Hormone Control Regions mediate steroid receptor-dependent genome organization.

François Le Dily[1,°], Enrique Vidal[1,°], Yasmina Cuartero[1,2], Javier Quilez[1,3], A. Silvina Nacht[1], Guillermo P. Vicent[1,4], José Carbonell-Caballero[1], Priyanka Sharma[1], José Luis Villanueva-Cañas[1], Roberto Ferrari[1], Lara Isabel De Llobet[1], Gaetano Verde[1,5], Roni H. G. Wright[1] and Miguel Beato[1,6,*].

[1] Gene Regulacion, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain
[2] CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Baldiri Reixac 4, Barcelona 08028, Spain
[3] Current address: Bioinformatics, Clarivate Analytics, Barcelona, Spain
[4] Current address: Centro de Biología Molecular Severo Ochoa, CSIC-UAM, Nicolás Cabrera 1, Cantoblanco, 28049 Madrid, Spain
[5] Current address: Department of Basic Sciences, Universitat Internacional de Catalunya, Barcelona, Spain
[6] Universitat Pompeu Fabra (UPF), Barcelona, Spain

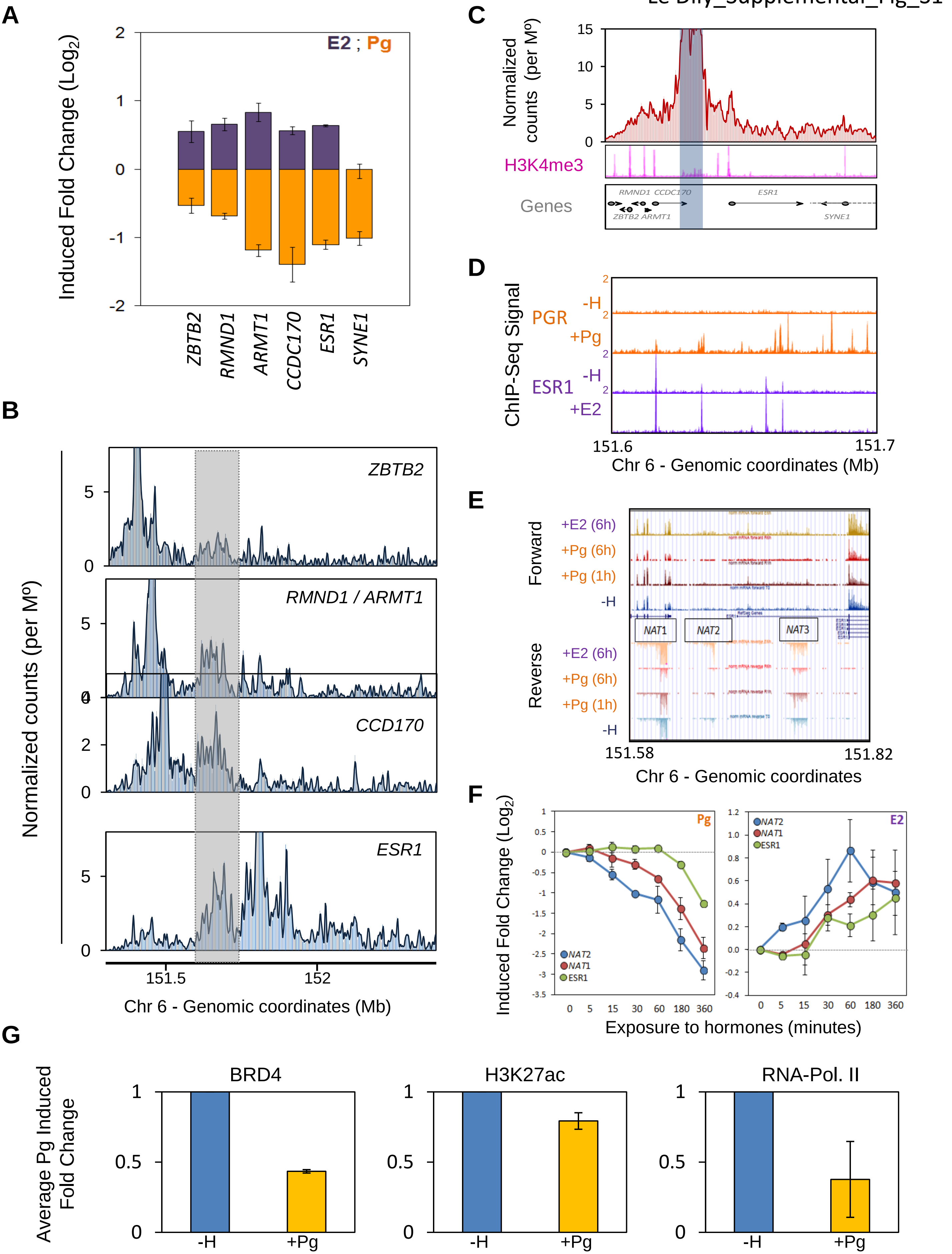
Supplemental Material:

Supplemental Table 1: Summary of 3C-derived datasets

				Raw counts in depicted regions	
Experiment	**Sample**	**Sequenced Read pairs**	**Valid Read pairs**	**ESR1-TAD**	**PGR-TAD**
In Situ Hi-C	**BT474 rep.1**	196,802,667	126,400,945	23,014	51,322
	BT474 rep.2	169,343,097	127,292,817	29,378	73,520
	MCF10A rep.1	306,018,506	187,513,575	48,500	31,102
	MCF10A rep.2	222,455,381	156,339,157	64,198	32,484
	MCF7 rep.1	320,056,659	166,607,495	43,758	12,056
	MCF7 rep.2	340,495,431	216,563,500	73,478	20,244
	SKBR3 rep.1	205,997,414	156,557,711	39,534	24,338
	SKBR3 rep.2	214,205,796	151,009,167	49,346	30,656
	T47D T0 rep.1	189,967,400	157,390,317	36,962	32,592
	T47D R30 rep.1	179,252,416	145,811,017	34,576	30,150
	T47D R180 rep.1	192,668,851	158,516,232	40,888	34,610
	T47D E30 rep.1	249,799,510	193,968,773	51,230	47,276
	T47D E180 rep.1	197,859,885	161,304,275	39,310	34,986
	T47D T0 rep.2	315,542,160	228,654,226	50,820	39,422
	T47D R30 rep.2	249,476,974	174,338,960	44,120	30,488
	T47D R180 rep.2	244,958,790	163,713,665	39,154	28,444
	T47D E30 rep.2	349,550,907	218,919,170	51,066	40,774
	T47D E180 rep.2	300,143,601	217,521,326	46,440	34,290
	T47D siRNA Control rep.1	206,292,967	116,641,257	26,752	26,378
	T47D siRNA Control rep.2	205,462,437	109,936,278	28,298	19,762
	T47D siRNA ER rep.1	209,173,554	118,085,021	33,092	24,388
	T47D siRNA ER rep.2	204,518,224	111,152,397	34,296	19,334
	T47D shEmpty	204,764,799	128,911,715	28,984	23,990
	T47D shPR	166,975,353	114,011,263	23,224	18,712
Capture-3C	**T47D T0**	325,166,775	57,414,832	1,564,728	ND
	T47D R60	283,013,228	43,446,033	1,594,758	ND
	T47D R180	281,176,954	47,767,277	1,953,130	ND
	T47D E60	282,812,094	26,517,206	935,620	ND
Capture-HIC	**T47D T0**	45,857,588	18,997,544	246,450	ND
	T47D R60	32,313,052	13,824,138	217,086	ND
	T47D R180	22,555,505	8,952,798	101,310	ND
	T47D E60	23,576,054	9,922,720	168,352	ND
	T47D E180	13,505,153	5,653,815	111,758	ND
	GM 12878	81,097,244	28,516,961	218,996	ND

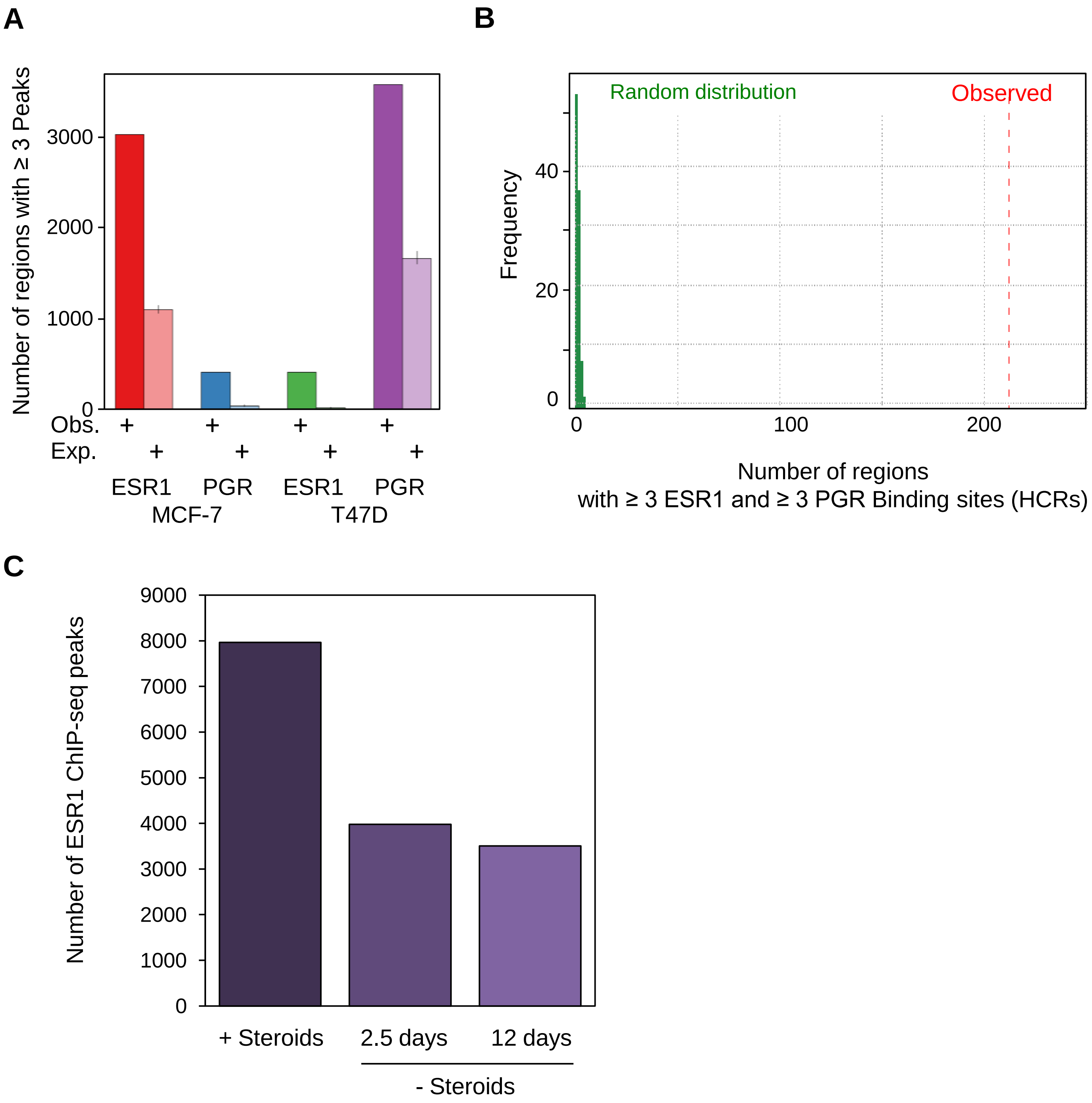
*Valid read pairs after mapping and filtering of: PCR duplicates, self-circle, dangling-ends, random breaks and errors. ND: Not Determined.

