [24] between samples with HE of the TF gene and Rest (Fig. 1C and Supplementary Fig. 1). In doing so, GSA identified associations between expression of all TF genes and six different KEGG pathways, including the SRP gene set (Fig. 1C, Supplementary Table 1). To evaluate the impact of altered expression of genes in the six gene sets on PCa disease progression, we performed a survival analysis using Cox proportional hazards (PH) models (see Methods). Of the genes within the six gene sets, we found that dysregulated SRP gene expression showed the strongest association with disease recurrence (HR= 25.5; 95% CI=14.6-44.5) (Fig. 1D and Supplementary Table 2). Additionally, using a SRP gene set score to stratify patients (see Methods), we observed a statistically significant difference in time to disease progression between patients with a high score as compared to those with a low score (p-value < 0.0001) (Fig. 1E), thereby highlighting the importance of SRP genes in the PCa disease phenotype.

To determine which of the five TFs may be most important for SRP gene regulation, we employed a linear regression modelling approach (See Methods). We found that the overexpression of *FOXA1* gave the best results in terms of determination coefficient (R^2=0.3, Fig. 1F) when modelling SRP gene expression using only one TF gene. Increasing the model complexity led to a closer fitting between TF overexpression and SRP gene expression levels, with the five variables giving the highest fitting (R^2=0.54, Fig. 1F). We next measured the relative importance of each regressor in the linear model with all five TF genes using the averaging over ordering method [25] (see Methods). We found that *FOXA1* expression is the most important regressor contributing to 36% of the fitting of the model (Fig. 1G). Collectively, these findings suggest that, among all tested TFs, the expression of FOXA1 shows the strongest correlation with the modulation of SRP gene expression in PCa.

The FOXA1 cistrome includes a subset of actively-transcribed SRP genes

To identify SRP candidate genes regulated by FOXA1, we performed differential expression analysis using three distinct approaches (see Methods): Firstly, we identified a total of 76 SRP genes that were significantly up- (n=54) or down- (n=22) regulated in samples with *FOXA1* HE as compared with *FOXA1* Rest (Supplementary Data 2 and Supplementary Fig. 2A). Secondly, we determined which of the 76 SRP genes had enrichment for the five TF binding sites within their promoter regions and gene bodies using the ReMap database [26] (Fig. 2A). Finally, we selected sites of active transcription by overlapping TF binding sites with the H3K27ac and H3K4me3 signatures. FOXA1 binding sites were most enriched within these genes compared to the other transcription factors (Fig. 2B), with 63/76 SRP genes containing FOXA1-binding sites (Supplementary Fig. 2A). Of the 63 SRP genes with FOXA1 binding sites, the majority (47/63) were up-regulated. These data suggest that FOXA1 might directly control expression of up-regulated SRP genes in PCa.

126 *The expression and function of FOXA1 and SRP genes are similar in human PCa cell line models*

127

128 To characterise human PCa cell line models for downstream validation, we profiled *FOXA1*
129 expression by qRT-PCR and western blotting in DU145, PC3, LNCaP and VCaP cells (Fig. 2C and
130 Supplementary Fig. 2B). We identified the highest level of FOXA1 expression in the AR- and ERG-
131 positive VCaP cells, and the lowest level of expression in DU145 cells (Fig. 2D). FOXA1 has been
132 identified as an essential PCa gene in a RNAi genome-wide loss of function screen [27], and in this
133 dataset VCaP and DU145 cells harboured the greatest and least dependency (DEMETER scores) on
134 FOXA1, respectively (Fig. 2E). To confirm this observation, we used two independent siRNA
135 duplexes (Supplementary Table 3) to deplete VCaP and DU145 cells of FOXA1 protein
136 (Supplementary Fig. 2C, upper panel). Following FOXA1 depletion, we observed a statistically
137 significant reduction in cell growth in VCaP cells as compared with NSI controls, but no statistically
138 significant change in cell growth in DU145 cells (Supplementary Fig. 2C, lower panel).

139 We hypothesised that FOXA1-overexpressing VCaP cells may also be dependent on FOXA1-
140 associated SRP genes. Of the 47 up-regulated SRP genes with FOXA1-binding sites identified from
141 the TCGA analysis (Supplementary Fig. 2A), we selected the top 10 ranked SRP genes by TPM and
142 top 10 ranked by fold change of median TPM between FOXA1 HE samples and Rest for further
143 analysis (Fig. 2E). Using the RNAi dataset [27], we ranked these 20 SRP genes by VCaP cell
144 dependency DEMETER score, and selected seven candidate SRP genes with a DEMETER score
145 smaller than -1 for qRT-PCR validation in four different PCa cell lines. We profiled expression of the
146 seven candidate SRP genes in the four PCa cell lines by qRT-PCR and observed the highest levels
147 of expression of SRP genes in FOXA1-overpressing VCaP cells as compared with DU145 cells (Fig.
148 2F and Supplementary Table 4). Taken together, our data identify a subset of seven up-regulated
149 and FOXA1-associated SRP genes in primary PCa that phenocopy FOXA1 in PCa cells that over-
150 express FOXA1.

151

152

153 *FOXA1 regulates SRP gene expression in vitro in human PCa cell line models*

154

155 To determine whether the above candidate SRPs are regulated by FOXA1 *in vitro*, we utilized two
156 independent siRNA duplexes to deplete PCa cell lines of FOXA1 (Fig. 3A-D, left panels and
157 Supplementary Table 5). siRNA-mediated depletion of FOXA1 protein in VCaP cells resulted in a
158 reduction in expression of six out of seven SRPs genes by qRT-PCR, where statistical significance
159 was observed with at least one siRNA duplex (p-value <0.05, Fig. 3A, right panel, Supplementary
160 Tables 6-7). A similar effect of FOXA1 depletion on SRP gene expression was observed in LNCaP
161 (Fig. 3B, right panel, 3/7 SRPs), PC3 (Fig. 3C, right panel, 4/7 SRPs) and DU145 (Fig. 3D right, 4/7
162 SRPs). We then tested whether ectopic expression of FOXA1 in PC3 cells, which harbour lower
163 FOXA1 protein level (Fig. 2C and D), would conversely result in an increase in SRP gene expression
164 (Fig. 3E). We observed a general increase in SRP gene expression, with four SRP genes reaching
165 statistical significance (Fig. 3E, right panel).

166 Overall, there was a statistically-significant reduction in *HNRNPK* expression levels of in all
167 four cell lines upon FOXA1 depletion, and a statistically-significant increase upon the overexpression
168 of FOXA1 in PC3 cells (Fig. 3F). With the exception of the DU145 cell line that expresses the lowest
169 level of FOXA1, we observed a consistent impact of *FOXA1* depletion and overexpression on
170 *HNRNPA1* and *HNRNPK* expression (Fig. 3F). These data demonstrate that FOXA1 can regulate
171 SRP gene expression in PCa cell lines, with the most consistent effect observed in FOXA1-
172 overexpressing VCaP cells.

173

174 *ERG and AR do not appear to regulate SRP gene expression in human PCa cell line models*

175

176 Of the candidate SRP genes, a number were bound by AR and ERG as well as FOXA1 (Fig. 2E).
177 Since FOXA1 has been implicated as a pioneer factor [28] and is co-opted by AR [29] and ERG [30] in PCa
178 cells, we hypothesised that the observed FOXA1 regulated SRP gene expression is mediated by AR
179 and ERG. To test this hypothesis, we used siRNA to deplete ERG in the VCaP cells, and AR in VCaP
180 and LNCaP cells (Supplementary Fig. 3, left panels). Surprisingly, following siRNA-mediated ERG

knockdown, we observed an increase in expression of two out of seven SRP genes by qRT-PCR (Supplementary Fig. 3A, right panel), which appeared to be due to an increase in FOXA1 protein expression as determined by western blotting (Supplementary Fig. 3A, left panel). siRNA-mediated AR knockdown in LNCaP and VCaP cells did not significantly impact on FOXA1 protein expression nor SRP gene expression (Supplementary Fig. 3B and C). Consistent with our findings *in silico* (Fig. 1), these data confirm that AR and ERG are not as important as FOXA1 in SRP gene regulation in PCa, and suggest that this role may be a novel AR-independent function of FOXA1.

