EstroGene2.0: A multi-omic database of response to estrogens, ER-modulators, and resistance to endocrine therapies in breast cancer

Supplementary Figures



Supplementary Figure S1 (related to Fig. 1)

A. Stacked histogram showing the metadata separated by technologies and three endocrine resistant model types. B and C. Stacked histogram summarizing the dose (B) and duration (C) of tamoxifen resistant model selection separated by technologies.

	Data Set II	Experiment ID	Modality	Cell model	Cell Line	Replicates	Cell Model Type	Variant Type	Ins
Modality	ERR7	MUTR1_1	RNA-seq	ESR1 Mutation	T47D	4	CRISPR/Cas9	Y537S, D538G	University
Multi-select	ERR7	MUTR1_2	RNA-seq	ESR1 Mutation	MCF7	4	AAV Genome-edited	Y537S, D538G	University
ER ChIP-seq	ERR8	MUTR2	RNA-seq	ESR1 Mutation	MCF7	3	CRISPR/Cas9	Y537S	Imperial C
Microarray RNA-seg	ERR9	MUTR3	RNA-seq	ESR1 Mutation	T47D	3	CRISPR/Cas9	Y537S, D538G	University
	ERR10	MUTR4_1	RNA-seq	ESR1 Mutation	T47D	3	Dox-inducible Overex	Y537S	Dana-Fart
Cell Line	ERR10	MUTR4_2	RNA-seq	ESR1 Mutation	MCF7	3	Dox-inducible Overex	Y537S, D538G, Y537N	Dana-Fart
Multi-select	ERR11	MUTR5	RNA-seq	ESR1 Mutation	MCF7	6	CRISPR/Cas9	Y537S, D538G, L536	Imperial C
MCF7	ERR12	MUTR6	RNA-seq	ESR1 Mutation	MCF7	2	CRISPR/Cas9	Y537S, D538G	UT Southw
SUM44	ERR13	MUTR7_1	RNA-seq	ESR1 Mutation	MCF7	2	CRISPR/Cas9	Y537S, D538G	University
SUM44PE	ERR13	MUTR7_2	RNA-seq	ESR1 Mutation	T47D	2	CRISPR/Cas9	Y537S, D538G	University
T47D	ERR15	MUTR8_1	RNA-seq	ESR1 Mutation/LTED	MCF7	3	Naturally Occurred in	Y537C	Institute o
	ERR15	MUTR8_2	RNA-seq	ESR1 Mutation/ITED	SUM44PE	3	Naturally Occurred in	Y537S	Institute o
Cell Model Type	ERR16	MUTR9	RNA-seq	ESR1 Mutation	T47D	1	CRISPR/Cas9	Y537S, D538G	Baylor Col
Multi-select	ERR17	MUTR10	RNA-seq	ESR1 Mutation	MCF7	2	TALEN Genome Editing	Y537S, D538G,E380Q	Dana-Fart
CRISPR/Cas9	ERM4	MUTM1	Microarray	ESR1 Mutation	MCF7	3	Stable Overexpression	Y537S, S463P, Y537	MSKCC
Dox-inducible	ERM5	MUTM2	Microarray	ESR1 Mutation	MCF7	3	Stable Overexpression	Y537S, D538G	H3 Biome
Overexpression	ERM6	MUTM3_1	Microarray	ESR1 Mutation	MCF7	3	Stable Overexpression	Y537S	Baylor Col
ITED condition	ERM6	MUTM3_2	Microarray	ESR1 Mutation	ZR751	3	Stable Overexpression	Y537S	Baylor Col
	ERM7	MUTM4	Microarray	ESR1 Mutation	T47D	3	Stable Overexpression	Y537S	University
F-		MUTC1 1	ER ChIP-seq	ESR1 Mutation	MCF7	1	AAV Genome-edited	Y537S, D538G	University
	ERC1	MUTC1_2	ER ChIP-seq	ESR1 Mutation	T47D	1	CRISPR/Cas9	Y537S, D538G	University
	ERC2	MUTC2_1	ER ChIP-seq	ESR1 Mutation	MCF7	2	CRISPR/Cas9	Y537S, D538G	University
	ERC2	MUTC2_2	ER ChIP-seq	ESR1 Mutation	T47D	2	CRISPR/Cas9	Y537S, D538G	University
	ERC3	MUTC3	ER ChIP-seq	ESR1 Mutation	MCF7	1	CRISPR/Cas9	Y537S	Imperial C