Le Dily_Supplemental_Fig_S1



Supplemental Figure S1: (A) Log$_2$ fold change of the transcripts level of protein-coding genes located in the *ESR1*-TAD after 6 hours of exposure to Pg (Orange) or E2 (Purple). The plot shows the average ± SEM expression of 3 independent RNA-seq experiments. **(B)** Virtual 4C profiles at 2 kb resolution using as baits the TSS of genes within the *ESR1*-TAD. Data were obtained from T47D cells grown in absence of hormones. A common region of contacts is highlighted in grey. **(C)** Results obtained by Capture-C (Figure 1) were confirmed in Capture-HiC. Virtual 4C profile at 5 kb resolution (expressed as normalized counts per Mº within the region depicted) using the 90 kb intergenic region (highlighted in blue) as bait. **(D)** Magnification of the Genome Browser tracks from Fig. 1C: ChIP-seq profiles obtained for PGR and ESR1 in absence (-H) or presence of Pg (+Pg) or E2 (+E2) within the region Chr 6: 151,600,000-151,700,000. **(E)** RNA-seq profiles obtained in T47D grown in absence of hormones (-H) or exposed 1h or 6h to Pg (+Pg) or 6h minutes to E2 (+E2) showing the expression of 3 non-annotated transcripts (*NAT1*, *NAT2* and *NAT3*). **(F)** Quantitative RT-PCR showing the kinetics of changes of *NAT1*, *NAT2* and *ESR1* after exposure to Pg (left panel) or E2 (right panel) showing that the non-annotated transcripts within the *ESR1*-TAD HCR are coordinately regulated by the hormones similarly to the protein coding genes. Plots correspond to the average ± SEM of two independent kinetics. **(G)** ChIP-qPCR showing the fold changes induced by Pg of BRD4, H3K27ac, and RNA-Pol. II levels at the *ESR1*-TAD HCR (Average ± SD of 3 different sites – Chr 6: 151,523,214-151,523,355 ; Chr 6:151,633,723-151,633,877 ; chr6:151,658,121-151,658,214).

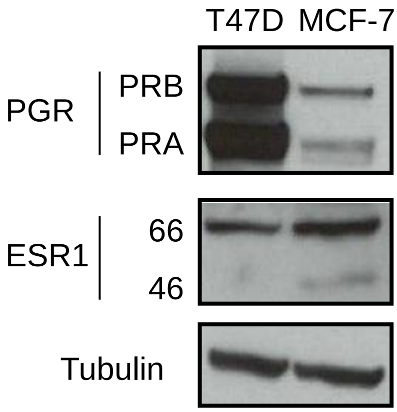
Le Dily_Supplemental_Fig_S2



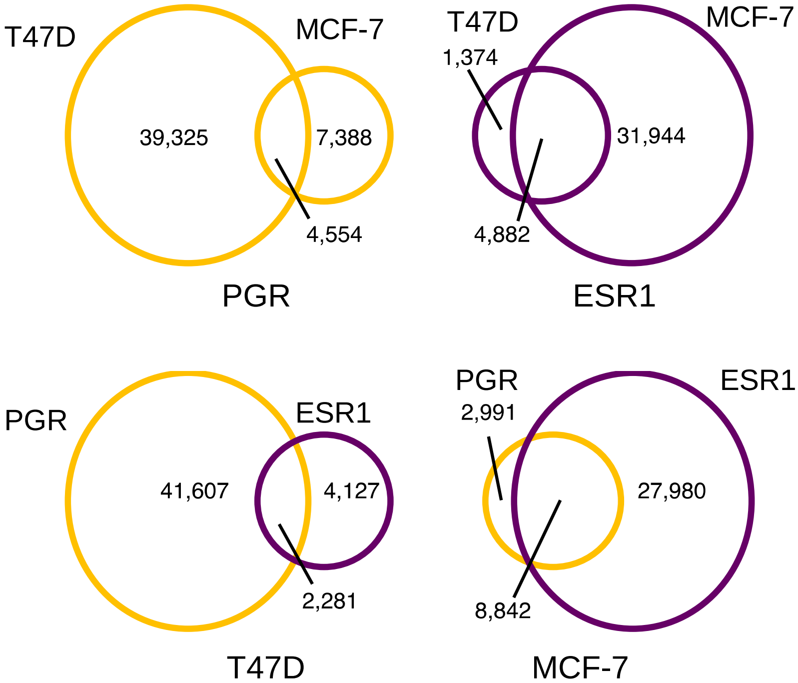
Supplemental Figure S2: (A) Histograms showing the observed and expected (randomization of sites) numbers of ESR1 or PGR clusters (≥ 3 consecutive sites separated by ≤ 25 Kb) in MCF-7 or T47D. **(B)** Plot showing the observed number of HCRs in T47D (red dashed line – n= 212) compared to the expected frequencies based on random distribution of ESR1 and PGR clusters. **(C)** Histograms showing the number of ESR1 binding sites detected by ChIP-seq in T47D grown in medium containing normal FBS (+ steroids) or after 2.5 and 12 days after steroid withdrawal.