FOXA1-associated SRP gene dysregulation is persistent in advanced treatment-relapsed PCa

We sought to determine whether the dysregulation of SRP genes persists in late-stage, metastatic, castration-resistant PCa (CRPCa) which inevitably develops following longstanding androgen deprivation therapy (ADT) [5], and whether this is associated with expression of *FOXA1*, *ERG*, *AR*, *HOXB13* and *MYC*. We utilised published RNA-Seq gene expression data of CRPCa samples from patients (n=118) included in the Stand Up to Cancer (SU2C) study [31]. As before, transcriptomes were stratified by expression levels (*i.e.* TPMs) of TFs (Fig. 4A) with a cut-off of the top 25% of gene expression by TPM defining HE. We did not detect a statistically significant difference in the levels of *FOXA1* expression between the *FOXA1* HE tumours in the TCGA and SU2C datasets (two-tailed Wilcoxon rank sum test, Supplementary Fig. 4A). Next, we repeated the GSA to determine the biological processes that are altered upon HE of TF genes by comparing the cumulative TPM values of genes in the 16 KEGG gene sets between samples with TF gene HE and Rest (Supplementary Fig. 4B). Compared with Rest, GSA identified significant associations of only *FOXA1* and *MYC* HE with the SRP gene set (Fig. 4B). It is worth noting that for both primary PCa and CRPCa, GSA provided consistent results only for sample stratification by *FOXA1* expression and to a minor extent by *MYC* expression. (Figs. 2C and 4B).

To determine whether FOXA1 regulates expression of the same SRP genes in both primary PCa and CRPCa, we compared significantly altered SRP genes associated with HE *FOXA1* in the TGCA and SU2C datasets (Fig. 4C, see Methods). We identified 16 SRP genes associated with HE

FOXA1 in both primary PCa and CRPCa (Fig. 4D). These data demonstrate the persistence of the FOXA1-SRP association in gene expression in treatment-relapsed CRPCa, which includes a subset of SRP genes (*HSPA8*, *HNRNPK*, *SF3B1* and *SNRNP200*) that both phenocopy FOXA1 and are regulated by FOXA1 *in vitro* (Figs. 2 and 3).

Discussion

FOXA1 is a member of a group of forkhead box domain-containing transcription factors which interacts with chromatin and binds DNA, causing nucleosome rearrangement and stabilising an open chromatin conformation [32]. FOXA1 has been described as a pioneer factor [28] for other transcription factors allowing the re-defining of cistromes [7,30,33,34], including the AR cistrome during prostate tumourigenesis [7,35]. *FOXA1* is one of the most commonly mutated genes in PCa [36,37] and disease susceptibility loci fall within the FOXA1 cistrome [10], thereby highlighting its dysregulation in PCa.

Little is known of the mechanisms underpinning the transcriptional regulation of genes encoding RNA-binding and other proteins involved in AS [11,16], which we term as SRPs. To date, only MYC has been shown to transcriptionally drive expression of SRPs [20-22], thereby presenting AS as a cancer therapeutic vulnerability for *MYC*-driven tumours [18,19]. Here, we show for the first time that FOXA1, but not AR or ERG, regulates a subset of SRP genes that phenocopy the *FOXA1* dependency of PCa cells, thereby expanding the AR-independent FOXA1 gene regulatory repertoire [33,35]. AR-independent FOXA1-driven transcriptional programmes have been shown to occur via genomic interactions with other transcriptional regulators such as MYBL2 and CREB1 in CRPCa [33], GATA-3 and the Estrogen Receptor (ER) in breast cancer [38], and PPARγ in bladder cancer [39]. Additionally, interactions between ER and glucocorticoid receptor (GR) and FOXA1 appear to be dynamic [40], hence our AR-independent observations may be context-dependent in PCa.

Considering the overlapping cistromes for FOXA1 and the ETS-family TF ERG [30], we were surprised that depletion of *ERG* and *FOXA1* had opposing effects on SRP gene expression. We concluded that this was due to an ERG-driven increase in *FOXA1* expression by the *ERG* siRNA. Overexpression of oncogenic ERG typically occurs as a result of a recurrent gene fusion between the

AR-regulated *TMPRSS* gene and *ERG* in ~50% PCa cases [41]. Hence, our findings may highlight differing mechanisms of FOXA1-mediated SRP gene regulation in *ERG* fusion-positive and -negative PCa. Since loss of the tumour suppressor *PTEN* appears to be required for widespread co-operative AR and ERG-driven transcriptomic changes [8], we may have failed to identify ERG-driven SRP gene expression changes in PTEN-proficient VCaP cells. However, our data in PTEN-deficient LNCaP cells suggest that AR does not regulate SRP gene expression in the absence of PTEN.

A recent genome-wide screen of PCa cell dependencies [42] identified the heterogeneous nuclear ribonuclear protein (hnRNP) family of SRPs, including the FOXA1-regulated SRP genes *HNRNPA1* and *HNRNPK*, which have been previously implicated in PCa [43-46] and other malignancies [47]. The SRP hnRNPA1 has been shown to regulate expression of the CRPCa- and FOXA1-associated AR splice variant AR-V7 [43,44,46,48,49]. Interestingly, we show that FOXA1 can control expression of *SF3B1*, which is the most-commonly altered SRP gene in PCa [3] and haematological malignancies [50]. Lethality can be induced in cancers with mutations affecting *SF3B1* and other SRP genes by therapeutic targeting the SF3b complex of the spliceosome[51], although the impact in cancers with up-regulated (but wild-type) *SF3B1* and other SRP genes is unknown. We also identify the SRP *HSPA8*, which encodes a heat shock protein (HSP) scaffold in the core spliceosome complex [52], as a FOXA1-regulated gene that exhibited a high dependency for PCa cells. Therapeutic targeting the HSP family member HSP90 has been shown to harbour anti-tumour activity and also modulate AS in CRPCa [53,54].

The role of FOXA1 in advanced PCa is still contradictory, with reports of AR-dependent [55] and -independent functions [33,35]. On one hand, FOXA1 protein expression is up-regulated in primary PCa [34,56,57], metastases and CRPCa [56], and is associated with metastasis [58], disease recurrence [56,57,59], and survival [34]. However, on the other hand, FOXA1 has also been described as an inhibitor of metastasis [35] and neuroendocrine differentiation [60], which is associated with a poor prognosis. We were unable to identify a difference in the levels of *FOXA1* expression amongst the FOXA1 HE tumours in the TCGA and SU2C datasets, however, we did observe a persistent association between *FOXA1* and SRP genes in CRPCa.

Our data demonstrate, for the first time in both primary PCa and CRPCa, that FOXA1 is associated with SRP gene expression, the dysregulation of which confers a poor patient prognosis. In a subset of FOXA1 binding and actively-transcribed SRP genes that phenocopy the *FOXA1* dependency of PCa cells, we confirm FOXA1-regulated SRP gene expression in PCa cell lines. Hence, we speculate that in both primary PCa and CRPCa, targeting SRPs may represent a therapeutic vulnerability for FOXA1-overexpressing PCa in an analogous way to *MYC*-driven cancers. This would need to be tested in future studies by therapeutic targeting of SRPs or upstream signaling cascades.

Methods

Gene set analysis of the human prostate cancer transcriptome

RNA sequencing (RNA-Seq) data were downloaded from The Cancer Genome Atlas (TCGA) Data Matrix portal (Level 3, https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm) and from cBioPortal [61,62] websites for 409 primary untreated and 118 metastatic (Stand Up to Cancer, SU2C [31]) PCa samples, respectively. The number of transcripts per million reads (TPM) was measured starting from the scaled estimate expression values provided for 20,531 genes as previously described [63]. For the SU2C dataset RPKM (Reads Per Kilobase of transcript per Million mapped reads) values were converted into TPM. For each TF, the distribution of expression levels across samples was measured. A TF was considered as highly expressed (HE) if its TPM value was greater or equal to the 75th percentile of the distribution [63] (Supplementary Data 1).

