

1 Conserved Role of FOXC1 in TNBC is parallel to FOXA1 in ER+ Breast
2 Cancer

3

4 Revathy Ramachandran¹, Shakhzada Ibragimova^{1†}, Laura M. Woods^{2†}, Tamader AlHouqani¹,
5 Roshna Lawrence Gomez¹, Fabrizio Simeoni⁵, Mahmood Y. Hachim¹, Tim C. P. Somerville⁵,
6 Anna Philpott^{2,3}, Jason S. Carroll⁴, and Fahad R. Ali^{1*}

7

8 ¹ College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences,
9 Dubai, United Arab Emirates.

10 ² Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre,
11 Cambridge Biomedical Campus, Cambridge, United Kingdom.

12 ³ Department of Oncology, University of Cambridge, Cambridge, United Kingdom.

13 ⁴ Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United
14 Kingdom.

15 ⁵ Cancer Research UK Manchester Institute, University of Manchester, Manchester, United
16 Kingdom.

17 [†]Authors contributed equally

18 *Corresponding author email address: fahad.ali@mbru.ac.ae

19

20 **Running title:** Parallel role of FOXC1 and FOXA1 in breast cancer

21 **Keywords:** Triple-negative breast cancer, FOXC1, gene regulation, ChIP-seq, heterogeneity,
22 oncogenes, tumor suppressors, FOXA1, NR2F2

23

24 SUMMARY

25 Triple-negative breast cancer (TNBC) is characterized by lack of estrogen (ER),
26 progesterone, and HER2 receptors, and standard receptor-targeted therapies are ineffective.
27 FOXC1, a transcription factor aberrantly overexpressed in many cancers, drives growth,
28 metastasis, and stem-cell-like properties in TNBC. However, the molecular function of
29 FOXC1 is unknown, partly due to heterogeneity of TNBC. Here, we show that although
30 FOXC1 regulates many cancer hallmarks in TNBC, its function is varied in different cell lines,
31 highlighted by the differential response to CDK4/6 inhibitors upon FOXC1 loss. Despite this
32 functional heterogeneity, we show that FOXC1 regulates key oncogenes and tumor
33 suppressors and identify a set of core FOXC1 peaks conserved across TNBC cell lines. We
34 identify the ER-associated and drug-targetable nuclear receptor NR2F2 as a cofactor of
35 FOXC1. Finally, we show that core FOXC1 targets in TNBC are regulated in parallel by the
36 pioneer factor FOXA1 and the nuclear receptor NR2F2 in ER+ breast cancer.

37

38 INTRODUCTION

39 Triple-negative breast cancer (TNBC) is defined by the absence of the estrogen receptor
40 ($ER\alpha$), progesterone receptor (PR), and the human epidermal growth factor receptor-2
41 (ERBB2/HER2) in tumor cells. TNBC accounts for around ~15% of all diagnosed breast
42 cancer cases and is aggressive, prone to metastasis, and has few targeted therapies available.
43 (1, 2). A compounding issue in TNBC treatment is its heterogeneous nature, reflected in the
44 high variability of TNBC cell line models (3). Sub-classification of TNBC based on
45 transcriptional or molecular properties has been attempted, mainly to identify
46 responsiveness to different therapies and afford patients the best available treatment

47 options (4). In the absence of hormone receptor-driven growth signalling, it is unclear what
48 drives the proliferation of tumour cells in TNBC (5). This is in contrast to the luminal breast
49 cancer subtype, where the transcriptional cascade spearheaded by ER α and its co-factors
50 stimulates proliferative pathways and are thus targeted by therapies (6).

51 Crucial to the development of targeted therapies in TNBC is the identification of key
52 drivers of proliferation and their detailed molecular function. Transcription factor (TF)
53 FOXC1 is over-expressed in TNBC and has diagnostic and prognostic value, with its
54 expression correlating with chemoresistance and metastasis in basal-like breast cancer (7-
55 10). Over-expression of FOXC1 in TNBC cell lines has generally been shown to increase
56 tumorigenic properties such as proliferation, invasion, and migration (9, 11, 12), although,
57 one study reported inhibition of metastasis upon over-expression of FOXC1 (13). FOXC1
58 belongs to the human Forkhead family of around 50 TFs that contain the highly conserved
59 Forkhead DNA binding domain. It is expressed during embryogenesis and is crucial for
60 mesenchymal lineage specification and organ development, including bone, cartilage, and eye
61 (14). FOXC1 is overexpressed in many cancers such as breast, bladder, colorectal, acute
62 myeloid leukaemia (AML), hepatocellular carcinoma, and non-small cell lung cancer (10) and
63 is a patented clinical marker for TNBC, particularly for the basal-like phenotype (7, 15).
64 FOXC1 has been shown to be involved with EGFR function, NF- κ B, canonical and non-
65 canonical Wnt, PI3K/Akt/mTOR and non-canonical Hedgehog signaling (reviewed in (10)).

66 Nonetheless, genome-wide analysis of FOXC1 binding sites and identification of
67 FOXC1-regulated genes in TNBC has not been reported. Furthermore, there is variation in
68 expression of the FOXC1 protein among TNBC cell lines, and it remains unknown whether
69 the level of FOXC1 protein expression correlates with its function. Since TNBC is a

70 heterogeneously classified subgroup of breast cancer, the identification of core conserved
71 FOXC1-regulated gene networks will elucidate the functional role of FOXC1 in TNBC and
72 could improve our understanding of TNBC pathogenesis. In this study, using transcriptomic,
73 epigenetic, proteomic, and bioinformatic analysis, we aim to identify an exhaustive list of
74 FOXC1 targets and cofactors and define, in greater detail, its conserved molecular function
75 across breast cancer.

76

77 RESULTS

78

79 FOXC1 regulates oncogenic phenotypes via varied gene targets in TNBC

80 To study the function of FOXC1 in TNBC, we generated CRISPR-edited FOXC1 gene
81 knockouts (FOXC1_KO) in four TNBC cell lines: BT-549, Hs578t, MDA-MB-231 and MDA-
82 MB-468 (Table S1). The cell lines were selected to represent variation in FOXC1 protein
83 expression, with BT-549 and Hs578t expressing higher amounts (Fig. 1a). Homozygous
84 clones successfully edited for FOXC1 KO were verified using Sanger sequencing, and
85 absolute loss of protein expression was confirmed using western blot (Fig. 1a).

86

87 Upon loss of FOXC1, growth of the FOXC1_KO clones was slower than their respective
88 parental cells (Fig. 1b). Morphological changes were observed in all cell lines, with the
89 FOXC1_KO clones adopting a flatter, cuboid appearance with decreased cell-cell adhesion
90 and in the case of Hs578t, losing the mesenchymal morphology altogether (Fig. S1a).
91 Invasion, migration and colony forming assays corroborate the role of FOXC1 as an
92 oncogene that regulates proliferative, migratory, invasive, and clonogenic properties in

93 TNBC, although the percent reduction in oncogenic phenotypes varied among cell lines and
94 did not correlate with the FOXC1 protein expression levels (Fig. S1b and S1c). The cell cycle
95 profile between the parental and FOXC1_KO clones was not significantly altered, except in
96 the case of Hs578t, where a higher proportion of cells were present in the S-phase in
97 FOXC1_KO (Fig. S1d), suggesting a cell line specific function of FOXC1.

98

99 In order to identify common genes and pathways regulated by FOXC1 across TNBC,
100 RNA-seq analysis was performed on the four parental TNBC cell lines ('WT') and matched
101 FOXC1-deleted CRISPR clones ('KO'). Principal component analysis and correlation of gene
102 expression profiles show that the cell line accounts for the greatest source of the variation in
103 the dataset, with a smaller effect of FOXC1 status (Fig. S2a and Fig. S2b, respectively).
104 Differentially expressed genes (DEGs) were identified using filtering criteria (False Discovery
105 Rate (FDR) <0.05) and no correlation was observed between the number of DEGs and
106 FOXC1 protein expression, as both BT-549 and MDA-MB-231 had a similar number of
107 DEGs even though BT-549 has higher FOXC1 protein expression (Fig. S2c). Enrichment of
108 pathways in each individual cell line revealed regulation of major cancer hallmarks such as
109 epithelial to mesenchymal transition, apoptosis, hypoxia and hormone response (Fig. 1c).
110 However, upset plots show a limited set of DEGs that overlap and are similarly regulated
111 (up- or down-) in all four cell lines (Fig. 1d). Although 172 genes were significantly (padj
112 value < 0.05) differentially expressed between WT and FOXC1_KO, of the four cell lines (Fig.
113 S2d), only 26 genes were similarly regulated, (either up- or down-regulated across all four
114 cell lines); these 26 genes represent a high stringency list of the FOXC1 regulon in TNBC
115 (Fig. 1e). Unsupervised k-means clustering employed to the 4 WT vs. KO pairs generated 13

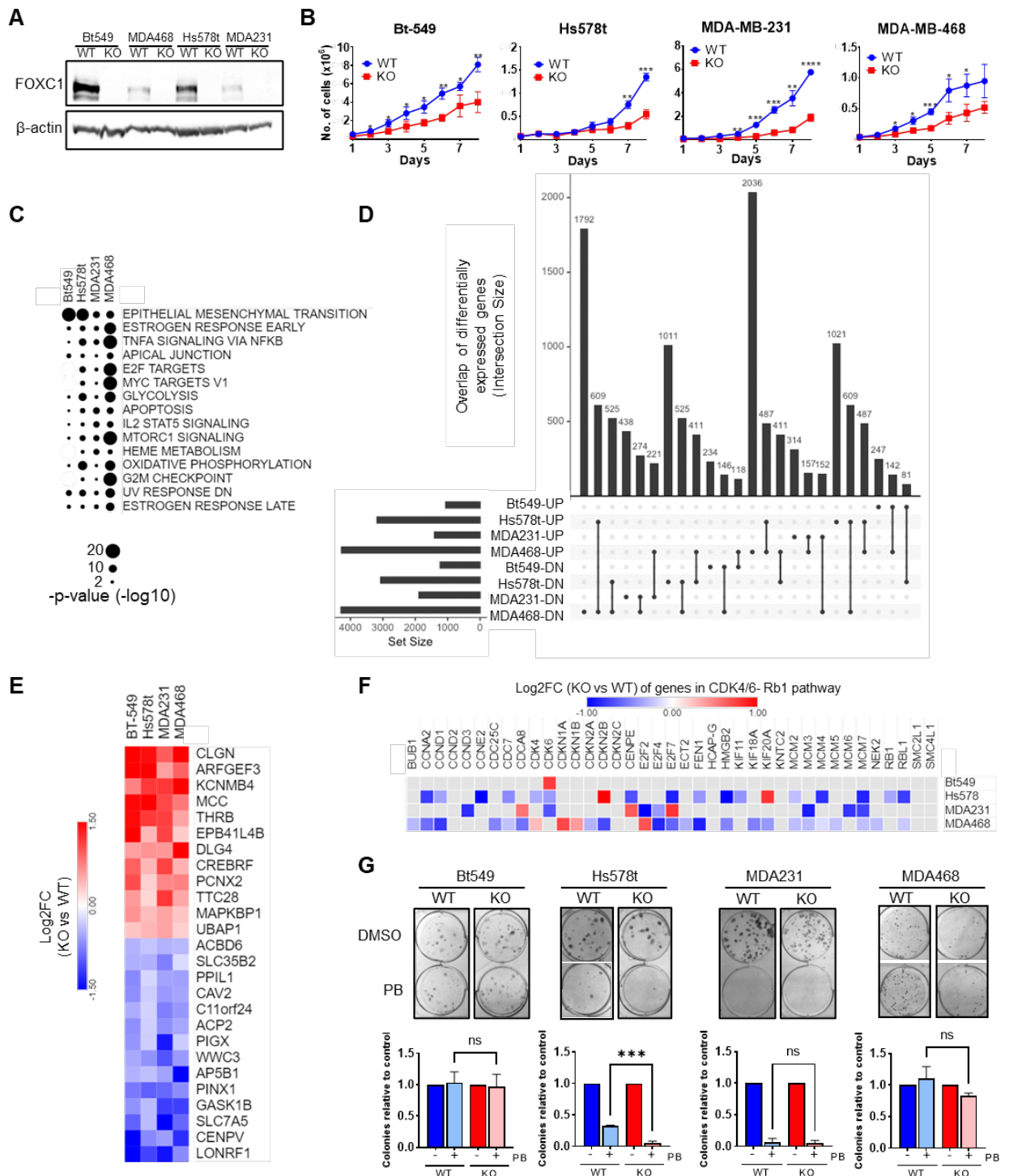
116 clusters based on the pattern of gene expression (Gene expression patterns Fig. S2e,
117 Heatmap Fig. S2f). Enrichment of GO terms biological processes reveal downregulation of
118 DNA replication upon loss of FOXC1 in Cluster 8 (similarly downregulated genes) and no
119 significantly enriched terms in the Cluster 13, (upregulated genes) (Fig. S2g). The lack of
120 majority in either cluster suggests that the functional FOXC1 regulon in TNBC is largely cell-
121 line dependent. Although FOXC1 has been implicated in the silencing of ER α (16), no
122 significant change in *ESR1* gene expression was observed (Fig. S2h).

123

124 In Hs578t, MDA-MB-231, and MDA-MB-468, targets of E2F were significantly enriched
125 (Fig. 1c), and many genes involved in the RB1 pathway were differentially expressed upon
126 loss of FOXC1, except in BT-549, where only *CDK6* was downregulated upon FOXC1_KO
127 (Fig. 1f). To validate the FOXC1-induced gene expression changes in the CDK4-6/Rb/E2F
128 pathway genes observed in RNA-seq data, clonogenic assays were performed to quantify
129 changes in sensitivity to Palbociclib, a CDK4/6 inhibitor that is used in treatment of ER+
130 tumors. Interestingly, only in Hs578t did the loss of FOXC1 further sensitize the cells to
131 Palbociclib, as seen by the reduction in the number of colony-forming units in the
132 Hs578t_KO (Fig. 1g). Both BT549_KO and MDA-MB-468_KO remained equally resistant to
133 Palbociclib as the parental cell lines, and MDA-MB-231_KO showed no further sensitivity
134 than the parental. FOXC1's role in modulating the sensitivity of Hs578t to Palbociclib, a
135 clinically relevant CDK4/6 inhibitor, necessitates the study of FOXC1 gene targets in more
136 detail while further emphasizing the diversity of FOXC1 function in TNBC.

