UC Office of the President
Research Grants Program Office (RGPO) Funded Publications

Title
RNA-Seq of Circulating Tumor Cells in Stage II—III Breast Cancer

Permalink
https://escholarship.org/uc/item/72m295qf

Journal
Annals of Surgical Oncology, 25(8)

ISSN
1068-9265

Authors
Lang, Julie E
Ring, Alexander
Porras, Tania
et al.

Publication Date
2018-08-01

DOI
10.1245/s10434-018-6540-4

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Ann Surg Oncol (2018) 25:2261–2270
https://doi.org/10.1245/s10434-018-6540-4

Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

CrossMark

ORIGINAL ARTICLE – BREAST ONCOLOGY

RNA-Seq of Circulating Tumor Cells in Stage II–III Breast Cancer

Julie E. Lang, MD[1], Alexander Ring, MD, PhD[2], Tania Porras, MS[1], Pushpinder Kaur, PhD[1], Victoria A. Forte, MD[3], Neal Mineyev, MD[1], Debu Tripathy, MD[4], Michael F. Press, MD, PhD[5], and Daniel Campo, PhD[6]

[1]Section of Surgical Oncology, Department of Surgery and University of Southern California Norris Cancer Center, University of Southern California, Los Angeles, CA; [2]Department of Oncology, University Hospital Zurich, Zurich, Switzerland; [3]Division of Medical Oncology, Department of Medicine, Maimonides Medical Center, New York, NY; [4]Department of Breast Medical Oncology, UT MD Anderson Cancer Center, Houston, TX; [5]Department of Pathology and University of Southern California Norris Cancer Center, University of Southern California, Los Angeles, CA; [6]Department of Biological Sciences, University of Southern California, Los Angeles, CA

ABSTRACT

Background. We characterized the whole transcriptome of circulating tumor cells (CTCs) in stage II–III breast cancer to evaluate correlations with primary tumor biology.

Methods. CTCs were isolated from peripheral blood (PB) via immunomagnetic enrichment followed by fluorescence-activated cell sorting (IE/FACS). CTCs, PB, and fresh tumors were profiled using RNA-seq. Formalin-fixed, paraffin-embedded (FFPE) tumors were subjected to RNA-seq and NanoString PAM50 assays with risk of recurrence (ROR) scores.

Results. CTCs were detected in 29/33 (88%) patients. We selected 21 cases to attempt RNA-seq (median number of CTCs = 9). Sixteen CTC samples yielded results that passed quality-control metrics, and these samples had a median of 4,311,255 uniquely mapped reads (less than PB or tumors). Intrinsic subtype predicted by comparing estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) versus PAM50 for FFPE tumors was 85% concordant. However, CTC RNA-seq subtype assessed by the PAM50 classification genes was highly discordant, both with the subtype predicted by ER/PR/HER2 and by PAM50 tumors. Two patients died of metastatic disease, both of whom had high ROR scores and high CTC counts. We identified significant genes, canonical pathways, upstream regulators, and molecular interaction networks comparing CTCs by various clinical factors. We also identified a 75-gene signature with highest expression in CTCs and tumors taken together that was prognostic in The Cancer Genome Atlas and Molecular Taxonomy of Breast Cancer International Consortium datasets.

Conclusion. It is feasible to use RNA-seq of CTCs in non-metastatic patients to discover novel tumor biology characteristics.

Circulating tumor cells (CTCs) are malignant cells that have shed into the bloodstream from primary or metastatic cancer tissue and are implicated in cancer metastasis.[1] The presence of CTCs in both metastatic and non-metastatic breast cancer patients has been shown to be associated with poorer survival.[2–4] In non-metastatic populations, detection of even a single CTC was correlated with poorer clinical outcomes and occurred in fewer than 25% of patients.[3–5] Some evidence suggests that CTC enumeration may allow detection of disease progression earlier than imaging studies.[6,7] Although CTC enumeration assays are prognostic, they are not predictive of what type of therapy a breast cancer patient requires.

Julie E. Lang and Alexander Ring served as co-first authors.

Electronic supplementary material The online version of this article (https://doi.org/10.1245/s10434-018-6540-4) contains supplementary material, which is available to authorized users.