A list of 16 manually curated gene sets representing Genetic Information Processing (2.1-2.4) from the Kyoto Encyclopedia of Genes and Genomes (KEGG) were downloaded from MSigDb version 5 [24]. A list of 66 additional RBPs available from the RNAcompete catalogue [14] were added to the KEGG spliceosome gene set (n=128). To further refine a list of genes encoding SRPs, a gene ontology (GO) analysis of biological processes was performed using the function *gprofiler* in the R

288 'gProfileR' [64] on the total set of 194 genes. A final gene set of 148 genes with GO terms related to
289 splicing (SRPs) was retained for further analyses.

290 For each TF *t* and each gene set *i*, the cumulative TPM values of genes in *i* were compared
291 between samples where the *t* was HE and the remaining ones (Rest) using a two-tailed Wilcoxon
292 rank sum test (Supplementary Fig. 1A). The resulting p-values were corrected for multiple tests using
293 the Bonferroni method. To control for false discoveries, Monte Carlo simulation was implemented as
294 previously described [65]. For 10,000 times, we randomly extracted 103 samples (corresponding to the
295 number of samples with HE TF) and the cumulative expression of each gene set was compared with
296 that of the remaining samples. Next, for each gene set, the empirical p-value is measured as the
297 number of times the p-value is smaller than the observed one over the total number of iterations.

298

299 *Patient survival analysis*

300 Clinical data were downloaded from the TCGA Data Matrix portal. Disease-free survival time was
301 defined as the interval between the date of treatment and disease progression, as defined by
302 biochemical or clinical recurrence, or until end of follow-up [2]. The relationship with disease recurrence
303 for genes within the top 6 gene sets *i* significantly associated with HE of all TFs was tested using a
304 multivariable Cox proportional hazards (PH) model, and coefficients α for each gene *j* were used to
305 calculate a patient gene set score as following:

306 $$S(i) = \sum_{j=1}^{n} \alpha_j * e_j$$

307 where e is the expression level of each gene.

308 For each gene set, patients were then stratified on the 75th percentile of the score distribution, and a
309 univariable Cox PH model was used to generate hazard ratios (HR) between patients with a high
310 score as compared to those with a low score. For the SRP gene set scores, event-time distributions

for the time to disease progression were compared using the log-rank test. The PH assumption and influential observations were met. All analyses were performed using the R 'survival' package [66].

Linear regression modelling of the SRP gene set

SRP gene expression was fitted on the expression of the TFs using a linear regression model. The search for the best subsets of regressor was performed using a branch-and-bound algorithm [67] implemented in the *regsubsets* function in th R 'leaps' package. For models using a different number of variables (i.e. from 1 to 5 TFs) the best model in terms of determination coefficient R^2 was reported. Relative importance of regressors in the linear regression model of five TFs was calculated using the function *calc.relimp* in the R 'reclaimpo' package [68]. This function divides the determination coefficient R^2 into the contribution of each regressor using the averaging over ordering method [25]. The confidence intervals of the contributions of the regressors were measured using a bootstrap procedure implemented in the function *boot.relaimp*. For 1,000 iterations the full observation vectors were resampled and the regressor contributions were calculated.

Selection of highly expressed SRPs

SRPs that were highly expressed between *FOXA1* HE (n=103 and 30 for TCGA and SU2C, respectively) and Rest (n=306 and 88 for TCGA and SU2C, respectively) were identified comparing the TPM distributions of the two groups with a one-tailed Kolmogorov-Smirnov (KS) test. The resulting p-values were corrected for multiple tests using the Bonferroni method. To control for size differences between the two cohorts, a Monte Carlo procedure was implemented. For 10,000 times, *FOXA1* HE and Rest samples were randomly selected and, for each SRP, the TPM distributions were compared using a one-tailed KS test. Next, for each SRP gene, the empirical p-value was measured as the proportion of tests with p-value smaller than the corresponding observed one over the total number of iterations. Differentially expressed genes (DEG) between *FOXA1* HE and Rest samples were detected using the R package 'DESeq2' and 'EdgeR' for the TCGA dataset, for which raw sequencing

counts were available. Briefly, read counts of 20,531 genes of each sample were used as input for DESeq2 and EdgeR. Genes with read count equal to zero across all samples were removed. For the TCGA dataset, a total of 76 SRP genes with a KS Bonferroni-corrected p-value and empirical p-value less than 0.05, a false discovery rate (FDR) ≤ 0.1 and an absolute $\log_2$ Fold-Change (FC) ≥ 0.2 measured by DESeq2 or EdgeR were considered as having an altered expression in *FOXA1* HE samples as compared to Rest. For the SU2C dataset, a total of 26 SRP genes with a KS Bonferroni-corrected p-value and empirical p-value less than 0.05 and an absolute $\log_2$ (FC) ≥ 0.2 were considered as altered (Supplementary Data 2).

Selection of TFs implicated in SRP gene expression

Enrichment analysis of TF binding sites within SRP genes was performed using chromatin immunoprecipitation sequencing (ChIP-Seq) data from the ReMap database and annotation tool [69]. Genomic coordinates of SRP genes were extended by 5,000 bp and used as input for the ReMap enrichment tools (http://tagc.univ-mrs.fr/remap/index.php?page=annotation). Binding sites of TFs that significantly overlapped (minimum 10%) with the input regions were collected. For all five TFs, binding sites in regions of active transcription as defined by H3K27ac and H3K4me3 epigenetic modification markers were further selected. In particular, H3K27ac ChIP-Seq replicated narrow peaks from prostate epithelial cells were collected from the ENCODE Data Matrix (https://www.encodeproject.org/files/ENCFF655JIF/) and mapped on hg19 using The University of California, Santa Cruz (UCSC) liftover software (http://genome-euro.ucsc.edu/cgi-bin/hgLiftOver). H3K4me4 ChIP-Seq narrow peaks from LNCaP cells were retrieved from the UCSC hg19 database (http://hgdownload.soe.ucsc.edu/goldenPath/hg19). TF binding site regions were retained if overlapping with at least 10% of their sequence with H3K27ac and H3K4me3 peaks (Supplementary Data 2).

Cell dependency analysis

Gene dependency data for 17,098 genes across 4 PCa cell lines (DU145, PC3, LNCaP, VCaP) [27] were downloaded from Project Achilles data portal (https://portals.broadinstitute.org/achilles). DEMETER inferred z-scores representing gene knockdown effects were extracted for *FOXA1* and SRP genes.

Cell lines, antibodies, plasmids, oligonucleotides

DU145 (HTB-81, ATCC), PC3 (CRL-1435, ATCC), LNCaP (CRL-1740, ATCC), and VCaP (CRL-2876, ATCC) cells were obtained from American Type Culture Collection (ATCC) and identities confirmed by Short Tandem Repeat (STR) profiling (DDC Medical). pcDNA3.1 FOXA1 was provided by Jason Carroll (Cancer Research UK Cambridge Institute). The following antibodies were used: anti-FOXA1 (Abcam: ab23738), anti-actin (Sigma: A1978), anti-AR (BD Bioscience: 554225), anti-ERG (Santa Cruz: sc-271048), anti-mouse IgG HRP-linked (Dako: P044701-2), anti-rabbit IgG HRP-linked (Dako: P044801-2). Sequences used to generate siRNA duplexes are as previously described [70] or commercially-designed (ON-TARGETPlus, Dharmacon Horizon Discovery) and are listed in Supplementary Table 3. Sequences used to generate oligonucleotide primers for PCR were designed by entering the Ensembl (http://www.ensembl.org) Transcript ID representing the principal isoform for each gene into the University Probe Library (UPL) Assay Design Centre (https://lifescience.roche.com/en_gb/brands/universal-probe-library.html#assay-design-center). Sequences were checked using the National Center for Biotechnology Information (NCBI) Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast) prior to synthesis (Integrated DNA Technologies). Primer sequences are listed in Supplementary Table 8.

