HORM CANC (2013) 4:222–232
DOI 10.1007/s12672-013-0140-3

ORIGINAL PAPER

# RNA Sequencing of MCF-7 Breast Cancer Cells Identifies Novel Estrogen-Responsive Genes with Functional Estrogen Receptor-Binding Sites in the Vicinity of Their Transcription Start Sites

**Ryonosuke Yamaga · Kazuhiro Ikeda · Kuniko Horie-Inoue · Yasuyoshi Ouchi · Yutaka Suzuki · Satoshi Inoue**

Received: 16 January 2013 / Accepted: 9 March 2013 / Published online: 23 March 2013


**Abstract** Estrogen receptor α (ERα) is a key transcription factor in breast cancer, which plays an essential role in the pathophysiology of the disease by regulating the expression of various target genes. In the present study, we performed deep RNA sequencing (RNA-seq) as an unbiased high-throughput technique for comprehensive transcriptome analysis in ERα-positive human breast cancer MCF-7 cells, to facilitate the elucidation of ERα regulatory gene networks. From the 17,336 mapped RefSeq genes from the sequenced fragments of the cell samples treated with estrogen time dependently, substantial numbers of sequence reads were observed in 3,386 genes (>100 tags per million reads per sample at any of the six time points studied). ERα occupancy within and in the proximal regions of the genes (<10-kb upstream and downstream regions) was significantly enriched in the subgroup of the 3,386 genes compared to the whole 17,336 RefSeq genes. Of the 3,386 genes, we focused on 29 genes, which included ERα occupancy adjacent to their transcription start sites and whose expression was estrogen dependently altered by >3-fold. Knockdown studies using siRNAs specific to the 29 genes validated that prototypic ERα targets V-myc myelocytomatosis viral oncogene homolog and cyclin D1 promote both proliferation and migration of MCF-7 cells and further identified novel candidate ERα targets EIF3A and tumor protein D52-like 1, which will also facilitate the proliferation or migration of MCF-7 cells. Taken together, the present findings provide a valuable dataset that will elucidate ERα regulatory mechanisms in breast cancer biology, based on the integrative analysis of RNA-seq combined with the genome-wide information for ERα occupancy.

**Abbreviations**

| | |
|---|---|
| CCND1 | Cyclin D1 |
| ChIP-chip | Chromatin immunoprecipitation analysis on tiling arrays |
| EIF3A | Eukaryotic translation initiation factor 3 subunit A |
| ER | Estrogen receptor |
| ERα | Estrogen receptor alpha |
| ERBS | Estrogen receptor-binding site |
| ERE | Estrogen-responsive element |
| $E_2$ | 17β-estradiol |
| FAM84B | Family with sequence similarity 84, member B |
| GREB1 | Growth regulation by estrogen in breast cancer 1 |
| IGFBP4 | Insulin-like growth factor binding protein 4 |
| MYC | V-myc myelocytomatosis viral oncogene homolog |

**Electronic supplementary material** The online version of this article (doi:10.1007/s12672-013-0140-3) contains supplementary material, which is available to authorized users.

R. Yamaga · Y. Ouchi · S. Inoue (✉)
Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bumkyo-ku, Tokyo 113-8655, Japan
e-mail: INOUE-GER@h.u-tokyo.ac.jp

R. Yamaga · S. Inoue
Department of Anti-Aging Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

R. Yamaga · K. Ikeda · K. Horie-Inoue · S. Inoue
Division of Gene Regulation and Signal Transduction, Research Center for Genomic Medicine, Saitama Medical University, Saitama, Japan

Y. Suzuki
Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

| | |
|---|---|
| MLL5 | Myeloid/lymphoid or mixed-lineage leukemia 5 |
| PGR | Progesterone receptor |
| qRT-PCR | Quantitative real-time reverse transcriptase polymerase chain reaction |
| RAB17 | Member RAS oncogene family 17 |
| RNA-seq | RNA sequencing |
| RNAi | RNA interference |
| SERBP1 | SERPINE1 mRNA binding protein 1 |
| siRNA | Small interfering RNA |
| SNP | Single nucleotide polymorphism |
| TPD52L1 | Tumor protein D52-like 1 |
| tpm | Tags per million |
| TP53 | Tumor protein p53 |
| TSS | Transcription start site |

## Introduction

Breast cancer is the most prevalent cancer in woman worldwide. The estimates of GLOBOCAN 2008 indicate that there are about 1.4 million breast cancer cases per year and about 458,000 breast cancer deaths per year in the world [1]. Breast cancer is especially common in more developed countries. In those areas, the cumulative incident risk and mortality risk of breast cancer in 75-year-old women are estimated to be 7.0 and 1.7 %, respectively [2]. Thus, overcoming breast cancer is a social problem.

The steroid hormone estrogen plays important roles in the proliferation and development of breast cancer [3, 4]. Estrogen functions by binding to its cognate receptor, estrogen receptors (ERs). Estrogen-stimulated ERs bind to estrogen-responsive elements (EREs) on the genome and regulate the transcription of their target genes that exert important roles in the proliferation and progression of breast cancer [5]. Since estrogen signaling is totally mediated by estrogen target genes that are involved in various cellular functions, including transcription, translation, cell cycle progression, and apoptosis, it is important to study those target genes to understand the mechanisms of estrogen-mediated cancer development.

Genome-wide transcriptome analysis is considered as one of the most effective approaches to elucidate the complicated estrogen signaling in breast cancer cells. Therefore, microarray analyses of human breast cancer cells have been performed to identify estrogen-responsive genes [6, 7]. However, these analyses have relied on hybridization with predesigned probes.

The recent advent of next-generation sequencing techniques has enabled high-throughput and low-biased transcriptome analysis [8–12]. Analysis of the sequencing results has yielded new insights into the genes or pathways that are involved in the development of cancers. For example, the up-regulation of ERBB2-induced signals was found in breast cancer using RNA sequencing (RNA-seq) technique [12].

In the present study, we performed RNA-seq analysis of estrogen-treated human breast cancer MCF-7 cells to reveal the time-dependent regulation of estrogen-responsive genes and to identify novel estrogen-responsive genes. By analyzing the expressional changes of RefSeq genes, we identified 869 candidate estrogen-responsive genes. Of them, 29 genes were studied because they were highly expressed, highly responsive to estrogen, and located near reported estrogen receptor-binding sites (ERBSs) [13]. These 29 genes included novel estrogen-responsive genes as well as already known ones such as growth regulation by estrogen in breast cancer 1 (GREB1) [14] and cyclin D1 (CCND1) [15, 16]. Next, we performed functional screening using small interfering RNAs (siRNAs) targeting each estrogen-responsive gene. The knockdown experiment revealed that v-myc myelocytomatosis viral oncogene homolog (MYC), CCND1, member RAS oncogene family 17 (RAB17), eukaryotic translation initiation factor 3 subunit A (EIF3A), and tumor protein D52-like 1 (TPD52L1) genes affected the proliferation or migration of MCF-7 cells. To our knowledge, RAB17, EIF3A, and TPD52L1 have been identified for the first time to be involved in estrogen regulation and function in breast cancer cells. This study provides a new insight into estrogen signaling that is associated with the proliferation or progression of breast cancer.

## Materials and Methods

### Cell Culture and Reagents

Estrogen receptor alpha (ERα)-positive human breast cancer MCF-7 cells were obtained from the American Type Culture Collection (Virginia, USA) and cultured in Dulbecco's modified Eagle's medium (DMEM; Nacalai Tesque, Kyoto, Japan) supplemented with 10 % fetal bovine serum (FBS; Nichirei Bioscience, Tokyo, Japan) at 37 °C in 5 % $CO_2$. Before ligand treatment, cells were starved under hormone-free conditions for 3 days. Hormone-free medium consisted of phenol red-free DMEM (Sigma-Aldrich, Missouri, USA) and 5 % dextran–charcoal stripped FBS. 17β-Estradiol ($E_2$) was purchased from Sigma-Aldrich. The siRNAs were purchased from Life Technologies (California, USA). Polymerase chain reaction (PCR) primers were designed using Primer Express 3.0 software (Life Technologies) and purchased from Life Technologies (Supplementary Table 1).