© Society of Surgical Oncology 2018

First Received: 10 January 2018;
Published Online: 4 June 2018

J. E. Lang, MD
e-mail: julie.lang@med.usc.edu

To fully realize the potential of CTCs as a useful liquid biopsy biomarker, detailed molecular profiling of CTCs must inform the discovery of therapeutic predictors and druggable targets. Our group focused on gene expression profiling (RNA-seq) of CTCs since not all DNA mutations are expressed, and detailed examination at the RNA level allows for determination of the signaling pathways that are potentially targetable. Few studies have characterized the whole transcriptome phenotype of either single or small pools of CTCs in stage IV breast cancer,[8–11] and, to our knowledge, none have reported this in non-metastatic breast cancer.

We have previously reported that EpCAM-based immunomagnetic enrichment followed by fluorescence-activated cell sorting (IE/FACS) in combination with whole transcriptome gene expression profiling at the single cell or picogram input level is feasible for rare CTCs, with high accuracy.[8,12–14] In this study, we conducted a prospective, observational study to determine the feasibility of EpCAM-based IE/FACS CTC capture followed by RNA-seq for the evaluation of known and novel biomarkers to predict treatment response.

METHODS

Study Subjects

Eligibility was limited to female patients with biopsy-proven stage II–III breast cancer without prior therapy. The study was approved by the Institutional Review Board of the University of Southern California (USC), and was compliant with the REMARK criteria.[15]

Blood Draws and Tumor Tissue Acquisition

All de-identified, annotated samples were collected at baseline (prior to therapy). Twenty milliliters of peripheral blood (PB) was drawn into EDTA tubes. An aliquot of 200 μL of PB was placed in 600 μL of RNAlater (Thermo Fisher Scientific, Carlsbad, CA, USA) and stored at −80 °C until further use.

Clinical pathology reports utilized the 2010 American Society of Clinical Oncology and College of American Pathologists breast cancer biomarker guidelines.[16] Formalin-fixed, paraffin-embedded (FFPE) primary tumors (PTs) were requested from the Pathology Department for each patient. A total of five 10 μM sections each were used for RNA extraction using the RNeasy FFPE kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. RNA was eluted in 15 μL of RNase-free water.

Circulating Tumor Cell (CTC) Isolation (Immunomagnetic Enrichment Followed by Fluorescence-Activated Cell Sorting)

CTCs were isolated and captured as previously reported using a rapid processing method yielding highly purified CTCs within 3 h of blood draw (assays began within 20 min of blood draw).[8,12] Briefly, our IE/FACS assay involved immunomagnetic separation using EpCAM (MJ37) monoclonal antibody (mAb)-coated magnetic beads followed by FACS with PE mouse anti-human EpCAM (EBA-1), Thioflavin T buffer dye, and PerCP-Cy5.5 mouse anti-human CD45 (2D1) for exclusion of leukocytes (BD Biosciences, San Jose, CA, USA) [electronic supplementary Table 1]. Analysis of each CTC specimen included BT474 as internal positive and negative gating controls, as well as fluorescence compensation controls using BD CompBeads. CTCs were sorted into 5 μL Prelude Direct Lysis Module (NuGEN, San Carlos, CA, USA). A threshold of a single cell meeting these criteria was qualified as a positive test result. All lysates were immediately stored at −80 °C.

RNA-Seq Library Preparation and Sequencing

For sequencing library preparation from all sorted CTC samples, duplicate aliquots of 1 μL cell lysate were used as direct input for sequencing library preparation using the Ovation Single Cell RNA-Seq System (NuGEN). For PB, total RNA isolation was performed using the QIAamp RNA Blood Mini Kit (QIAGEN). RNA-seq library preparation for PB and PTs was performed using the NuGEN Ovation RNA System V2 and NuGEN Ultra Low Library System V2. The quality and quantity of amplified libraries were evaluated using Qubit (Invitrogen, Carlsbad, CA, USA) and Agilent Bioanalyzer 2100 analysis (Agilent Technologies, Santa Clara, CA, USA).

All libraries were sequenced using an Illumina HiSeq 2500 at the University of California, Los Angeles, Clinical Microarray Core with 100 bp paired end reads. FASTQ files were stored on the High Performance Computer Cluster of USC.

