SUPPLEMENTAL MATERIAL

Estrogen Receptor 1 chromatin profiling in human breast tumors reveals high inter-patient heterogeneity with enrichment of risk SNPs and enhancer activity at most-conserved regions

Stacey E.P. Joosten[1,2,#], Sebastian Gregoricchio[1,2,#], Suzan Stelloo[2,3], Elif Yapıcı[4,5], Chia-Chi Flora Huang[6], Kerim Yavuz[6], Maria Donaldson Collier[1,2], Tunç Morova[6], Umut Berkay Altintaş[6], Yongsoo Kim[7], Sander Canisius[8,9], Cathy B. Moelans[10], Paul J. van Diest[10], Gozde Korkmaz[4,5], Nathan A. Lack[4,5,6], Michiel Vermeulen[2,3,11], Sabine C. Linn[8,9,13] and Wilbert Zwart[1,2,14,*]

[1] Division of Oncogenomics, the Netherlands Cancer Institute, Amsterdam, The Netherlands

[2] Oncode Institute, The Netherlands

[3] Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands.

[4] Koç University School of Medicine, Istanbul, Türkiye

[5] Koç University Research Center for Translational Medicine (KUTTAM), Istanbul, Türkiye

[6] Vancouver Prostate Centre, University of British Columbia, Vancouver, Canada

[7] Department of Pathology, Amsterdam University Medical Center, Cancer Center Amsterdam, Amsterdam, the Netherlands.

[8] Division of Molecular Pathology, the Netherlands Cancer Institute, Amsterdam, The Netherlands

[9] Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[10] Department of Pathology, University Medical Centre, Utrecht, The Netherlands

[11] Division of Molecular Genetics, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[12] Department of Medical Oncology, Antoni van Leeuwenhoek hospital, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[13] Department of Pathology, Utrecht University Medical Centre, Utrecht, The Netherlands

[14] Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

[#] Authors equally contributed to this work

* To whom correspondence should be addressed. Email: w.zwart@nki.nl

SUPPLEMENTAL MATERIAL AND METHODS

ChIP-seq library preparation

Five newly generated ESR1 samples were generated as previously described (Zwart et al. 2013; Singh et al. 2019). Sample details are described in Supplemental Table S1. Fresh frozen tumor material was cryosectioned, collected in Eppendorf tubes and stored at -80°C until processing. An H&E slide was assessed by pathologist to confirm tumor cell content. For ChIP, tissue was defrosted on ice and cross-linked in solution A (50 mM Hepes, 100 mM NaCl, 1mM EDTA, 0.5 mM EGTA, pH = 7.4) containing 2 mM DSG (Sigma-Aldrich) and incubated at room temperature for 25 min while rotating. Next, formaldehyde was added to 1% final concentration and rotation was continued for 20 min. Reaction was quenched by 0.2 M glycine. Samples were pelleted, washed with cold PBS and tissue architecture was disrupted by a pellet pestle (Sigma-Aldrich) and subsequently sonicated (Diagenode PicoBioruptor). IP was performed by overnight incubation with 5 µg ESR1 antibody (SC-543, Santa Cruz) to 50 µL dynabeads (Invitrogen) in blocking buffer (5% BSA in PBS), then washed ten times with RIPA (50 mM HEPES, 500 mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-DOC, pH = 7.6), washed in TBS, reverse cross-linked at 65°C in Elution Buffer (50 mM Tris, 10 mM EDTA, 1% SDS) and DNA eluted. ChIP DNA was then prepared for Illumina multiplex-sequencing with 10 samples per lane at 65 bp single end and sequenced on Illumina HiSeq 2500 (Illumina).

Publicly available data access. ESR1 ChIP-seq on 30 male breast cancer samples are available at Gene Expression Omnibus (GEO) (Barrett et al. 2013) under accession number GSE104399 (Severson et al. 2018). The female cohort raw data can be found at GEO under accession numbers GSE104399 (Severson et al. 2018), GSE32222 (Ross-Innes et al. 2012) and GSE40867 (Jansen et al. 2013).

ESR1 STARR-seq library preparation

MCF-7 cells (>2 × 10^8 cells/replica for 3 biological replica) were grown for 48 hrs in DMEM (low glucose, pyruvate, no glutamine, and no phenol red) (Gibco) supplemented with 5% Charcoal Striped Serum (Biowest) and 1% penicillin-streptomycin (Gibco) before being transfected with the cloned ESR1-focused STARR-seq capture library using polyethylenimine (Polysciences). Following 24 hrs incubation, cells were treated with estradiol (E2) (10nM, MedChemExpress) for 6 hrs. Total RNA was extracted using TRIzol reagent (Invitrogen). The poly(A) mRNA was isolated using the Oligo (dT)25 Dynabeads (Thermo Fisher Scientific), digested with Turbo DNase I (Thermo Fisher Scientific), and reverse transcribed into cDNA with the gene-specific primer (5′-CTCATCAATGTATCTTATCATGTCTG-3′). After the treatment with RNase A (Thermo Fisher Scientific), the synthesized ESR1 STARR-seq cDNA was amplified by a junction PCR (17 cycles) with the RNA_jPCR_Fw primer (5′-TCGTGAGGCACTGGGCAG*G*T*G*T*C-3′) and the jPCR_Rv primer (5′-CTTATCATGTCTGCTCGA*A*G*C-3′). The ESR1-focused STARR-seq capture library plasmid DNA was PCR-amplified (12 cycles) with the DNA-specific junction PCR primers (DNA_jPCR_Fw, 5′-CCTTTCTCTCCACAGGT*G*T*C-3′) and jPCR_Rv primers. After purification with AmpureXP beads (Beckman Coulter), all final Illumina compatible ESR1 STARR-seq and ESR1-focused STARR-seq capture libraries were prepared by PCR amplification (7 cycles) with NEBNext universal and single indexing primers (NEB), and were sequenced on Illumina NovaSeq 6000 (150bp Paired-End).