### RNA Extraction and RNA-seq

The hormone-starved MCF-7 cells were treated with 100 nM $E_2$ for 0, 2, 4, 8, 12, and 24 h, and the total RNA

was isolated using ISOGEN (Nippon Gene, Toyama, Japan) in accordance with the manufacturer's instructions. RNA samples were Poly(A) selected, and libraries for RNA-seq were prepared using mRNA-Seq Sample Prep Kit (Illumina, California, USA) according to the manufacturer's instructions. Single-end RNA-seq of 36-bp read length was performed using Illumina GAIIx according to the standard protocol. FASTQ sequence files were obtained, and the RNA-seq tags were aligned to the human reference genome (Human Build 36).

### Quantitative Real-Time Reverse Transcriptase PCR

Total RNA was reverse transcribed using SuperScript II reverse transcriptase (Life Technologies) as recommended by the supplier. Quantitative real-time (qRT)-PCR was carried out on StepOnePlus (Life Technologies) using the FAST SYBR Green Master Mix (Life Technologies) and 150 nM of each gene-specific forward and reverse primers (Supplementary Table 1). The cycling conditions were 95 °C for 2 min, followed by 40 cycles at 95 °C for 2 s and 60 °C for 30 s. The relative differences in PCR product amounts were determined by the comparative cycle threshold method by using glyceraldehyde-3-phosphate dehydrogenase as an internal control [17]. The experiments were carried out in triplicate.

### Cell Growth Assay

Cell growth was estimated using the MTS assay by using the CellTiter 96 AQueous One Solution Cell Proliferation (Promega, Wisconsin, USA). For this, 1,500 MCF-7 cells were seeded in 96-well plates containing 200 μL DMEM with 10 % FBS. After the cells were incubated for 24 h, 5 nM siRNAs targeting each gene or control siRNA (siControl) [18] were added to the medium by using Lipofectamine 2000 (Life Technologies) as per the manufacturer's instructions. Next, 10 μL of MTS solution was added to each well on the indicated times (0, 24, 48, 72, and 96 h). Plates were incubated for an additional 2 h at 37 °C, and the absorbance at 490 nm was read using Multiskan FC (Thermo Fisher Scientific, Massachusetts, USA). The effects of each siRNA on cell growth were estimated by comparing with the siControl. The experiments were performed in triplicate.

### Cell Migration Assay

Cell migration assay was performed using Cell Culture Insert with 8.0 μm pore size PET filter (Becton Dickinson, New Jersey, USA). Before the assay, MCF-7 cells were treated with 5 nM siRNAs with Lipofectamine 2000 for 24 h. The lower surface of the filter was immersed for 30 min in 10 μg/mL fibronectin (Sigma-Aldrich) diluted in phosphate-buffered saline (PBS). Next, 700 μL DMEM with 10 % FBS was added to the lower chamber. Subsequently, $5\times10^4$ cells were suspended in 300 μL DMEM with 10 % FCS and added to the upper chamber. After incubation for 48 h at 37 °C, the cells on the upper surface of the filter were completely removed by wiping with cotton swabs. The cells on the lower surface of the filter were fixed in methanol for 30 min, washed with PBS, and stained with Giemsa stain solution (Sigma-Aldrich) for 30 s. The filters were washed three times with PBS and mounted on a glass slide. The cells on the lower surface were counted from photographs of at least three fields obtained at a magnification of ×200 under a microscope. The effects of each siRNA on migration were estimated by comparing with siControl. The experiments were performed in triplicate.

### Luciferase Assay

Luciferase assay was performed as described previously [19] with some modifications. Briefly, 200 ng ptk-ERE-Luc plasmids and 20 ng pRL-CMV control plasmids (Promega) were used to measure the transcriptional activity of ERs. Transient transfections were carried out using Lipofectamine 2000 with 5 nM siRNAs in MCF-7 cells. After the cells were incubated for 24 h with hormone-free medium containing 100 nM $E_2$ or vehicle, the luciferase activity was measured. The experiments were performed in triplicate.

## Results

### Identification of Estrogen-Responsive Genes Using RNA-seq

To identify novel estrogen-responsive genes associated with the biology of breast cancer cells, we applied transcriptome sequencing for breast cancer MCF-7 cells before and after the treatment with $E_2$. From billions of sequence fragments from six lanes of single-read sequencing performed on the Genome Analyzer IIx, 91 million mapped reads were recovered after performing short-read gapped alignment to the Human Genome NCBI Build 36, with a median of 9.9 million mapped reads per sample (Fig. 1a). Of the 91 million mapped reads, the majority (86.5 %) were located within RefSeq genes, including 73.9 % for exonic regions and 12.6 % for intronic regions. A substantial percentage of reads (13.5 %) lacked any overlap with RefSeq genes and were designated intergenic. The relative ratios for sequence reads mapped to exons, introns, and intergenic regions of RefSeq genes were almost maintained for every sample during the time course of estrogen treatment (Fig. 1a).

The expression levels of genes were quantified by counting the number of reads mapped to each RefSeq gene