NanoString PAM50

NanoString nCounter gene expression assays using the PAM50 CodeSet (research use only) were performed as per the manufacturer's specifications (NanoString Technologies, Seattle, WA, USA).[17,18] Breast cancer molecular subtype classification was predicted using the genefu package in R for both the NanoString and RNA-seq assays.[19]

Bioinformatics Analysis of the RNA-Seq Data

Initial read quality and adaptor content of FASTQ files was assessed using FastQC.[20] Reads were then trimmed based on quality score, and adaptor sequences removed using Trimmomatic.[21] After filtering, surviving reads were checked again in FastQC to ensure that only high-quality transcriptome reads were put into the analysis pipeline. These high-quality reads were mapped to the human genome (ver. GRCh38.p7) using the ultra-fast aligner STAR;[22] the same software was used to obtain uniquely mapping read counts for each gene feature included in a Gene Transfer Format (GTF) file. Both the genome and the GTF file were downloaded from the GENCODE database (https://www.gencodegenes.org). Samples with a very low number of uniquely mapping reads (< 300,000) were omitted from any downstream analysis.

No background subtraction was performed for the CTC samples as our CTC gates were ultrapure, given the stringent gating approach.[8,14]

Differential gene expression analysis and RPKM value estimation was performed using the R/Bioconductor package edgeR,[23] using the raw counts obtained in the previous step as input. Sequencing data were deposited in the Gene Expression Omnibus at GSE111842.

To estimate the relative purity of our samples, we performed a cell enrichment analysis using the online tool xCell, which compares the gene expression profile of each sample with the gene expression signature of 64 immune and stroma cell types.[24] The xCell heatmap was created using GraphPad Prism.[24]

For Ingenuity Pathway Analysis (IPA; Qiagen), the differential gene expression lists comparing samples either based on sample type (CTCs, PTs, and PB) or clinical parameters (triple-negative breast cancer, ER-positive, human epidermal growth factor receptor 2 [HER2]-positive, pathologic complete response [pCR]) were used and a core analysis was run using default settings. For the Kyoto Encyclopedia of Genes and Genomes (KEGG) breast cancer pathway overlay with the differential gene expression data from our data analysis, the KEGG breast cancer curated gene set ($n = 151$) was used to create a new pathway in the IPA.[25]

For survival analysis, the top 75 upregulated genes, after differential gene expression analysis comparing sample groups, were used in cBioPortal to create survival tables using The Cancer Genome Atlas (TCGA; $n = 817$) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC; $n = 2509$) breast cancer data sets.[26,27] Kaplan–Meier curves were plotted using GraphPad Prism (La Jolla, CA, USA).

RESULTS

Enumeration of CTCs and Clinical Factors

Thirty-three newly diagnosed stage II–III breast cancer patients were accrued, along with 23 healthy female controls used to establish the thresholds for negative gating by FACS. Twenty-nine of 33 patients (88%) were positive for CTCs (at least one CTC per 20 mL) using our rapid processing strategy. The median number of CTCs present in this cohort was seven cells (range 0–65). Clinicopathologic and treatment factors are shown in electronic supplementary Table 2. Sixteen of 33 patients (48%) received neoadjuvant chemotherapy. We examined the association of tumor size, nodal status, estrogen receptor (ER), progesterone receptor (PR) and HER2 status, as well as the presence/absence of pCR with CTC status. None of these factors were associated with the detection of CTCs (electronic supplementary Table 3), which is driven by the fact that 88% of patients were CTC-positive in our assay.