Overlaps with STARR-seq regions and MCF-7 ESR1 peaks (Ross-Innes et al. 2012) were performed using *Rseb*'s (v0.3.3) (Gregoricchio et al. 2022) function *intersect.regions*.

STARR-seq differential analyses

Raw FASTQ reads were mapped onto hg19/GRCh37 genome build using *BWA* (v0.7.17) aligner (Li and Durbin 2009). Aligned fragments were filtered for mapping quality greater than 30 (MAPQ>30) using SAMtools (v1.14). BAM files converted into BEDPE files and fragment counts at target regions have been computed using BEDTools (v2.30).

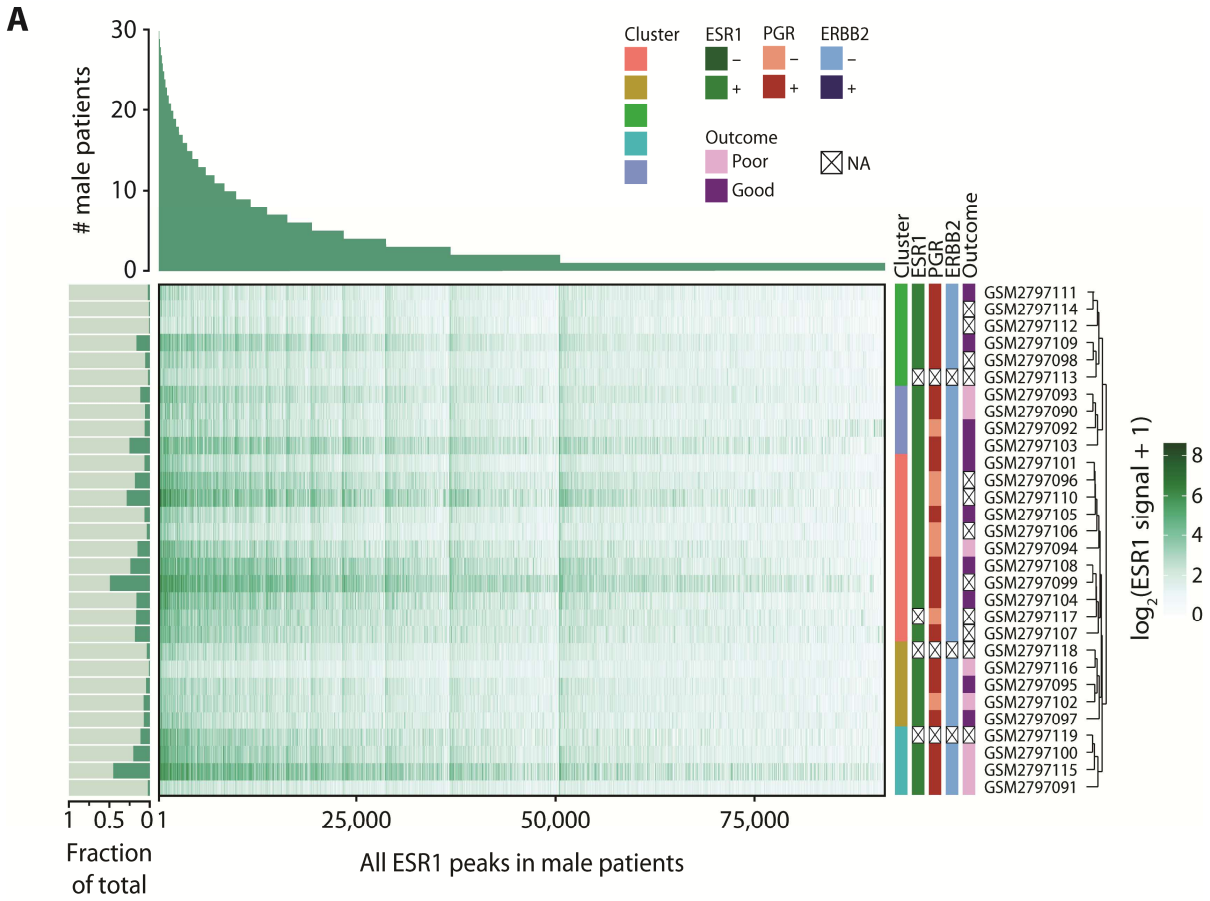
Differential analyses of STARR-seq data were performed using *DESeq2* (v1.30.1) (Love et al. 2014). If fold change over DNA in both E2 and DMSO was lower than 1, the signal was considered background (*inactive*). Sites were considered *non-induced* if the linear fold change DMSO/E2 was lower than 2. If the fold change DMSO/E2 exceeded 2 and the accompanying adjusted *p*-value was lower than 0.00001, sites were considered as *E2-induced*.

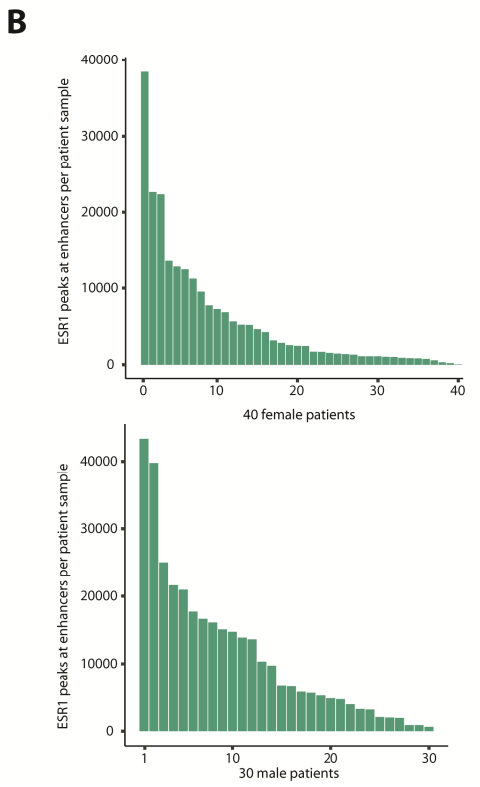
Raw counts, *DESeq2*-normalized counts and differential analyses results are available in Supplemental Table S6.