137

Figure 1



138

139

140 Figure 1: Effect of FOXC1 knockout on phenotype and gene expression

141 (A) Western blot of cell lysate (75 μ g) from four parental TNBC cell lines ('WT') and the

142 corresponding FOXC1 knockout CRISPR clones ('KO'), probed using anti-FOXC1 and anti- β -actin

143 antibodies. **(B)** Proliferation assays showing growth of FOXC1_KO clones (red) when compared to
144 the parental cell lines (blue). Viable cells were counted every 24 hours. Data represent mean and
145 standard deviation (SD) of three biological replicates. Asterisk denotes significant digits in p-value
146 derived from unpaired t-tests with * < 0.05, ** < 0.005 and *** < 0.0005. **(C)** Enrichment using
147 Hallmark gene sets (MSigDB (17)) showing significantly enriched cancer hallmarks among the DEGs
148 in each of the four TNBC cell lines. **(D)** Upset plot depicting intersection between up- or down-
149 regulated genes among the four cell lines, showing a limited set of DEGs that overlap and are
150 similarly regulated (up- or down-) in all four cell lines. **(E)** Heatmap depicting log₂(fold change) of
151 the 26 genes significantly regulated upon the loss of FOXC1 in all four cell lines, representing genes
152 repressed by FOXC1 (red) and genes activated by FOXC1 (blue). **(F)** Log₂(fold change) of selected
153 genes in the RB1/CDK4/6 pathway (FOXC1_KO vs WT). **(G)** Colony-forming assays in the presence
154 of either DMSO (top row) or 0.5 μM Palbociclib (bottom row) for 14 days prior to crystal violet
155 staining. Bar graphs (bottom panel) display relative reduction in the number of colonies compared to
156 the DMSO control in the parental (WT, blue) or the FOXC1_KO (KO, red) groups. Data represents
157 mean ± SD from three biological replicates each. Asterisk denotes significant digits in p-value derived
158 from unpaired t-tests. *** < 0.005.

159

160

161 Genome-wide mapping of FOXC1 binding sites in four TNBC cell lines

162 To identify genes directly regulated by FOXC1, we mapped genome-wide FOXC1 binding
163 sites using CHIP-seq in four TNBC cell lines. A high correlation was observed among the
164 peaks identified in the three biological replicates within each cell line (Fig. S3a), and
165 subsequently, the three replicates were merged before peak calling. Nearly half of the
166 FOXC1 binding sites were observed between promoters, 5' UTR, and first intronic sequences,
167 with about 25% present in distal intergenic, likely enhancer, regions (Fig. S3b). This was
168 supported by the average peak distribution relative to TSS, with the majority of peaks at 10-
169 100 kb from TSS (Fig. S3c). In sum, 863 peaks were common between all four cell lines, with
170 MDA-MB-468 containing the highest number of peaks (Fig. S3d). There was higher
171 conservation among FOXC1 peaks than FOXC1-associated DEGs across the four cell lines
172 (863 common peaks versus 26 DEGs in Fig. 1e) suggesting cell-line context dominates gene
173 expression changes.

174
175 In each cell line, we identified differentially expressed genes (FOXC1_KO vs WT) within 50
176 kb of a peak and identified the functional, directly regulated gene targets of FOXC1. Thirty-
177 eight genes were associated with peaks and significantly ($p_{adj} < 0.05$) differentially
178 expressed upon FOXC1 knockout in all four cell lines. We believe these stringently selected
179 38 genes are conserved FOXC1 direct targets in TNBC (Fig. 2a). Within this list, FOXC1
180 represses expression of tumor suppressors (*CREBRF*, *THRB*, *INPP4B*, *ARHGAP24*, *CYBRD1*)
181 and activates the expression of oncogenes (*CRABP2*, *SLC7A5*, *PDE7B*, *ASAP1*, *WWC3*),
182 (illustrated in Fig. 2b, c and Fig. S4). The expression of the majority of these FOXC1 targets
183 (21/38) is prognostic (p -value < 0.05) in predicting relapse-free survival of patients with

184 ER/PR/HER2 negative breast cancer (Fig. S5a-d). The expression of a panel of 17 genes
185 upregulated by FOXC1 correlates with poor survival (Fig. 2d), underscoring the clinical
186 relevance of FOXC1 regulated gene targets in TNBC. The expression of a panel of 12
187 downregulated genes did not significantly correlate with survival (Fig. S5b).

188 The top *de novo* consensus binding motif for FOXC1 in high confidence peaks was
189 similar in all four cell lines (Fig. S3e). The consensus binding sites were overlaid with that
190 identified from FOXC1-ChIP-seq binding data in an acute myeloid leukaemia cell line (Fujioka
191 cells) (18) and revealed a slightly different consensus binding motif (Fig. S3e), suggesting a
192 degree of tissue/disease-specificity in FOXC1 binding sites (AML versus TNBC). TF motifs
193 enriched within 100 bp of a FOXC1 peak using Homer revealed similarly enriched families in
194 all four TNBC cell lines (Fig. S3f). The highest enrichment was seen for 2 main families of
195 motifs, the Forkhead/winged helix factor motif and the basic leucine zipper factor motif
196 (bZIP), which includes TFs such as FOSL2, JUNB, and AP1. Enrichment of KEGG pathways
197 from among genes associated with FOXC1 peak in each cell line revealed Rap1, PI3-AKT,
198 MAPK, Hippo, and Ras signalling pathways as conserved between all four cell lines, followed
199 by TGF β and Wnt signalling pathway as conserved in at least 2 cell lines (Fig. S3g).

200

201 Using publicly available H3K27ac ChIP-seq data, we were able to observe that FOXC1
202 binding sites are present largely in active promoters and accessible enhancer regions, as
203 these are enriched for H3K27ac histone modifications (Fig. 2e). Quantification of the FOXC1
204 and H3K27ac signal intensity at each overlapped site revealed that MDA-MB-468 had the
205 smallest fraction of FOXC1 peaks associated with H3K27ac signal (Fig. 2f). Thus, although
206 MDA-MB-468 had the highest number of FOXC1 peaks (Fig. S3d), most peaks did not

207 associate with active chromatin regions. In contrast, in MDA-MB-231, which had the
208 smallest number of FOXC1 peaks, a majority associated with H3K27ac histone signal,
209 suggesting that FOXC1 has a more 'active' role in that cell line.

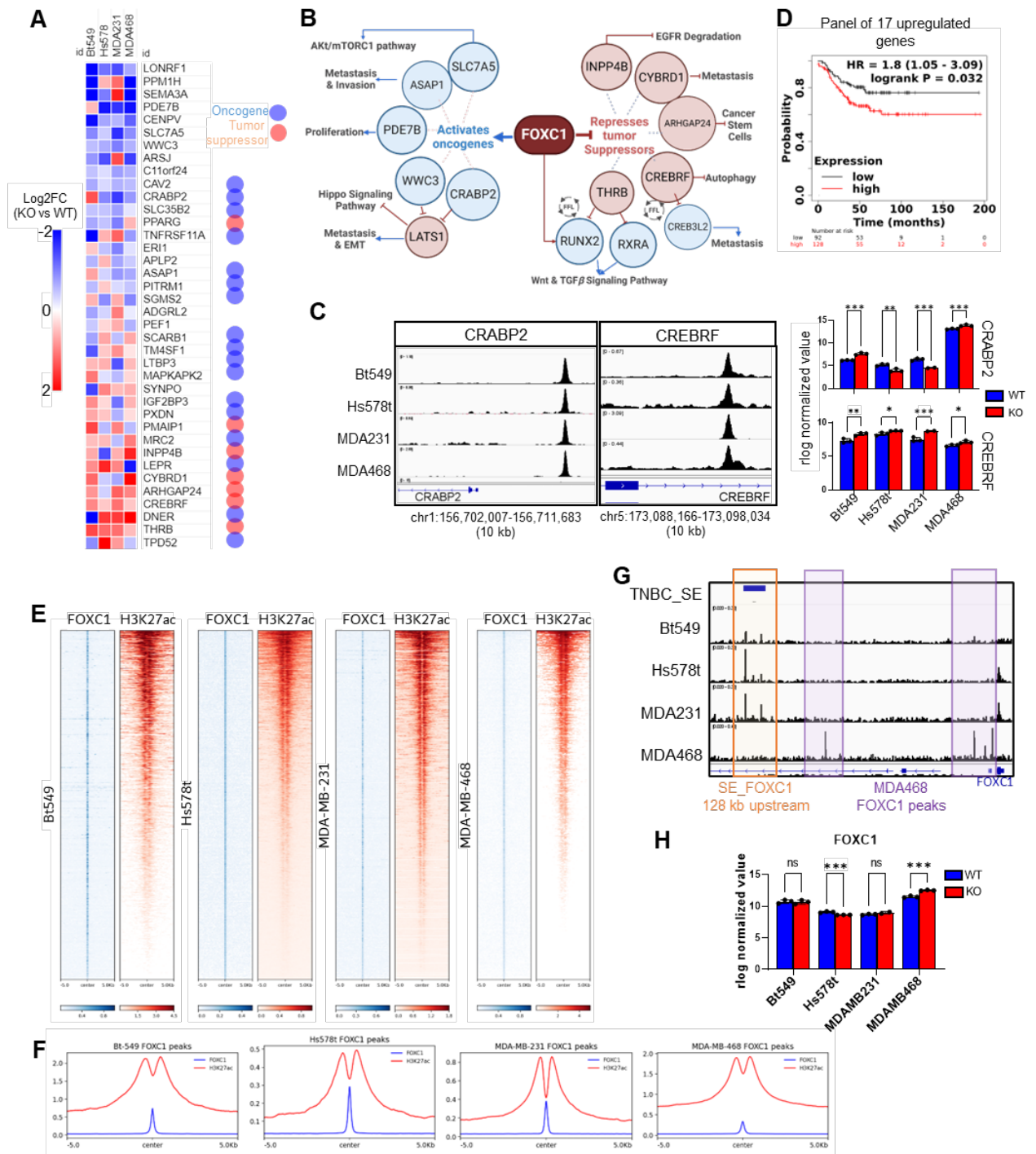
210

211 Super-enhancers (SEs) are large clusters of transcriptionally active enhancers that drive
212 oncogene expression(19) (20). In order to identify TNBC-critical SEs that are bound by
213 FOXC1, we overlapped our ChIP-seq data with the publicly available list of 331 TNBC-
214 specific SEs that are near genes upregulated in TNBC (19, 20). In total, FOXC1 peaks were
215 observed at fifteen TNBC specific SEs in all four TNBC cell lines (Fig. S3h). However, MDA-
216 MB-468 has been suggested to have an epigenetic landscape that is similar to that of non-
217 TNBC breast cancer cell lines and lacks many TNBC-specific SEs (19). Thus, excluding MDA-
218 MB-468, 19 SEs were common between Bt-549, Hs578t, and MDA-MB-231 and include
219 important SEs, such as those associated with oncogenes *IL6*, *CXCL8*, *FOXC1* and *CDK6*.
220 Although not a comprehensive analysis of all SEs associated with FOXC1 peaks, these 19
221 TNBC-specific SEs highlight FOXC1's role in regulating the respective key target genes that
222 control the TNBC phenotype.

223 A SE 128 kb upstream of the *FOXC1* gene that has been shown to activate the
224 expression of FOXC1 was absent in MDA-MB-468 (Fig. 2g), supporting a previous study
225 (19). Further, FOXC1 seems to repress its own gene expression in MDA-MB-468, as the
226 normalized FOXC1 mRNA levels increase 2-fold in the absence of FOXC1 protein (Fig. 2h). A
227 possible repressive FOXC1 binding site is visible upstream of the *FOXC1* gene in MDA-MB-
228 468 and not in the other cell lines (Fig. 2g), which could indicate the presence of negative
229 feedback of FOXC1 mRNA in MDA-MB-468. Moreover, only in MDA-MB-468 does FOXC1

230 directly bind and repress expression of receptor *ERBB2* (HER2), *EGFR* and *FOXA1*, genes
231 important for hormone/growth signalling (Fig. S6). EGFR expression, in particular, decreases
232 two-fold upon loss of FOXC1, suggesting that FOXC1 drives proliferation in MDA-MB-468
233 via the upregulation of EGFR expression. MDA-MB-468 is the only basal-like/Basal A TNBC
234 subtype in this analysis, and together with our observation that MDA-MB-468 has a distinct
235 gene expression (Fig. S2a, S2b) and genome-wide binding pattern, the data suggests that
236 FOXC1's function in MDA-MB-468 is unique and that the cell line has characteristics
237 atypical of TNBC and reminiscent of luminal subtypes. Cumulatively, the diversity of FOXC1
238 function in TNBC, likely due to the miscellaneous nature of TNBC classification, necessitates
239 identification of the core, conserved gene targets that can inform selection of appropriate
240 targeted therapies.