Whole Transcriptome RNA-Seq Gene Expression Profiling

We selected 21 CTC samples to evaluate the feasibility of RNA-seq; 76.1% (16/21) passed quality-control metrics (> 300,000 uniquely mapping reads). CTC samples had a median of 4,311,255 uniquely mapped reads (less than PB or tumors) [$p = 0.002$]. We collected FFPE PT specimens from 75% (12/16) of these cases; six PB samples were also sequenced as a negative control. Figure 1a shows a principle component analysis demonstrating that, at the whole transcriptome level, the majority of CTC specimens separate to form a cohesive group that is distinct from PB or PTs without any background subtraction. Figure 1b presents a transcriptome heatmap of unsupervised hierarchical clustering (top 10,000 most variable genes) of the CTCs, PB, and PTs, based on normalized read counts, which shows that each specimen type forms a distinct group. Figure 1c is a Venn diagram for intergroup comparison (CTCs versus PB, CTCs versus PTs, PTs versus PB), showing the numbers of differentially expressed genes (p value ≤ 0.05 after false discovery rate [FDR] correction). Figure 1d is a volcano plot showing the $\log_2$ fold change in expression for every gene in the CTC versus PT comparison on the x-axis, versus statistical significance ($-\log_{10}$ of the FDR-corrected p-value) on the y-axis (the top five most significantly upregulated genes in CTCs versus PTs were *HBB, HAND2, OR52H1, CATSPER4*, and *CLRN1*).