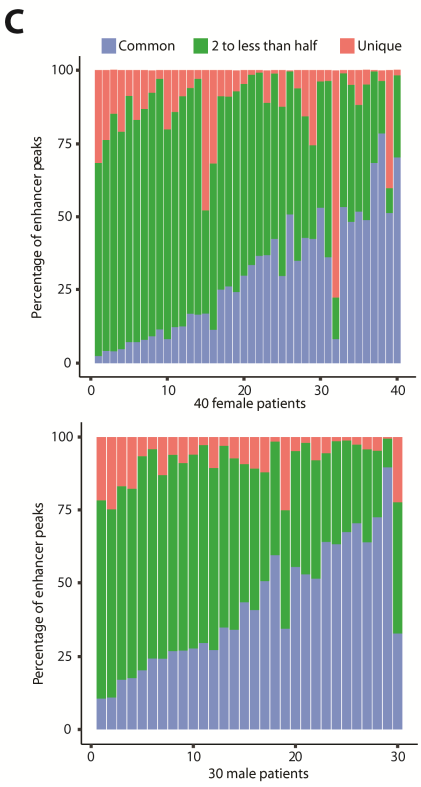
Luciferase reporter assay

For luciferase assays the regions of interest (WT) were PCR amplified from pooled male human genomic DNA (Promega). The amplified regions were cloned by Gibson assembly into a STARR luciferase vector ORI empty plasmid (Addgene #99298) (Muerdter et al. 2018) using the NEBuilder HiFi DNA Assembly master mix (NEB). Variants were either introduced by site-directed mutagenesis PCR or found endogenously in the genomic DNA pool (EREluc_WT_Fw: 5′-AActcgagACTGACAgatCTTTCCTCCCCTCGTTCACA-3′; EREluc_WT_Rv: 5′-GTctcgagAGTAGTCgatCCAGGAGAACAGGGAAGGGA-3′; FRKluc_SNP_Fw 3′-GAGGTGTTATgTGTTTATGTATTTTTCAG-5′; FRKluc_SNP_Rv 5′-TGCAGAGGCAGTGGTTTTTAA-3′). All inserts were verified by Sanger sequencing. MCF-7 cells were seeded in a 96-well plate (5×10^4/well) in low glucose, phenol red-free DMEM (Gibco), supplemented with 10% charcoal dextran-stripped FBS (Gibco, US origin), 1% penicillin-streptomycin (Gibco), and 2mM L-Glutamine (Gibco). After seeding (24 hrs), cells were transfected using 100 ng of reporter DNA per well in 10 µL of Opti-MEM (Gibco) along with 5 ng of pRL-CMV *Renilla* reporter plasmid as an internal control, using Mirus TransIT-2020 transfection reagent (Mirus Bio) at a 1:3 DNA:transfection reagent ratio according to the manufacturer's protocols. STARR luciferase validation vector bearing CMV enhancer (Addgene #99312) (Muerdter et al. 2018) was used to monitor transfection efficiency. 48 hrs post-transfection, cells were treated with 10 nM of E2 or 100% ethanol (vehicle control) for 24 h prior to harvest using 50 µL of 1× passive lysis buffer (Promega) per well. 20 µL lysate was used for each assay with the Dual-Luciferase® Reporter Assay kit (Promega) according to the manufacturer's instructions using the M200Pro TECAN Luminometer in technical duplicate with three biological replicates. For data analysis, Firefly luciferase values were normalized to the Renilla luciferase values after which technical duplicates were averaged and then normalized to the vehicle treated-empty vector sample.