Cell Culture, DNA and RNA transfections

Cells were incubated at 37°C, 5% CO_2 in a humidified incubator. Cells were maintained at sub-confluency in RPMI-1640 medium (21875-034, Gibco) (DU145, PC3 and LNCaP) or DMEM (41966-029, Gibco) (VCaP) containing 2 mM L-glutamine, supplemented with 10% foetal calf serum (FCS)

(Gibco), 100 units/ml penicillin and 100 µg/ml streptomycin (15140-122, Gibco) and regularly tested for the presence of mycoplasma. Transfections with plasmid DNA and siRNA duplexes were carried out as detailed in the figure legends using ViaFect (E4981, Promega) and RNAiMax (13778-075, Thermo Fisher Scientific), respectively, according to manufacturers' instructions.

Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis (SDS-PAGE) and Western blotting

Whole cell lysate protein samples were obtained by lysis of cells in RIPA (Radio-Immunoprecipitation Assay) buffer for 30 minutes at 4°C followed by lysate clearing by centrifugation. Protein concentration was calculated using the bicinchoninic acid (BCA) assay (10678484, Thermo Fisher Scientific) method and samples adjusted to equal concentrations of total protein. Samples were denatured in a 2-Mercapto-ethanol- based SDS sample buffer. Proteins were then separated by SDS -PAGE, transferred onto PVDF (polyvinylidene difluoride) membrane (000000003010040001, Sigma) using the wet transfer method, blocked in 5% milk in TBST (Tris-Buffered Saline and Polysorbate 20) and then placed in primary antibodies diluted in 5% BSA (Bovine Serum Albumin) in TBST over-night at 4°C. Membranes were washed and incubated with relevant HRP-conjugated secondary antibodies for 1 hour at room temperature. For signal detection, membranes were washed and incubated for 3 minutes each in Luminata Crescendo Western HRP substrate (10776189, Thermo Fisher Scientific) before bands were visualised on a Chemidoc system (Amersham Imager 600, Amersham). Antibody concentrations were as follows: anti-FOXA1 (1:1000), anti-actin (1:100,000), anti-AR (1:1000), anti-ERG (1:1000); HRP-linked secondaries (1:5000). Where indicated, densitometric assessments of protein bands were performed using Image Studio Lite Ver 5.2 (LI-COR), and signal intensities used to calculate relative normalised FC in protein expression.

RNA extraction and quantitative reverse transcription polymerase chain reaction (qRT-PCR)

Total RNA was isolated from cells by direct lysis in TRIzol Reagent (15596026, Thermo Fisher Scientific) according to the manufacturer's instructions and contaminating genomic DNA removed using DNase I (AMPD1-1KT, Sigma). Reverse transcription (RT) to cDNA was achieved using the

High-Capacity cDNA Reverse Transcription Kit (4368813, Thermo Fisher Scientific). qRT-PCR was performed on the 7500 Fast Real-Time PCR machine (Applied Biosystems, 4351106) using triplicate cDNA templates with the FastStart Universal Probe Master with ROX (4913949001, Roche) and UPL set (04683633001, Roche) according to the manufacturer's instructions. Only primers within a 10% efficiency range from 90-100% were included (Supplementary Table 8). Reaction conditions were as follows: 20 s at 50 °C, 10 min at 95 °C, and 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Relative gene expression was determined by the $2^{-\Delta\Delta CT}$ method using the geometric mean of two validated endogenous control genes (*ACTB* and *B2M*) to ensure the reliability and reproducibility of observed effects. Data shown are from three independent biological experimental replicates with two technical replicates.

Cell viability assays

Cell viability assays were performed using (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (MTT) (L11939.06, Alfa Aesar) according to the manufacturer's instructions. Briefly, 4000–10 000 cells were seeded into each well of a 96-well plate and grown to ~20–30% confluence prior to transfection with siRNA. After 72 h, MTT was added to each well to a final concentration of 0.67 mg/ml and incubated at 37°C, 5% CO_2 in a humidified incubator for 2 h. Subsequently, MTT reagent was removed, 100µl dimethyl sulfoxide (DMSO) (10213810, Thermo Fisher Scientific) added to each well and agitated at room temperature for 15 mins. Absorbance was measured at 560nm and 630nm (SpectraMax Plus384 Absorbance Microplate Reader, Molecular Devices), and normalised by subtracting the 630nm value from the 560nm value, and percentage viability calculated as follows: Treatment absorbance ÷ DMSO control absorbance ×100. All siRNA data were normalized to a non-silencing control. Results shown are the means ± SEM of at least three independent experiments with at least 3 technical replicates.

442 *Statistical analysis for in vitro data*

443 Graphical data shown represent the means ± standard error of the mean (SEM) of independent
444 experiments. The one-tailed independent sample t-test was employed to identify differences in means
445 between groups with p-value < 0.05 taken to indicate statistical significance.

446

447 ## Figure Legends

448 **Fig. 1: FOXA1 expression is independently associated with SRP expression in primary PCa** (A)
449 Boxplot distributions of normalized gene expression levels (transcript per million reads, TPMs) for
450 each of the five TF genes in primary PCa samples. Green lines and dots refer to samples with TF
451 expression greater than the 75th percentile of its distribution across samples. PCa-related TFs are
452 reported in red. (B) Heat map showing the stratification of tumour samples accordingly to the
453 expression levels of each of the TF genes. Green cells depict samples exhibiting the high expression
454 (HE) of the corresponding TF. The last row reports samples undergoing ERG gene fusion. (C) Results
455 of the gene set analysis (GSA) for 16 biological KEGG processes related to the HE of each TF. The
456 altered processes are hierarchically clustered on the basis of their statistical significance (i.e. adjusted
457 p-value). Non-statistically significant processes are depicted in white dashed boxes. The side panel
458 reports the corresponding KEGG pathway category. (D) Hazard Ratios (HR) from univariable Cox
459 proportional hazards (PH) model using gene set scores for the six biological processes associated
460 with FOXA1 HE. (E) Kaplan-Meier plot of survival probabilities over time for patients stratified on the
461 75th percentile of the SRP gene set score distribution with log rank test p-value. (F) Coefficient of
462 determination (R2) of the linear regression model of the SRP gene expression for increasing number
463 of regressors in the model (i.e. the TFs included in the model). F=FOXA1; M=MYC; E=ERG;
464 H=HOXB13; A=AR. (G) Relative importance of each predictor (i.e. TF) to the R2 measured by the
465 linear regression model including all TFs.

466

467 **Fig. 2: FOXA1 and SRP gene expression correlate in PCa cell line models** (A) Enrichment
468 analysis of TF binding sites within SRP genes was performed using chromatin immunoprecipitation

sequencing (ChIP-Seq) data from the ReMap database and annotation tool, as described in the Materials and Methods section. (B) Enrichment values for the five selected TFs generated from the ReMap analysis. Red indicates TFs that are significantly up-regulated in FOXA1 HE compared to Rest. (C) Representative Western blotting images of whole cell lysates from PCa cell lines using antibodies to AR, ERG and FOXA1, and actin. (D) Densitometric band quantitation was performed to calculate $\log_{10}$ relative normalized fold change (FC) in expression relative to DU145. Data from at least three independent experiments were used to calculate the means ± SEM. (E) Heatmap depicts TPM values of candidate SRPs in FOXA1 HE and rest PCa samples (first block), fold change of expression (second block), number of TF binding sites in region of active transcription (third block), DETEMER Z-score for SRP dependency in VCaP, LNCaP, PC3 and DU145 cell lines (forth block). DEMETER Z-score for FOXA1 in the four cell lines is reported. Selected SRPs for *in vitro* validation in PC cell lines are reported in black. Missing values or zeros are depicted by white dashed blocks. In the upper block SRPs are sorted in decreasing order of TPM values in FOXA1 samples. In the lower block SRPs are sorted in decreasing order of FC. (F) qRT-PCR was performed on cDNAs from different PCa cells lines, and levels of SRP transcript expression were normalized to a geometric mean of *ACTB* and *B2M* levels to calculate $\log_{10}$ relative normalized FC in expression compared to DU145 cells. Data from at least three independent experiments were used to calculate the means ± SEM.