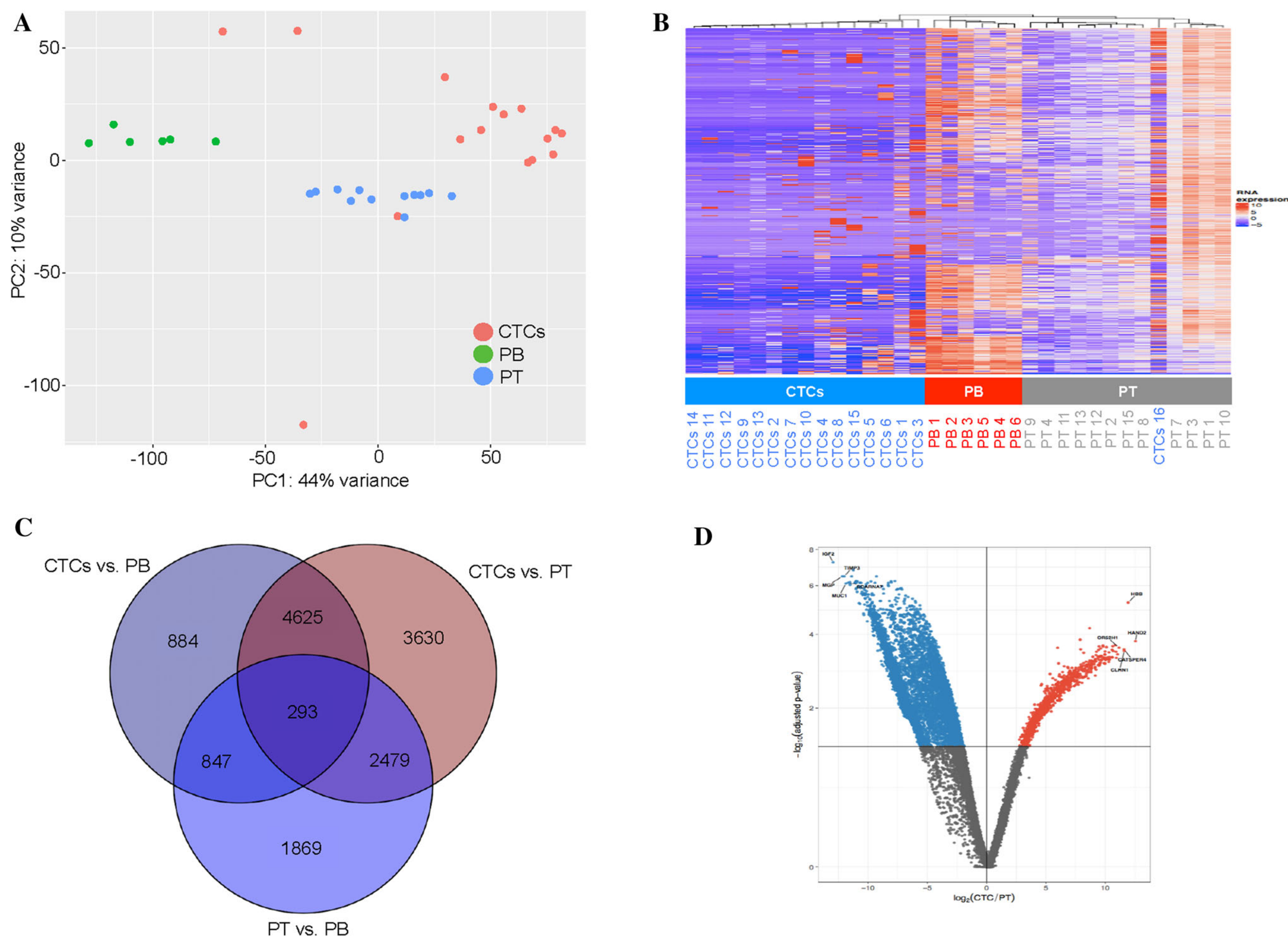

FIG. 1 Intergroup comparison of RNA-seq gene expression. **a** PCA plot demonstrating separation based on sample type: CTCs (n = 16), FFPE tumors (PTs; n = 12), and PB (n = 6). **b** Heatmap of the 10,000 most variable genes with unsupervised hierarchical clustering of samples based on gene expression in RNA-seq: CTCs (n = 16), FFPE tumors (n = 12), and PB (n = 6). **c** Venn diagram for intergroup comparison (CTCs vs. PB, CTCs vs. PTs, PTs vs. PB) overlap of differentially expressed genes. **d** Volcano plot of differential gene expression in CTCs vs. PTs. The five most differentially expressed genes upregulated in PTs were *IGF2, TIMP3, MGP, MUC1, and* SCARNA7, while the most upregulated genes in CTCs were *HBB, HAND2, OR52H1, CATSPER4, and CLRN1*. *PCA* principal component analysis, *CTCs* circulating tumor cells, *FFPE* formalin-fixed, paraffin-embedded, *PTs* primary tumors, *PB* peripheral blood

Molecular Subtyping of CTCs

Table 1 shows the intrinsic molecular subtyping of CTCs and PTs based on NanoString PAM50 assays, RNA-seq results for the PAM50 genes, and the molecular subtype predicted by ER/PR/HER2 testing. Eighty-five percent (11/13) of evaluable specimens showed concordance between the subtype predicted by ER/PR/HER2 testing and NanoString PAM50 of PTs. For all other comparisons, concordance was poor (7.1–28%). Two patients died of metastatic disease, both of whom had high risk of recurrence (ROR) scores based on the PT and high CTC counts.

Genes of Relevance to Breast Cancer and Signaling Pathways

Figure 2 presents selected genes of relevance to breast cancer, many of which are potentially clinically actionable, including ER, PR, HER2, androgen receptor (AR), Ki67, EGFR/RAF/MEK, JAK, IGF-1/PI3 K/AKT/mTOR, cancer stem cell, epithelial, mesenchymal, proliferation, immune, and DNA mismatch repair genes. Marked heterogeneity in gene expression was observed across all samples. Strikingly, no single marker was universally present in all CTCs, and there was significant overlap in expression of biomarkers with both PTs and PB for many genes.

Figure 3a shows the marked heterogeneity of CTCs for PAM50 genes, as shown in a heatmap comparing CTCs