Le Dily_Supplemental_Fig_S3

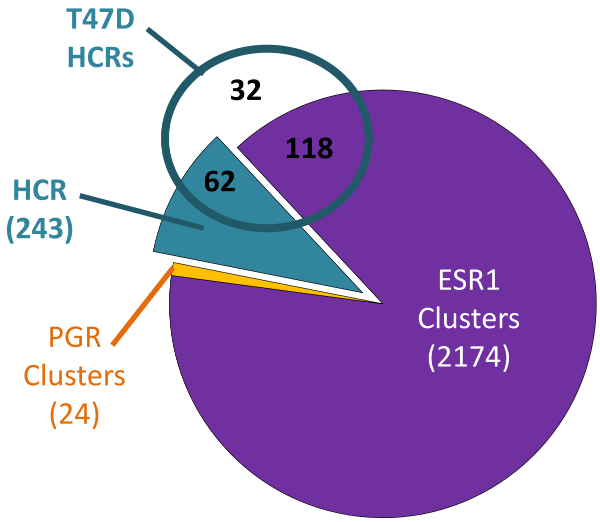
A



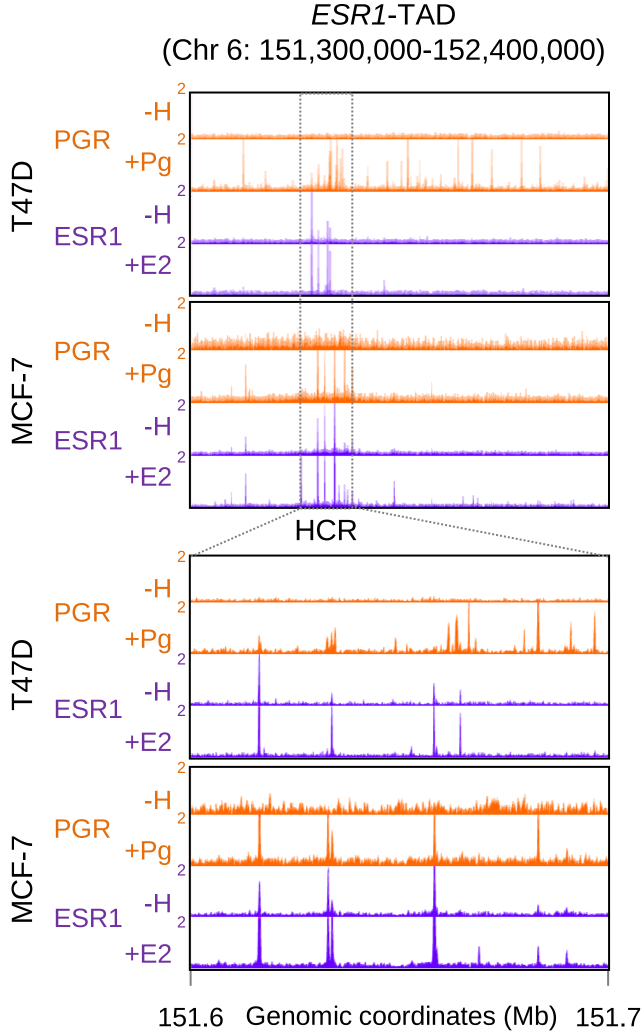
B



C



D

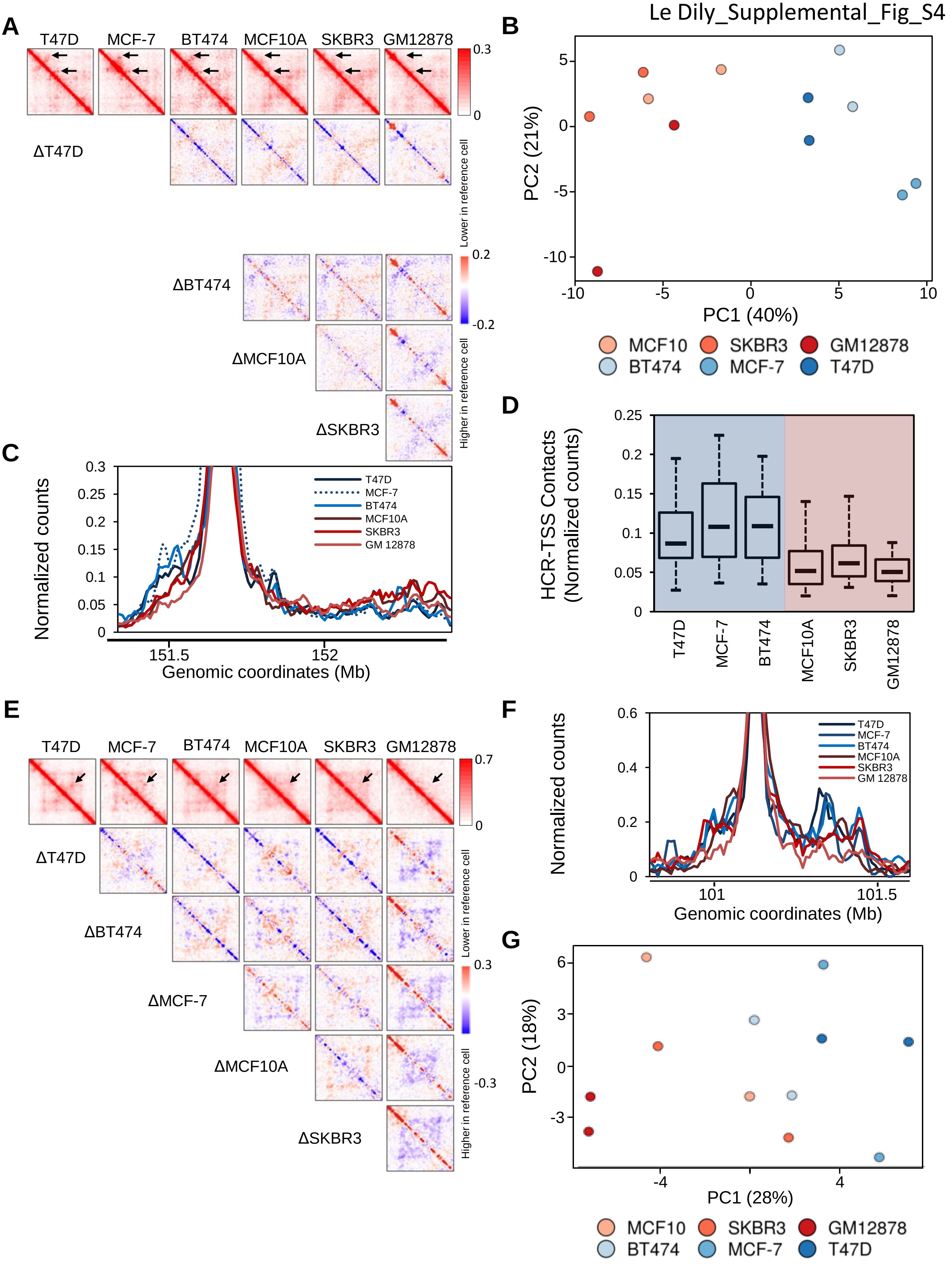


E

Overlapping between ESR1 and PGR (Jaccard Index)		
T47D	Inside HCRs	0.24
	Outside HCRs	0.18
MCF-7	Inside HCRs	0.65
	Outside HCRs	0.58

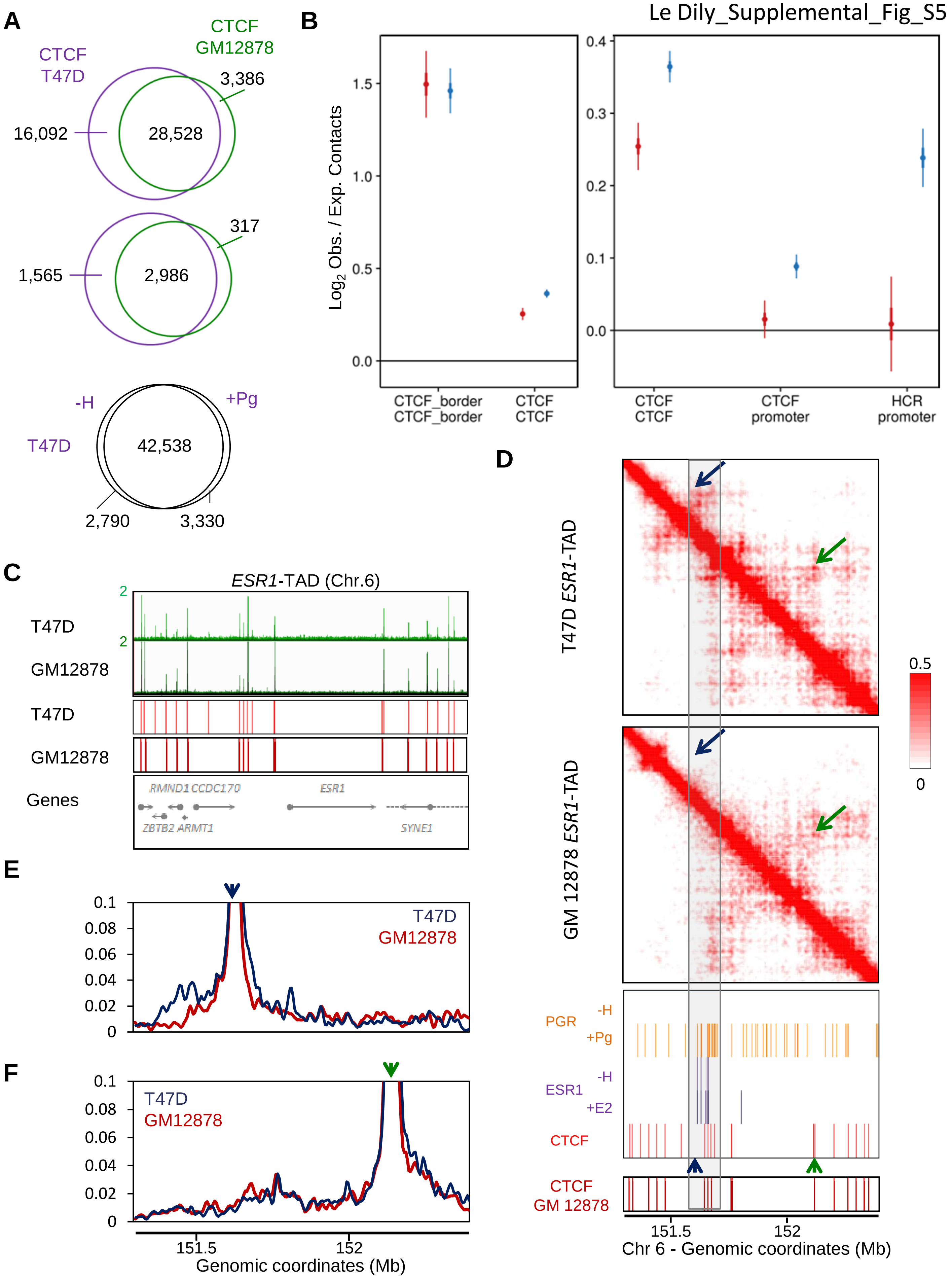
<u>Supplemental Figure S3</u>: (A) Western Blots showing the levels of ESR1 (isoforms of 46 and 66 KDa), PGR (isoforms PRA and PRB) expressed in T47D and MCF-7. Tubulin was used as loading control. **(B)** Venn diagrams showing the number of PGR (top left panel) and ESR1 (top right panel) binding sites overlapping in T47D and MCF-7 exposed to 60 minutes of Pg or E2, respectively. The bottom panels show the overlap between ESR1 and PGR binding sites in T47D (left) or MCF-7 (right). **(C)** The pie chart indicates the number of clusters (PGR-predominant; ESR1-predominant or HCR) identified in MCF-7 cells using the same strategy described in the main text for T47D. The overlapping with HCRs identified in T47D is shown. **(D)** Genome browser view of ChIP-seq profiles of PGR and ESR1 in T47D (same as in Fig. 1) and MCF-7 exposed or not to Progestin (+Pg) or E2 (+E2) for 60 minutes over the *ESR1*-TAD (top panel). The panel below corresponds to a magnification within the region defined as the ESR1-TAD HCR for T47D (as in Supplemental Fig. S1D) or MCF-7. **(E)** To estimate the overlapping between ESR1 and PGR binding sites, a Jaccard index between the two types of binding sites was computed for binding sites located within or outside HCRs in T47D and MCF-7.