Fig. 3: FOXA1 regulates SRP expression in PCa cell lines *in vitro* (A) DU145, (B) PC3, (C) LNCaP, and (D) VCaP were transfected with two siRNA duplex sequences to FOXA1 (FOXA1 si1 or si2), or non-silencing (NSI) control to final concentration of 40 nM. (A-D, Left Panels) After 72 h, total cell lysates were harvested and subjected to western blotting with antibodies to FOXA1 and actin. Western blotting images shown are representative of three independent experiments, from which densitometric band quantitation was performed to calculate the mean relative normalized fold change (FC) in protein expression (shown below FOXA1 blot images). (A-D, Right Panels) qRT-PCR was performed on cDNAs and levels of SRP transcript expression were normalized to a geometric mean of *ACTB* and *B2M* levels to calculate relative normalised FC in expression compared to NSI. Data

from at least three independent experiments were used to calculate the means ± SEM. (E) PC3 cells were transfected with expression vectors for pcDNA3.1-FOXA1 or vector only (VO) control (1 µg) as indicated. After 72 h, total cell lysates were harvested and subjected to western blotting with antibodies to FOXA1 and actin (E, left panel). Western blotting images shown are representative of three independent experiments, from which densitometric band quantitation was performed to calculate the mean relative normalized FC in protein expression (shown below FOXA1 blot images). (E, right panel) qRT-PCR was performed on cDNAs and levels of SRP transcript expression were normalized to a geometric mean of ACTB and B2M levels to calculate relative normalised FC in expression compared to VO. Data from at least three independent experiments were used to calculate the means ± SEM. Unpaired two-tailed T-test was used to compared groups: *p-value <0.05, **p-value <0.01, ***p-value <0.001, ****p-value <0.0001. (F) Heatmap showing significant expression changes of SRP genes (black cells) in the different PCa cell lines upon FOXA1 silencing and overexpression.

Fig. 4: FOXA1 expression is independently associated with SRP expression in metastatic PCa (A) Boxplot distributions of normalized gene expression levels (transcript per million reads, TPMs) for each of the five TFs in metastatic CRPCa samples (SU2C dataset). Green lines and dots refer to sample with TF expression greater than the 75th percentile of its distribution across samples. PCa-related TFs are reported in red. (B) Statistically significant results from the gene set analysis (GSA) for 16 biological KEGG processes related to the high expression (HE) of each TF. The altered processes are hierarchically clustered on the basis of their statistical significance (i.e. adjusted p-value). Non-statistically significant processes are depicted in white. The side panel reports the corresponding KEGG pathway category. Gene sets that were significantly altered by the five TFs in *FOXA1* HE PCa samples are reported in cyan. (C) Venn diagram showing the overlap between 76 and 26 significantly altered SRPs in PC and CRPCa samples, respectively, upon *FOXA1* HE. (D) Heatmap showing TPM values of altered SRPs in both primary PCa (TGCA) and CRPCa (SU2C) samples upon *FOXA1* HE and relative fold changes (FC). SRPs validated *in vitro* in PCa cell lines are reported in black. Missing values are depicted by white dashed blocks.

Acknowledgements

The following reagents were generously gifted: VCaP cells (from Y-J. Lu, Barts Cancer Institute, UK), and pcDNA3-FOXA1 (from J. Carroll, Cancer Research UK Cambridge Institute, UK). This work was funded by a joint Royal College of Surgeons of England/Cancer Research UK Clinician Scientist Fellowship in Surgery (C19198/A15339 to PR), The Urology Foundation and John Black Charitable Foundation (to PR), the Barts Charity, the Orchid Charity (to JDK and NL) and by the Italian Association for Cancer Research (AIRC MFAG 20566 to MC and IG 20240 to SO).

Author Contributions

JGF, MC and PR designed research. JGF, RA, MDG, CA, AL, and MC performed research. JDK, NRL and SO contributed new reagents or analytic tools. JGF, MDG, CA, MC, and PR analysed data. JGF, MC and PR wrote the paper.

Materials & Correspondence

Correspondence and material requests should be addressed to either j.foster@qmul.ac.uk, matteo.cereda@iigm.it or p.rajan@qmul.ac.uk

References

1 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66**, doi:10.3322/caac.21332.

2 Network, C. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011-1025, doi:10.1016/j.cell.2015.10.025 (2015).

3 Armenia, J. *et al.* The long tail of oncogenic drivers in prostate cancer. *Nat Genet* **50**, 645-651, doi:10.1038/s41588-018-0078-z (2018).

4 Wedge, D. C. *et al.* Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat Genet* **50**, 682-692, doi:10.1038/s41588-018-0086-z (2018).

5 Cornford, P. *et al.* EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part II: Treatment of Relapsing, Metastatic, and Castration-Resistant Prostate Cancer. *Eur Urol* **71**, 630-642, doi:10.1016/j.eururo.2016.08.002 (2017).

6 Singh, I. *et al.* Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat Commun* **9**, 1716, doi:10.1038/s41467-018-04112-z (2018).

7 Pomerantz, M. M. *et al.* The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat Genet* **47**, 1346-1351, doi:10.1038/ng.3419 (2015).

8 Chen, Y. *et al.* ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat Med* **19**, 1023-1029, doi:10.1038/nm.3216 (2013).

9 Whitington, T. *et al.* Gene regulatory mechanisms underpinning prostate cancer susceptibility. *Nat Genet* **48**, 387-397, doi:10.1038/ng.3523 (2016).

10 Dadaev, T. *et al.* Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. *Nat Commun* **9**, 2256, doi:10.1038/s41467-018-04109-8 (2018).

11 Kechavarzi, B. & Janga, S. C. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol* **15**, R14, doi:10.1186/gb-2014-15-1-r14 (2014).

12 Seiler, M. *et al.* Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep* **23**, 282-296 e284, doi:10.1016/j.celrep.2018.01.088 (2018).

13 Wang, Z. L. *et al.* Comprehensive Genomic Characterization of RNA-Binding Proteins across Human Cancers. *Cell Rep* **22**, 286-298, doi:10.1016/j.celrep.2017.12.035 (2018).

14 Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172-177, doi:10.1038/nature12311 (2013).

15 Dvinge, H., Kim, E., Abdel-Wahab, O. & Bradley, R. K. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* **16**, 413-430, doi:10.1038/nrc.2016.51 (2016).

16 Munkley, J., Livermore, K., Rajan, P. & Elliott, D. J. RNA splicing and splicing regulator changes in prostate cancer pathology. *Hum Genet* **136**, 1143-1154, doi:10.1007/s00439-017-1792-9 (2017).

17 Lee, S. C. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat Med* **22**, 976-986, doi:10.1038/nm.4165 (2016).

18 Iwai, K. *et al.* Anti-tumor efficacy of a novel CLK inhibitor via targeting RNA splicing and MYC-dependent vulnerability. *EMBO Mol Med* **10**, doi:10.15252/emmm.201708289 (2018).

19 Hsu, T. Y. *et al.* The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **525**, 384-388, doi:10.1038/nature14985 (2015).

20 Koh, C. M. *et al.* MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature* **523**, 96-100, doi:10.1038/nature14351 (2015).

21 Das, S., Anczukow, O., Akerman, M. & Krainer, A. R. Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. *Cell Rep* **1**, doi:10.1016/j.celrep.2011.12.001.

22 David, C. J., Chen, M., Assanah, M., Canoll, P. & Manley, J. L. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**, 364-368, doi:10.1038/nature08697 (2010).

23 Cereda M., L. A., Caselle M., Oliviero S. Characterization of biological processes in heterogeneous cohorts by discretization of expression profiles. *Nature Communications* **(Under revision)** (2018).