TABLE 1 Specimen biomarkers

			Receptor status of primary tumor (pathology report)				NanoString				RNA-Seq CTC		RNA-Seq FFPE Tumor	
							PAM50		PAM50		PAM50		PAM50	
Patient #	#CTCs	Metastasis	ER	PR	HER2	Predicted subtype	Tumor subtype	Probability	CTC subtype	Probability	CTC subtype	Probability	Tumor subtype	Probability
Patient 1	39	Y	N	N	Y	**HER2**	**HER2**	0.67	***LumA***	0.6	***LumA***	0.4	**HER2**	0.6
Patient 2	11	N	N	N	N	**TNBC**	**Basal**	0.72	HER2	0.61	**Basal**	0.52	**Basal**	0.75
Patient 3	2	N	N	N	N	**TNBC**	**Basal**	1	***LumB***	0.51	***LumB***	0.69	HER2	0.62
Patient 4	12	N	N	N	N	**TNBC**	**Basal**	0.94	LumB	0.42	HER2	0.61	**Basal**	0.51
Patient 5	1	N	Y	Y	N	**LumA**	NA		NA		**LumA**	0.57	NA	
Patient 6	9	N	Y	Y	N	**LumA**	**LumA**	0.61	***LumB***	0.75	***LumB***	0.47	NA	
Patient 7	9	N	N	Y	N	LumB	**Basal**	0.7	HER2	0.89	LumA	0.77	**Basal**	0.77
Patient 8	65	Y	N	N	N	**TNBC**	**Basal**	0.71	Normal	0.71	LumA	0.59	**Basal**	0.85
Patient 9	11	N	Y	Y	N	**LumA**	**LumA**	0.74	LumB	0.54	**LumA**	0.67	**LumA**	0.56
Patient 10	7	N	Y	Y	N	**LumA**	**LumA**	0.87	NA		**LumA**	0.46	Basal	0.43
Patient 11	2	N	Y	Y	N	**LumA**	**LumA**	0.53	***Basal***	0.48	***Basal***	NA	**LumA**	0.56
Patient 12	2	N	Y	Y	Y	**LumB**	**LumB**	0.5	Normal	0.61	Basal	NA	**LumB**	0.41
Patient 13	9	N	Y	Y	N	**LumA**	LumB	0.81	**LumA**	0.56	HER2	0.87	**LumA**	0.52
Patient 14	2	N	N	Y	Y	LumB	NA		HER2	0.52	LumA	0.65	NA	
Patient 15	65	N	Y	Y	N	**LumA**	**LumA**	0.53	Basal	0.89	Normal	0.73	**LumA**	0.51
Patient 16	12	N	N	N	Y	HER2	NA		Normal	0.59	LumA	0.66	NA	

Concordance between pathology ER/PR/HER2 predicted subtype and/or Nanostring PAM50 of FFPE with other types of samples are bold. Concordance between all other specimen types are in bold italics