Le Dily_Supplemental_Fig_S4

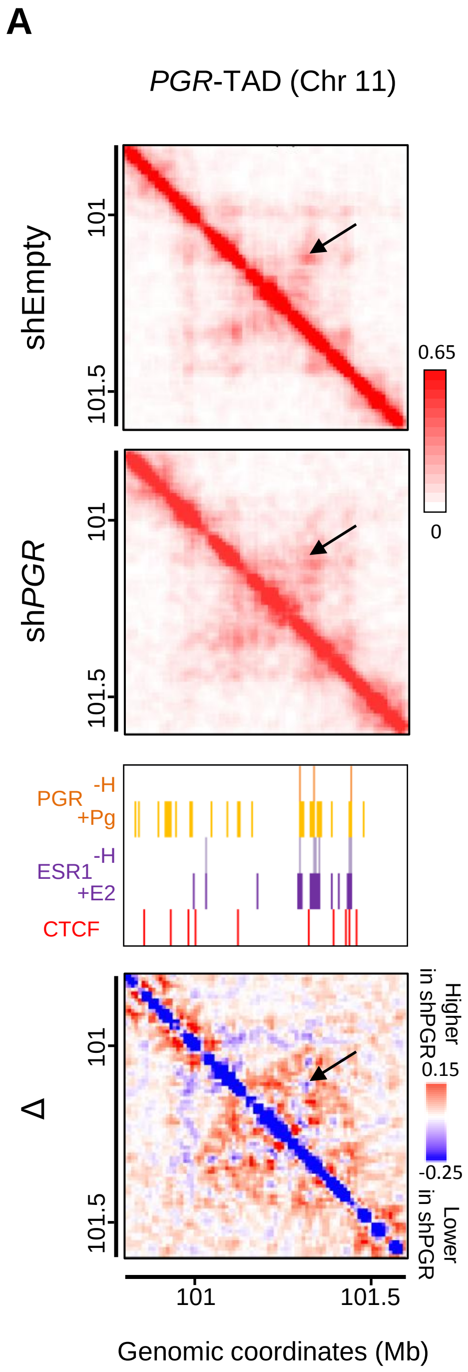


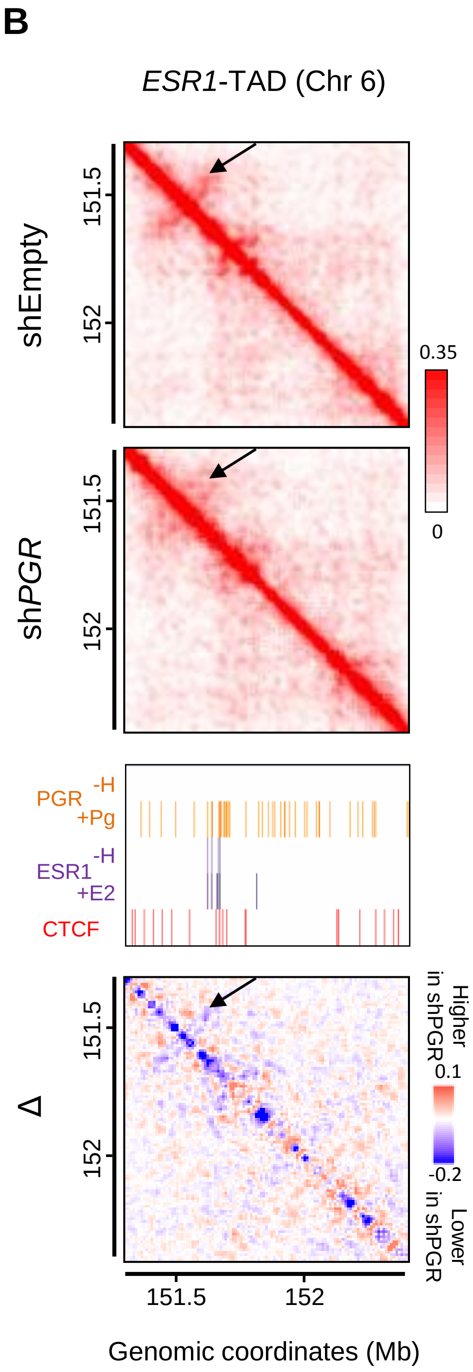
Supplemental Figure S4: (A) Normalized *In Situ* Hi-C contact matrices at 10 kb resolution of the *ESR1*-TAD (Chr 6: 151,300,000-152,400,000) obtained in the different cell lines grown in the absence of hormones. Arrows highlight loops between HCR and promoters detected in T47D cells. The matrices below show the differences between the contact matrices of the cell lines at the top minus the reference cell line on the left (blue and red corresponding to contacts higher and lower in the reference cell line, respectively). MCF-7 cells exhibit a duplication of the region corresponding to the HCR. The differences for this line are not shown. **(B)** Principal Component Analysis (PCA) of the contacts profiles of the region corresponding to *ESR1*-TAD HCR of 2 biological replicates of Hi-C performed in each of the cell lines. **(C)** Virtual 4C profiles at 10 kb resolution using the *ESR1*-TAD HCR as bait in the different ESR1$^+$-PGR$^+$ (blue lines) and ESR1$^-$-PGR$^-$ (red lines) grown in absence of hormones. MCF-7 cells are represented with a dashed blue line. **(D)** Distribution of contacts between genomic bins within the HCR with the TSS of protein-coding genes within the ESR1-TAD. **(E)** Normalized *In Situ* Hi-C contact matrices at 10 kb resolution of the PGR-TAD (Chr11: 101,800,000-101,600,000) obtained in the cell lines listed grown in absence of hormones. The matrices below show the differences between the contact matrices of the cell lines at the top minus the reference cell line on the left (blue and red corresponding to contacts higher and lower in the reference cell line, respectively). **(F)** Virtual 4C profiles at 10 kb resolution (expressed as normalized counts per thousand within the region depicted) using the TSS of *PGR* as bait in the different ESR1$^+$-PGR$^+$ (blue lines) and ESR1$^-$-PGR$^-$ (red lines) grown in the absence of hormones. **(G)** PCA of the contacts profiles of the region corresponding to *PGR*-TAD HCR of 2 biological replicates of Hi-C performed in each of the cell lines.