Figure 2



241

242 Figure 2: Genome-wide binding site analysis of FOXC1 in TNBC

243 (A) Log2(fold change) of normalized mRNA expression of 38 core, conserved gene targets that

244 associate with a FOXC1 peak and are differentially regulated upon FOXC1 KO in all four TNBC cell

245 lines. Dots represent literature-based evidence of the gene's function, either as an oncogene (blue

246 dots) or a tumor suppressor (red dots) in breast cancer. Nine genes (*THRB*, *CREBRF*

247 *SLC35B2*, *CAV2*, *C11orf24*, *WWC3*, *SLC7A5*, *CENPV* and *LONRF1*) overlap with those in Fig. 1e. **(B)**
248 Schema illustrating the tumor suppressors and oncogenes directly regulated by FOXC1 and their
249 respective role in tumorigenesis in TNBC. Created with BioRender. **(C)** ChIP-seq traces
250 showing FOXC1 peaks near oncogene *CRABP2* and tumor-suppressor *CREBRF*. Bar plots
251 depict normalized gene expression value in parental (WT, yellow) or FOXC1_KO (KO, blue)
252 cell lines. Statistical significance determined using DESeq2 (* = padj < 0.05, ** = padj <
253 0.005, ***padj = < 0.0005). **(D)** Kaplan-Meier survival analysis of a panel of 17 core
254 upregulated targets of FOXC1 of ER- (IHC), PR-(IHC) and HER2 negative (array) in 220
255 breast cancer patient samples from kmplot.com (21). **(E)** Heatmap of FOXC1 peaks
256 associated with a H3K27ac signal in each of the four TNBC cell lines. **(F)** Averaged
257 quantification of FOXC1 and H3K27ac signal intensity in heatmap in panel D. **(G)** FOXC1
258 ChIP-seq peak traces near the super-enhancer 128 kb upstream of FOXC1 gene (orange
259 box), and FOXC1 peaks unique to MDA-MB-468 (purple boxes). **(H)** Bar plots showing the
260 normalised expression of FOXC1 mRNA in parental (Blue, WT) and FOXC1_KO (Red, KO) in
261 TNBC cell lines. Statistical significance determined using DESeq2 (ns = not significant, *** =
262 padj < 0.0005).
263
264
265

266 Identifying core, conserved FOXC1 gene targets in TNBC

267 Unsupervised k-means clustering was used to group FOXC1-binding peaks, leading to the
268 classification of peaks into five clusters with Cluster 1 (binding sites present in all four cell
269 lines) followed by four cell-line dominant clusters, where FOXC1 peak intensity was
270 dominant in that respective cell line (Table S2) (Fig. 3a). The cluster analysis indicates that
271 the majority of FOXC1 binding events are largely cell-line specific since Cluster 1 was the
272 smallest cluster (203 peaks) (Fig. 3b). Nonetheless, Cluster 1 contained peaks with overall
273 the highest intensity, as seen by the quantification of the normalised FOXC1 signal in each
274 cluster (Fig. 3c), indicating that the strongest FOXC1 binding sites are conserved across all
275 four cell lines. FOXC1 peak intensity was also high within a respective cell line's dominant
276 cluster, except in the case of MDA-MB-231, where all cell lines displayed equal FOXC1
277 intensity at peaks (Fig 3c). This suggests that the MDA-MB-231 cluster is least exclusive, i.e,
278 MDA-MB-231 FOXC1 peaks are most likely to be present in the other 3 cell lines as well.
279 Cluster 5, consisting of FOXC1 peaks that were enriched in MDA-MB-468, was the largest
280 (~10,000 peaks) and most unique cluster. Furthermore, core FOXC1 peaks in Cluster 1 were
281 over-represented at promoter features when compared to the other clusters (Fig. 3d),
282 suggesting that conserved FOXC1 binding sites are more likely to be found at promoters and
283 cell-line specific FOXC1 binding sites are likely at enhancer regions. Lastly, these results
284 suggest that, even in cell lines where FOXC1 is only moderately expressed, such as MDA-
285 MB-231, the core peaks remain conserved. High-FOXC1 expressing cell lines did not possess
286 an additional cluster of unique FOXC1 peaks (such as a cluster of peaks present in only
287 Hs578t and Bt549), but rather possessed individual cell-line specific clusters.

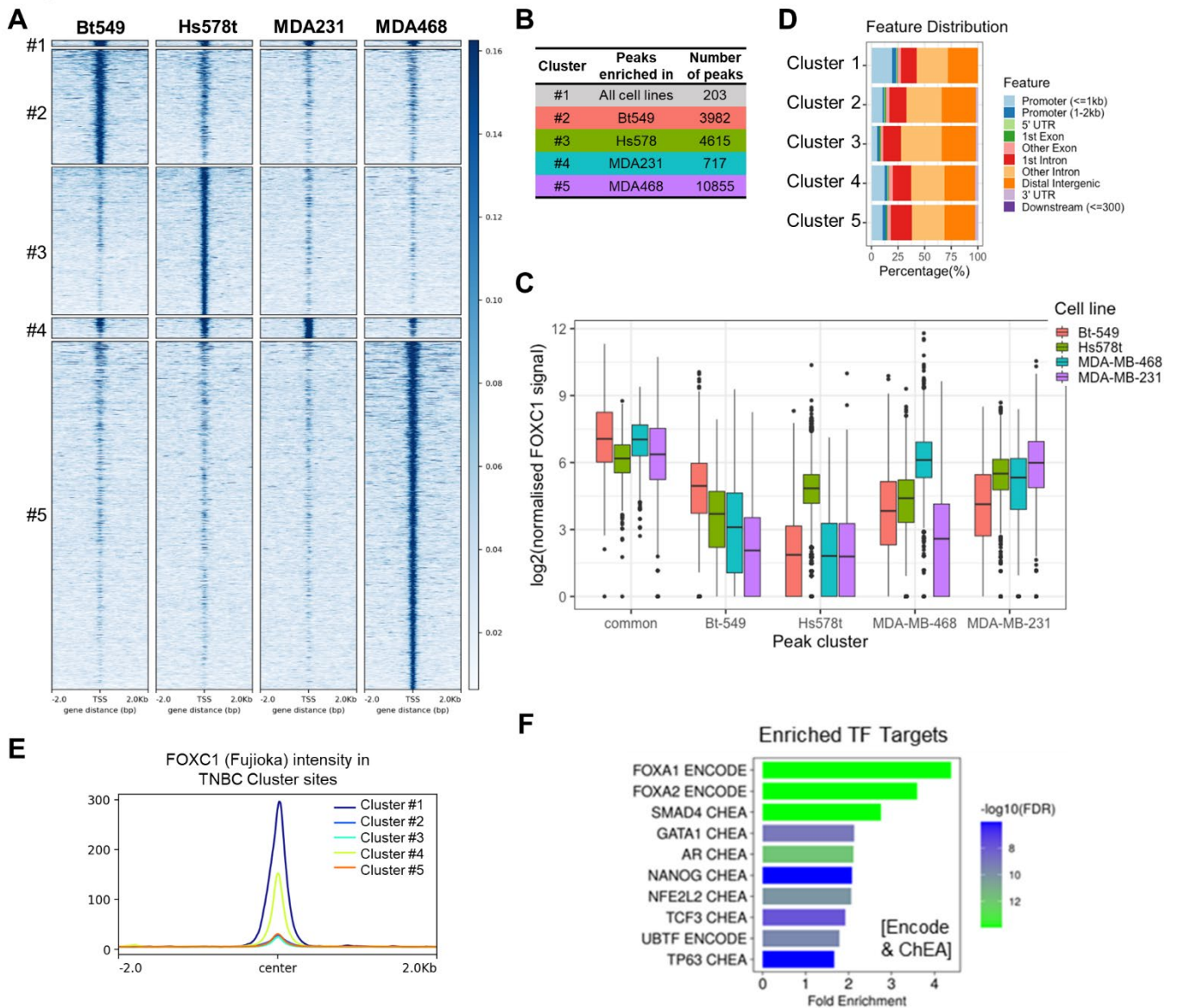
288 We compared our FOXC1 binding profile in TNBC to that published in an acute
289 myeloid leukaemia (AML) cell line model (Fujioka cell line) (18). FOXC1 peak intensity in
290 Fujioka (AML) cell lines was grouped along the five TNBC Clusters in Fig. 3a and revealed
291 highest intensity over Cluster 1 binding sites, followed by those of the MDA-MB-231
292 dominant cluster 4 (Fig. 3e, Fig. S7). In summary, the conserved, highest-intensity FOXC1
293 binding sites in TNBC are also bound by FOXC1 in AML. Genes regulated by FOXC1 binding
294 sites at Cluster 1 and Cluster 4 peaks could thus represent the core gene targets of FOXC1
295 across disease types.

296

297 Annotation of peaks with genes present within 50 kb in Cluster 1 or Cluster 4 yielded a list
298 of 1646 genes. To identify other TFs that bind upstream of the same genes, we performed
299 enrichment analysis using the TF Target database (consensus of ChIP-seq experiments in
300 ENCODE and ChEA). The highest fold enrichment, nearly 4-fold, was observed for FOXA1
301 targets (Fig. 3f), followed by FOXA2 and SMAD4. FOXA1 is a pioneer TF that establishes the
302 ER α transcriptional complex along with other cofactors, such as GATA3, NR2F2 (6, 22).
303 Shared gene targets between FOXC1 and FOXA1 suggests that FOXC1 could regulate many
304 crucial ER α targets in TNBC and motivated us to identify possible co-factors of FOXC1.

305

Figure 3



306

307 Figure 3: Core, conserved FOXC1 gene targets in TNBC

308 **(A)** Unsupervised k-means clustering of peaks among the four TNBC cell lines into 5

309 clusters: Cluster 1, peaks in all cell lines, Cluster2-5, peaks dominant in a particular cell line.

310 **(B)** Number of peaks in each cluster identified in previous panel. **(C)** Quantification of the

311 FOXC1 signal in each of the five clusters. One-way ANOVA followed by Tukey's post-hoc

312 test, comparing cell lines within each peak cluster. **(D)** Location of FOXC1 peaks near

313 genome features in each of the five clusters. **(E)** The average intensity of FOXC1 peaks in

314 Fujioka (AML) cell lines overlaid onto the five clusters of FOXC1 binding sites in TNBC from

315 panel A. **(F)** Top enriched transcription factors that have similar gene targets as core targets
316 of FOXC1, using the TF target (Encode and ChEA consensus from CHIP-X experiments),
317 generated using ShinyGO (23).
318

319 [FOXC1 interacts with NR2F2](#)

320 To identify protein cofactors of FOXC1 in TNBC, we immunoprecipitated proteins
321 crosslinked with FOXC1 and analysed interacting partners via mass spectrometry (Rapid
322 Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME), (24) in BT-549 &
323 Hs578t, as these two cell lines have higher FOXC1 protein levels. Fifty and 26 unique
324 proteins were identified to interact with FOXC1 in BT-549 and Hs578t respectively, in two
325 biological replicates, after filtering for non-specific proteins that were immunoprecipitated
326 using an IgG negative control (Fig. 4a, Table S3).

327

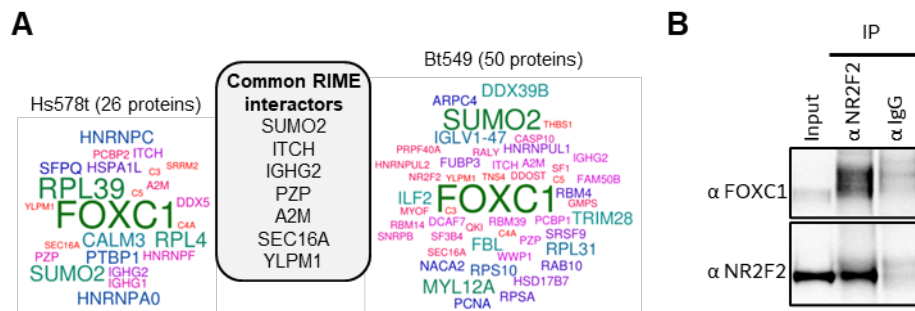
328 Ten proteins were common between the two cell lines, excluding FOXC1, which was the most
329 significantly enriched protein with 12 unique peptides, validating the specificity of the
330 antibody used and the method for isolating associated proteins. The strongest interactor
331 was SUMO2, a small ubiquitin-like modifier, which has been shown to SUMOylate FOXC1
332 and regulate its activity (25). Other high-confidence co-factors identified include TRIM28,
333 which interacts with FOXC1 at the promoter of the Wnt5A gene (26), and ITCH, an E3
334 ubiquitin ligase.

335 Interestingly, NR2F2, a nuclear receptor involved in the ER α transcriptional complex,
336 was identified as one of the interacting partners of FOXC1 in BT-549, albeit with low
337 peptide coverage. This prompted us to test whether FOXC1 and NR2F2 interact. We
338 overexpressed FOXC1 and NR2F2 in BT-549, co-immunoprecipitated proteins bound to
339 NR2F2 using anti-NR2F2, and assayed for the presence of FOXC1. FOXC1 was enriched
340 specifically in the NR2F2 immunoprecipitated reaction above that of the IgG negative

341 control (Fig. 4b), suggesting that the two proteins interact and thus could cooperatively
 342 function to regulate gene expression in TNBC.

343

Figure 4



344

345 Figure 4: Identification of FOXC1 cofactors

346 **(A)** Protein co-interactors of FOXC1 identified in Hs578t and BT-549 via RIME in two
 347 biological replicates, after subtracting those identified in negative IgG control. **(B)** Western
 348 blot depicting co-immunoprecipitation of FOXC1 using antibody against NR2F2 with input
 349 and IgG negative controls.