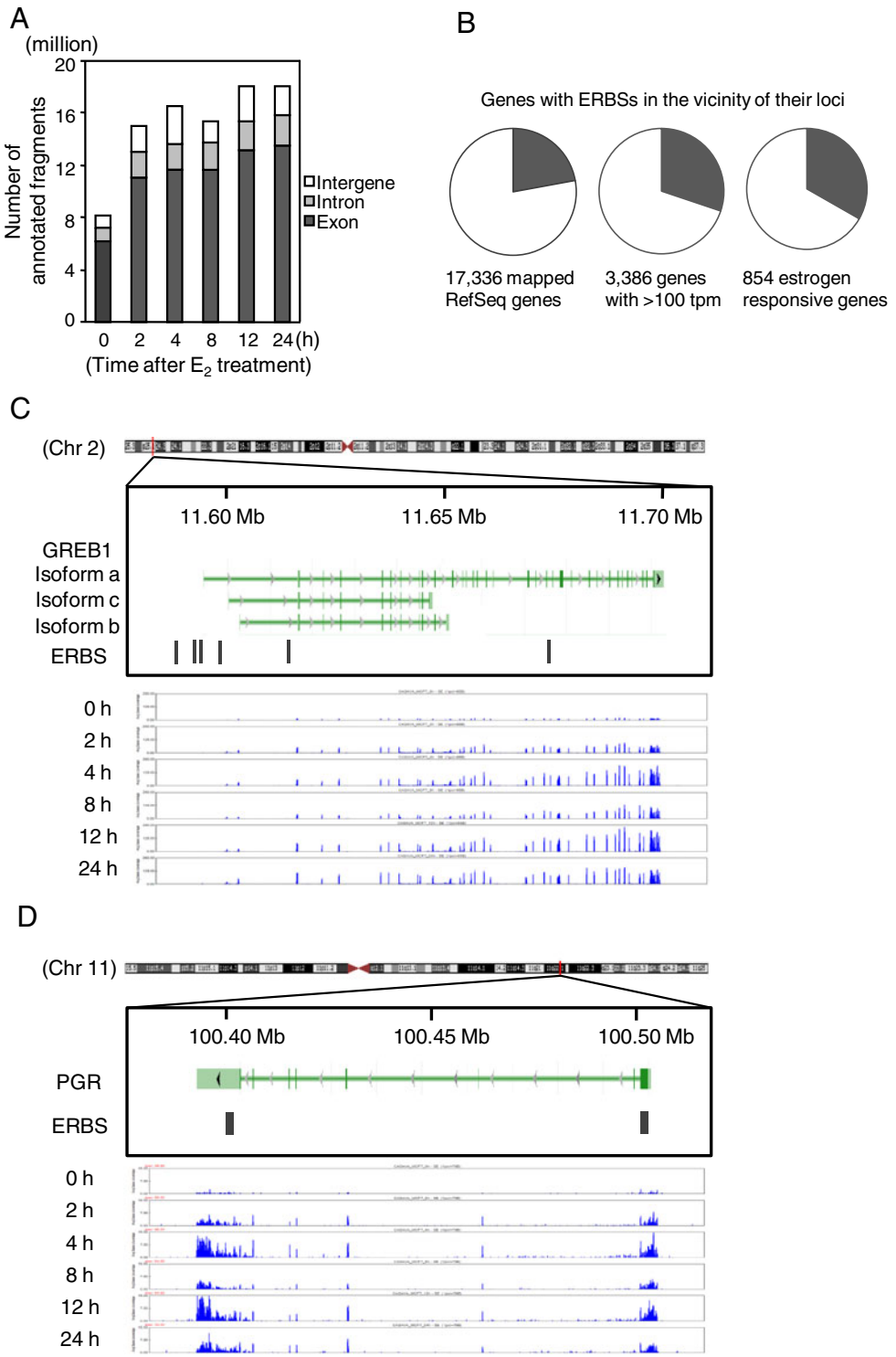


**Fig. 1** Alteration in the gene expression level by 17β-estradiol ($E_2$). **a** Annotation of RNA fragments at each time point after $E_2$ treatment. RNA-seq was performed using RNAs prepared from MCF-7 cells that were treated with $E_2$ for 0, 2, 4, 8, 12, and 24 h. Of all mapped fragments, $6.73\times10^7$ (73.9 %) fragments were mapped onto exons, $11.5\times10^7$ (12.6 %) were mapped onto introns, and $1.22\times10^7$ (13.4 %) were mapped onto the regions where no RefSeq genes were detected (Intergene). **b** Venn diagram of $E_2$-regulated RefSeq genes (fold change, >2). Genes whose expression levels are low (less than 1 tpm at any time point) were excluded and, of the 3,439 genes, 776 (22.6 %) were up-regulated and 92 (2.7 %) were down-regulated by $E_2$. Only one gene was both up- and down-regulated. **c** GREB1, a known estrogen-responsive gene, was found to be up-regulated by $E_2$ treatment. *Black bars* represent reported estrogen receptor-binding sites (*ERBSs*) [13], which are located within 10 kb from GREB1. **d** PGR, a known estrogen-responsive gene, was also found to be up-regulated by $E_2$ treatment. *Black bars* represent reported ERBSs located within 10 kb from PGR

and determining "tags per million (tpm)" for each gene by normalization to the total mapped reads for each sample. Of the total 17,336 mapped RefSeq genes, 3,386 genes (19.5 %) were abundantly expressed in the MCF-7 cells [>100 tpm as maximal expression (max-tpm) for at least any of the six samples]. Of the 3,386 genes, 854 were identified as estrogen-responsive genes, whose expressions were altered by >2-fold at least at any time point after estrogen treatment compared to the sample at 0 h. Of the estrogen-responsive genes, 765 genes were up-regulated, including prototypic estrogen-responsive GREB1 and progesterone receptor (PGR; Fig. 1c), and 90 were down-regulated in response to estrogen. One of the 765 genes was both up-regulated and down-regulated by >2-fold.

In terms of ERα binding in the vicinity of RefSeq genes, 3,833 of the 17,335 annotated genes (22.1 %) included significant ERBSs determined by chromatin immunoprecipitation analysis on tiling arrays (ChIP-chip, at a threshold of $P<$ 1e−3 by MAT algorithm) [13] within the proximal regions and gene loci (<10 kb from the both ends of the genes and within the gene loci) of the genes. Of the 3,386 genes with max-tpm of >100 tpm, 1,021 genes (30.2 %) included ERBSs within and in the proximal regions of the genes, the percentage of which was significantly higher than that of the 17,335 mapped RefSeq genes ($P$=2.3e−14 by Fisher's exact test). Of the 854 estrogen-responsive genes, 284 genes (33.3 %) included ERBSs within and in the proximal regions of the genes, the percentage of which was also significantly higher than that of the 17,335 mapped RefSeq genes ($P$=2.3e−8; Fig. 1d).

## Selection of Estrogen Target Genes with Substantial Signals and ERα Binding in Their Vicinities

Of the 855 genes, 37 were selected as prominent ERα-regulated genes, by the criteria of >100 max-tpm, estrogen-dependent change of gene expression by >3-fold, and the involvement of ERBSs [13] within the gene loci or <10 kb from both ends of the genes. Of the 37 genes, 29 showed substantial responsiveness to estrogen as revealed by quantitative real-time reverse transcriptase polymerase chain reaction (qRT-PCR). Of the 29 genes, most of the genes were up-regulated by 24 h after $E_2$ treatment, whereas PKP1 was once down-regulated at 4–12 h and then up-regulated at 24 h after $E_2$ treatment (Fig. 2). Within the proximal regions and the gene loci of the 29 ERα-regulated genes (<10 kb from both ends of the genes and within the gene loci), 63 ERBSs were involved (Table 1) as per the published ChIP-chip dataset [13]. Notably, more than one third of the genes included ≥3 ERBSs within the regions. The average number of ERBSs involved per gene was 2.2 of the subgroup, which is substantially larger than the expected number of ERBSs divided by the total number of transcripts involved in RefSeq database, since the latter will be 0.61 per transcript ($P$=4.0e−9 by Fisher's exact test). All the 29 genes notably included at least one estrogen receptor-binding site (ERBS) in their proximal 5′ regions or within their genes from the transcription start sites (TSSs) to their second exons. Therefore, functional ERα binding in the vicinity of TSSs might contribute to the substantial estrogen responsiveness of the genes. Examples for the expressions of sequence reads and ERBS mapping are shown for myeloid/lymphoid or mixed-lineage leukemia 5 (MLL5), Family with sequence similarity 84, member B (FAM84B), insulin-like growth factor binding protein 4 (IGFBP4), and CCND1, which include ERBSs in the 5′ region, exon 2, intron 1, and multiple locations, respectively (Fig. 3).

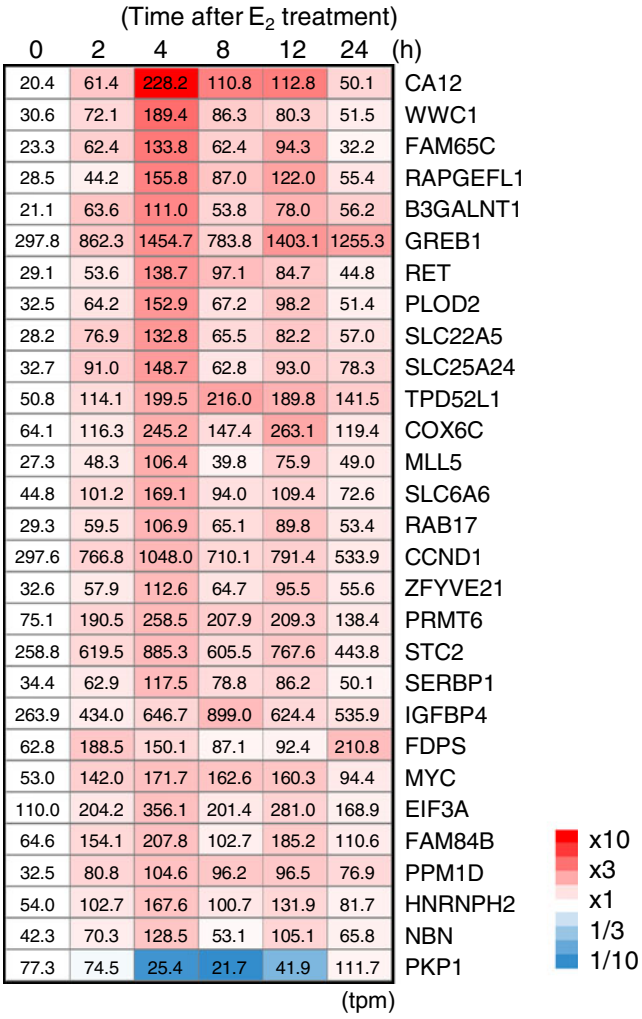


**Fig. 2** Estrogen-regulated genes validated by qRT-PCR. Of the 869 genes that were detected as candidate estrogen-responsive genes in RNA-seq analysis, 29 were located within 10 kb from reported ERBSs and validated for their estrogen responsiveness by qRT-PCR. *Numbers* represent tpm at each time point, and *colors* represent relative expression levels compared to those at 0 h in RNA-seq