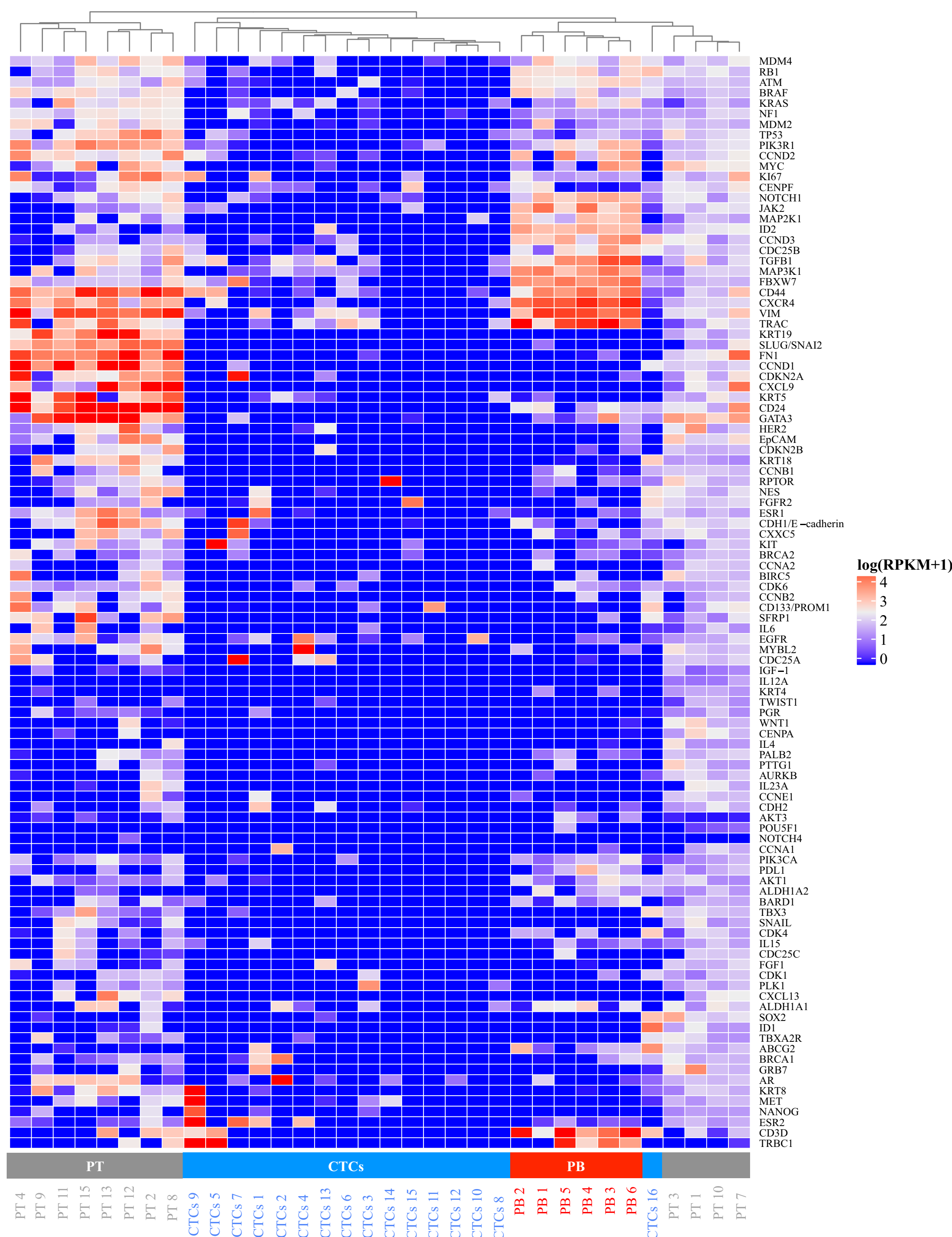

FIG. 2 Absolute normalized expression (RPKM) of selected genes in CTCs, PTs, and PB. The heatmap shows expression as log (RPKM + 1) for breast cancer-related genes for hormone receptors, proliferation markers, epithelial and mesenchymal markers, and stem cell (CSC)-related genes, including potentially clinically actionable genes of relevance to breast cancer (red indicates high expression, blue indicates low expression). *CTCs* circulating tumor cells, *PTs* primary tumors, *PB* peripheral blood, *CSC* circulating stem cells

versus PTs versus PB. Figure 3b shows differentially expressed KEGG curated breast cancer pathway genes for CTCs versus PTs. Notably, CTCs were upregulated in *WNT8A, FGF8, FZD5, PIK3CB*, and *ESR2* compared with PTs. Electronic supplementary Table 4 presents a pathway analysis listing the top five significant canonical pathways