24 Subramanian, A. *et al.* in *Proc Natl Acad Sci USA* Vol. 102 15545-15550 (2005).

25 Lindeman, R. H. Introduction to bivariate and multivariate analysis. (1980).

26 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

27 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516, doi:10.1016/j.cell.2017.06.010 (2017).

28 Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227-2241, doi:10.1101/gad.176826.111 (2011).

29 Zhao, J. C. *et al.* FOXA1 acts upstream of GATA2 and AR in hormonal regulation of gene expression. *Oncogene* **35**, 4335-4344, doi:10.1038/onc.2015.496 (2016).

30 Kron, K. J. *et al.* TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nature genetics*, doi:10.1038/ng.3930 (2017).

31 Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228, doi:10.1016/j.cell.2015.05.001 (2015).

32 Yang, Y. A. & Yu, J. Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. *Genes Dis* **2**, 144-151, doi:10.1016/j.gendis.2015.01.003 (2015).

33 Zhang, C. *et al.* Definition of a FoxA1 Cistrome that is crucial for G1 to S-phase cell-cycle transit in castration-resistant prostate cancer. *Cancer Res* **71**, 6738-6748, doi:10.1158/0008-5472.CAN-11-1882 (2011).

34 Sahu, B. *et al.* Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J* **30**, 3962-3976, doi:10.1038/emboj.2011.328 (2011).

35 Jin, H.-J. J., Zhao, J. C., Wu, L., Kim, J. & Yu, J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nature communications* **5**, 3972, doi:10.1038/ncomms4972 (2014).

36 Carver, B. S. *et al.* Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nature genetics* **41**, 619-624, doi:10.1038/ng.370 (2009).

37 Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* **44**, 685-689, doi:10.1038/ng.2279 (2012).

38 Lee, J. Y., Park, Y. J., Oh, N., Kwack, K. B. & Park, K. S. A transcriptional complex composed of ER(alpha), GATA3, FOXA1 and ELL3 regulates IL-20 expression in breast cancer cells. *Oncotarget* **8**, 42752-42760, doi:10.18632/oncotarget.17459 (2017).

39 Warrick, J. I. *et al.* FOXA1, GATA3 and PPAR Cooperate to Drive Luminal Subtype in Bladder Cancer: A Molecular Analysis of Established Human Cell Lines. *Sci Rep* **6**, 38531, doi:10.1038/srep38531 (2016).

40 Swinstead, E. E. *et al.* Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell* **165**, 593-605, doi:10.1016/j.cell.2016.02.067 (2016).

41 Ayala, G. *et al.* Expression of ERG protein in prostate cancer: variability and biological correlates. *Endocr Relat Cancer* **22**, 277-287, doi:10.1530/ERC-14-0586 (2015).

42 Fei, T. *et al.* Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1617467114 (2017).

43 Nadiminty, N. *et al.* NF-kappaB2/p52:c-Myc:hnRNPA1 Pathway Regulates Expression of Androgen Receptor Splice Variants and Enzalutamide Sensitivity in Prostate Cancer. *Mol Cancer Ther* **14**, 1884-1895, doi:10.1158/1535-7163.MCT-14-1057 (2015).

44 Capaia, M. *et al.* A hnRNP K(-)AR-Related Signature Reflects Progression toward Castration-Resistant Prostate Cancer. *Int J Mol Sci* **19**, doi:10.3390/ijms19071920 (2018).

45 Ciarlo, M. *et al.* Regulation of neuroendocrine differentiation by AKT/hnRNPK/AR/beta-catenin signaling in prostate cancer cells. *Int J Cancer* **131**, 582-590, doi:10.1002/ijc.26402 (2012).

46 Tummala, R., Lou, W., Gao, A. C. & Nadiminty, N. Quercetin Targets hnRNPA1 to Overcome Enzalutamide Resistance in Prostate Cancer Cells. *Mol Cancer Ther* **16**, 2770-2779, doi:10.1158/1535-7163.MCT-17-0030 (2017).

47 Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* **135**, 851-867, doi:10.1007/s00439-016-1683-5 (2016).

48 Li, H. *et al.* Prognostic Value of Androgen Receptor Splice Variant 7 in the Treatment of Castration-resistant Prostate Cancer with Next generation Androgen Receptor Signal Inhibition: A Systematic Review and Meta-analysis. *Eur Urol Focus*, doi:10.1016/j.euf.2017.01.004 (2017).

49 Hsu, T. Y. *et al.* The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **525**, doi:10.1038/nature14985.

50 Inoue, D. & Abdel-Wahab, O. Modeling SF3B1 Mutations in Cancer: Advances, Challenges, and Opportunities. *Cancer Cell* **30**, 371-373, doi:10.1016/j.ccell.2016.08.013 (2016).

51 Seiler, M. *et al.* H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nat Med* **24**, 497-504, doi:10.1038/nm.4493 (2018).

52 van Maldegem, F. *et al.* CTNNBL1 facilitates the association of CWC15 with CDC5L and is required to maintain the abundance of the Prp19 spliceosomal complex. *Nucleic Acids Res* **43**, 7058-7069, doi:10.1093/nar/gkv643 (2015).

53 Moses, M. A. *et al.* Targeting the Hsp40/Hsp70 Chaperone Axis as a Novel Strategy to Treat Castration-Resistant Prostate Cancer. *Cancer Res* **78**, 4022-4035, doi:10.1158/0008-5472.CAN-17-3728 (2018).

54 Ferraldeschi, R. *et al.* Second-Generation HSP90 Inhibitor Onalespib Blocks mRNA Splicing of Androgen Receptor Variant 7 in Prostate Cancer Cells. *Cancer Res* **76**, 2731-2742, doi:10.1158/0008-5472.CAN-15-2186 (2016).

55 Robinson, J. L. L. *et al.* Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. *Oncogene* **33**, 5666-5674, doi:10.1038/onc.2013.508 (2014).

56 Gerhardt, J. *et al.* FOXA1 promotes tumor progression in prostate cancer and represents a novel hallmark of castration-resistant prostate cancer. *Am J Pathol* **180**, 848-861, doi:10.1016/j.ajpath.2011.10.021 (2012).

57 Robinson, J. L. *et al.* Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. *Oncogene* **33**, 5666-5674, doi:10.1038/onc.2013.508 (2014).

58 Jain, R. K., Mehta, R. J., Nakshatri, H., Idrees, M. T. & Badve, S. S. High-level expression of forkhead-box protein A1 in metastatic prostate cancer. *Histopathology* **58**, 766-772, doi:10.1111/j.1365-2559.2011.03796.x (2011).

59 Tsourlakis, M. C. *et al.* FOXA1 expression is a strong independent predictor of early PSA recurrence in ERG negative prostate cancers treated by radical prostatectomy. *Carcinogenesis* **38**, 1180-1187, doi:10.1093/carcin/bgx105 (2017).

60 Kim, J. *et al.* FOXA1 inhibits prostate cancer neuroendocrine differentiation. *Oncogene* **36**, 4072-4080, doi:10.1038/onc.2017.50 (2017).

61 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404, doi:10.1158/2159-8290.CD-12-0095 (2012).

62 Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1, doi:10.1126/scisignal.2004088 (2013).

63 Cereda, M. *et al.* Patients with genetically heterogeneous synchronous colorectal cancer carry rare damaging germline mutations in immune-related genes. *Nat Commun* **7**, 12072, doi:10.1038/ncomms12072 (2016).

64 Reimand, J. *et al.* g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res* **44**, W83-89, doi:10.1093/nar/gkw199 (2016).

65 Cereda, M. *et al.* RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing. *Genome Biol* **15**, R20, doi:10.1186/gb-2014-15-1-r20 (2014).

66 Therneau, T. M. & Grambsch, P. M. *Modeling survival data: extending the Cox model*. (Springer Science & Business Media, 2013).

67 Miller, A. *Subset selection in regression*. (Chapman and Hall/CRC, 2002).

68 Gromping, U. Relative importance for linear regression in R: The package relaimpo. *J Stat Softw* **17** (2006).

69 Cheneby, J., Gheorghe, M., Artufel, M., Mathelier, A. & Ballester, B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res* **46**, D267-D275, doi:10.1093/nar/gkx1092 (2018).