## Identification of Estrogen-Responsive Genes from the siRNA-Mediated Knockdown Study

Next, loss-of-function study was conducted for the 29 estrogen-regulated genes in MCF-7 cells. The individual knockdown effects of the 29 genes together with ERα (ESR1) and tumor protein p53 (TP53) were evaluated by gene expression, cell growth, cell migration, and ERα-dependent luciferase activity by using MCF-7 cells (Fig. 4). The reduction of gene expression levels was achieved by >40 % with the siRNA treatment (Fig. 4a). Under the condition where siRNAs targeting ESR1 and TP53 exhibited negative and positive effects on the cell growth, respectively, siRNAs targeting MYC, EIF3A, and CCND1 significantly attenuated the growth of MCF-7 cells (Fig. 4b). In terms of trans-well migration assay, siRNAs specific for RAB17, TPD52L1,

**Table 1** Estrogen-regulated genes nearby estrogen receptor-binding sites (ERBSs) detected by RNA-Seq

| Gene symbol | Chromosome | Gene; start–end | ERBS; start–end | Location of ERBS |
|---|---|---|---|---|
| CA12 | 15 | 61,402,783–61,461,128 | 61,460,991–61,461,147 | 1st intron |
| WWC1 | 5 | 167,651,643–167,831,886 | 167,674,684–167,676,601 | 1st intron |
| WWC1 | 5 | 167,651,643–167,831,886 | 167,676,966–167,677,566 | 1st intron |
| WWC1 | 5 | 167,651,643–167,831,886 | 167,725,157–167,725,757 | 1st intron |
| WWC1 | 5 | 167,651,643–167,831,886 | 167,761,427–167,762,028 | 5th intron |
| FAM65C | 20 | 48,636,053–48,686,833 | 48,695,439–48,696,039 | 5′ |
| FAM65C | 20 | 48,636,053–48,686,833 | 48,674,415–48,675,016 | 2nd intron |
| RAPGEFL1 | 17 | 35,587,768–35,605,432 | 35,590,047–35,591,038 | 1st intron |
| B3GALNT1 | 3 | 162,284,365–162,305,854 | 162,300,922–162,302,197 | 5′ |
| B3GALNT1 | 3 | 162,284,365–162,305,854 | 162,307,822–162,308,422 | 5′ |
| GREB1 | 2 | 11,591,693–11,700,363 | 11,581,710–11,582,310 | 5′ |
| GREB1 | 2 | 11,591,693–11,700,363 | 11,588,198–11,588,845 | 5′ |
| GREB1 | 2 | 11,591,693–11,700,363 | 11,589,526–11,590,641 | 5′ |
| GREB1 | 2 | 11,591,693–11,700,363 | 11,596,942–11,598,236 | 1st intron |
| GREB1 | 2 | 11,591,693–11,700,363 | 11,616,036–11,617,283 | 2nd intron |
| GREB1 | 2 | 11,591,693–11,700,363 | 11,675,420–11,676,021 | 22th intron |
| RET | 10 | 42,892,523–42,945,803 | 42,897,495–42,898,272 | 1st intron |
| RET | 10 | 42,892,523–42,945,803 | 42,901,726–42,902,327 | 1st intron |
| RET | 10 | 42,892,523–42,945,803 | 42,920,697–42,921,297 | 3rd intron |
| RET | 10 | 42,892,523–42,945,803 | 42,925,909–42,926,923 | 6th intron |
| PLOD2 | 3 | 147,269,918–147,361,972 | 147,360,648–147,361,249 | 1st intron |
| PLOD2 | 3 | 147,269,918–147,361,972 | 147,340,190–147,340,956 | 1st intron |
| PLOD2 | 3 | 147,269,918–147,361,972 | 147,315,702–147,316,486 | 3rd intron |
| PLOD2 | 3 | 147,269,918–147,361,972 | 147,294,311–147,294,911 | 7th intron |
| SLC22A5 | 5 | 131,733,300–131,759,205 | 131,734,834–131,735,434 | 1st intron |
| SLC22A5 | 5 | 131,733,300–131,759,205 | 131,735,877–131,736,478 | 1st intron |
| SLC22A5 | 5 | 131,733,300–131,759,205 | 131,741,672–131,742,948 | 1st intron |
| SLC22A5 | 5 | 131,733,300–131,759,205 | 131,751,572–131,752,172 | 4th intron |
| SLC25A24 | 1 | 108,478,867–108,544,503 | 108,543,328–108,544,120 | 1st exon |
| TPD52L1 | 6 | 125,516,578–125,626,343 | 125,517,463–125,518,564 | 1st intron |
| TPD52L1 | 6 | 125,516,578–125,626,343 | 125,560,464–125,561,450 | 1st intron |
| TPD52L1 | 6 | 125,516,578–125,626,343 | 125,564,620–125,565,541 | 1st intron |
| TPD52L1 | 6 | 125,516,578–125,626,343 | 125,566,118–125,566,843 | 1st intron |
| TPD52L1 | 6 | 125,516,578–125,626,343 | 125,567,624–125,568,902 | 1st intron |
| COX6C | 8 | 100,959,399–100,975,418 | 100,974,747–100,975,347 | 1st exon |
| MLL5 | 7 | 104,441,873–104,541,768 | 104,434,034–104,434,634 | 5′ |
| SLC6A6 | 3 | 14,419,110–14,505,861 | 14,421,920–14,423,302 | 1st intron |
| SLC6A6 | 3 | 14,419,110–14,505,861 | 14,429,405–14,430,006 | 1st intron |
| SLC6A6 | 3 | 14,419,110–14,505,861 | 14,448,627–14,449,673 | 2nd intron |
| RAB17 | 2 | 238,147,704–238,164,508 | 238,163,797–238,164,398 | 1st exon |
| CCND1 | 11 | 69,165,054–69,178,423 | 69,162,760–69,163,409 | 5′ |
| CCND1 | 11 | 69,165,054–69,178,423 | 69,172,899–69,173,500 | 4th intron |
| CCND1 | 11 | 69,165,054–69,178,423 | 69,177,825–69,179,657 | 3′ |
| ZFYVE21 | 14 | 103,251,834–103,269,758 | 103,246,745–103,247,346 | 5′ |
| ZFYVE21 | 14 | 103,251,834–103,269,758 | 103,261,809–103,262,410 | 1st intron |
| ZFYVE21 | 14 | 103,251,834–103,269,758 | 103,272,261–103,273,347 | 3′ |
| PRMT6 | 1 | 107,400,790–107,403,439 | 107,401,263–107,401,863 | 1st exon |
| STC2 | 5 | 172,674,332–172,689,112 | 172,689,289–172,690,644 | 5′ |
| STC2 | 5 | 172,674,332–172,689,112 | 172,687,670–172,688,679 | 1st intron |