350

351 [FOXC1 and FOXA1/NR2F2 bind similar sites in respective breast cancer subtypes](#)
352 In luminal breast cancer, ER α is the major driver of estrogen-responsive proliferation [26].
353 Inaccessible chromatin is first bound by the pioneer factor FOXA1, which then recruits
354 chromatin modifiers to open up the chromatin. ER α is recruited onto a subset of FOXA1
355 sites which are also bound by GATA3 and/or NR2F2. Approximately 85% of ER α bound
356 sites are bound by FOXA1, GATA3, and NR2F2 as part of the ER α transcriptional complex,
357 and as such, these TFs share many of the binding sites [21]. We analysed whether FOXC1
358 binds and regulates similar targets in TNBC as those regulated by the ER α -associated TFs in
359 luminal breast cancer. ESR1 (ER α), FOXA1, and GATA3 are not expressed in TNBC, and
360 NR2F2 is variably expressed, as seen by mRNA expression data and normalized protein
361 expression (Fig. S8a, (27)). Thus, we utilized publicly available ChIP-seq data for ESR1,
362 FOXA1, GATA3, or NR2F2 generated in luminal MCF-7 cell line and clustered the respective
363 binding sites to our FOXC1 ChIP-seq data (Fig. 5a). Some overlap was observed between
364 FOXC1 peaks in TNBC and the four TFs in MCF-7, with highest overlap in Cluster 1,
365 consisting of core FOXC1 peaks that were common across all four TNBC cell lines. Peak
366 intensity of each TF in the respective five clusters was quantified (Fig. 5b). While the
367 intensity of FOXA1, NR2F2, GATA3 and ESR1 remained relatively equal in the Bt-549 and
368 Hs578t clusters 2 and 3 respectively, intensity of FOXA1 and NR2F2 was significantly
369 higher in Cluster 1 and 4 when compared to either ESR1 or GATA3 (Fig. 5b). This suggests
370 that FOXA1 and NR2F2 preferentially bind at the identified 'core peaks' of FOXC1,
371 represented by Cluster 1 and 4, in the luminal cell line they are expressed in.
372

373 Next, we did the reverse analysis and quantified FOXC1 intensity to binding sites
374 assigned a FOXA1, GATA3, or NR2F2 identity depending on whether that site was bound by
375 the TFs in MCF-7 ('cognate sites'). Indeed, intensity of FOXC1 peaks in MDA-MB-231 was
376 observed to be significantly higher in sites bound by FOXA1+NR2F2 in MCF-7 versus those
377 bound by GATA3 (Fig. 5c). Since cluster 1+ cluster 4 peaks in Fig. 5a account for high-
378 intensity FOXC1 binding sites in MDA-MB-231, this result independently shows that FOXC1
379 preferentially binds at sites that are bound by FOXA1 and NR2F2 in MCF-7. In the other cell
380 lines as well, similar patterns of FOXC1 binding preferences emerged, where FOXC1 binding
381 intensity was significantly higher at cognate sites that were bound by FOXA1, NR2F2, or a
382 combination of all TFs in luminal MCF-7 when compared to sites bound by GATA3 alone
383 (Fig. S8b).

384

385 The above analysis compared FOXC1 ChIP-seq data in TNBC cell lines to that of
386 other TFs in luminal MCF-7 cell line. To validate whether exogenous expression of a luminal
387 TF would lead to co-binding at sites bound by FOXC1, we compared our FOXC1 ChIP-seq
388 data from MDA-MB-231 with publicly available ChIP-seq data in MDA-MB-231 with
389 exogenous expression of GATA3 (28). About 25% of FOXC1 in MDA-MB-231 were bound
390 by GATA3 as well, and most GATA3 peaks within 50 kb of a FOXC1 peak were present
391 within 100 bp of the FOXC1 peak (Fig. 5d), suggesting that the two proteins can bind to
392 similar binding sites, at least in MDA-MB-231. Of the 445 DEGs in MDA-MB-231 that were
393 associated with FOXC1 peaks, a GATA3 peak was seen at 145 genes, suggesting co-
394 regulation (Fig. S8c). The average intensity of GATA3 was equally intense in cluster 1
395 FOXC1 peaks, and cluster 4, the MDA-MB-231 cluster (Fig. 5e), again emphasizing the

396 importance of the identified core FOXC1 peaks. However, our data also suggests that
397 FOXC1 represses expression of GATA3 via a binding site in the first intron, in all cell lines
398 except MDA-MB-231, where an increase in GATA3 expression was observed (Fig. S8d, S8e).
399 GATA3 has been previously shown to repress FOXC1 expression (16, 29), suggesting that
400 the two proteins are mutually exclusively expressed, and that exogenous expression as in the
401 case of MDA-MB-231 ChIP-seq data above, could result in abnormal function.

402

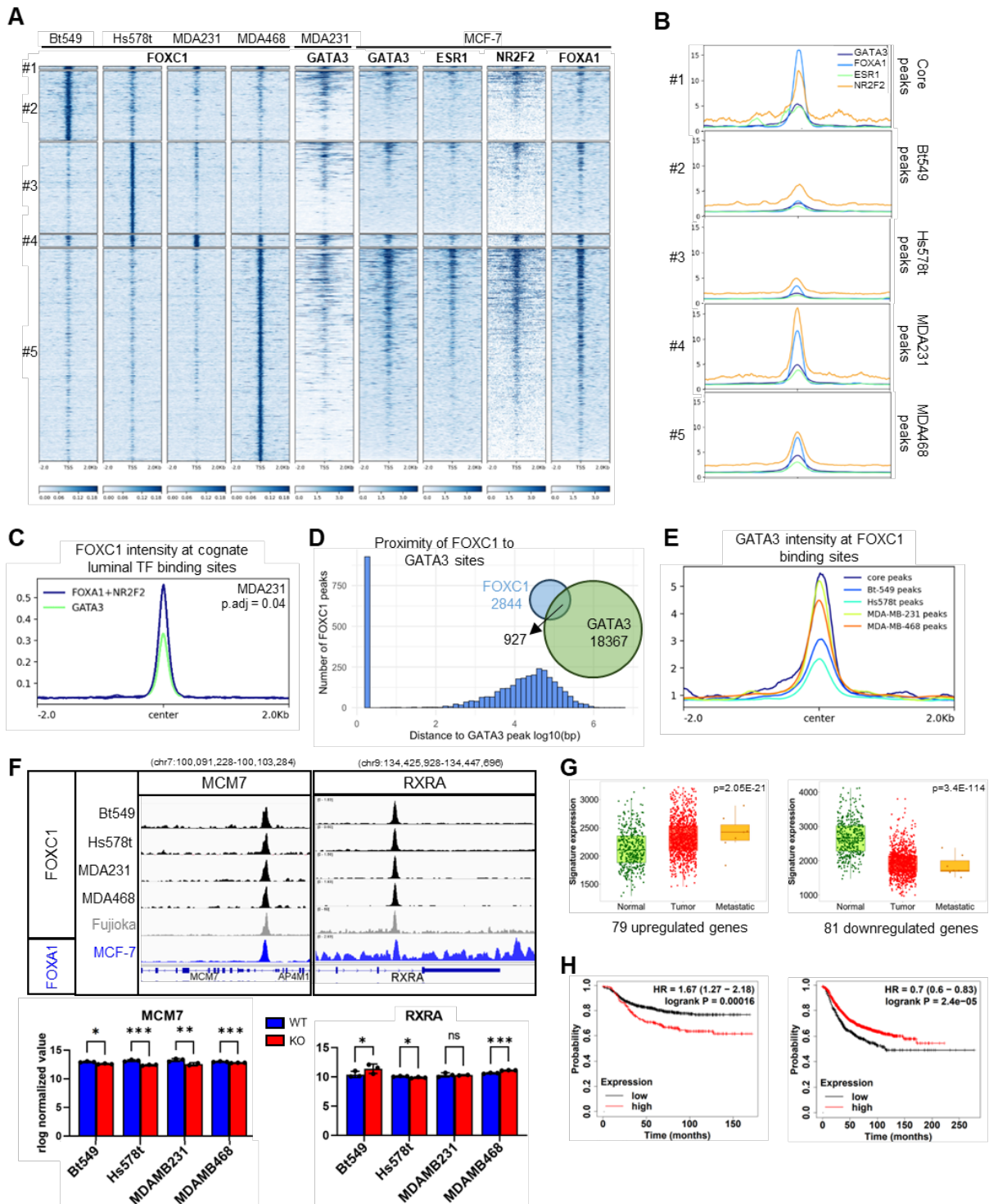
403 In summary, along with the observation that FOXC1 and NR2F2 co-interact, it is likely that
404 core binding sites of FOXC1 in TNBC identified here and represented by Cluster 1 and
405 Cluster 4 peaks, are sites where FOXA1/NR2F2 are bound in luminal breast cancer (and vice
406 versa). To identify genes functionally regulated by the identified peaks, we assembled a list
407 of genes within 50 kb of a peak in Cluster 1 and Cluster 4, that were differentially expressed
408 upon FOXC1 knockout in at least 3 cell lines. This yielded an exhaustive list of 164 genes
409 that we termed as the 'core targets of FOXC1' (Fig. S9a, Table S4). Pathway enrichment
410 using Hallmark gene sets (MSigDB) revealed enrichment of key cancer hallmarks such as
411 EMT, IL-2 STAT5 signalling and Early response to estrogen (Fig. S9b). Among the genes are
412 key prognostic TNBC-specific biomarkers, drug-targetable receptors, kinases, lncRNA and
413 crucial TFs (Fig. 5f, S9c, d) as well as novel ligands that were previously unknown to be
414 under direct regulation by FOXC1 (Fig. S10).

415

416 Clinical relevance for the above identified core targets was evaluated by visualizing
417 mean expression of either the upregulated or downregulated genes in normal, tumor or
418 metastasized patient samples (TNMplot.com (30)). As expected, the panel of 81 genes

419 repressed by FOXC1 were seen to have significantly lower expression in tumor and
420 metastatic breast invasive carcinoma samples, inclusive of all subtypes (Fig. 5g). Likewise,
421 mean expression of 79 upregulated genes was greater in tumor compared to normal
422 samples. Moreover, the above up- and down-regulated gene panels were seen to have
423 significant prognostic value in breast cancer patient samples, regardless of subtype (Fig. 5h).
424 High expression of genes activated by FOXC1 was seen to correlate with poor survival
425 whereas high expression of genes repressed by FOXC1 correlated with better survival
426 (Kaplan-Meier Plotter (21)). Together, the correlation of FOXC1 regulation with the mean
427 expression and prognostic value of the up- and down-regulated targets in all breast invasive
428 carcinomas, irrespective of subtype, indicates that the identified 164 core FOXC1-regulated
429 genes are important targets of transcriptional regulation in breast cancer.

Figure 5



430

431 Figure 5: Core FOXC1 binding sites in TNBC are bound by FOXA1/NR2F2 in luminal breast
432 cancer

433 **(A)** Heatmaps depicting overlap of ChIP-seq data of transcription factors (GATA3, ESR1,

434 NR2F2 and FOXA1) from MCF-7 luminal breast cancer cell lines with the five TNBC clusters

435 of FOXC1 peaks identified in Fig. 3A. **(B)** Average ChIP-seq tag intensity of transcription
436 factors (GATA3, ESR1, NR2F2 and FOXA1) in MCF-7 over TNBC peak clusters from
437 previous panel. **(C)** Average intensity of FOXC1 peaks in MDA-MB-231 at sites that are
438 bound by FOXA1+NR2F2 versus GATA3 (cognate sites) in MCF-7. Adjusted p-value
439 calculated using a Tukey post-hoc test. **(D)** Proximity of FOXC1 to GATA3 binding sites in
440 MDA-MB-231 and the number of peaks overlapping between FOXC1 and over-expressed
441 GATA3 in MDA-MB-231. **(E)** Average intensity of GATA3 over FOXC1 peaks in each of the
442 five clusters in Fig. 3a. **(F)** ChIP-seq traces (above) showing FOXC1 peaks in TNBC and AML
443 cell lines and FOXA1 peaks in MCF-7 near MCM7 and RXRA genes, two critical genes in
444 breast cancer and barplots (below) show changes in normalized gene expression from RNA-
445 seq of the genes in WT (Blue) or FOXC1_KO (Red). Statistical significance determined using
446 DESeq2 (ns = not significant, * = padj < 0.05, **** = padj < 0.00005). **(G)** Change in mean
447 expression of up- or down-regulated core genes in all breast invasive carcinomas based on
448 RNA-seq data from the GEO database in tumor, adjacent normal or metastatic samples
449 (TNMplot.com (30)). P-value shown calculated using Kruskal-Wallis tests. **(H)** Kaplan-Meier
450 survival analysis of upregulated (left) and downregulated (right) core gene targets of FOXC1
451 on 2032 breast cancer patient samples from kmplot.com (21).

452

453

454 DISCUSSION

455

456 In this study, we first establish that, although phenotypically FOXC1 affects the major
457 tumour-associated properties in TNBC, only 26 genes are differentially expressed upon loss
458 of FOXC1 across four TNBC cell lines with varying levels of FOXC1 expression. The lack of
459 conserved DEGs in RNA-seq suggests cell-line specific and context dependent function for
460 FOXC1 in TNBC. The heterogeneity of FOXC1 function in TNBC is reflective of the
461 heterogeneity of TNBC and emphasizes the obstacles in identifying unified targeted
462 therapies. Given that FOXC1 is over-expressed in many cancers (10) and given its role as a
463 clinically useful biomarker and potential therapeutic target, identification of core target
464 genes regulated by FOXC1 and its co-factors sheds light on its function and elucidates
465 mechanisms of tumor progression.

466 By analysing genome-wide FOXC1 binding sites in four cell lines, we identify conserved
467 gene targets of FOXC1 in TNBC and we show that these core targets are critical for TNBC
468 pathogenesis (Fig. 2a) and have clinical relevance (Fig. 2d). Within the conserved TNBC
469 regulon, we show that FOXC1 represses expression of the tumor suppressors *CREBRF*,
470 *THRB*, *ARHGAP24*, and *INPP4B* and activates expression of the oncogenes *CRABP2*,
471 *WWC3*, *SLC7A5*, *ASAP1*, *PDE7B*, and *IGF2BP3* (illustrated in schematic Fig. 2b, Fig. S4). To
472 illustrate an example, the FOXC1 target CREBRF is a tumor suppressor that regulates
473 autophagy (31). CREBRF facilitates the degradation of CREB3 family TFs, which promote
474 metastasis (32). In our data FOXC1 directly regulates both CREBRF and the downstream
475 CREB3-family TF, CREB3L2, resulting in an apparent feed-forward regulatory loop (Fig. 2b).
476 Likewise, an oncogene directly upregulated by FOXC1 is the amino acid transporter SLC7A5

477 (LAT1), which mediates uptake of leucine in exchange for glutamine, and is typically over-
478 expressed in TNBC, with higher expression correlating with worse prognosis (33). FOXC1
479 activates expression of SLC7A5 nearly 2-fold in all four cell lines via binding site within the
480 first intron (Fig. S4d). As leucine is a regulator of mTORC1, SLC7A5 has been speculated to
481 promote proliferation via the Akt/mTORC1 pathway and is a target in preclinical trials (34).
482 Lastly, many ligands and proteins regulated by FOXC1 such as DDIT4, LONRF1, CHI3L2, and
483 SYNPO (Fig. S10) pose as attractive targets for further functional studies in TNBC, that
484 although have no studied role in TNBC, their prognostic value and function in other
485 malignancies suggests an important function relevant to TNBC.