SUPPLEMENTAL FIGURES



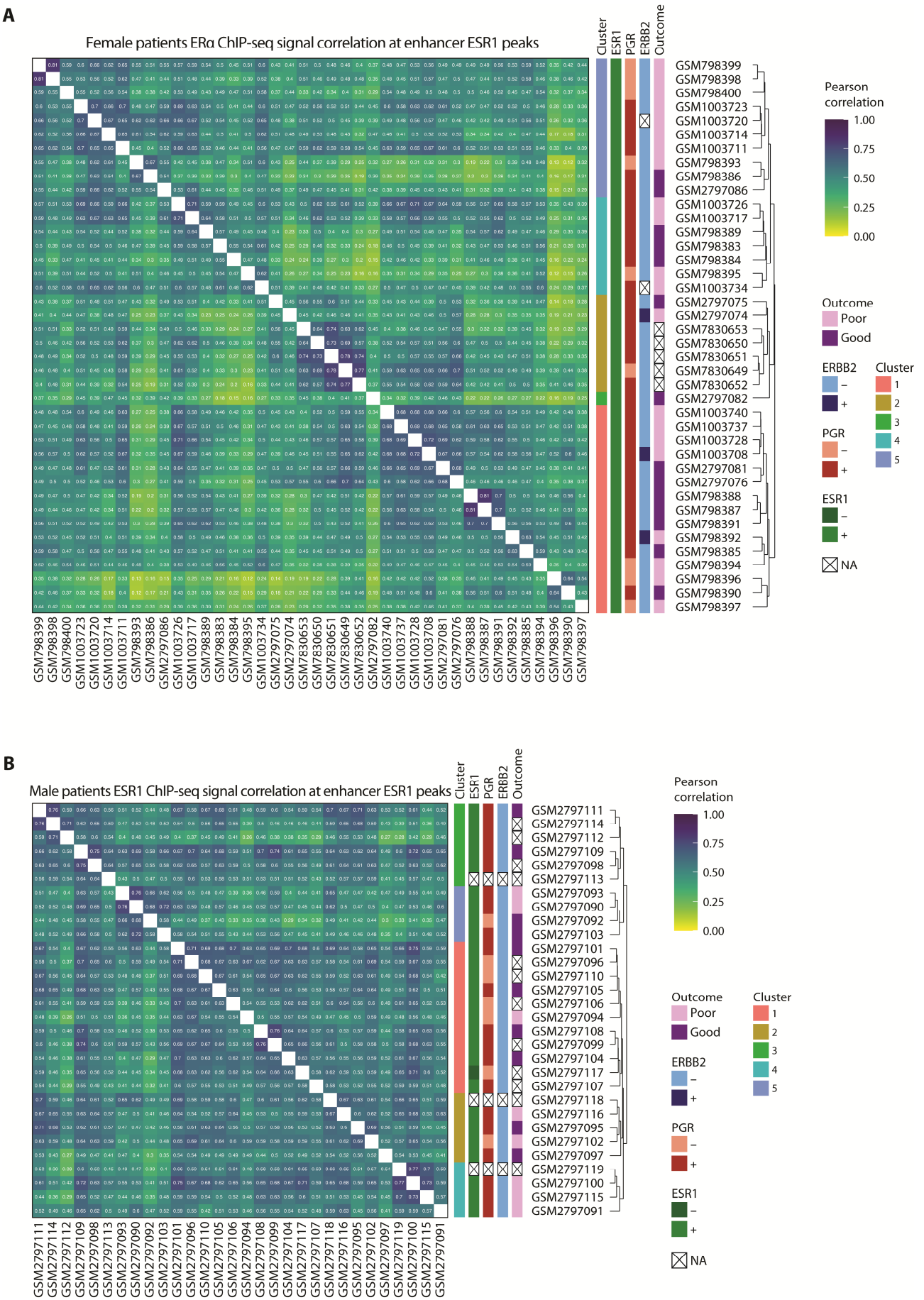




Supplemental Figure S1

Supplemental Figure S1 | ESR1 ranking in males and additional features of ESR1 sites in females

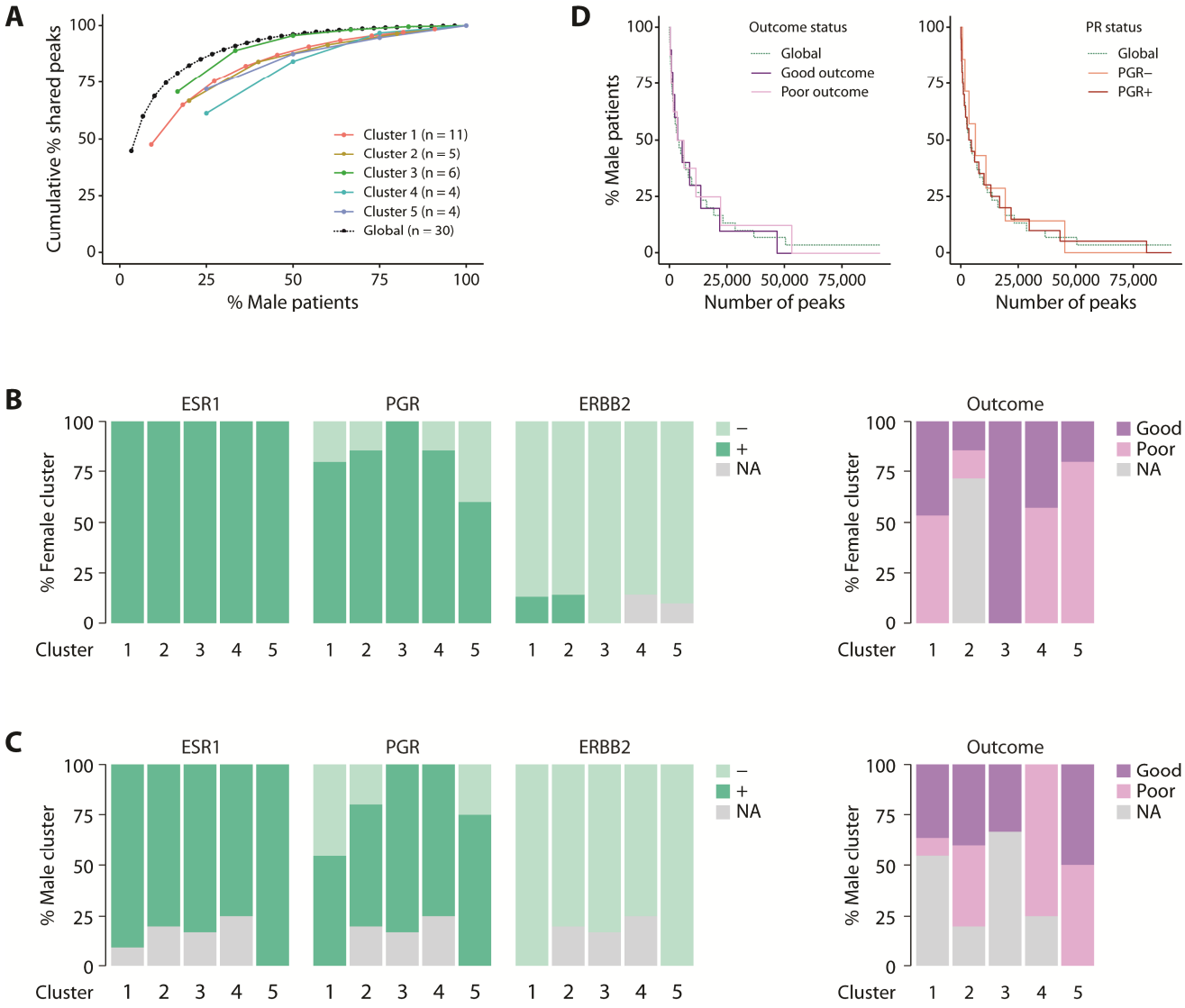
(**A**) A ranked overview of 91,712 distal ESR1 peaks showing the representation of each peak per tumor sample, in a cohort of 30 male patients. Heatmap shows the ESR1 ChIP-seq score at a specific peak for each sample. The bar plot on the left indicates the fraction of peaks found in each patient, relative to the total number of peaks found. Clustering is based on the Pearson correlation coefficient at ESR1 peaks for the ESR1 ChIP-seq signal as defined in Supplemental Fig. S2A. (**B**) The absolute number of distal peaks, per patient sample, in the female (top) and male (bottom) cohort. (**C**) Within each patient sample, the percentage of distal peaks considered common, less common (< half of patients) and patient unique in the ranking of all distal ESR1 peaks for that cohort.