**Table 1** (continued)

| Gene symbol | Chromosome | Gene; start–end | ERBS; start–end | Location of ERBS |
|---|---|---|---|---|
| SERBP1 | 1 | 67,646,081–67,668,711 | 67,668,770–67,669,370 | 5′ |
| IGFBP4 | 17 | 35,853,202–35,867,508 | 35,858,051–35,859,092 | 1st intron |
| FDPS | 1 | 153,545,163–153,557,081 | 153,546,058–153,546,658 | 1st intron |
| MYC | 8 | 128,817,497–128,822,862 | 128,819,702–128,820,399 | 2nd exon |
| EIF3A | 10 | 120,784,531–120,830,324 | 120,830,505–120,831,106 | 5′ |
| FAM84B | 8 | 127,633,865–127,639,893 | 127,638,788–127,639,388 | 2nd exon |
| PPM1D | 17 | 56,032,326–56,098,422 | 56,031,582–56,032,271 | 5′ |
| HNRNPH2 | X | 100,549,777–100,555,784 | 100,549,141–100,549,988 | 5′ |
| NBN | 8 | 91,014,740–91,066,075 | 91,065,110–91,065,711 | 1st intron |
| NBN | 8 | 91,014,740–91,066,075 | 91,035,534–91,036,517 | 10th intron |
| PKP1 | 1 | 199,519,203–199,568,744 | 199,517,666–199,518,436 | 5′ |
| PKP1 | 1 | 199,519,203–199,568,744 | 199,521,801–199,522,695 | 1st intron |
| PKP1 | 1 | 199,519,203–199,568,744 | 199,525,081–199,525,991 | 1st intron |
| PKP1 | 1 | 199,519,203–199,568,744 | 199,543,550–199,544,385 | 2nd intron |

Estrogen-regulated genes were detected within 10 kb from reported ERBSs [13]. Positions of the genes and ERBSs were represented in hg18

MYC, EIF3A, and CCND1 significantly repressed the numbers of migrated cells (Fig. 4c). The ERα-dependent luciferase activity was reduced by ESR1 siRNA, whereas the rest of the tested siRNAs did not exhibit any significant effects on ERα-dependent luciferase activity in the MCF-7 cells (Fig. 4d). Similar to the prototypic ERα target genes such as MYC and CCND1, the data indicate that RAB17, TPD52L1, and EIF3A are new candidate genes that promote the growth or migration of ERα-positive breast cancer cells. Further detailed results for EIF3A on chromosome 10q26 and TPD52L1 on chromosome

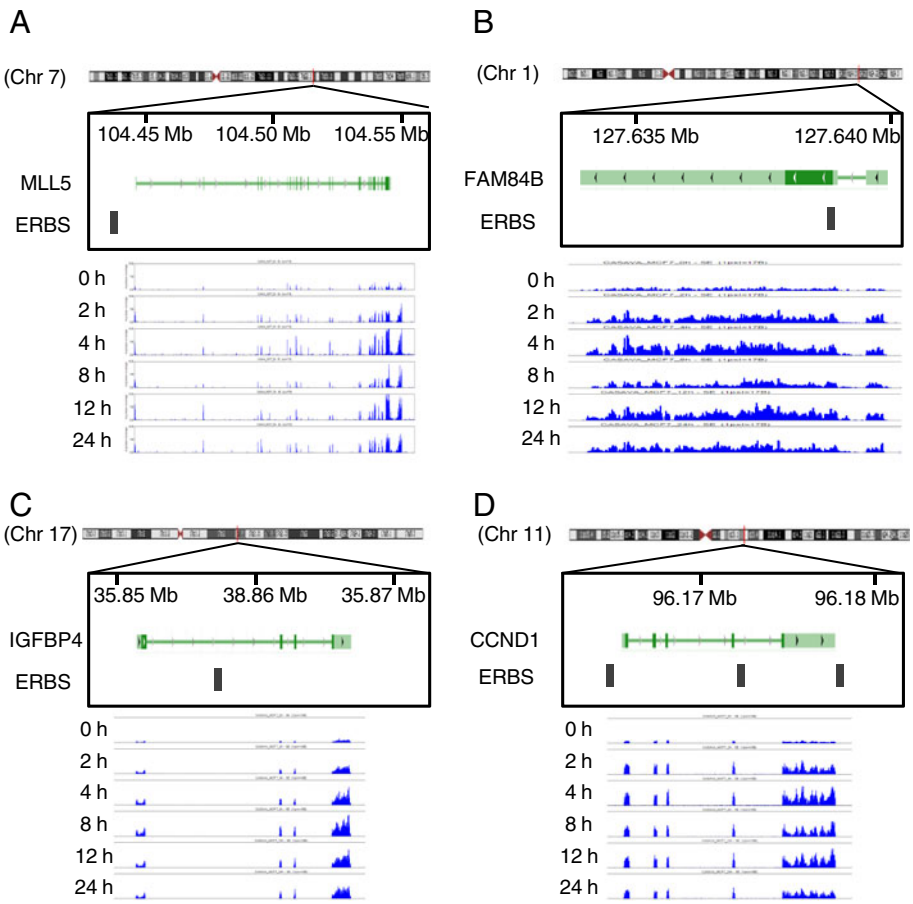


**Fig. 3** Locational relationship between estrogen-responsive genes and ERBS. **a** An ERBS was located in the upstream of MLL5. *Lower panel* shows the result of RNA-seq, indicating that MLL5 is up-regulated by $E_2$. **b** An ERBS was located in the second exon of the FAM84B. *Lower panel* shows the result of RNA-seq, indicating that FAM84B is up-regulated by $E_2$. **c** An ERBS was located in the first intron of insulin-like growth factor binding protein 4 (*IGFBP4*). *Lower panel* shows the result of RNA-seq, indicating that IGFBP4 is up-regulated by $E_2$. **d** ERBSs were located in the upstream, fourth intron, and downstream of cyclin D1 (*CCND1*). *Lower panel* shows the result of RNA-seq, indicating that CCND1 is up-regulated by $E_2$