486

487 Among the co-factors of FOXC1 identified in this study are SUMO and TRIM28, which
488 have been shown to previously interact with FOXC1 (25, 26), as well as two newly identified
489 protein interactors, the E3 ubiquitin ligase ITCH and nuclear receptor NR2F2. FOXC1
490 protein levels are regulated by post-translational modifications via the ubiquitin 26S
491 proteasomal degradation pathway (35), suggesting that ITCH could be the E3 ligase
492 responsible for ubiquitinating FOXC1. Targeting E3 ubiquitin enzymes for cancer therapy
493 has long been considered an attractive mechanism-based drug discovery approach and the
494 possibility of regulating FOXC1 activity via targeting ITCH remains to be seen (36).

495 The interaction between FOXC1 and NR2F2 was confirmed using co-
496 immunoprecipitation. NR2F2 is an orphan nuclear receptor in the steroid hormone receptor
497 family that influences cell invasion, migration and EMT in breast cancer (37). Among breast
498 cancer subtypes, its expression is highest in the Luminal A subtype, although in the TNBC
499 cell lines Bt549 and MDA-MB-231 used in this study, NR2F2 expression is comparable to

500 that in MCF-7, a luminal A cell line (Fig. S8a). In MCF-7, NR2F2 has been shown to be
501 present in the ER α transcriptional complex along with FOXA1 and GATA3, where it
502 influences ER α binding on genes driving proliferation and metastasis (22). A small molecule
503 inhibitor of NR2F2 was recently shown to reduce prostate cancer tumor growth, specifically
504 by interrupting the interaction of NR2F2 with FOXA1, a master regulator in prostate
505 tumorigenesis (38). Whether NR2F2 and FOXC1 influence their respective binding in TNBC
506 remains to be determined, but the possibility of using the NR2F2-inhibitor to disrupt the
507 interaction with FOXC1 in TNBC presents a potential targeted-therapy approach for basal-
508 like breast cancer.

509

510 By analysing genome-wide binding across TNBC cell lines with varying FOXC1 protein
511 expression, we were able to identify the core set of conserved FOXC1 peaks. Intriguingly, we
512 found that high-FOXC1 expressing cell lines did not have a common set of peaks, suggesting
513 that apart from the core sites identified in this study, FOXC1 binding sites are largely cell-
514 line specific. Next, using publicly available binding site data of ER α -associated TFs, we show
515 that the identified core sites bound by FOXC1 in TNBC are bound by FOXA1 and NR2F2 in
516 ER+ breast cancer. FOXA1 is a pioneer factor that directs ER α binding in ER+ breast cancer
517 (39) and controls expression of critical genes. Knockdown of FOXA1 in ER+ breast cancer
518 reduces luminal-lineage proliferation (40, 41). That many FOXA1 targets are regulated by
519 core FOXC1 binding sites in TNBC emphasizes the importance of these targets in TNBC and
520 highlights the need for further functional studies. The 164 core targets of FOXC1 includes
521 many important ligands, receptors, kinases, lncRNA and TFs such as *MCM7*, *RXRA*, *CDK6*,
522 *RUNX2*, *NEAT1* and *ABL1* (Fig. S9). To highlight, FOXC1 was seen to regulate expression of

523 *RXRA*, that encodes retinoid X receptor alpha (Fig. 5f). *RXRA* in turn forms heterodimers
524 with $TR\beta$, encoded by *THRB*, which is strongly repressed by *FOXC1* as well (Fig. 2b, Fig. S4a).
525 *THRB* is a known tumor suppressor in TNBC, whose high expression correlates with better
526 survival [33]. Upon binding the agonist T3, $TR\beta$ form heterodimers with RXR receptors and
527 promotes tumor suppressive pathways [34]. Interestingly, RXR motifs are enriched near
528 *FOXA1* peaks, but only in an *ESR1*-mutated, endocrine resistant background (42). Moreover,
529 the same study reported stimulation of RXR: $TR\beta$ activity using T3 in the context of breast
530 cancer. $TR\beta$ inhibits the expression of *RUNX2*, an oncogene in TNBC [35], whose expression
531 is in turn upregulated by *FOXC1*, in another example of a coherent feed forward loop (Fig.
532 2b, Fig. S9d). Lastly, both *FOXC1* and *FOXA1* have multiple binding sites within the *CDK6*
533 gene, that overlaps with a TNBC-specific SE (Fig. S9d). *CDK6* is a clinically important target,
534 since *CDK4/6* inhibitors are used in treatment of ER+ breast cancer, and clinical trials are
535 currently evaluating their effectiveness in TNBC (43). Here we show that *FOXC1* drives
536 sensitivity to Palbociclib, a *CDK4/6* inhibitor in Hs578t, by regulating key genes in the
537 CyclinD1/*CDK4/6* Rb1/ pathway such as *CDK6*, *E2F7*, *MCM7* and *HMGB2* (Fig. 1f, Fig. 1g).
538 Nonetheless, many TNBC cell lines, except those of LAR subtype, are resistant to *CDK4/6*
539 inhibitors via mechanisms not clearly defined but involving *CDK2* and/or AR expression, and
540 p53, PTEN, or Rb1 status (44) (45, 46).

541

542 We have delineated the core, conserved function of *FOXC1* in TNBC, and show that these
543 core sites are parallelly regulated by *FOXA1* in ER+ breast cancer. The core gene targets of
544 *FOXC1* identified in this study form a conserved list of genes important in breast cancer, a
545 molecular signature for *FOXC1*-overexpressing cancers and could present as attractive

546 targets for subtype-independent breast cancer therapies. To our knowledge, this is the first
547 global genomic analysis of FOXC1's molecular role in TNBC, and broadens considerably the
548 known regulon of FOXC1, a prognostic marker and crucial transcription factor that is over-
549 expressed in TNBC/basal-like tumors (10).

550

551 LIMITATIONS OF THE STUDY

552 A crucial facet of TNBC is its heterogenous nature, which leads to selective growth of
553 resistant-sub populations upon drug treatment (2). Cell line models used in this study do
554 not fully encapsulate the heterogeneity of TNBC, but rather represent the heterogeneity in
555 FOXC1 expression. The small number of core conserved FOXC1 targets identified in the
556 transcriptomic analysis could be due to the inherent noise in RNA-seq data, considering the
557 minimal three replicates performed in this study that afford very narrow margins for
558 significance. Another limitation of this study is the use of single FOXC1_KO CRISPR clones
559 from each cell line for the RNA-seq analysis, which is compensated by the fact that the
560 CRISPR KO from the four cell lines were grouped and used for comparative analysis and
561 generation of core gene target lists, in effect functioning as four independent CRISPR clones.
562 CHIP-seq datasets reproducibility is impeded by cell line variation, antibodies, or downstream
563 analysis, and our study achieves robustness by using three biological replicates from four
564 different TNBC cell lines. Future work will focus on validating the core targets and the
565 biological processes these targets are involved in, as well as expanding FOXC1's role to
566 patient tumor samples that are more complex and heterogenous compared to cell lines using
567 single-cell omics. In addition, therapies that potentially target FOXC1 via modulating its
568 interaction with cofactors can be investigated, as discussed in the case of NR2F2 and ITCH

569 above. Overall, this rigorous study adds to the available public data on transcription factors
570 binding data, allowing for comprehensive and global analysis in the future, that can take into
571 account the network of transcription factors that define cell fate and control oncogenesis.

572

573

574 AUTHOR CONTRIBUTIONS

575 RR was responsible for designing the study, performing the experiments, analyzing the data,
576 preparing the figures and writing the manuscript. SI was responsible for performing the
577 experiments and writing the manuscript. LW was responsible for bioinformatic analysis of
578 the RNA-seq and CHIP-seq data, preparing the figures, and reviewing the manuscript. TAH
579 performed the Co-immunoprecipitation experiment. RLG performed phenotypic assays. FS,
580 TS, generated the FOXC1 antibody, generated the AML CHIP-seq data and reviewed the
581 manuscript. MM was responsible for conceptualization of the study and reviewing the
582 manuscript. AP, JC reviewed the manuscript and were responsible for funding. FA was
583 responsible for designing the study, analyzing the data, writing the manuscript and funding.
584 All authors read and approved the final manuscript.

585

586 ACKNOWLEDGEMENTS

587 We would like to thank Ahmed Abou Tayoun, Sathish Ramaswamy and Maha ELNaofal (Al
588 Jalila Genomics Center of Excellence, Al Jalila Children's Specialty Hospital, Dubai, United
589 Arab Emirates) for sequencing support. We would like to acknowledge the support of the
590 Mohammed Bin Rashid University of Medicine and Health Sciences and Al Jalila Foundation.

591 We would also like to thank Eva Papachristou and Clive DeSantos from the Cambridge CRUK
592 core proteomic facility for their support with the RIME experiments.

593

594 FUNDING

595 This work is supported by Research Grants from Mohammed Bin Rashid University of
596 Medicine and Health Sciences (MBRU-CM-RG2021–3) (FA), Al Jalila Foundation (RR, FA),
597 and Sheikh Hamdan Bin Rashid Al Maktoum Award for Medical Sciences (MRG/25/2020)
598 (FA). Part of this work was supported by C5759/A20971 (FS, TS) and core support from
599 the Wellcome Trust and the MRC Cambridge Stem Cell Institute. AP was funded in whole, or
600 in part, by the Wellcome Trust (203151/Z/16/Z, 203151/A/16/Z; 212253/Z/18/Z);
601 Cancer Research UK (A25636) and the UKRI Medical Research Council (MC_PC_17230).
602 For the purpose of open access, the author has applied for a CC BY public copyright license
603 to any Author Accepted Manuscript version arising from this submission.

604

605 DECLARATION OF INTERESTS

606 The authors declare no competing interests.

607

608 FIGURE LEGENDS

609 Figure 1: Effect of FOXC1 knockout on phenotype and gene expression

610 **(A)** Western blot of cell lysate (75 µg) from four parental TNBC cell lines ('WT') and the
611 corresponding FOXC1 knockout CRISPR clones ('KO'), probed using anti-FOXC1 and anti-β-actin
612 antibodies. **(B)** Proliferation assays showing growth of FOXC1_KO clones (red) when compared to
613 the parental cell lines (blue). Viable cells were counted every 24 hours. Data represent mean and

614 standard deviation (SD) of three biological replicates. Asterisk denotes significant digits in p-value
615 derived from unpaired t-tests with * < 0.05, ** < 0.005 and *** < 0.0005. **(C)** Enrichment using
616 Hallmark gene sets (MSigDB (17)) showing significantly enriched cancer hallmarks among the DEGs
617 in each of the four TNBC cell lines. **(D)** Upset plot depicting intersection between up- or down-
618 regulated genes among the four cell lines, showing a limited set of DEGs that overlap and are
619 similarly regulated (up- or down-) in all four cell lines. **(E)** Heatmap depicting log₂(fold change) of
620 the 26 genes significantly regulated upon the loss of FOXC1 in all four cell lines, representing genes
621 repressed by FOXC1 (red) and genes activated by FOXC1 (blue). **(F)** Log₂(fold change) of selected
622 genes in the RB1/CDK4/6 pathway (FOXC1_KO vs WT). **(G)** Colony-forming assays in the presence
623 of either DMSO (top row) or 0.5 μM Palbociclib (bottom row) for 14 days prior to crystal violet
624 staining. Bar graphs (bottom panel) display relative reduction in the number of colonies compared to
625 the DMSO control in the parental (WT, blue) or the FOXC1_KO (KO, red) groups. Data represents
626 mean ± SD from three biological replicates each. Asterisk denotes significant digits in p-value derived
627 from unpaired t-tests. *** < 0.005.

628

629 Figure 2: Genome-wide binding site analysis of FOXC1 in TNBC

630 **(A)** Log₂(fold change) of normalized mRNA expression of 38 core, conserved gene targets that
631 associate with a FOXC1 peak and are differentially regulated upon FOXC1 KO in all four TNBC cell
632 lines. Dots represent literature-based evidence of the gene's function, either as an oncogene (blue
633 dots) or a tumor suppressor (red dots) in breast cancer. Nine genes (*THRB*, *CREBRF*
634 *SLC35B2*, *CAV2*, *C11orf24*, *WWC3*, *SLC7A5*, *CENPV* and *LONRF1*) overlap with those in Fig. 1e. **(B)**
635 Schema illustrating the tumor suppressors and oncogenes directly regulated by FOXC1 and their
636 respective role in tumorigenesis in TNBC. Created with BioRender. **(C)** ChIP-seq traces
637 showing FOXC1 peaks near oncogene *CRABP2* and tumor-suppressor *CREBRF*. Bar plots

638 depict normalized gene expression value in parental (WT, yellow) or FOXC1_KO (KO, blue)
639 cell lines. Statistical significance determined using DESeq2 (* = padj < 0.05, ** = padj <
640 0.005, ***padj = < 0.0005). **(D)** Kaplan-Meier survival analysis of a panel of 17 core
641 upregulated targets of FOXC1 of ER- (IHC), PR-(IHC) and HER2 negative (array) in 220
642 breast cancer patient samples from kmplot.com (21). **(E)** Heatmap of FOXC1 peaks
643 associated with a H3K27ac signal in each of the four TNBC cell lines. **(F)** Averaged
644 quantification of FOXC1 and H3K27ac signal intensity in heatmap in panel D. **(G)** FOXC1
645 ChIP-seq peak traces near the super-enhancer 128 kb upstream of FOXC1 gene (orange
646 box), and FOXC1 peaks unique to MDA-MB-468 (purple boxes). **(H)** Bar plots showing the
647 normalised expression of FOXC1 mRNA in parental (Blue, WT) and FOXC1_KO (Red, KO) in
648 TNBC cell lines. Statistical significance determined using DESeq2 (ns = not significant, *** =
649 padj < 0.0005).