В

GSM ID	Cell Line	Sample ID	ESR1 Genotype	GREB1
GSM2392590	T47D	MUTR1_1_DG1	D538G	6.710
GSM2392591	T47D	MUTR1_1_DG2	D538G	6.505
GSM2392592	T47D	MUTR1_1_DG3	D538G	6.628
GSM2392593	T47D	MUTR1_1_DG4	D538G	6.365
GSM2392582	T47D	MUTR1_1_WT1	WT	5.806
GSM2392583	T47D	MUTR1_1_WT2	WT	5.689
GSM2392584	T47D	MUTR1_1_WT3	WT	5.776
GSM2392585	T47D	MUTR1_1_WT4	WT	5.770
GSM2392586	T47D	MUTR1_1_YS1	Y537S	7.385
GSM2392587	T47D	MUTR1_1_YS2	Y537S	7.332
GSM2392588	T47D	MUTR1_1_YS3	Y537S	7.697
GSM2392589	T47D	MUTR1_1_YS4	Y537S	7.420

C							
	Comparison	log2 Fold Change	Adjusted P-value				
	D538G over WT	0.8002	3.18e-41				
	Y537S over WT	1.713	5.43e-79				
Е							
	MUTR1_1_D538G_over_WT_DEseq_output.cs						
	MUTR1_1_log2CPM_Expression_Matrix.csv #						
	B MUTR1_	1_Y537S_over_WT_DEs	eq_output.csv 🚢				



SRR ID	Cell Line	Sample ID	ESR1 Genotype	Mapped Reads(M)	Total Peak
SRR8444288	MCF7	MCF7_D538G_Vehicle	D538G	17.36	399
SRR8444282	MCF7	MCF7_WT_Vehicle	WT	19.59	26
SRR8444285	MCF7	MCF7_Y537S_Vehicle	Y537S	18.18	265
	MUTC1	_1_MCF7_D538G_Ve	nicle_peaks.narr e_peaks.narrowF	owPeak 🚢 Peak 🚢	
	MUTC1	_1_MCF7_Y537S_Veh	icle_peaks.narro	owPeak 🚢	
	MUTC1	_1_read.counts.rpkm	.csv 🏔		



### Supplementary Figure S2 (Continued)

Supplementary Figure S2. (related to Fig. 2)

A. A screen shot from EstroGene2.0 browser of "METADATA" tab of ESR1 mutant models. Red dots indicate the two example data sets elaborated in the following panels.

B-E. Screen shot from EstroGene2.0 browser of searchable "Gene Expression Matrix" panel (B), "Principal Component Analysis" (C) and "Download" tabs for MUTR1\_1 RNA-seq experiment, as an example of gene expression profiling. F-H. Screen shot from EstroGene2.0 browser of "Experimental Metadata" (F), "Genomic Track Veiw" (G) and "Download" tabs for MUTC1\_1 ER ChIP-seq experiment, as an example of ChIP-seq profiling.

#### Please choose the model (choose a model to start):

#### Mode1: Single gene trans- and cis-level visualization

- Regulation intensity and consistency within each model for a user's input gene.
- Expressional changes from RNA-seq and microarray compared to corresponding controls.
- ER binding profiles at TSS regions with genomic track view.