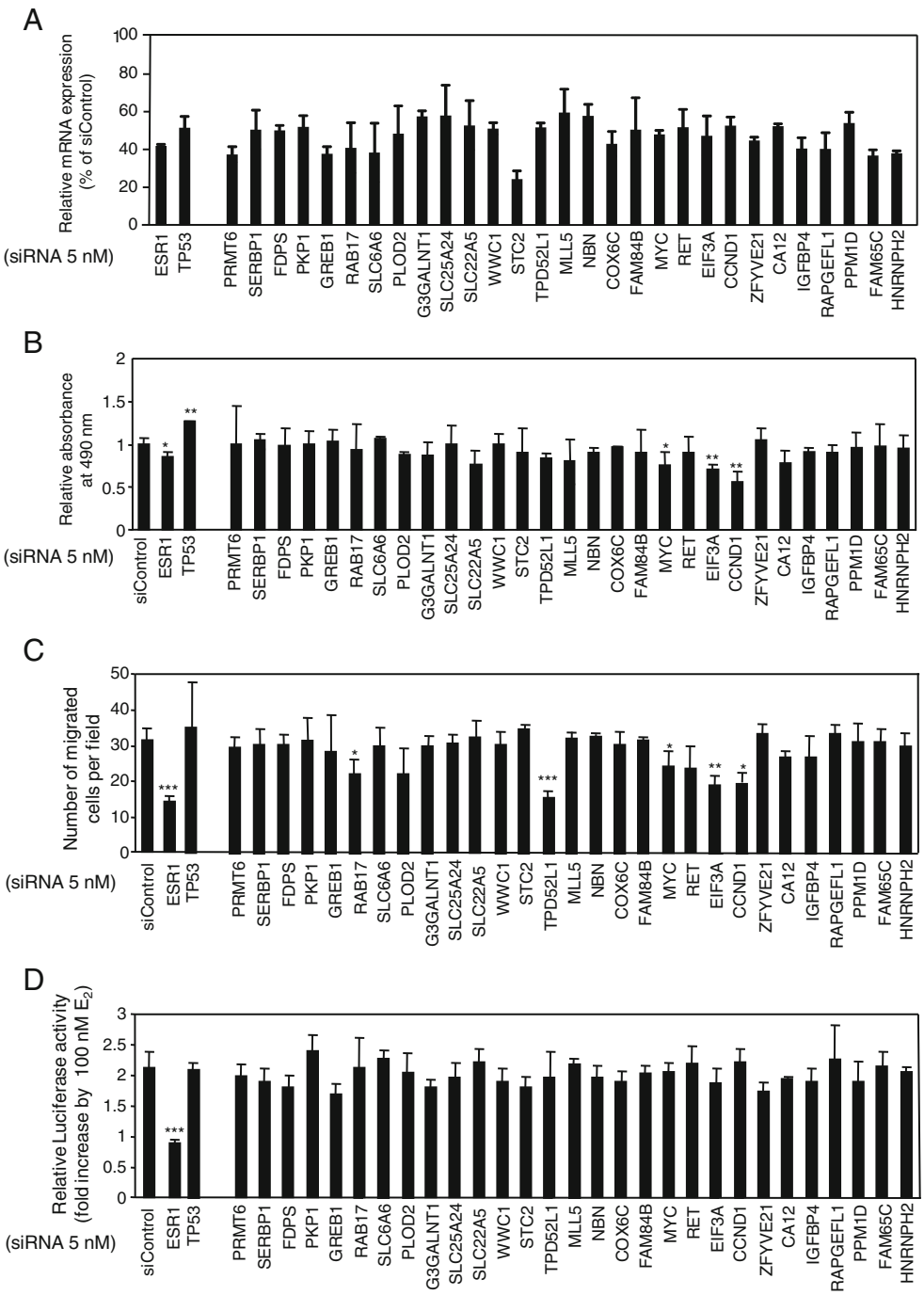


**Fig. 4** Knockdown analysis of novel estrogen-responsive genes by siRNA. **a** Knockdown effect of each siRNA against targeting genes. MCF-7 cells were transfected with 5 nM siRNA that target the $E_2$-reponsive candidate genes and siControl with Lipofectamine 2000 for 48 h. Gene expression levels with siRNA were detected by qRT-PCR and normalized with siControl. **b** Assessment of cell growth using MTS assay. Cell viability was determined using MTS assay. MCF-7 cells were transfected with 5 nM siRNA specific for the indicated genes and control siRNA for 4 days. Results are represented as mean ±SD of three experiments. Fold change of absorbance at 490 nm of each siRNA sample was normalized with that of siControl sample. **c** Trans-well migration assay. MCF-7 cells transfected with each siRNA were incubated for 24 h and plated to uncoated filters for 48 h. The migrating cells were fixed, stained, and counted for three fields of the filter. **d** Assay for ER-mediated transcription. ERE-luciferase reporter assay. ERE-luciferase reporter plasmid and 5 nM siRNA were transfected in MCF-7 cells with or without 100 nM $E_2$ for 24 h. Fold change of luciferase activity in response to $E_2$ was calculated for each siRNA. Student's *t* test was performed between each siRNA sample and siControl sample (*$P$<0.05, **$P$<0.01, ***$P$<0.005)

6q22–q23 are shown in terms of mRNA expression, cell growth, and cell migration (Figs. 5 and 6). EIF3A functions positively in both cell proliferation and migration (Fig. 5d, e), whereas TPD52L1 functions positively in cell migration, but not in cell proliferation (Fig. 6d, e).

## Discussion

To elucidate the mechanism underlying the estrogen-mediated development and progression of breast cancer, we explored the genome-wide transcriptional network of MCF-7 cells by RNA-seq using next-generation sequencer and combining functional experiments using siRNAs. Our screening system using RNA-seq and siRNAs effectively identified the genes significantly involved in the growth/migration of breast cancer cells. These genes included *MYC* and *CCND1*. MYC is a transcription factor that is directly up-regulated by $E_2$ and ERα and modulates the transcription of target genes, including members of the cyclin family, and exerts oncogenic action [15, 16]. Knockdown of MYC by RNAi has been reported to inhibit MCF-7

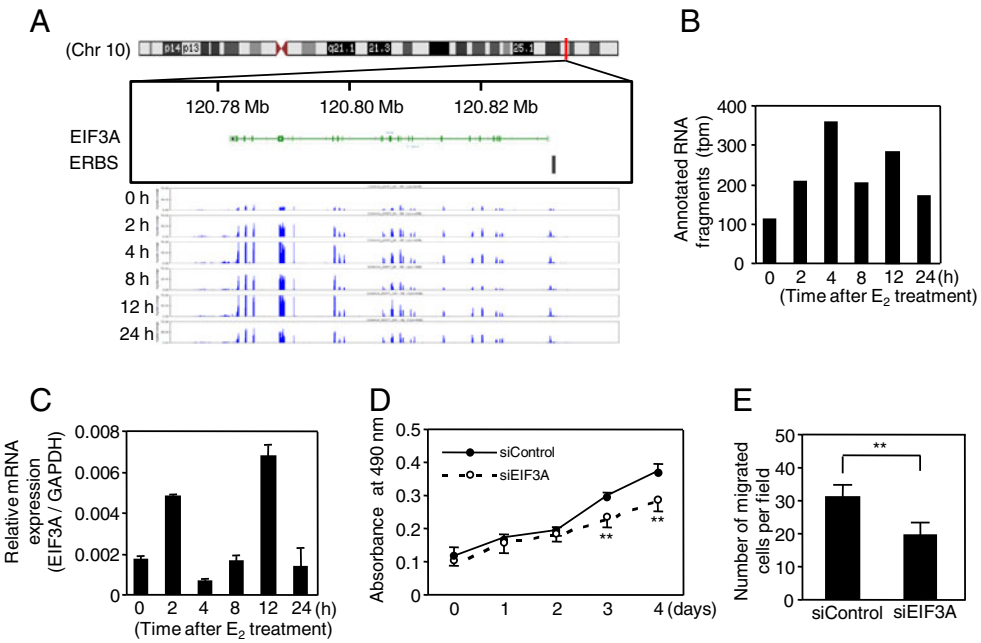


**Fig. 5** EIF3A is a novel estrogen-responsive gene that affects the growth and migration of MCF-7 cells. **a** Genome view shows RNA fragments mapped on EIF3A. **b** Quantification of mapped fragments onto EIF3A indicated the $E_2$ responsiveness. **c** qRT-PCR confirmed the $E_2$ responsiveness of EIF3A. **d** MTS assay indicated that the growth of MCF-7 cells was decreased by the knockdown of EIF3A. **e** Trans-well assay indicated that the migration of MCF-7 cells was decreased by the knockdown of EIF3A. Student's *t* test was performed between each siRNA sample and siControl sample (**$P<0.01$)

cell growth in vitro and in vivo [20]. Transcription of *CCND1* is also directly up-regulated by $E_2$ and ER$\alpha$. CCND1 is known to be an oncogene that activates the cyclin-dependent kinase 4/6 and controls the G1-S transition in the cell cycle [21, 22]. The expression level of CCND1 has been reported to be associated with poor prognosis in estrogen receptor (ER)-positive breast cancer [23]. Therefore, our screening system had a high degree of accuracy in identifying significant genes associated with the growth and/or migration of breast cancer cells.