650

651 Figure 3: Core, conserved FOXC1 gene targets in TNBC

652 **(A)** Unsupervised k-means clustering of peaks among the four TNBC cell lines into 5
653 clusters: Cluster 1, peaks in all cell lines, Cluster2-5, peaks dominant in a particular cell line.
654 **(B)** Number of peaks in each cluster identified in previous panel. **(C)** Quantification of the
655 FOXC1 signal in each of the five clusters. One-way ANOVA followed by Tukey's post-hoc
656 test, comparing cell lines within each peak cluster. **(D)** Location of FOXC1 peaks near
657 genome features in each of the five clusters. **(E)** The average intensity of FOXC1 peaks in
658 Fujioka (AML) cell lines overlaid onto the five clusters of FOXC1 binding sites in TNBC from
659 panel A. **(F)** Top enriched transcription factors that have similar gene targets as core targets

660 of FOXC1, using the TF target (Encode and ChEA consensus from CHIP-X experiments),
661 generated using ShinyGO (23).

662

663 Figure 4: Identification of FOXC1 cofactors

664 **(A)** Protein co-interactors of FOXC1 identified in Hs578t and BT-549 via RIME in two
665 biological replicates, after subtracting those identified in negative IgG control. **(B)** Western
666 blot depicting co-immunoprecipitation of FOXC1 using antibody against NR2F2 with input
667 and IgG negative controls.

668

669 Figure 5: Core FOXC1 binding sites in TNBC are bound by FOXA1/NR2F2 in luminal breast
670 cancer

671 **(A)** Heatmaps depicting overlap of CHIP-seq data of transcription factors (GATA3, ESR1,
672 NR2F2 and FOXA1) from MCF-7 luminal breast cancer cell lines with the five TNBC clusters
673 of FOXC1 peaks identified in Fig. 3A. **(B)** Average CHIP-seq tag intensity of transcription
674 factors (GATA3, ESR1, NR2F2 and FOXA1) in MCF-7 over TNBC peak clusters from
675 previous panel. **(C)** Average intensity of FOXC1 peaks in MDA-MB-231 at sites that are
676 bound by FOXA1+NR2F2 versus GATA3 (cognate sites) in MCF-7. Adjusted p-value
677 calculated using a Tukey post-hoc test. **(D)** Proximity of FOXC1 to GATA3 binding sites in
678 MDA-MB-231 and the number of peaks overlapping between FOXC1 and over-expressed
679 GATA3 in MDA-MB-231. **(E)** Average intensity of GATA3 over FOXC1 peaks in each of the
680 five clusters in Fig. 3a. **(F)** CHIP-seq traces (above) showing FOXC1 peaks in TNBC and AML
681 cell lines and FOXA1 peaks in MCF-7 near MCM7 and RXRA genes, two critical genes in
682 breast cancer and barplots (below) show changes in normalized gene expression from RNA-
683 seq of the genes in WT (Blue) or FOXC1_KO (Red). Statistical significance determined using

684 DESeq2 (ns = not significant, * = padj < 0.05, **** = padj < 0.00005). **(G)** Change in mean
685 expression of up- or down-regulated core genes in all breast invasive carcinomas based on
686 RNA-seq data from the GEO database in tumor, adjacent normal or metastatic samples
687 (TNMplot.com (30)). P-value shown calculated using Kruskal-Wallis tests. **(H)** Kaplan-Meier
688 survival analysis of upregulated (left) and downregulated (right) core gene targets of FOXC1
689 on 2032 breast cancer patient samples from kmplot.com (21).
690

691 STAR METHODS

692 KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
FOXC1 (ChIPseq and RIME)	(18)	
Recombinant Anti-FOXC1 antibody [EPR20685] (For WB, Co-IP)	Abcam	ab227977, RRID:AB_2916124
Mouse Anti-Actin, beta Monoclonal Antibody, Unconjugated	Abcam	ab6276, RRID:AB_2223210
Goat Anti-Rabbit IgG - H&L Polyclonal antibody, Hrp Conjugated	Abcam	ab6721, RRID:AB_955447
Anti-NR2F2 (ARP-1 (B-2)) (Co-IP)	Santa Cruz Biotech	sc-271265X, RRID:AB_10608838
Recombinant Anti-NR2F2 antibody [EPR18443] (WB)	Abcam	RRID:AB_2895604
Chicken polyclonal Secondary Antibody to Mouse IgG - H&L	Abcam	ab6706, RRID:AB_956003
Chemicals and Recombinant Proteins		
DSG (disuccinimidyl glutarate)	Thermoscientific	Catalog number: 20593, 50MG
TrueCut™ Cas9 Protein v2	Invitrogen	Catalog number: A36499
Lipofectamine CRISPRmax	ThermoFisher	Catalog number: CMAX00003
Propidium Iodide	Thermoscientific	Catalog number: P3566
GelTRex LDEV	Thermoscientific	Catalog number: A1413201
Commercially available kits		
GeneArt™ Genomic Cleavage Detection Kit	Invitrogen	Catalog number: A24372
Guide-it™ Genotype Confirmation Kit	Takara	Catalog number: 632611
Pierce™ BCA Protein Assay Kit	ThermoFisher Scientific	Catalog number: 23225
RNeasy Mini Kit	Qiagen	Catalog number: 74106
RNase-Free DNase Set	Qiagen	Catalog number: 79254
Illumina stranded mRNA library prep kit	Illumina	Catalog number: 20040534

TruSeq Stranded mRNA Library Prep Kit	Illumina	Catalog number: 20020594
Qubit BR RNA kit	Invitrogen	Catalog number: Q10211
Qubit dsDNA BR kit	Invitrogen	Catalog number: Q32850
KAPA quantification kit	Roche	Catalog number: 07960140001
Cell Lines		
MDA-MB-231	AddexBio	C0006002, RRID:CVCL_0062
MDA-MB-468	AddexBio	C000600, RRID:CVCL_0419
Bt549	AddexBio	C0006017, RRID:CVCL_1092
Hs578t	AddexBio	C0006016, RRID:CVCL_0332
MDA-MB-231_KO	This manuscript	
MDA-MB-468_KO	This manuscript	
Bt549_KO	This manuscript	
Hs578t_KO	This manuscript	
Oligonucleotides		
Forward PCR Primer for FOXC1 (RR042)	5'CCATGAGCGTGTAC TCGCA	This manuscript
Reverse PCR Primer for FOXC1 (RR043)	5'GCGCATCCAGGACA TCAAGA	This manuscript
Synthetic guide RNA for FOXC1 knockout (CRISPR684819_SGM)	GAACGCCCCGGACAA GAAGA	Synthego, ThermoFisher Scientific
Synthetic guide RNA for FOXC1 knockout (CRISPR684834_SGM)	GACAAGAAGATCACC CTGAA	Synthego, ThermoFisher Scientific
Software and Algorithms		
ImageJ	https://imagej.nih.gov/ij/	RRID: SCR_003070 https://imagej.nih.gov/ij/
FlowJo 10 v10.6.1	Becton Dickinson	RRID: SCR_008520 https://www.flowjo.com
GraphPad Prism v8.3.0	GraphPad Software, LLC	RRID: SCR_002798 https://www.graphpad.com
DESeq2	Bioconductor	RRID:SCR_015687
Bowtie2		RRID:SCR_016368
Gene set enrichment analysis		RRID:SCR_003199

ClusterProfiler		RRID:SCR_016884
Metascape	https://metascape.org/	RRID:SCR_016620
TrimGalore	https://github.com/FelixKrueger/TrimGalore	RRID:SCR_011847
Deposited Data		
Raw and processed ChIP-seq and RNA-seq sequencing data generated in this study	NCBI's Gene Expression Omnibus https://www.ncbi.nlm.nih.gov/geo/	GEO Series accession number GSE213841; Use token uqhksuefvivlez to access
Publicly available datasets used in this study		
GATA3 ChIP-seq in MDA-MB-231	NCBI's Gene Expression Omnibus	GSE162003
ESR1 ChIP-seq in MCF-7	NCBI's Gene Expression Omnibus	GSE170139
FOXA1 ChIP-seq in MCF-7	NCBI's Gene Expression Omnibus	GSE105305
GATA3 ChIP-seq in MCF-7	NCBI's Gene Expression Omnibus	GSE127656
NR2F2 ChIP-seq in MCF-7	NCBI's Gene Expression Omnibus	GSM1010837
H3K27ac ChIP-seq in BT549 & MDA-MB-468	NCBI's Gene Expression Omnibus	GSE69107
H3K27ac ChIP-seq in Hs578t.	European Nucleotide Archive, https://www.ebi.ac.uk/ena/browser/home	PRJEB33558
H3K27ac ChIP-seq in MDA-MB-231	NCBI's Gene Expression Omnibus	GSE38548
Other Reagents		
DMEM/F-12	Gibco	Catalog number: 31330095
RPMI-1640	Gibco	Catalog number: 61870044
Fetal Bovine Serum (FBS)	Gibco	Catalog number: 11573397
Penicillin-Streptomycin	Gibco	Catalog number: 15140122
Trypsin EDTA (0.25%)	Gibco	Catalog number: 25200072

Opti-MEM Reduced Serum media	ThermoFisher	Catalog number: 31985062
4-12 % Bolt BisTris SDS-PAGE gel	ThermoFisher Scientific	Catalog number: NW04120BOX
LDS Sample Buffer	ThermoFisher Scientific	Catalog number: BP0007
MES SDS Running Buffer	ThermoFisher Scientific	Catalog number: NP0002
NuPAGE Reducing Agent	ThermoFisher Scientific	Catalog number: NP0009
iBright Prestained Protein Ladder	ThermoFisher Scientific	Catalog number: LC5615
iBlot™ 2 Transfer Stack, nitrocellulose	ThermoFisher Scientific	Catalog number: IB23001
Pierce Fast Blocking Buffer	ThermoFisher Scientific	Catalog number: 37575
PBS	Gibco	Catalog number: AM9625
TWEEN 20	Millipore-Aldrich	Catalog number: CAS-9005-64-5
SuperSignal™ West Pico PLUS Chemiluminescent Substrate	Thermo Fisher Scientific	Catalog number: 34580
Dynabeads Protein G	ThermoFisher	Catalog number: 10004D

693

694 RESOURCE AVAILABILITY

695 Lead contact

696 Further information and requests for resources and reagents should be directed to and will
697 be fulfilled by the Lead Contact, Dr. Fahad Ali (fahad.ali@mbru.ac.ae).

698

699 Materials availability

700 This study generated new FOXC1 KO cell lines in four parental TNBC cell lines which can be
701 requested by contacting Dr. Fahad Ali. Mohammed Bin Rashid University would require an
702 MTA to be signed for sharing of these cell lines.

703

704 Data and code availability

- 705 • All raw and processed ChIP-seq and RNA-seq sequencing data generated in the
706 current study have been submitted to the NCBI Gene Expression Omnibus (GEO);

707 <https://www.ncbi.nlm.nih.gov/geo/>) and are available under accession number
708 GSE213841. All other data needed to evaluate the conclusions in the paper are
709 present in the paper, in the above GEO project and/or the Supplementary Materials
710 and tables.

711 • Publicly available data used in this study was downloaded from GEO. The accession
712 numbers for the datasets are listed in the key resources table.

713 • This paper does not report original code.

714 • Any additional information required to reanalyze the data reported in this paper is
715 available from the lead contact upon request.

716

717 EXPERIMENTAL MODEL AND SUBJECT DETAIL

718

719 Cell lines

720 Cell lines used in this study include four female TNBC cell lines BT549 (RRID:CVCL_1092),
721 Hs578t (RRID:CVCL_), MDA-MB-231 (RRID:CVCL_0062) and MDA-MB-468
722 (RRID:CVCL_0419). Cell lines generated in this study include four FOXC1_Knockout clones
723 generated from the above parental TNBC cell lines (BT549_FOXC1_KO,
724 Hs578t_FOXC1_KO, MDA-MB-231_FOXC1_KO, MDA-MB-468_FOXC1_KO) isolated from
725 single colonies and verified to possess homozygous, frameshifting indels in the FOXC1 open
726 reading frame. Refer to the key resources table and method details for information on the
727 source and generation of each cell line and to the method details below for culture
728 conditions.

729

730 METHOD DETAIL

731

732 Cell lines and culture maintenance

733 TNBC cell lines, BT-549 (HTB-122), Hs578t (HTB-126), MDA-MB-231 (HTB-26), and
734 MDA-MB-468 (HTB-132) were purchased from AddexBio (San Diego, CA). Hs578t and
735 MDA-MB-231 were maintained in DMEM/F-12 (Gibco, ThermoFisher Scientific), whereas
736 BT-549 and MDA-MB-468 were maintained in RPMI 1640 (Gibco) in a 5% CO₂ incubator at
737 37°C. The medium was supplemented with 10% fetal bovine serum (South American origin,
738 Gibco) and 1% Penicillin-streptomycin (Gibco).

739

740 CRISPR gene editing of cell lines to create FOXC1 knockout

741 Knockout of FOXC1 was performed using CRISPR/Cas9. Two predesigned TrueGuide™
742 synthetic sgRNA (CRISPR684819_SGM: GAACGCCCGGACAAGAAGA or
743 CRISPR684834_SGM: GACAAGAAGATCACCTGAA) (ThermoFisher Scientific, Waltham,
744 USA) targeting the single exon in FOXC1 gene were complexed with Truecut Cas9 protein
745 v2 (ThermoFisher Scientific) and used to transfect cells. The efficiency of genomic cleavage
746 at the FOXC1 gene loci was determined using the GeneArt™ Genomic cleavage detection kit
747 (ThermoFisher Scientific), and wells with the highest cleavage efficiency were selected.
748 Single cells were plated into 6x 96 wells and incubated for 4-6 weeks in an Incucyte Zoom
749 system (Essen Biosciences, Ann Arbor, USA) with images taken every 24 hours. Clonal
750 populations derived from single cells were selected following limited dilution of CRISPR-
751 edited libraries. Potential clones were verified to have originated from a single cell and
752 passaged. An aliquot of cell suspension was used to verify homozygous editing at the FOXC1
753 locus using the Takara Guide-It™ Genotype Confirmation kit (Takara, Kusatsu, Japan).