Estrogen Treatment

Anti-estrogen Treatment



Long-term Estradiol Deprivation

4

ESR1 Mutation

Α

#### Mode2: Gene Signature **Enrichment Analysis**



Calculate and visualize user's input gene signature expressional change and consistency in each model

#### 3

#### Mode3: Inter-model pattern and similarity analysis



4.5

4.0

3.0

ER peak distribution at -200 to +200 kb of GREB1 TSS from 16 ChIP-seq profiles

#### Visualize regulation pattern across all five models at one time. Derive other gene

with highly similar regulation patterns.

### Mode4: ER Interactome search from **RIME** profiling

1



 Visualize ER interaction partners in 16 ER+ cancer cell lines.

Β





0.5 0.0 0.5 1.0 1.5 2.0 2.5 3 log2 Fold Change (ESR1 Mutant/WT)

20

-2.0

Modality Multi-select Cell Types Multi-select del Type Multi-select el Type Ge Multi-select Variant Type Multi-select

Ε

С

Penlic



2

Distribution of GREB1 regulatory percentile across 45 comparisons



D



Model Types CRISPR\_Cas9 Naturally\_Occurred Maior Cell Types MCF7 **T47D** Log2 Peak Intensity Scor SUM44PE -200 -150 -10 TSS Position (kb)

		#1444 1 - 48 E2F6					
		#10100 # CH + + + + E2F6					
				3 items too smail - zoom in	to view. (Dismiss)		
TepeeMasker	1.0	a a ta 11 ka 1,0000 ka 000 ka 000 ta 10 ka 10 ka 10	e di kala di kala di kala di kala di mada	in a differenti ne si di datti ne i	n it de likelen die derektet fal	d till te offentlik holdet i	entited bid it within a law bi
	0.0			347 items too smail - 200m	in to view. (Diamise)		
MOT7_WT_Vev	90						
55E126117_exp1_	-						
st.MCF7_YSS75_ dehicle							
SSE125117_exp1_ i2_MCF7_D8380_	34		1				
SEL2117_exp2.	14		المحالف معرجة مراجعه ال				
* *	٥.						
552125117_exp2_ v1_TKYD_Y5379_ fehicle	-						
102126117 exp2_	9		1				Analida
fahada			Lange and the second		alashi an ang sa kata sa	Cale and accorded, char	
T470_WT_cone DMBO_CNP-see							
55E148277_orp1_ +1_TETD_V5376_e	517		0				
ono1_DMSO_ChiP Heg SSE148277_exe1	0		L	· · · · · · · · · · · · · · · · · · ·	·····		
id TKYD D5380 tenet_DMSD_ON							
Tero WT cone	30						
55E168277_exp2_	56						
unad_DMSD_ChiP Http://www.chiP			H.				
1021148277 emp2_ 12_T47D_D55963_ iterad2_DM82_Ch4	29						
7560 55E 148277, oxp3_	250						· · · · · · · · · · · · · · · · · · ·
1_DMSO_DHPse	0		i				
552148277_eep3_ x1_MCF7_15375 Sena1_DMS0_ON	290 "						
Paug 552(48277_eep)_ d_MCF7_06380	280*					-	
Senet_DMSD_CN	0		الملك السنا				
MOT7 MT.con d DMED DiPee							
SELLEDTT ment.	260		4				
Iluna2_03800_014 7580 5562148277_0404	0.		L			-	
d MCF7 D8360 Jone2 DMS0_Chi Fae							

112008

118008

## Supplementary Figure S3 (Continued)









#### Supplementary Figure S3. (related to Fig. 2)

A. A screen shot from EstroGene2.0 browser of "ANALYSIS" tab. Red dots indicate the sections elaborated in the following panels. B-E. Screen shots from EstroGene2.0 browser of "Volcano Plot" (B) and "Percentile Plot" (C) from gene entry "GREB1" in ESR1 mutation section, and ER ChIP-seq data visualization of "TSS Region View" (D) and "Genomic Track View" (E) from from gene entry "GREB1" in ESR1 mutation section.

F. A screen shot of output from EstroGene2.0 browser of Mode2: Gene Signature Enrichment Analysis, using EstroGene metasignature as an input and the enrichment score in estrogen treatment experiments were plotted.