Le Dily_Supplemental_Fig_S5



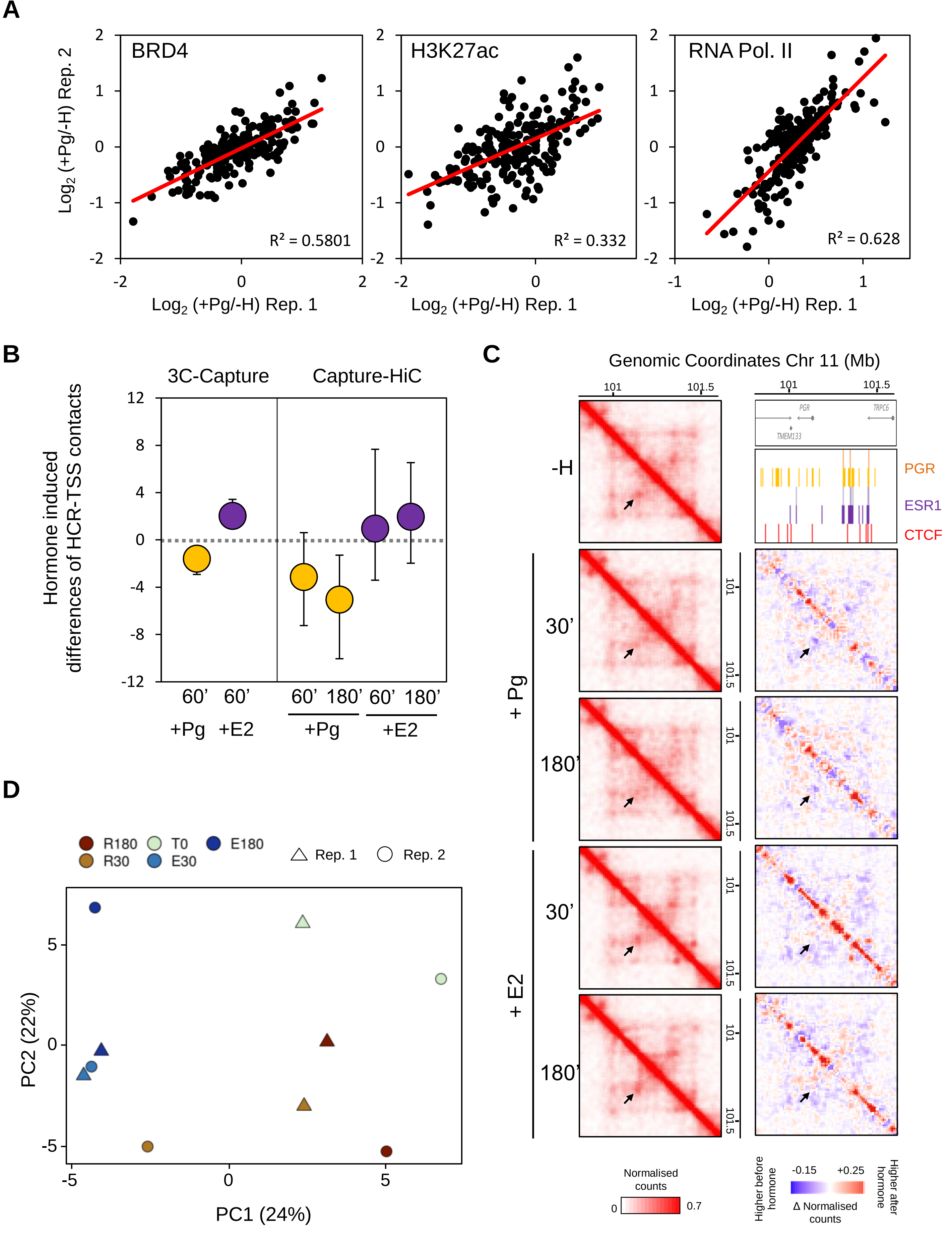
Supplemental Figure S5: (A) Venn diagrams showing the overlapping between CTCF sites detected by ChIP-seq in T47D and in GM12878. Top: genome-wide CTCF binding sites; Middle: CTCF binding sites in TADs containing HCRs. The venn diagram at the bottom shows the overlapping between CTCF sites detected before (-H) and after exposure of T47D to progestins (+Pg). **(B)** Plot showing the contact frequencies between CTCF sites located at the border of TADs (CTCF border) and between CTCF sites located within the same TAD (CTCF-CTCF) in T47D or GM12878 cells (left panel). The plot on the right shows the contact frequencies between CTCF sites located within the same TAD (CTCF-CTCF), between CTCF sites and sites marked by H3K4me3 (CTCF-promoter) as well as between HCR and sites marked by H3K4me3 (HCR-promoter) in T47D or GM12878 cells. **(C)** Genome browser tracks correspond to ChIP-seq profiles for CTCF in T47D and GM12878 cells in the *ESR1*-TAD. Middle panels show the position of significant CTCF binding sites detected in the 2 lines. **(D)** Normalized Capture-HiC contact matrices at 5 kb resolution of the *ESR1*-TAD in T47D (top) and GM12878 (bottom) cells. Purple arrows point to the loops established between the HCR (highlighted in grey) and the upstream promoters in T47D, which are absent in GM12878. Green arrows point to loops established between CTCF sites, which are conserved in both cell lines. Positions of CTCF in T47D and GM12878 as well as ESR1 in T47D are shown below the matrices. Blue and green arrows at the bottom show the position of the baits used in S5E and S5F, respectively**(E)** Virtual 4C profiles at 5 kb resolution using a ESR1 binding site within the HCR (purple arrow at the bottom of **(D)**) as bait in T47D (blue line) and GM12878 (red line) grown in absence of hormones. **(F)** Virtual 4C profiles at 5 kb resolution using a CTCF sites (green arrow at the bottom of **(D)**) as bait in T47D (blue line) and GM12878 (red line) grown in absence of hormones.


A
PGR-TAD (Chr 11)
shEmpty
sh*PGR*
Δ
101
101.5
0.65
0
PGR -H
+Pg
ESR1 -H
+E2
CTCF
Higher in shPGR
0.15