Supplemental Figure S2

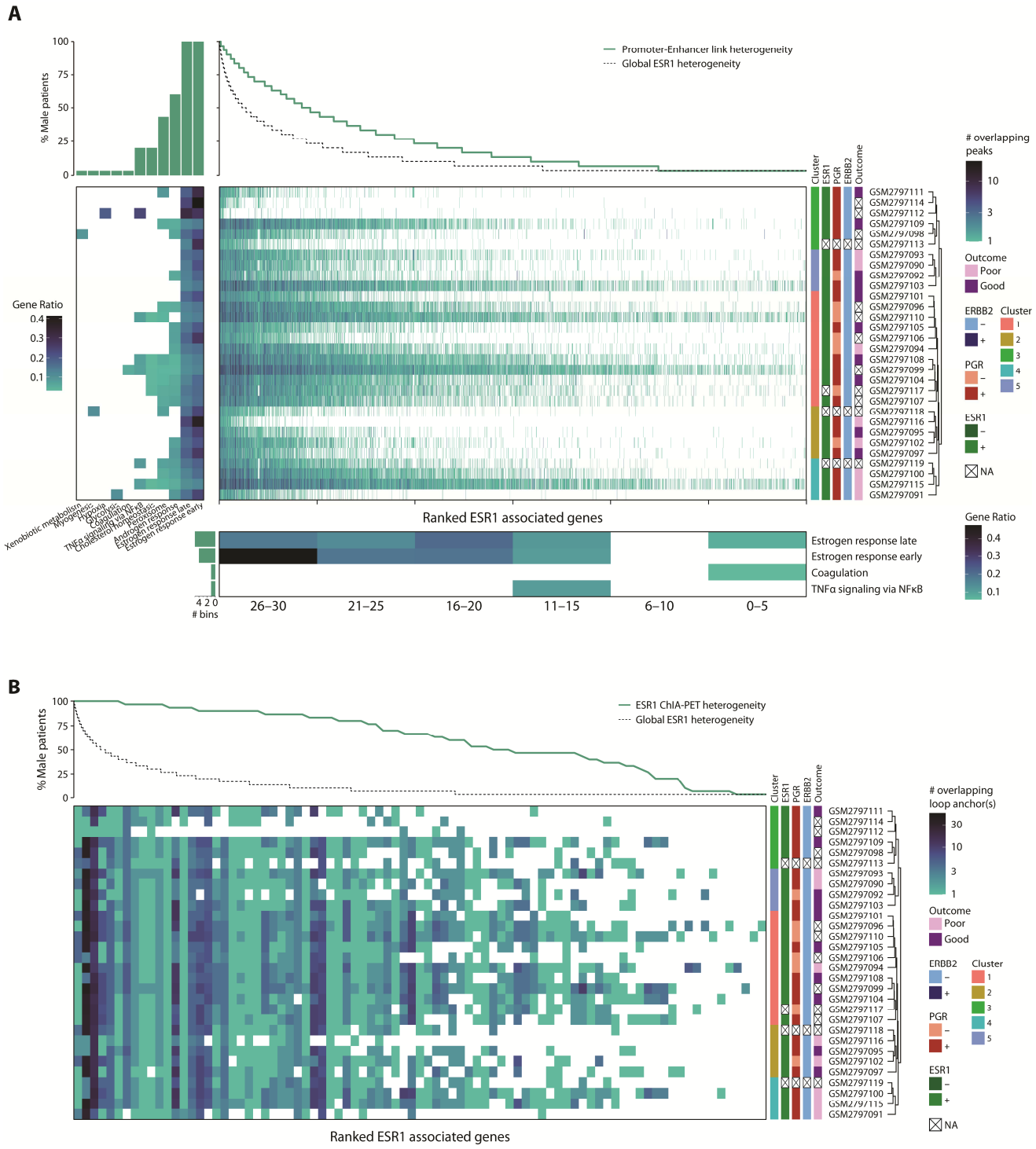
Supplemental Figure S2 | Patients correlation based on ESR1 ChIP-seq signal

(**A-B**) Correlation heatmap of the Pearson correlation coefficients for the ESR1 ChIP-seq signal at enhancer peaks in female (**A**) and male (**B**) breast cancer patients. Side bars on the right indicate hierarchical clustering, pathological information (ESR1, PGR and ERBB2 status) and outcome of the patients.