In this study, we identified three novel estrogen-responsive genes, *EIF3A*, *TPD52L1*, and *RAB17*. EIF3A is a subunit of the eukaryotic translation initiation factor 3 that plays a role in translational regulation and cell growth. EIF3A is suspected to be required for cell proliferation and tissue development, and the expression of EIF3A is reported to be elevated in many human cancers, in particular, cancer of the breast, cervix, esophagus, lung, and stomach [24, 25]. An SNP in the *EIF3A* gene is reported to be associated with a risk of breast cancer [26]. In our study, *EIF3A* was recognized as an estrogen-responsive gene and found to be involved in the growth and/or migration of MCF-7cells. This result is consistent with those of previous reports suggesting that EIF3A may act as an oncogenic in human breast cancer [25]. These findings provide a new insight into the ER-mediated signals, and analysis of the function of EIF3A in

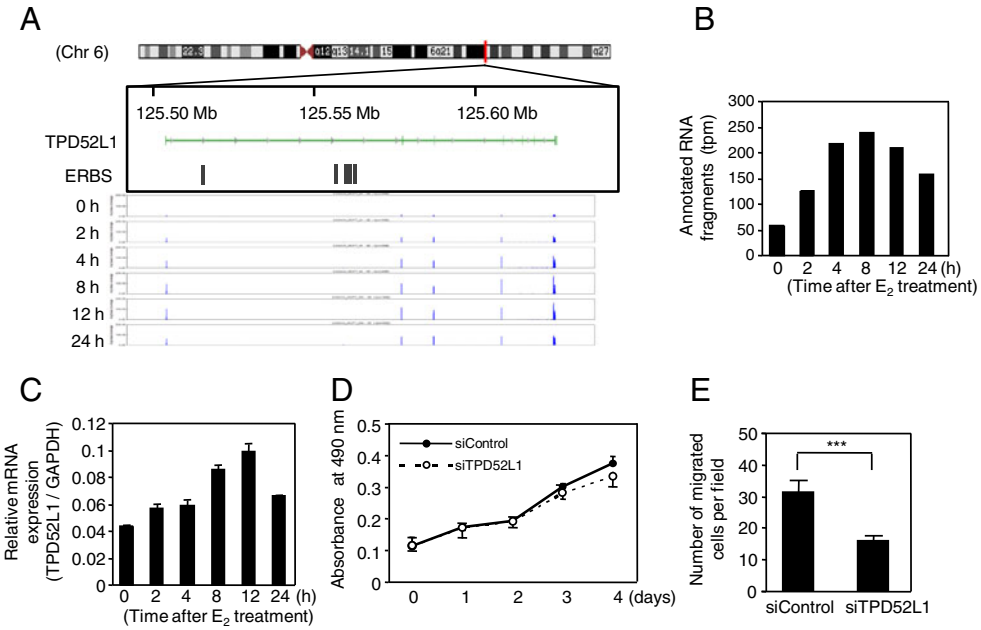


**Fig. 6** TPD52L1 is also a novel estrogen-responsive gene that affects the migration of MCF-7 cells. **a** Genome view shows RNA fragments mapped on TPD52L1. **b** Quantification of mapped fragments onto TPD52L1 indicated the $E_2$ responsiveness. **c** qRT-PCR confirmed the $E_2$ responsiveness of TPD52L1. **d** MTS assay indicated that the growth of MCF-7 cells was not affected by the knockdown of TPD52L1. **e** Trans-well assay indicated that the migration of MCF-7 cells was decreased by the knockdown of TPD52L1. Student's *t* test was performed between each siRNA sample and siControl sample (***$P<0.005$)

breast cancer might clarify the mechanisms of estrogen-mediated tumor development. *TPD52L1* encodes a member of tumor protein D52 family that contains a coiled-coil domain. The protein may form a homo- or hetero-dimer with TPD52 family members and is reported to be involved in cell proliferation. TPD52L1 is supposed to be involved in cell cycle, since it binds to the cell cycle-related 14-3-3 family proteins and is expressed in the G2-M phase [27, 28]. TPD52L1 has been reported to likely be positively associated with lymph node metastasis in human breast cancers [29]; however, the mechanisms underlying this association are not known. In the present study, TPD52L1 was identified as an estrogen-responsive gene and was found to be concerned with the migration of MCF-7 cells. These data suggest that TPD52L1 mediates the migration of breast cancer cells. Furthermore, we found that RAB17, a small GTPase, is transcriptionally regulated by estrogen and is involved in the migration of breast cancer cells. RAB17 was previously reported as a regulator of intracellular transport and was thought to regulate membrane trafficking through the apical recycling endosome [30]. However, the association between RAB17 and cancers is not known. The members of the RAS oncogene family, RAB11a, RAB25, and RAB27b, have already been shown to have roles in the development or progression in breast cancer [31–33]. In our study, RAB17 was suggested to have a role in breast cancer cells.

We compared the estrogen-regulated profiles (>2- or <0.5-folds) in MCF-7 cells identified by our RNA-seq analysis and by a previous microarray study [34]. In terms of up-regulated genes, 765 and 132 were identified by our RNA-seq and the microarray analyses, respectively. Only 11 up-regulated genes were overlapped between the studies including CCND1, proliferating cell nuclear antigen (PCNA), IGFBP4, stanniocalcin 2 (STC2), retinoic acid receptor, alpha (RARA), RAB31, adenylate cyclase 9, GADD45B (growth arrest and DNA-damage-inducible, beta), nuclear receptor interacting protein 1 (NRIP1), MYC, and MYB. Besides prototypic estrogen-regulated genes CCND1 and MYC, IGFBP4 [35, 36], PCNA [37], STC2 [38], RARA [39], RAB31 [40], NRIP1 [41], and MYB [42] have been reported in conjunction with the estrogen-regulated gene expression and/or breast cancer biology. Nevertheless, it is interesting that the rest of estrogen-regulated genes determined either by RNA-seq or microarray are unique even in a single breast cancer cell line. No overlapped gene was observed in comparison between 90 and 306 down-regulated genes by RNA-seq and microarray studies, respectively. The difference of profiles may be due to the sensitivity of each technique, as RNA-seq would be more powerful to identify the alteration of gene expression at lower level compared to microarray study. Taken together, deep sequencing technology will contribute to the discovery of various estrogen-regulated genes in breast cancer cells, which have not been previously revealed by microarray study or other conventional expression analyses.

In summary, RNA-seq and subsequent knockdown screening using siRNA were useful for clarifying the estrogen signaling throughout the genome in breast cancer cells. This approach can provide a new insight into the transcriptional regulation and physiologically significant genes and would help in the diagnosis and treatment of breast cancer.

**Acknowledgments** This work was supported by Grants of the Cell Innovation Program (to S.I.), Grants-in-Aid (to S.I.), and Support Project of Strategic Research Center in Private Universities (to S.I.) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan; by Grants (to S.I., T.U., and S.T.) from the Japan Society for the Promotion of Science, Japan; by Grants-in-Aid (to S.I.) from the Ministry of Health, Labour, and Welfare, Japan; and by the Program for Promotion of Fundamental Studies in Health Sciences (to S.I.), the Advanced Research for Medical Products Mining Programme of the National Institute of Biomedical Innovation, Japan.

## References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127:2893–2917
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. CA Cancer J Clin 61:69–90
3. Feigelson HS, Henderson BE (1996) Estrogens and breast cancer. Carcinogenesis 17:2279–2284
4. Colditz GA (1998) Relationship between estrogen levels, use of hormone replacement therapy, and breast cancer. J Natl Cancer Inst 90:814–823
5. Klinge CM (2001) Estrogen receptor interaction with estrogen response elements. Nucleic Acids Res 29:2905–2919
6. Mo R, Tony Zhu Y, Zhang Z, Rao SM, Zhu YJ (2007) GAS6 is an estrogen-inducible gene in mammary epithelial cells. Biochem Biophys Res Commun 353:189–194
7. Inoue A, Omoto Y, Yamaguchi Y, Kiyama R, Hayashi SI (2004) Transcription factor EGR3 is involved in the estrogen-signaling pathway in breast cancer cells. J Mol Endocrinol 32:649–661
8. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63
9. Timmermann B, Kerick M, Roehr C, Fischer A, Isau M, Boerno ST, Wunderlich A, Barmeyer C, Seemann P, Koenig J, Lappe M, Kuss AW, Garshasbi M, Bertram L, Trappe K, Werber M, Herrmann BG, Zatloukal K, Lehrach H, Schweiger MR (2010) Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. PLoS One 5(12):e15661
10. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordóñez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen L, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia M, Kingsbury Z, Leroy C, Marshall J, Menzies A, Mudie LJ, Ning Z, Royce T, Schulz-Trieglaff OB, Spiridou A, Stebbings LA, Szajkowski L, Teague J, Williamson D, Chin L, Ross MT, Campbell PJ, Bentley DR, Futreal PA, Stratton MR (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463:191–196