754 FOXC1 genomic region was amplified using primers (RR042: 5' CCATGAGCGTGTACTCGCA
755 and RR043: 5' GCGCATCCAGGACATCAAGA) and sent for sanger sequencing (Macrogen,
756 Seoul, South Korea). Sanger sequencing data was analysed using ICE Analysis (2019. V3.0.
757 Synthego, Redwood City, USA) to identify clones that were successfully edited for
758 homozygous deletion in both FOXC1 alleles. Confirmation of FOXC1 knockout was
759 performed using western blot analysis.

760

761 [Western Blot Analysis of FOXC1 protein expression](#)

762 Western blotting was performed as described previously to confirm FOXC1 knockout in
763 CRISPR-edited clones (47). A total of 75 ug of cell lysate (quantified using Pierce BCA Protein
764 Assay Kit, ThermoFisher Scientific) was used to ensure the absence of FOXC1 protein
765 expression in the KO clones. Protein lysates were mixed with LDS sample buffer and
766 Reducing Agent (ThermoFisher Scientific) and heated at 85 °C for 10 minutes, before being
767 centrifuged for 2 minutes at 13,000 rpm. Soluble lysate fractions were loaded onto precast
768 4-12 % Bolt BisTris SDS-PAGE gel (ThermoFisher) and electrophoresed using Bolt MES
769 SDS Running Buffer (ThermoFisher) and transferred onto a nitrocellulose membrane using
770 the iBlot dry transfer system (ThermoFisher). The membrane was blocked using Pierce Fast
771 Blocking buffer (ThermoFisher) supplemented with 5% non-fat milk for 1 hour at room
772 temperature and incubated overnight at 4°C with anti-FOXC1 (1:1000; ab227977 Abcam,
773 Cambridge, UK) or anti-β-actin (1:10000; ab6276, Abcam) as an internal control, diluted in
774 blocking buffer with 2% non-fat milk. After four washes in 1x PBST, blots were incubated
775 for 1 h with HRP-conjugated secondary anti-rabbit or anti-mouse (1:10000; Abcam),

776 respectively. Blots were exposed to Pierce SuperSignal™ West Pico PLUS Chemiluminescent
777 Substrate and visualized using the Azure 600 Imager (Azure Biosystems, Dublin, CA, USA).

778

779 [Growth and proliferation assay](#)

780 An equal number of cells (5000-15,000 depending on the cell line) were seeded onto 24-well
781 tissue culture-treated plates, and three wells were trypsinized and counted every 24 hours.
782 Cell viability was analysed by mixing equal volumes of resuspended cells with 0.4% trypan
783 blue (Biorad, Hercules, CA, USA). After the addition of trypan blue, the cells were
784 immediately analyzed, and live cells were quantified using the CellDrop™ automated cell
785 counter (Denovix, Wilmington, USA).

786

787 [Colony forming assay](#)

788 Cells were seeded into 6-well tissue culture plates (150-300 cells/well depending on cell
789 line) and grown for 14 days. The cells were fixed with acidified methanol and stained with
790 0.05% crystal violet solution. Colonies were imaged in white light using the A600 Imager
791 (Azure Biosystem). Colonies were manually counted in three biological replicates, each with
792 three technical replicates.

793

794 [Invasion and Migration Assays](#)

795 Migration and invasion assays were performed in Transwell chambers containing 8-µm pore
796 membranes (6.5 mm Transwell® with 8.0 µm Pore Polyester Membrane Insert, Costar,
797 Corning Incorporated, Corning, USA). In brief, approximately 100,000-200,000 cells
798 resuspended in 200 µl serum-free media were seeded into the upper chamber, uncoated or

799 coated with a matrix, Geltrex™ LDEV-Free, hESC-Qualified, Reduced Growth Factor
800 Basement Membrane Matrix (Gibco) using 100 µl of 1:5 diluted matrix per insert and
801 allowed to set for 1 hr at 37°C prior to adding the cell suspension. Growth medium
802 containing 10% FBS was added to the lower chamber, and the plate was incubated for 20
803 hours at 37°C with 5% CO₂. The inner chamber of the Transwell was gently wiped using a
804 moist Q-tip, and cells on the bottom of the Transwell insert were fixed with formaldehyde,
805 stained with 0.05% crystal violet (dissolved in a 7:1::Methanol: Acetic acid), and imaged
806 using an inverted microscope (Olympus) at 10X magnification. Four representative images
807 were taken for each well, and the percentage of stained cells present was calculated using
808 ImageJ. Two biological replicates were performed in four technical replicates for each sample
809 along with a no FBS control.

810

811 [Propidium Iodide staining and cell cycle analysis using flow cytometry](#)

812 Cells were harvested from 10 cm tissue culture treated plates and fixed in 70% ethanol,
813 washed twice with cold 1 × PBS, and then incubated at 37°C in a staining solution containing
814 100 mM Tris-HCl, 150 mM NaCl, 0.1% Igepal, 1 mM CaCl₂, 0.5 mM MgCl₂ and 20 µg/ml
815 RNase A (ThermoFisher). Propidium Iodide (ThermoFisher) was added to a final
816 concentration of 1.5 µM to an equal number of RNase-treated cells and incubated overnight
817 at 4°C. Samples were strained using a 40 µM cell strainer prior to being detected on a
818 FACSaria III flow cytometer (Becton-Dickinson, Franklin Lakes, USA), and the percentages of
819 cells within each phase of the cell cycle were analyzed using FlowJo cell cycle module.

820

821 RNA-seq

822 RNA was isolated from the following eight samples in three biological replicates: BT-549,
823 BT-549_FOXC1_KO, Hs578t, Hs578t_FOXC1_KO, MDA-MB-231, MDA-MB-
824 231_FOXC1_KO, MDA-MB-468 and MDA-MB-468_FOXC1_KO using RNeasy mini kit
825 (Qiagen, Hilden, Germany), with on-column DNase treatment to remove genomic DNA.
826 Briefly, total RNA was isolated from the following eight samples in three biological replicates:
827 BT-549, BT-549_FOXC1_KO, Hs578t, Hs578t_FOXC1_KO, MDA-MB-231, MDA-MB-
828 231_FOXC1_KO, MDA-MB-468 and MDA-MB-468_FOXC1_KO using RNeasy mini kit
829 (Qiagen, Hilden, Germany), with on-column DNase treatment to remove genomic DNA.

830 In detail, RNA (1 ug) from BT-549, BT-549_FOXC1_KO, MDA-MB-468, and MDA-
831 MB-468_FOXC1_KO was used to prepare libraries using the Illumina stranded mRNA library
832 prep kit. RNA was quantified using Qubit BR RNA kit (Invitrogen), and the integrity of RNA
833 was determined using Tape station (Agilent). Strand-specific mRNA sequencing libraries
834 were prepared using Illumina Stranded mRNA Library Prep Kit (Illumina), following the
835 manufacturer's procedure. The fragments were amplified and quantified with KAPA
836 quantification kit (Roche), and the library size was determined using Tape station (Agilent).
837 Libraries were sequenced on the NovaSeq 6000 in a 150 bp paired-end run to a depth of at
838 least 20 million reads per library. RNA from Hs578t, Hs578t_FOXC1_KO, MDA-MB-231,
839 and MDA-MB-231_FOXC1_KO were similarly sequenced at Macrogen (Seoul, Korea) with
840 libraries generated using the TruSeq Stranded mRNA Library Prep Kit and sequenced on an
841 Illumina platform using 101 bp paired-end reads on the Illumina platform. Approximately
842 40-60 million reads were achieved for each sample. Reads were trimmed using TrimGalore
843 0.6.4 (<https://github.com/FelixKrueger/TrimGalore>) to remove sequencing adaptors and

844 poor-quality base calls, using a minimum Phred score cut-off of 20. Trimmed reads were
845 aligned to the hg38 genome with STAR_2.6.1d (Anders and Huber, 2010) and quantified
846 using the quantMode option. Genes with read counts of at least 10 in all biological replicates
847 per condition were retained for downstream processing. Read counts were normalised with
848 DESeq2 rlog transformation for PCA plots, selecting the 500 most variable genes.
849 Differentially expressed genes were identified using DESeq2 1.30.0 (Love et al., 2014) $p_{adj} <$
850 0.05 and no logfold change cut-off, with the contrasts FOXC1 KO vs FOXC1 WT for each cell
851 line. Pearson's correlation coefficients were calculated based on gene counts for each
852 replicate set, and biological replicates with a correlation < 0.95 were excluded from further
853 analysis, resulting in 3 biological replicates per group, except for MDA-MB-231_KO, which
854 was left with 2 replicates after quality control.

855

856 [Chromatin Immunoprecipitation sequencing \(ChIP-seq\)](#)

857 ChIP-seq was performed as previously described (47) using FOXC1 antibody generated
858 against a recombinant FOXC1 protein deleted for the forkhead domain ($\Delta 69-178$) (18), in
859 order to reduce non-specific binding by other forkhead-domain containing proteins. Briefly,
860 20×10^6 cells were fixed with 1% formaldehyde for 10 mins and quenched with 1.5 mM
861 Glycine. The cell pellet was harvested, lysed, and the DNA was sheared to 100-300 bp
862 fragments using the Bioruptor pico sonicator (Diagenode, Denville, NJ, USA). FOXC1
863 antibody (kindly gifted by Tim Somervaille) was used to precipitate bound DNA from 300-
864 500 μg of sheared chromatin. Cross-linking was reversed, and ChIP DNA was purified using
865 the Qiagen PCR purification kit (Qiagen, Hilden, Germany) and quantified with the Qubit HS
866 assay kit (Invitrogen). ChIP sample libraries were generated from three biological replicates,

867 and the respective input samples were pooled into one sequencing library, generated using
868 the NEBNext® Ultra™ II DNA Library Prep Kit (Illumina, San Diego, CA, USA). The amplified
869 libraries were quantified using the KAPA quantification kit (Roche, Basel, Switzerland), and
870 the library size was determined using Tape station (Agilent, Santa Clara, CA, USA). Libraries
871 were sequenced on the NovaSeq 6000 in a 150 bp paired-end run to a depth of at least 25
872 million mapped reads per sample. Reads were trimmed using TrimGalore 0.6.4 ([https://](https://github.com/FelixKrueger/TrimGalore)
873 github.com/FelixKrueger/TrimGalore) to remove sequencing adaptors and poor-quality base
874 calls, using a minimum Phred score cut-off of 20. Trimmed reads were aligned to the hg38
875 genome with Bowtie2 2.4.1 (Langmead and Salzberg, 2012). Unmapped reads, improperly
876 paired reads, reads with > 4 mismatches, or insert size >2000 bp were removed. BigWig files
877 were generated by combining bam files from individual biological replicates and subsampling
878 to 160 million reads before converting to BigWig format. FOXC1 peaks were called for
879 pooled replicates-input pairing, using Macs2 2.2.7.1 (Zhang et al., 2008) in narrowpeak
880 mode, using the following options: -f BAMPE -g hs --SPMR --qvalue 0.05 --keep-dup all.
881 Peaks with a less than 5-fold enrichment over background, peaks in centromere regions, and
882 peaks in hg38 blacklisted regions were removed, to generate a list of high-confidence FOXC1
883 peaks for each cell line. FOXC1 peaks were linked to predicted target genes within 50 kb,
884 based on significant differential expression (FOXC1 KO vs FOXC1 WT) of the potential
885 target gene from the RNA-seq data. If no genes within 50 kb were significantly differentially
886 expressed, the closest gene was assigned as the regulatory target. Significantly enriched
887 motifs were identified using Homer findMotifsGenome.pl with high-confidence peaks for
888 each cell line as the input and the options -size 100 -mask. Consensus motifs were plotted
889 with motifStack in R.

890

891 [Bioinformatic analysis](#)

892 Principal component analysis (PCA) and sample-to-sample distance heatmaps were plotted
893 using DiffBind normalized count data using default DiffBind settings for CHIP-seq and VST
894 normalized data in DESeq2 RNA-seq data. Gene set enrichment analysis (GSEA) was
895 performed with fgsea 1.160 in R using Reactome Pathway Database gene sets, or Gene
896 Ontology (GO) Biological Process (BP) gene sets. Pathways with an adjusted p-value < 0.05
897 were considered significantly enriched. Gene set overrepresentation analysis was performed
898 using ClusterProfiler 3.18.1 in R. GO BP or Reactome gene sets reaching adjusted p-value
899 (BenjaminiHochberg adjusted) < 0.05 were considered significant. Pathway enrichment from
900 various gene lists was performed using Metascape (48) or ShinyGO (23). Kaplan–Meier
901 survival curves were generated using gene expression data from public microarray datasets
902 using kmplotter (21). All tumors (Fig. 5h), or tumors classified as ER- (IHC), PR-(IHC) and
903 HER2 negative (array) (Fig. 2d and Fig. S5) were evaluated for relapse free survival benefits,
904 split using auto-selected cut-off values for gene expression and Jetset best probe set
905 options.

906

907 [Rapid Immunoprecipitation of endogenous proteins \(RIME\)](#)

908 RIME was performed as previously described (24) from two biological replicates using
909 FOXC1 antibody (18) and matched IgG negative controls (ab6706, Abcam). Briefly, 10×10^6
910 cells were double crosslinked with DSG and formaldehyde, and protein interactors were
911 immunoprecipitated using either anti-FOXC1 or anti-IgG antibody conjugated to Pierce
912 Protein A magnetic beads (ThermoFisher Scientific). On-bead tryptic digest and mass spec

913 identification of peptides was performed by the proteomics core facility at Cancer Research
914 UK Cambridge Institute. Proteins were considered as interactors when high-confident
915 peptides were identified in both replicates and when none of these peptides were identified
916 in matched IgG-negative controls.

917

918 [Cloning and over-expression of FOXC1 and NR2F2 in Bt-549](#)

919 The coding sequence of FOXC1 was amplified from FLAG-FOXC1 (a gift from Stefan Koch,
920 Addgene plasmid # 153114) using primers RR022 (5'
921 CCCTCGTAAAGAATTCGCCACCATGGATTACAAGGACGACG 3') and RR023 (5'
922 GAGGTGGTCTGGATCCTCAAACTTGCTACAGTCGTAGA3') and ligated into pLVX-
923 TetOne-Puro (Takara) using the In-Fusion HD Cloning Kit (Takara). Plasmid was verified by
924 sequencing. Plasmid pLenti-CMVSP6-NR2F2-SV40-PURO was a gift from Nathan Lawson
925 (Addgene plasmid # 138361). Plasmids were purified using EndoFree Plasmid Maxi Kit
926 (Qiagen, Germany). Transfection was performed using 14 µg of each plasmid and
927 Lipofectamine™ 3000 (Invitrogen), as per manufacturers protocol, in 10 cm petri dish with
928 70% confluent Bt549. FOXC1 expression was induced using 1 µg/ml of Doxycyclin for 48
929 hours prior to harvesting the cell pellet.