Supplemental Figure S3 | Evaluation of ESR1 inter-patient heterogeneity depending on molecular and pathological features

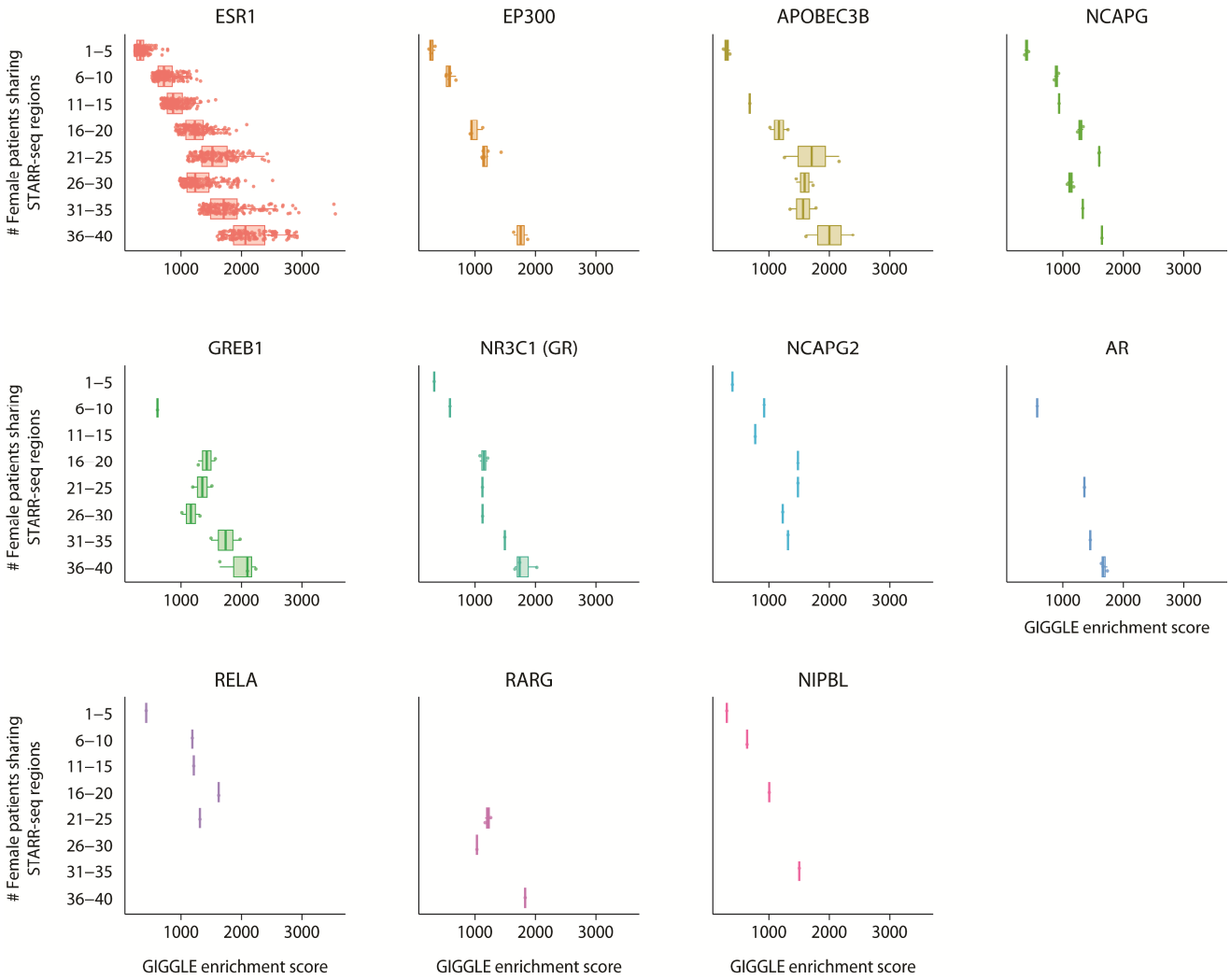
(**A**) Cumulative percentage of ESR1 peaks shared among female patients within the 5 clusters defined in Supplemental Fig. S2A. (**B-C**) Fraction of ESR1/PGR/ERBB2 status or outcome per patient cluster defined in Supplemental Fig. S2A, for female (B) and male (C) patients. (**D**) ESR1 peak heterogeneity in male tumors, between patient groups separated on outcome and PGR status. The global distribution over all the peaks is depicted by a dashed black line.



Supplemental Figure S4 | ESR1 male peaks converge to redundant enhancers regulating estrogen responsive genes