11. Huang Q, Lin B, Liu H, Ma X, Mo F, Yu W, Li L, Li H, Tian T, Wu D, Shen F, Xing J, Chen ZN (2011) RNA-Seq analyses generate comprehensive transcriptomic landscape and reveal complex transcript patterns in hepatocellular carcinoma. PLoS One 6(10):e26168
12. Carraro DM, Ferreira EN, de Campos MG, Puga RD, Abrantes EF, Trapé AP, Eckhardt BL, Nunes DN, Brentani MM, Arap W, Pasqualini R, Brentani H, Dias-Neto E, Brentani RR (2011) Poly (A) + transcriptome assessment of ERBB2-induced alterations in breast cell lines. PLoS One 6(6):e21022
13. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, Fox EA, Silver PA, Gingeras TR, Liu XS, Brown M (2006) Genome-wide analysis of estrogen receptor binding sites. Nat Genet 38:1289–1297
14. Deschênes J, Bourdeau V, White JH, Mader S (2007) Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences. J Biol Chem 282:17335–17339
15. Nass SJ, Dickson RB (1997) Defining a role for c-Myc in breast tumorigenesis. Breast Cancer Res Treat 44:1–22
16. Liao DJ, Dickson RB (2000) c-Myc in breast cancer. EndocrRelat Cancer 7:143–164
17. Takayama K, Kaneshiro K, Tsutsumi S, Horie-Inoue K, Ikeda K, Urano T, Ijichi N, Ouchi Y, Shirahige K, Aburatani H, Inoue S (2007) Identification of novel androgen response genes in prostate cancer cells by coupling chromatin immunoprecipitation and genomic microarray analysis. Oncogene 26:4453–4463
18. Ueyama K, Ikeda K, Sato W, Nakasato N, Horie-Inoue K, Takeda S, Inoue S (2010) Knockdown of Efp by DNA-modified small interfering RNA inhibits breast cancer cell proliferation and in vivo tumor growth. Cancer Gene Ther 17:624–632
19. Watanabe T, Inoue S, Ogawa S, Ishii Y, Hiroi H, Ikeda K, Orimo A, Muramatsu M (1997) Agonistic effect of tamoxifen is dependent on cell type, ERE-promoter context, and estrogen receptor subtype: functional difference between estrogen receptors alpha and beta. Biochem Biophys Res Commun 236:140–145
20. Wang YH, Liu S, Zhang G, Zhou CQ, Zhu HX, Zhou XB, Quan LP, Bai JF, Xu NZ (2005) Knockdown of c-Myc expression by RNAi inhibits MCF-7 breast tumor cells growth in vitro and in vivo. Breast Cancer Res 7:R220–R228
21. Arnold A, Papanikolaou A (2005) Cyclin D1 in breast cancer pathogenesis. J Clin Oncol 23:4215–4224
22. Roy PG, Thompson AM (2006) Cyclin D1 and breast cancer. Breast 15:718–727
23. Aaltonen K, Amini RM, Landberg G, Eerola H, Aittomäki K, Heikkilä P, Nevanlinna H, Blomqvist C (2009) Cyclin D1 expression is associated with poor prognostic features in estrogen receptor positive breast cancer. Breast Cancer Res Treat 113:75–82
24. Dong Z, Liu Z, Cui P, Pincheira R, Yang Y, Liu J, Zhang JT (2009) Role of eIF3a in regulating cell cycle progression. Exp Cell Res 315:1889–1894
25. Saletta F, SuryoRahmanto Y, Richardson DR (2010) The translational regulator eIF3a: the tricky eIF3 subunit! Biochim Biophys Acta 1806:275–286
26. Olson JE, Wang X, Goode EL, Pankratz VS, Fredericksen ZS, Vierkant RA, Pharoah PD, Cerhan JR, Couch FJ (2010) Variation in genes required for normal mitosis and risk of breast cancer. Breast Cancer Res Treat 119:423–430
27. Boutros R, Bailey AM, Wilson SH, Byrne JA (2003) Alternative splicing as a mechanism for regulating 14-3-3 binding: interactions between hD53 (TPD52L1) and 14-3-3 proteins. J Mol Biol 332:675–687
28. Boutros R, Byrne JA (2005) D53 (TPD52L1) is a cell cycle-regulated protein maximally expressed at the G2-M transition in breast cancer cells. Exp Cell Res 310:152–165
29. Abba MC, Sun H, Hawkins KA, Drake JA, Hu Y, Nunez MI, Gaddis S, Shi T, Horvath S, Sahin A, Aldaz CM (2007) Breast cancer molecular signatures as determined by SAGE: correlation with lymph node status. Mol Cancer Res 5:881–890
30. Zacchi P, Stenmark H, Parton RG, Orioli D, Lim F, Giner A, Mellman I, Zerial M, Murphy C (1998) Rab17 regulates membrane trafficking through apical recycling endosomes in polarized epithelial cells. J Cell Biol 140:1039–1053
31. Palmieri D, Bouadis A, Ronchetti R, Merino MJ, Steeg PS (2006) Rab11a differentially modulates epidermal growth factor-induced proliferation and motility in immortal breast cells. Breast Cancer Res Treat 100:127–137
32. Cheng JM, Volk L, Janaki DK, Vyakaranam S, Ran S, Rao KA (2010) Tumor suppressor function of Rab25 in triple-negative breast cancer. Int J Cancer 126:2799–2812
33. Hendrix A, Maynard D, Pauwels P, Braems G, Denys H, Van den Broecke R, Lambert J, Van Belle S, Cocquyt V, Gespach C, Bracke M, Seabra MC, Gahl WA, De Wever O, Westbroek W (2010) Effect of the secretory small GTPase Rab27B on breast cancer growth, invasion, and metastasis. J Natl Cancer Inst 102:866–880
34. Frasor J, Danes JM, Komm B, Chang KC, Lyttle CR, Katzenellenbogen BS (2003) Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. Endocrinology 144:4562–4574
35. Mita K, Zhang Z, Ando Y, Toyama T, Hamaguchi M, Kobayashi S, Hayashi S, Fujii Y, Iwase H, Yamashita H (2007) Prognostic significance of insulin-like growth factor binding protein (IGFBP)-4 and IGFBP-5 expression in breast cancer. Jpn J Clin Oncol 37:575–582
36. Qin C, Singh P, Safe S (1999) Transcriptional activation of insulin-like growth factor-binding protein-4 by 17beta-estradiol in MCF-7 cells: role of estrogen receptor-Sp1 complexes. Endocrinology 140:2501–2508
37. Wang C, Yu J, Kallen CB (2008) Two estrogen response element sequences near the PCNA gene are not responsible for its estrogen-enhanced expression in MCF7 cells. PLoS One 3:e3523
38. Bouras T, Southey MC, Chang AC, Reddel RR, Willhite D, Glynne R, Henderson MA, Armes JE, Venter DJ (2002) Stanniocalcin 2 is an estrogen-responsive gene coexpressed with the estrogen receptor in human breast cancer. Cancer Res 62:1289–1295
39. Laganière J, Deblois G, Giguère V (2005) Functional genomics identifies a mechanism for estrogen activation of the retinoic acid receptor alpha1 gene in breast cancer cells. Mol Endocrinol 19:1584–1592
40. Jin C, Rajabi H, Pitroda S, Li A, Kharbanda A, Weichselbaum R, Kufe D (2012) Cooperative interaction between the MUC1-C oncoprotein and the Rab31 GTPase in estrogen receptor-positive breast cancer cells. PLoS One 7:e39432
41. Augereau P, Badia E, Fuentes M, Rabenoelina F, Corniou M, Derocq D, Balaguer P, Cavailles V (2006) Transcriptional regulation of the human NRIP1/RIP140 gene by estrogen is modulated by dioxin signalling. Mol Pharmacol 69:1338–1346
42. Guérin M, Sheng ZM, Andrieu N, Riou G (1990) Strong association between c-myb and oestrogen-receptor expression in human breast cancer. Oncogene 5:131–135