930

931 [Co-Immunoprecipitation of FOXC1 and NR2F2](#)

932 Bt-549 cells transfected with plasmids overexpressing FOXC1 and NR2F2 in 10 cm petri
933 dish was harvested and the cell pellet was resuspended in 100 µl of RIPA lysis buffer
934 (ThermoFisher Scientific) containing 100 mM PMSF and 1X Protease Inhibitor cocktail
935 (ThermoFisher Scientific). Cell lysate was sonicated for 10 cycles (Bioruptor Pico,

936 Diagenode), centrifuged and the supernatant was incubated overnight with Protein-G beads
937 coupled with 5 µg of anti-NR2F2 (sc271265x, Santacruz Biotech, Dallas, Texas) or anti-IgG
938 (Ab6706, Abcam). Unbound protein was washed using RIPA Buffer and protein complexes
939 eluted using 4X Bolt LDS Sample buffer supplemented with 1X Reducing agent
940 (ThermoFisher Scientific). Western blotting was performed as described above.

941

942 QUANTIFICATION AND STATISTICAL ANALYSIS

943 Barplots in this study represent mean ± standard deviation from three independent
944 biological replicates, unless otherwise stated. P-value was computed using multiple unpaired
945 t tests in Prism GraphPad software and adjusted p-values representing differential
946 regulation in RNA-seq data were calculated using DESeq2. P-value <0.05 was considered
947 significant. Details of statistical analysis can be found in the figure legends.

948

949 REFERENCES

950

- 951 1. 1. Won KA, Spruck C. Triple-negative breast cancer therapy: Current and future
952 perspectives (Review). *Int J Oncol.* 2020;57(6):1245-61.
- 953 2. 2. Marra A, Trapani D, Viale G, Criscitiello C, Curigliano G. Practical classification of
954 triple-negative breast cancer: intratumoral heterogeneity, mechanisms of drug
955 resistance, and novel therapies. *NPJ Breast Cancer.* 2020;6:54.
- 956 3. 3. Chavez KJ, Garimella SV, Lipkowitz S. Triple negative breast cancer cell lines: one tool
957 in the search for better treatment of triple negative breast cancer. *Breast Dis.*
958 2010;32(1-2):35-48.
- 959 4. 4. Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping
960 and treatment progress. *Breast Cancer Res.* 2020;22(1):61.
- 961 5. 5. Gupta GK, Collier AL, Lee D, Hoefler RA, Zheleva V, Siewertsz van Reesema LL, et al.
962 Perspectives on Triple-Negative Breast Cancer: Current Treatment Strategies, Unmet
963 Needs, and Potential Targets for Future Therapies. *Cancers (Basel).* 2020;12(9).
- 964 6. 6. Carroll JS. Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer.
965 *Eur J Endocrinol.* 2016;175(1):R41-9.
- 966 7. 7. Han B, Bhowmick N, Qu Y, Chung S, Giuliano AE, Cui X. FOXC1: an emerging marker
967 and therapeutic target for cancer. *Oncogene.* 2017;36(28):3957-63.
- 968 8. 8. Jensen TW, Ray T, Wang J, Li X, Naritoku WY, Han B, et al. Diagnosis of Basal-Like
969 Breast Cancer Using a FOXC1-Based Assay. *J Natl Cancer Inst.* 2015;107(8).
- 970 9. 9. Ray PS, Wang J, Qu Y, Sim MS, Shamonki J, Bagaria SP, et al. FOXC1 is a potential
971 prognostic biomarker with functional significance in basal-like breast cancer. *Cancer Res.*
972 2010;70(10):3870-6.
- 973 10. 10. Ray T, Ryusaki T, Ray PS. Therapeutically Targeting Cancers That Overexpress
974 FOXC1: A Transcriptional Driver of Cell Plasticity, Partial EMT, and Cancer Metastasis.
975 *Front Oncol.* 2021;11:721959.
- 976 11. 11. Wang J, Ray PS, Sim MS, Zhou XZ, Lu KP, Lee AV, et al. FOXC1 regulates the
977 functions of human basal-like breast cancer cells by activating NF-kappaB signaling.
978 *Oncogene.* 2012;31(45):4798-802.
- 979 12. 12. Han B, Qu Y, Jin Y, Yu Y, Deng N, Wawrowsky K, et al. FOXC1 Activates
980 Smoothed-Independent Hedgehog Signaling in Basal-like Breast Cancer. *Cell Rep.*
981 2015;13(5):1046-58.

- 982 13. Du J, Li L, Ou Z, Kong C, Zhang Y, Dong Z, et al. FOXC1, a target of polycomb, inhibits
983 metastasis of breast cancer cells. *Breast Cancer Res Treat.* 2012;131(1):65-73.
- 984 14. Gilding LN, Somerville TCP. The Diverse Consequences of FOXC1 Deregulation in
985 Cancer. *Cancers (Basel).* 2019;11(2).
- 986 15. Ray PS, Bagaria SP, Wang J, Shamonki JM, Ye X, Sim MS, et al. Basal-like breast
987 cancer defined by FOXC1 expression offers superior prognostic value: a retrospective
988 immunohistochemical study. *Ann Surg Oncol.* 2011;18(13):3839-47.
- 989 16. Yu-Rice Y, Jin Y, Han B, Qu Y, Johnson J, Watanabe T, et al. FOXC1 is involved in
990 ERalpha silencing by counteracting GATA3 binding and is implicated in endocrine
991 resistance. *Oncogene.* 2016;35(41):5400-11.
- 992 17. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The
993 Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*
994 2015;1(6):417-25.
- 995 18. Simeoni F, Romero-Camarero I, Camera F, Amaral FMR, Sinclair OJ, Papachristou EK,
996 et al. Enhancer recruitment of transcription repressors RUNX1 and TLE3 by mis-
997 expressed FOXC1 blocks differentiation in acute myeloid leukemia. *Cell Rep.*
998 2021;36(12):109725.
- 999 19. Huang H, Hu J, Maryam A, Huang Q, Zhang Y, Ramakrishnan S, et al. Defining super-
1000 enhancer landscape in triple-negative breast cancer by multiomic profiling. *Nat Commun.*
1001 2021;12(1):2242.
- 1002 20. Raisner R, Bainer R, Haverty PM, Benedetti KL, Gascoigne KE. Super-enhancer
1003 acquisition drives oncogene expression in triple negative breast cancer. *PLoS One.*
1004 2020;15(6):e0235343.
- 1005 21. Lanczky A, Gyorffy B. Web-Based Survival Analysis Tool Tailored for Medical
1006 Research (KMplot): Development and Implementation. *J Med Internet Res.*
1007 2021;23(7):e27633.
- 1008 22. Jiang G, Wang X, Sheng D, Zhou L, Liu Y, Xu C, et al. Cooperativity of co-factor
1009 NR2F2 with Pioneer Factors GATA3, FOXA1 in promoting ERalpha function.
1010 *Theranostics.* 2019;9(22):6501-16.
- 1011 23. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and
1012 plants. *Bioinformatics.* 2020;36(8):2628-9.
- 1013 24. Mohammed H, Taylor C, Brown GD, Papachristou EK, Carroll JS, D'Santos CS. Rapid
1014 immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of
1015 chromatin complexes. *Nat Protoc.* 2016;11(2):316-26.

- 1016 25. 25. Danciu TE, Chupreta S, Cruz O, Fox JE, Whitman M, Iniguez-Lluhi JA. Small ubiquitin-
1017 like modifier (SUMO) modification mediates function of the inhibitory domains of
1018 developmental regulators FOXC1 and FOXC2. *J Biol Chem.* 2012;287(22):18318-29.
- 1019 26. 26. Han B, Zhou B, Qu Y, Gao B, Xu Y, Chung S, et al. FOXC1-induced non-canonical
1020 WNT5A-MMP7 signaling regulates invasiveness in triple-negative breast cancer.
1021 *Oncogene.* 2018;37(10):1399-408.
- 1022 27. 27. Lawrence RT, Perez EM, Hernandez D, Miller CP, Haas KM, Irie HY, et al. The
1023 Proteomic Landscape of Triple-Negative Breast Cancer. *Cell Rep.* 2015;11(6):990.
- 1024 28. 28. Tokheim C, Wang X, Timms RT, Zhang B, Mena EL, Wang B, et al. Systematic
1025 characterization of mutations altering protein degradation in human cancers. *Mol Cell.*
1026 2021;81(6):1292-308 e11.
- 1027 29. 29. Tkocz D, Crawford NT, Buckley NE, Berry FB, Kennedy RD, Gorski JJ, et al. BRCA1
1028 and GATA3 corepress FOXC1 to inhibit the pathogenesis of basal-like breast cancers.
1029 *Oncogene.* 2012;31(32):3667-78.
- 1030 30. 30. Bartha A, Gyorffy B. TNMplot.com: A Web Tool for the Comparison of Gene
1031 Expression in Normal, Tumor and Metastatic Tissues. *Int J Mol Sci.* 2021;22(5).
- 1032 31. 31. Mascia F, Mazo I, Alterovitz WL, Karagiannis K, Wu WW, Shen RF, et al. In search of
1033 autophagy biomarkers in breast cancer: Receptor status and drug agnostic
1034 transcriptional changes during autophagy flux in cell lines. *PLoS One.*
1035 2022;17(1):e0262134.
- 1036 32. 32. Sampieri L, Di Giusto P, Alvarez C. CREB3 Transcription Factors: ER-Golgi Stress
1037 Transducers as Hubs for Cellular Homeostasis. *Front Cell Dev Biol.* 2019;7:123.
- 1038 33. 33. El Ansari R, Craze ML, Miligy I, Diez-Rodriguez M, Nolan CC, Ellis IO, et al. The amino
1039 acid transporter SLC7A5 confers a poor prognosis in the highly proliferative breast
1040 cancer subtypes and is a key therapeutic target in luminal B tumours. *Breast Cancer Res.*
1041 2018;20(1):21.
- 1042 34. 34. Hafliger P, Charles RP. The L-Type Amino Acid Transporter LAT1-An Emerging
1043 Target in Cancer. *Int J Mol Sci.* 2019;20(10).
- 1044 35. 35. Berry FB, Mirzayans F, Walter MA. Regulation of FOXC1 stability and transcriptional
1045 activity by an epidermal growth factor-activated mitogen-activated protein kinase
1046 signaling cascade. *J Biol Chem.* 2006;281(15):10098-104.
- 1047 36. 36. Yin Q, Wyatt CJ, Han T, Smalley KSM, Wan L. ITCH as a potential therapeutic target
1048 in human cancers. *Semin Cancer Biol.* 2020;67(Pt 2):117-30.

- 1049 37. Xia B, Hou L, Kang H, Chang W, Liu Y, Zhang Y, et al. NR2F2 plays a major role in
1050 insulin-induced epithelial-mesenchymal transition in breast cancer cells. *BMC Cancer*.
1051 2020;20(1):626.
- 1052 38. Teng M, Zhou S, Cai C, Lupien M, He HH. Pioneer of prostate cancer: past, present
1053 and the future of FOXA1. *Protein Cell*. 2021;12(1):29-38.
- 1054 39. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key
1055 determinant of estrogen receptor function and endocrine response. *Nat Genet*.
1056 2011;43(1):27-33.
- 1057 40. Yang L, Kumegawa K, Saeki S, Nakadai T, Maruyama R. Identification of lineage-
1058 specific epigenetic regulators FOXA1 and GRHL2 through chromatin accessibility
1059 profiling in breast cancer cell lines. *Cancer Gene Ther*. 2024.
- 1060 41. Yamaguchi N, Ito E, Azuma S, Honma R, Yanagisawa Y, Nishikawa A, et al. FoxA1 as a
1061 lineage-specific oncogene in luminal type breast cancer. *Biochem Biophys Res Commun*.
1062 2008;365(4):711-7.
- 1063 42. Wu Y, Li Z, Wedn AM, Casey AN, Brown D, Rao SV, et al. FOXA1 Reprogramming
1064 Dictates Retinoid X Receptor Response in ESR1-Mutant Breast Cancer. *Mol Cancer Res*.
1065 2023;21(6):591-604.
- 1066 43. Alvarez-Fernandez M, Malumbres M. Mechanisms of Sensitivity and Resistance to
1067 CDK4/6 Inhibition. *Cancer Cell*. 2020;37(4):514-29.
- 1068 44. Asghar US, Barr AR, Cutts R, Beaney M, Babina I, Sampath D, et al. Single-Cell
1069 Dynamics Determines Response to CDK4/6 Inhibition in Triple-Negative Breast Cancer.
1070 *Clin Cancer Res*. 2017;23(18):5561-72.
- 1071 45. Hu Y, Gao J, Wang M, Li M. Potential Prospect of CDK4/6 Inhibitors in Triple-
1072 Negative Breast Cancer. *Cancer Manag Res*. 2021;13:5223-37.
- 1073 46. Robinson TJ, Liu JC, Vizeacoumar F, Sun T, Maclean N, Egan SE, et al. RB1 status in
1074 triple negative breast cancer cells dictates response to radiation treatment and selective
1075 therapeutic drugs. *PLoS One*. 2013;8(11):e78641.
- 1076 47. Gomez RL, Woods LM, Ramachandran R, Abou Tayoun AN, Philpott A, Ali FR. Super-
1077 enhancer associated core regulatory circuits mediate susceptibility to retinoic acid in
1078 neuroblastoma cells. *Front Cell Dev Biol*. 2022;10:943924.
- 1079 48. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al.
1080 Metascape provides a biologist-oriented resource for the analysis of systems-level
1081 datasets. *Nat Commun*. 2019;10(1):1523.