G and H. Screen shots of output from EstroGene2.0 browser of Mode3: Inter-model pattern and similarity analysis, using gene GREB1 as input. Output indicates the box plot view of overall regulatory percentile of each comparisons across five different sections for *GREB1* and *CCNG2* (G) and significantly similar genes shared across all four sections (H) for *GREB1*.

I. A screen shot of output from EstroGene2.0 browser of Mode4: ER Interactome search from RIME profiling, using FOXA1 as a default input. Volcano plot depicting the log2FC and -log10padj normalized to IgG control in 16 ER+ cancer cell lines.



Supplementary Figure S4 (related to Fig. 3)

A and B. Scattered plots representing the Pearson correlation of regulatory percentile between 100 nM and 1 µM fulvestrant (A) and tamoxifen (B) treatment. For tamoxifen treatment experiments, genes that more pronouncedly repressed by high and low dose were highlighted in pink and red respectively. (|delta regulatory percentile| >100 between the two models).

C. Bar plot showing the distribution of regulatory percentile of TFF3 and IRX4 in all the comparisons from 100nM and 1 µM

tamoxifen treatment. Percentiles are ranged between -100 to +100 and larger number indicates stronger regulation. D. Bar graph showing the significantly enriched Hallmark pathways in high and low tamoxifen-preferentially repressed genes indicated in B.

E. Bar plot representing the detected ER interactor numbers by RIME (log2FC>5, padj<0.05 to IgG) in number of breast cancer cell lines cumulative from 1-15.

F. Venn diagram showing the overlapping of ER interactors from RIME experiment (log2FC>5, padj<0.05 to IgG) among CAMA1, MCF7, T47D, BT483, BT474 and EFM19 cells.

G. Scattered plots representing the Pearson correlation of regulatory percentile between CAMA1 and average of MCF7, T47D, BT483, BT474 and EFM19 cells under tamoxifen treatment. Genes that are differentially regulated by CAMA1 and other five cell lines were highlighted in blue and purple respectively. (|delta regulatory percentile| >100 between the two subgorups).

H. Scattered plot showing the correlation of -log10 p values of LISA predicted regulators from differentially regulated genes between CAMA1 and MCF7, T47D, BT483, BT474 and EFM19 cells in F. Only significantly enriched regulators were shown and top targets skewed to each side were labelled.





Supplementary Figure S5 (related to Fig. 4)

A. Bar plot showing the distribution of regulatory percentile of IFI27 and IFIT1 in all the comparisons from TamR, LTED and ESR1 mutations respectively. Percentiles are ranged between -100 to +100 and larger number indicates stronger regulation. B. Bar graph showing the the alteration of GSVA enrichment scores of mean of interferon response  $\alpha$  and  $\gamma$  signatures in all the comparisons from endocrine resistant models. Color code indicate specific cell lines used.



Supplementary Figure S6 (related to Fig. 5)

A-E. Bar plot showing the distribution of regulatory percentile of indicated genes in the selected comparisons from endocrine resistant cells. Percentiles are ranged between -100 to +100 and larger number indicates stronger regulation.
F. Scattered plot showing Pearson correlation between Hallmark signature enrichment score alteration in Y537S and D538G models normalized to the corresponding WT controls.





Supplementary Figure S7 (related to Fig. 6)

A. Venn diagram showing the overlap of top consistent upregulated genes in ESR1 mutant cell models cell models (from Fig. 7A) and significantly upregulated genes in ESR1 mutant tumors from four ER+ metastatic cohorts.

B. Bar plot showing the distribution of regulatory percentile of NPY1R in the selected comparisons from ESR1 mutation cell models. Percentiles are ranged between -100 to +100 and larger number indicates stronger regulation.

C. Box plots showing the relative expression of NPY1R in genome-edited MCF7 and T47D ESR1 mutant cells and WT cells in the presence of scramble and NPY1R siRNA transfection .  $\Delta \Delta Ct$  method was used and p values were calculated using one-way ANOVA.