-0.25
Lower in shPGR
Genomic coordinates (Mb)


B
ESR1-TAD (Chr 6)
shEmpty
sh*PGR*
Δ
151.5
152
0.35
0
PGR -H
+Pg
ESR1 -H
+E2
CTCF
Higher in shPGR
0.1
-0.2
Lower in shPGR
Genomic coordinates (Mb)

Supplemental Figure S6: (A, B) Normalized *In Situ* Hi-C contact matrices at 10 kb resolution of the region over the *PGR*-TAD (**A**) and *ESR1*-TAD (**B**) obtained in T47D expressing control shRNA (shEmpty) or shRNA against *PGR* (sh*PGR*). Arrows highlight the position of loops between HCRs and promoters. Positions of protein coding genes, position of significant CTCF, ESR1 and PGR binding sites in presence or absence of hormones as well as genomic coordinates are shown at the bottom. The difference between the contact matrices is shown at the bottom with blue and red corresponding to contacts higher and lower in T47D shEmpty than in T47D sh*PGR*, respectively.

Le Dily_Supplemental_Fig_S7



Supplemental Figure S7: (A) Plot showing the correlations between ChIP-seq replicates (Rep. 1 and Rep. 2) of the fold changes induced by Pg at the HCRs for BRD4, H3K27ac and RNA-Pol. II. **(B)** Distribution of hormone induced differences of contacts between genomic bins of the HCR with the TSS of protein-coding genes in *ESR1*-TAD after treatment of T47D cells with Pg (orange) or E2 (purple) in 3C- and Hi-C capture experiments. **(C)** Normalized *In Situ* Hi-C contact matrices at 10 kb resolution of the *PGR*-TAD obtained in T47D grown in absence of hormone (-H) or after exposure to 30 and 180 minutes to Pg or E2. The matrices on the right show the differences between the contact matrices compared to the one obtained in absence of hormone (blue and red corresponding to contacts higher and lower in the absence of hormone, respectively). Positions of protein coding genes, position of significant CTCF, ESR1 and PGR binding sites in presence or absence of hormones as well as genomic coordinates are shown on the top. **(D)** PCA of the Hi-C profiles obtained in 2 independent kinetics of exposure of T47D cells to Pg or E2 for 0, 30 and 180 minutes showing that the two hormones induces reproducible hormone-specific changes in 3D organization of HCR containing TADs.