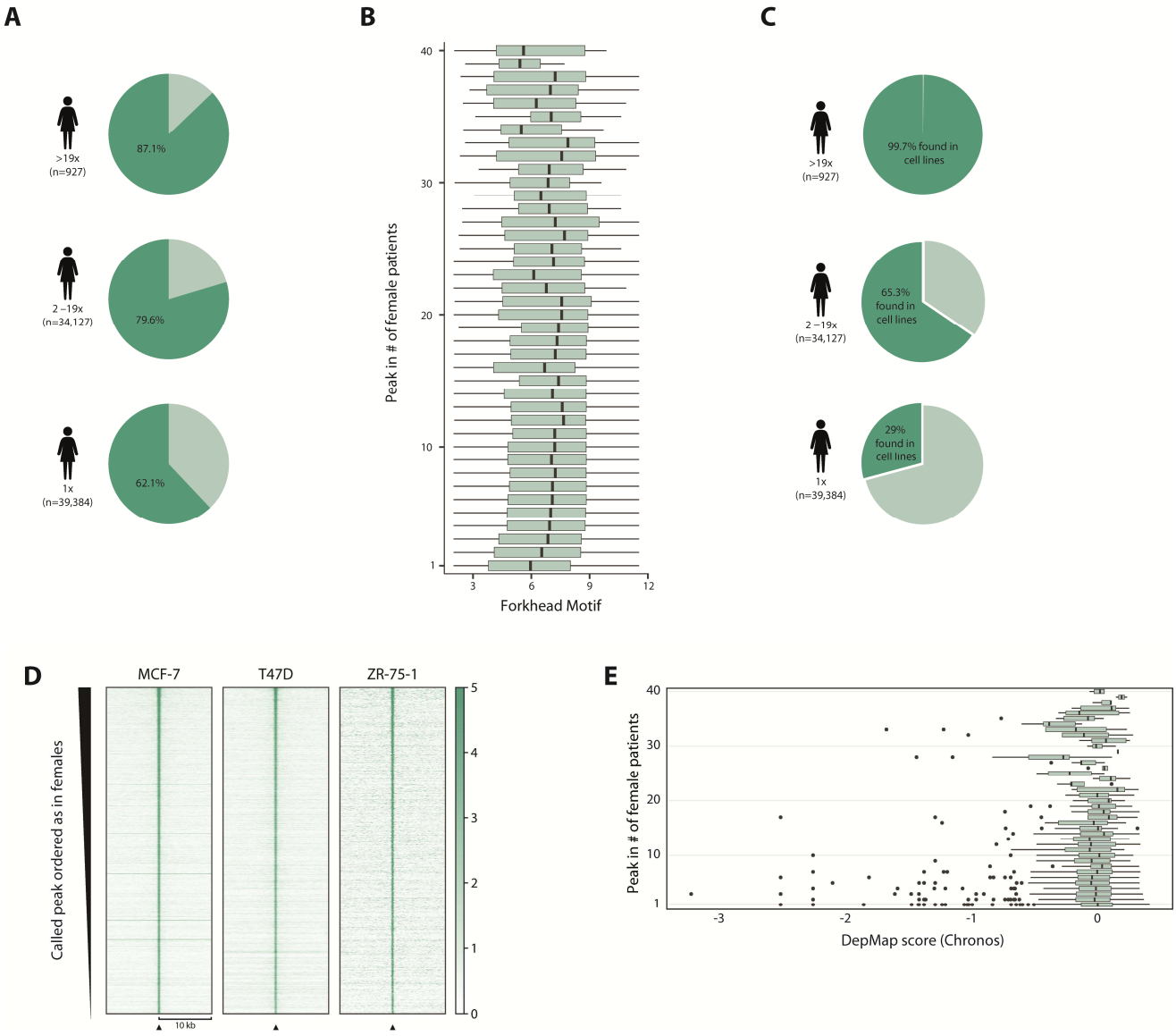
(**A**) Heatmap shows the number of ESR1 peaks overlapping with a region associated to a gene (Corces et al. 2018) (x-axis) per male patient (y-axis). Each gene is ranked by decreasing number of patients carrying ESR1 peaks associated to that specific gene. The number of patients sharing a gene is showed by the line above the heatmap. The global distribution of ESR1 peak conservation among samples is depicted by a black dashed line. Ranked genes are grouped in 6 bins depending on the degree of co-regulation among patients. For each bin are shown the statistically significantly enriched cancer hallmarks gene sets

(bottom-heatmap), and the bar plot on the bottom-left shows the number of bins sharing a given hallmark. On the left-sided heatmap are depicted the cancer hallmarks enriched in each patient; above the heatmap a bar plot indicates the percentage of patients showing the enrichment of each hallmark. (**B**) Same heatmap as in a, but in this case ESR1 peak-to-gene association is based on chromatin loops identified by ESR1 ChIA-PET in MCF7 (Fullwood et al. 2009; The ENCODE Project Consortium 2012).



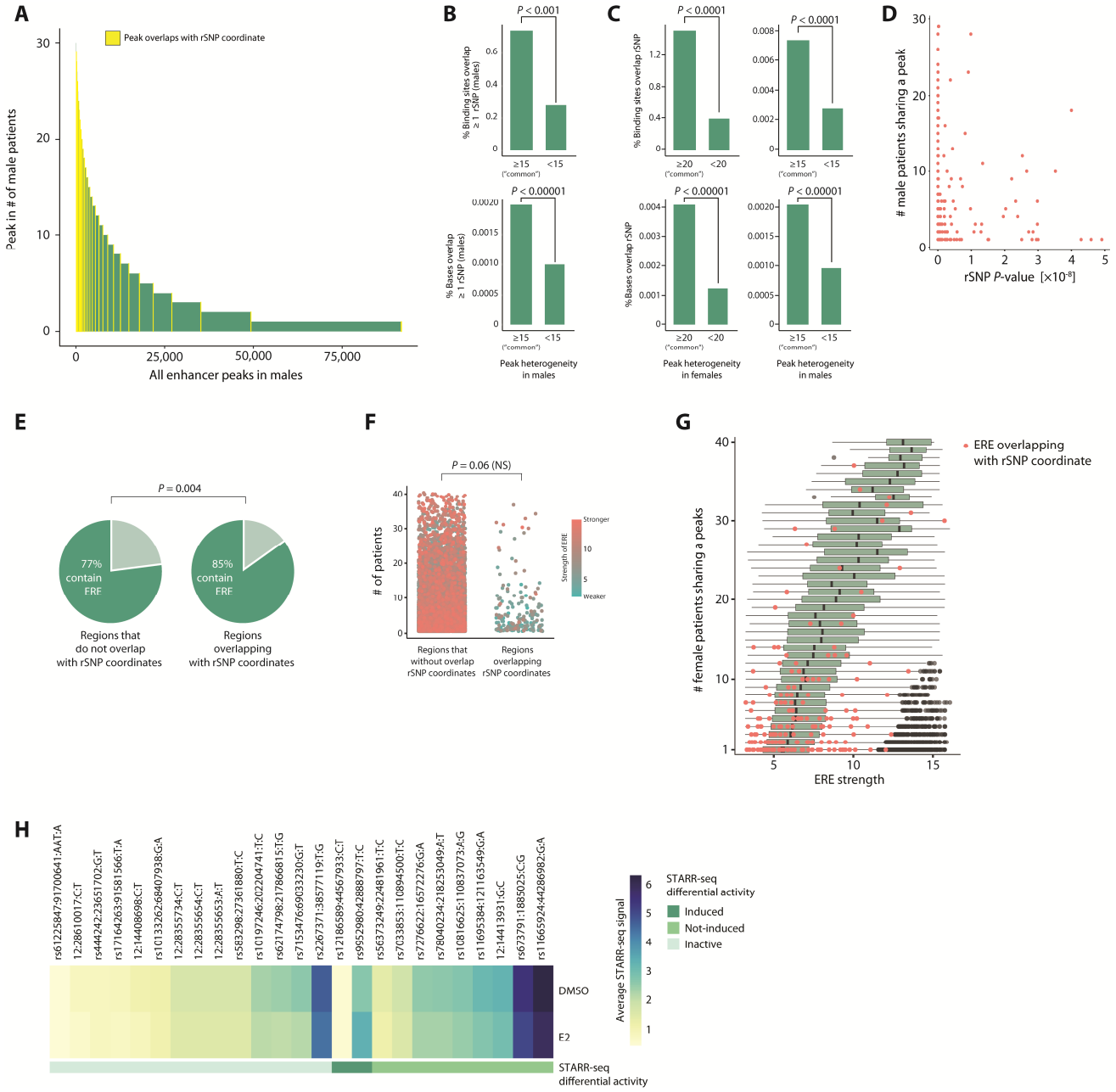
Supplemental Figure S5 | Transcription factor binding enrichment at female ESR1 peaks