70 Zheng, L. *et al.* FOXA1 positively regulates gene expression by changing gene methylation status in human breast cancer MCF-7 cells. *Int J Clin Exp Pathol* **8**, 96-106 (2015).

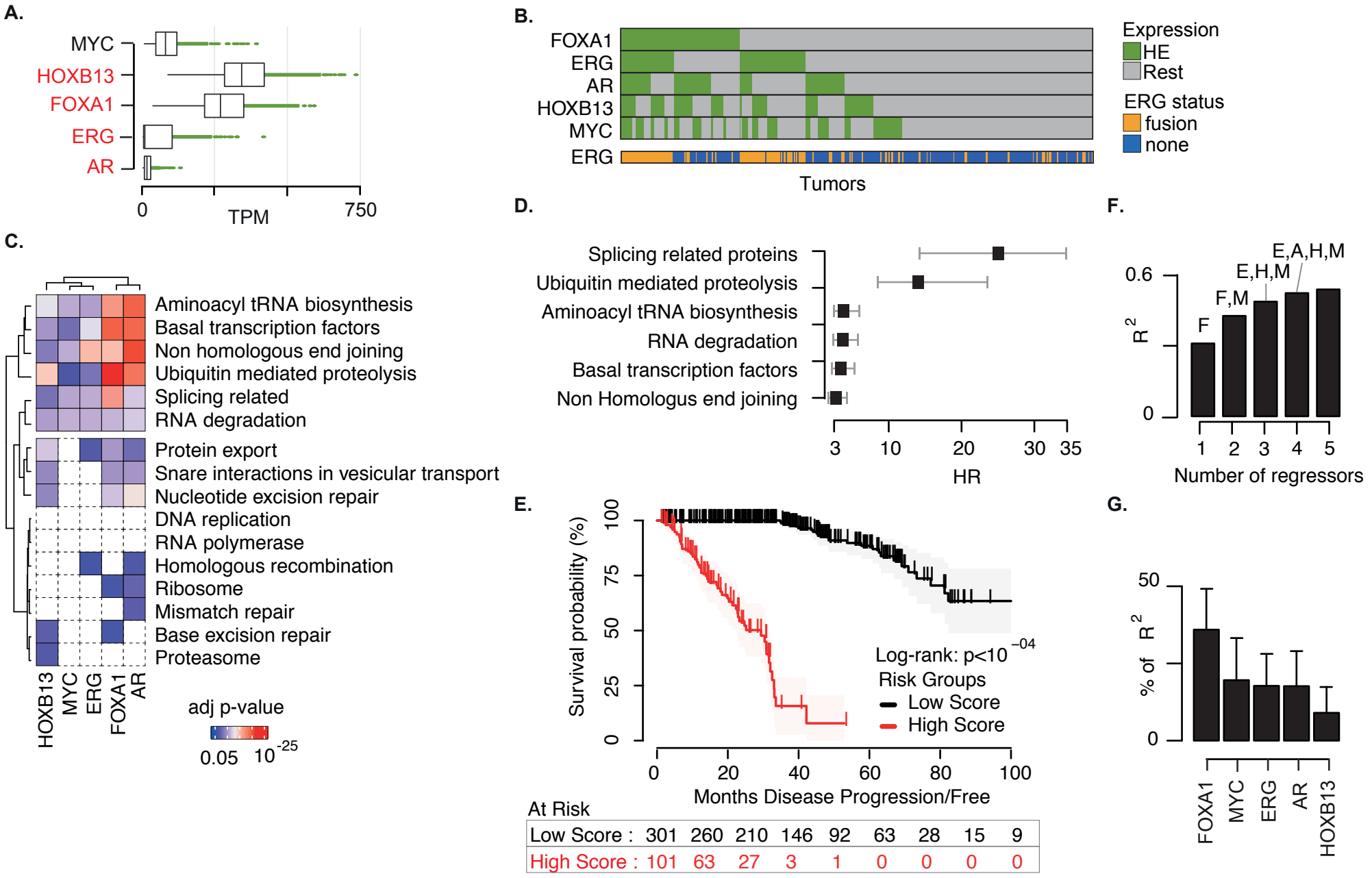

Fig. 1

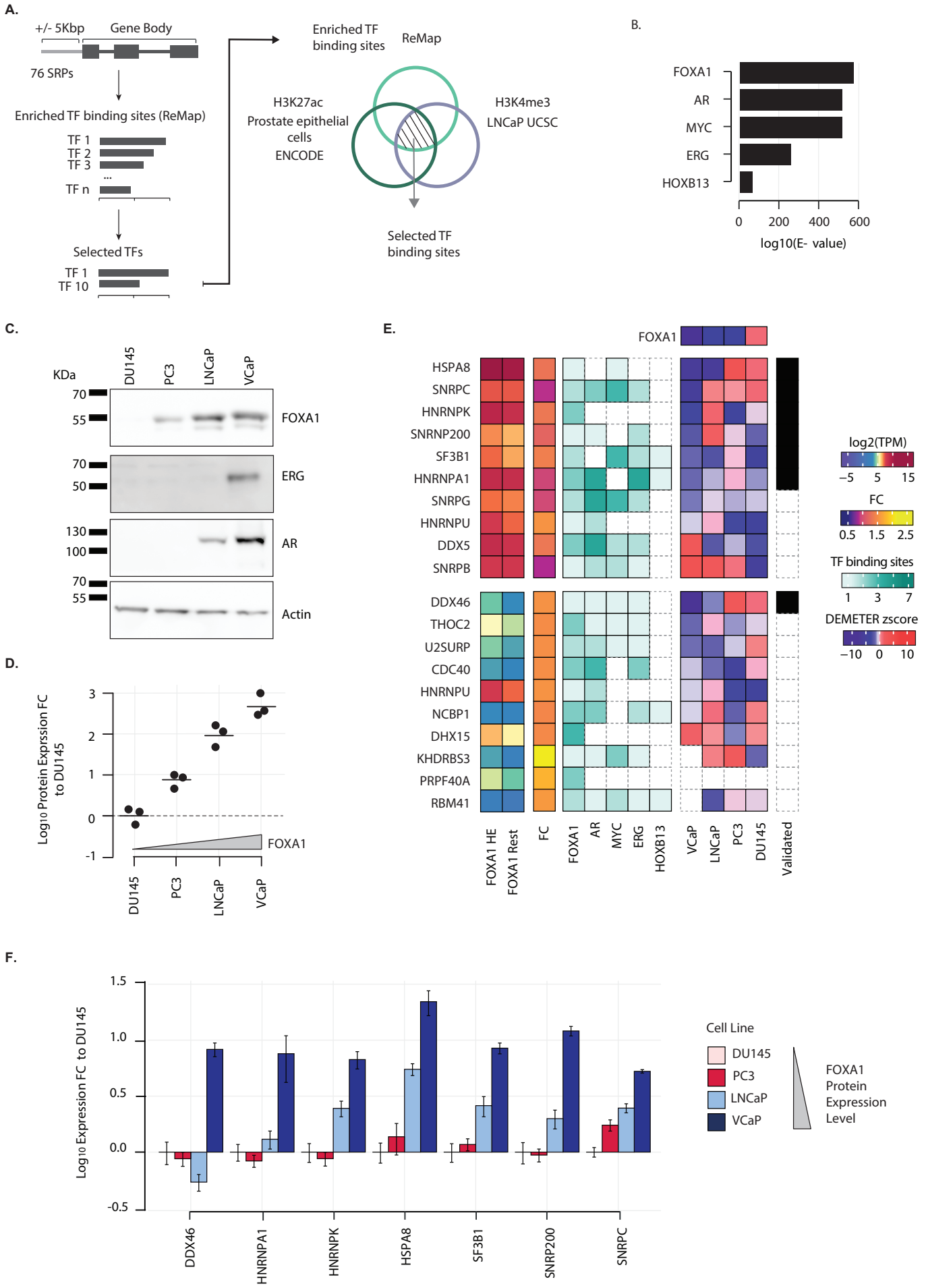

Fig. 2

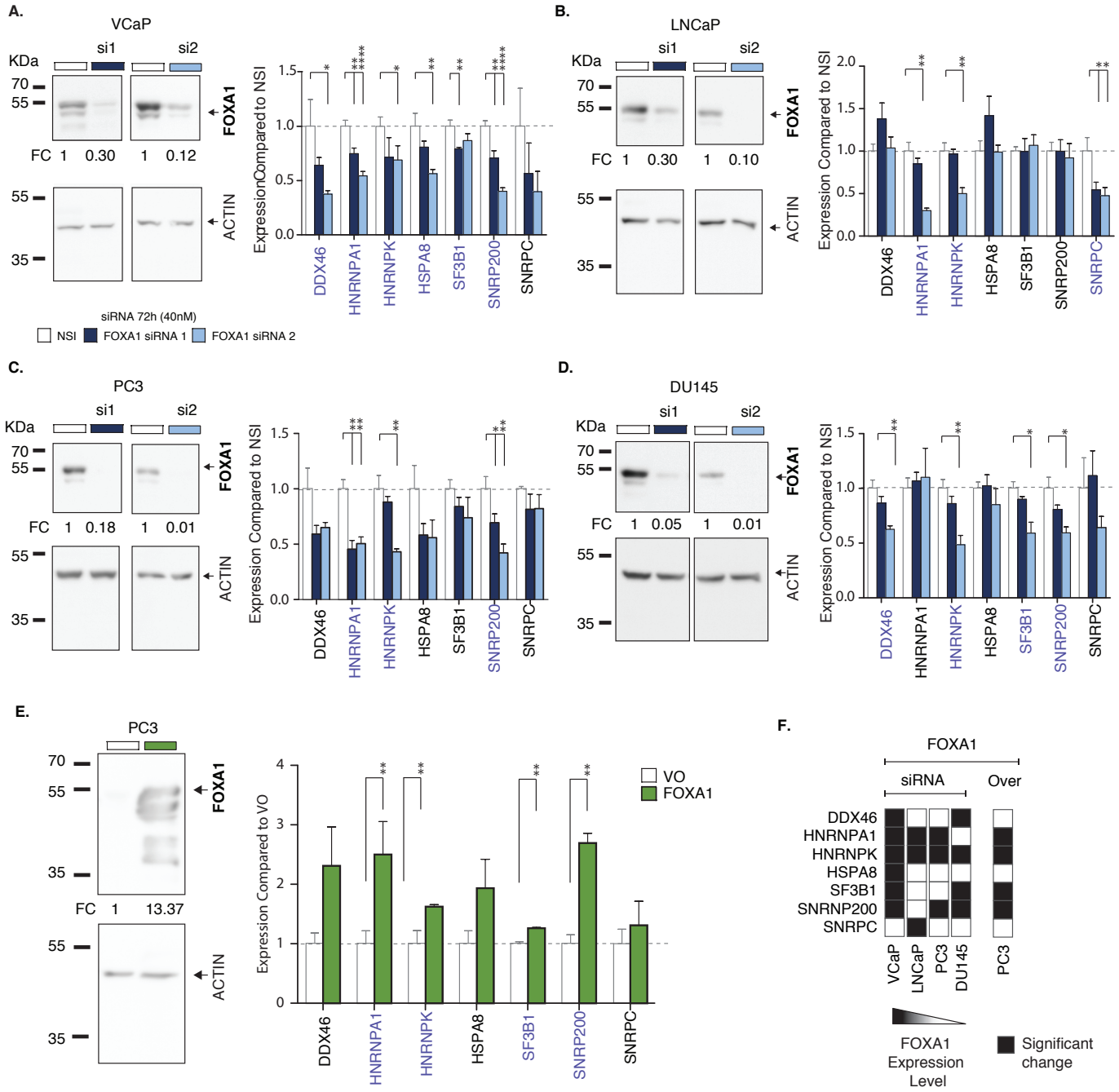

Fig. 3

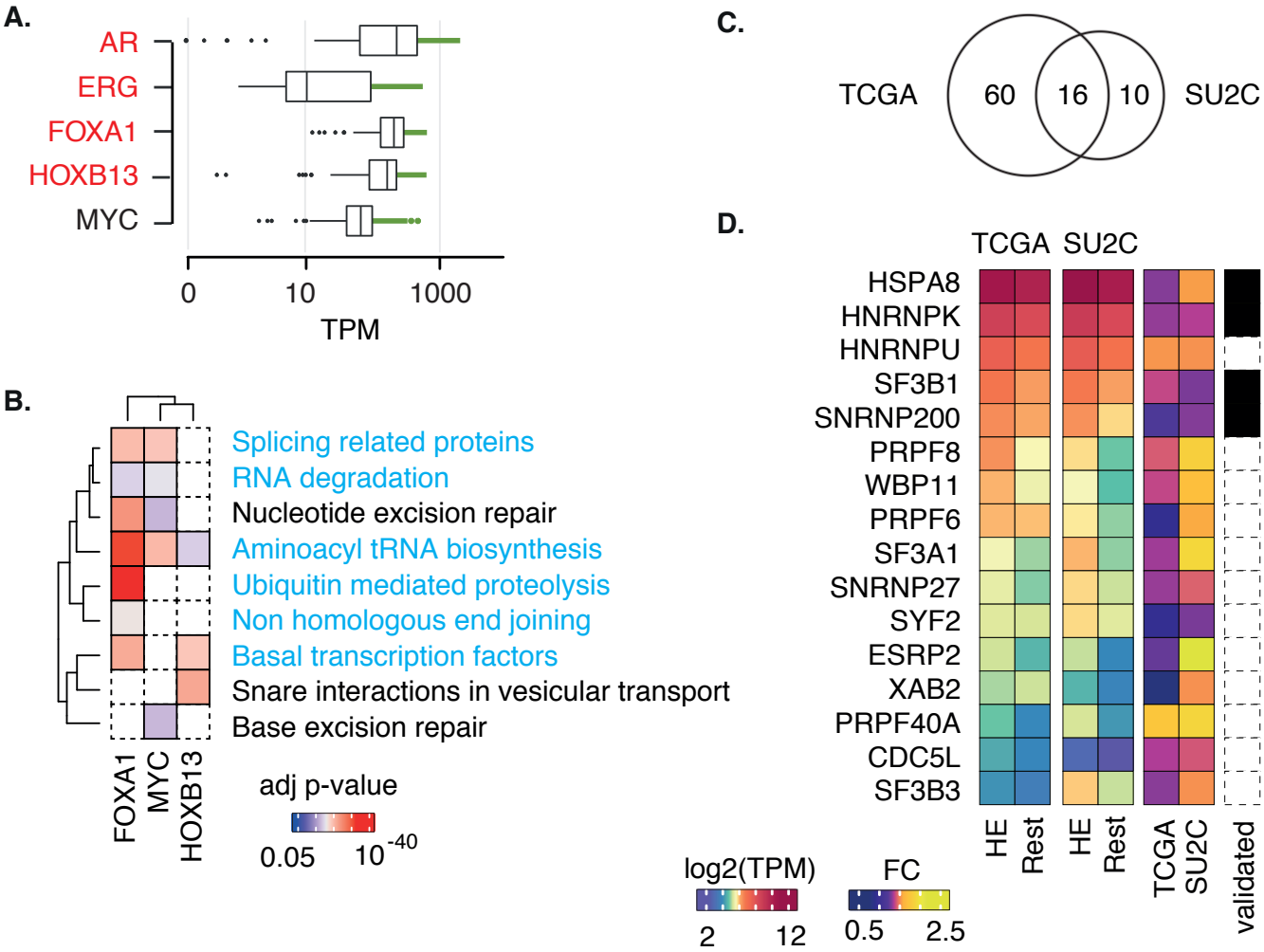

Fig. 4