Boxplots depicting the GIGGLE (Layer et al. 2018) transcription factor binding enrichment scores at ESR1 peaks over different bins of peaks shared among patients.



Supplemental Figure S6 | FORKHEAD motif analyses and ESR1 binding comparison with breast cancer cell lines

(**A**) The percentage of common, less common and patient unique ESR1 peaks in females containing a Forkhead motif. (**B**) Forkhead motif strength, ranked ESR1 peak heterogeneity in females. (**C**) The percentage of common, less common and patient unique ESR1 peaks in females that can also be found in the ESR1+ cell lines MCF7, T-47D and ZR-75-1. (**D**) Heatmaps showing ESR1 ChIP-seq signal as identified in MCF7, T-47D and ZR-75-1, ranked on peak conversation among patients (**E**) DepMap score of genes that couple to the distal ESR1 peaks found in 40 female breast cancer patients. Negative values indicate decreases cellular fitness upon gene knockout.



Supplemental Figure S7 | Extended details on rSNPs intersecting ESR1 peaks in female and male breast cancer patients

(**A**) Ranking of distal ESR1 peaks in the male cohort, with highlighted peaks that intersect with the coordinates of ESR1+ breast cancer rSNPs. (**B**) *Top*: Comparison (Fischer's exact test) of *percentage of* ESR1 *peaks* of which coordinates overlap with at least once rSNP coordinate, for common (≥15 patients) and less common (<15) ESR1 peaks. *Bottom*: Comparison (Fischer exact test) of the *percentage of bases* present in common or less common ESR1 peaks, that overlap with at least once rSNP coordinate. Unique peaks have been excluded. (**C**) *Top*: Comparison (Fischer's exact test) of *percentage of* all ESR1 *peaks* of which coordinates overlap with at least once rSNP coordinate, for common (≥15 patients) and less common (<15) ESR1 peaks. *Bottom*: Comparison (Fischer exact test) of the *percentage of bases* present in all common or less common ESR1 peaks, that overlap with at least once rSNP coordinate. Analyses for females on the left, and for males on the right. (**D**) For males, relationship between *p*-value of rSNPs with

coordinates that overlap with ESR1 peaks (x-axis) and the ranking position of those ESR1 peak in the male cohort (y-axis). (**E**) Comparison (Chi-square test) of ERE presence between enhancers peaks in females of which coordinates do not (left) and do overlap with at least one rSNP coordinate (right). (**F**) Comparison of ERE strength at ESR1 sites in females overlapping and not overlapping rSNP coordinates. (**G**) Boxplot depicting the ERE motif score distribution at ESR1 peaks as function of the degree of conservation among female patients. Individual EREs overlapping with a rSNP are indicated by red dots. (**H**) Average STARR-seq signal at 25 regions at which rSNPs coordinates intersect with an ERE.

REFERENCES

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M et al. 2013. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **41**: D991-995.

Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong CK, Cho SW et al. 2018. The chromatin accessibility landscape of primary human cancers. *Science* **362**.

Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH et al. 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**: 58-64.

Gregoricchio S, Polit L, Esposito M, Berthelet J, Delestré L, Evanno E, Diop MB, Gallais I, Aleth H, Poplineau M et al. 2022. HDAC1 and PRC2 mediate combinatorial control in SPI1/PU.1-dependent gene repression in murine erythroleukaemia. *Nucleic Acids Research* **50**: 7938-7958.

Jansen MP, Knijnenburg T, Reijm EA, Simon I, Kerkhoven R, Droog M, Velds A, van Laere S, Dirix L, Alexi X et al. 2013. Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. *Cancer Res* **73**: 6632-6641.

Layer RM, Pedersen BS, DiSera T, Marth GT, Gertz J, Quinlan AR. 2018. GIGGLE: a search engine for large-scale integrated genome analysis. *Nature Methods* **15**: 123-126.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Muerdter F, Boryń Ł M, Woodfin AR, Neumayr C, Rath M, Zabidi MA, Pagani M, Haberle V, Kazmar T, Catarino RR et al. 2018. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods* **15**: 141-149.

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR et al. 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**: 389-393.

Severson TM, Kim Y, Joosten SEP, Schuurman K, van der Groep P, Moelans CB, Ter Hoeve ND, Manson QF, Martens JW, van Deurzen CHM et al. 2018. Characterizing steroid hormone receptor chromatin binding landscapes in male and female breast cancer. *Nat Commun* **9**: 482.

Singh AA, Schuurman K, Nevedomskaya E, Stelloo S, Linder S, Droog M, Kim Y, Sanders J, van der Poel H, Bergman AM et al. 2019. Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. *Life Sci Alliance* **2**: e201800115.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Zwart W, Koornstra R, Wesseling J, Rutgers E, Linn S, Carroll JS. 2013. A carrier-assisted ChIP-seq method for estrogen receptor-chromatin interactions from breast cancer core needle biopsy samples. *BMC Genomics* **14**: 232.