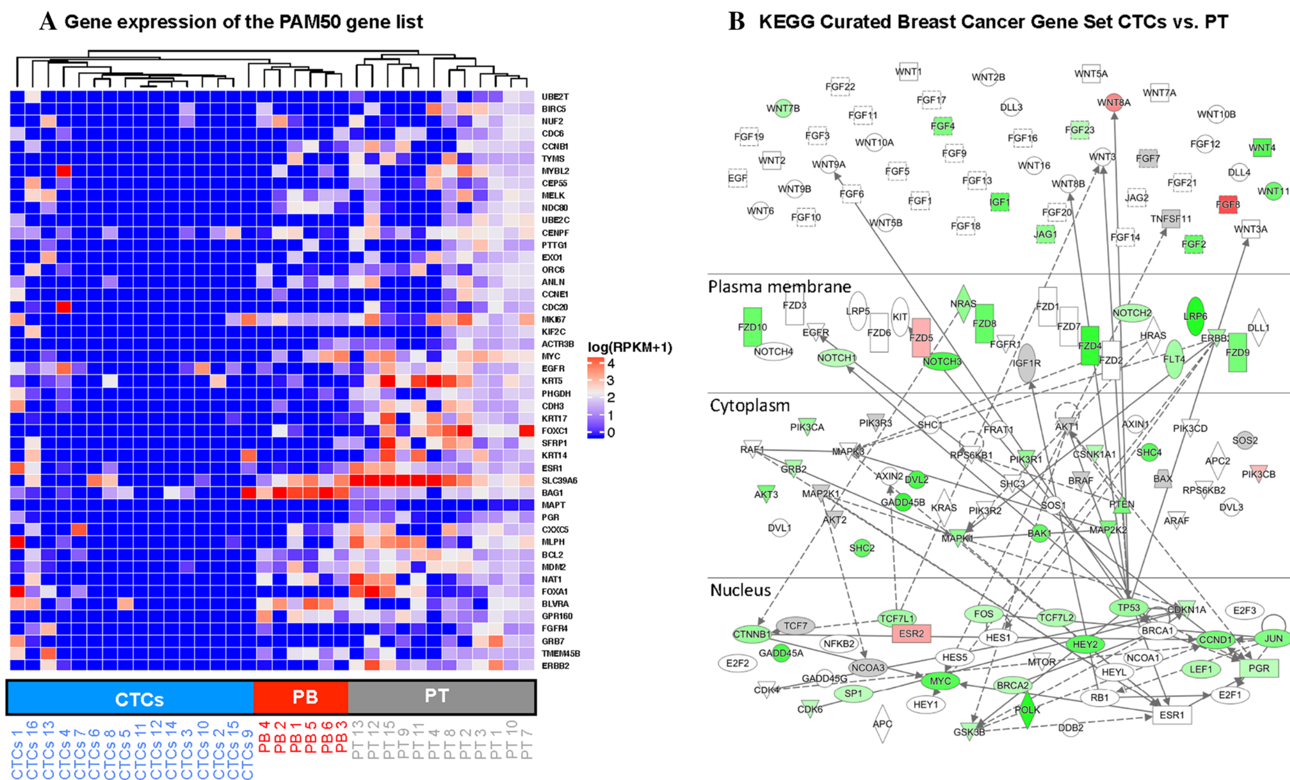

FIG. 3 RNA-seq gene expression of PAM50 genes and KEGG breast cancer pathway analysis. **a** Absolute normalized expression (RPKM) of PAM50 genes in all samples (*teal* indicates CTCs, $n = 16$; grey indicates FFPE PTs, $n = 12$; red indicates PB, $n = 6$). **b** The KEGG breast cancer genes ($n = 151$) were used to create a pathway representing all genes by cellular localization (extracellular, membrane bound, cytoplasmic, and nuclear). Differential gene expression analysis of CTCs vs. PTs was overlaid to show global differences in gene expression in CTCs compared with PTs of breast cancer-related genes (green indicates downregulated, red indicates upregulated, gray indicates the differential expression cut-off of 1.5-fold was not met, white indicates genes not expressed in the data set). Symbol legend: dotted square indicates growth factor, solid square indicates cytokine, upright rectangle indicates G-coupled receptor, horizontal rectangle indicates nuclear receptor, horizontal ellipse indicates transcription regulator, upright ellipse indicates transmembrane receptor, diamond indicates enzyme, triangle indicates kinase, trapezoid indicates transporter, *solid circle* indicates 'other'). *KEGG* Kyoto Encyclopedia of Genes and Genomes, *CTCs* circulating tumor cells, *FFPE* formalin-fixed, paraffin-embedded, *PTs* primary tumors, *PB* peripheral blood

and upstream regulators (FDR < 0.05) for each of the three comparison groups. Electronic supplementary Fig. 1 and electronic supplementary Table 5 present differential expression in CTCs versus PTs based on ER/PR/HER2 status, and presence or absence of pCR listing the top five significant canonical pathways and upstream regulators.

Validation with Publically Available Datasets

Figure 4 presents Kaplan–Meier curves comparing the gene signature of PTs versus PB, CTCs versus PB, and using TCGA and METABRIC cohorts[28,29]. In both cohorts, the gene signature of the shared top 75 upregulated genes in common between CTCs and PTs versus PB was prognostic of worse overall survival, while the signature found in CTCs alone versus PB was not prognostic.

Analysis of Cellular Composition

Electronic supplementary Fig. 2 presents the results for an analysis of the cellular heterogeneity based on the whole transcriptome profiles of each specimen type.[24] Based on enrichment analysis for 64 known immune and stromal cell types, CTC lysates have a distinct cellular composition that appears to be intermediate between PTs and PB for epithelial versus immune markers. Electronic supplementary Fig. 3 shows the results of NanoString assays of FFPE tumors for clinical-grade biomarkers.

DISCUSSION

In this pilot study, we found that the PB of nearly 90% of stage II–III breast cancer patients harbors CTCs when CTCs are freshly isolated immediately after blood draw. RNA-seq profiling of CTCs allows for a liquid biopsy

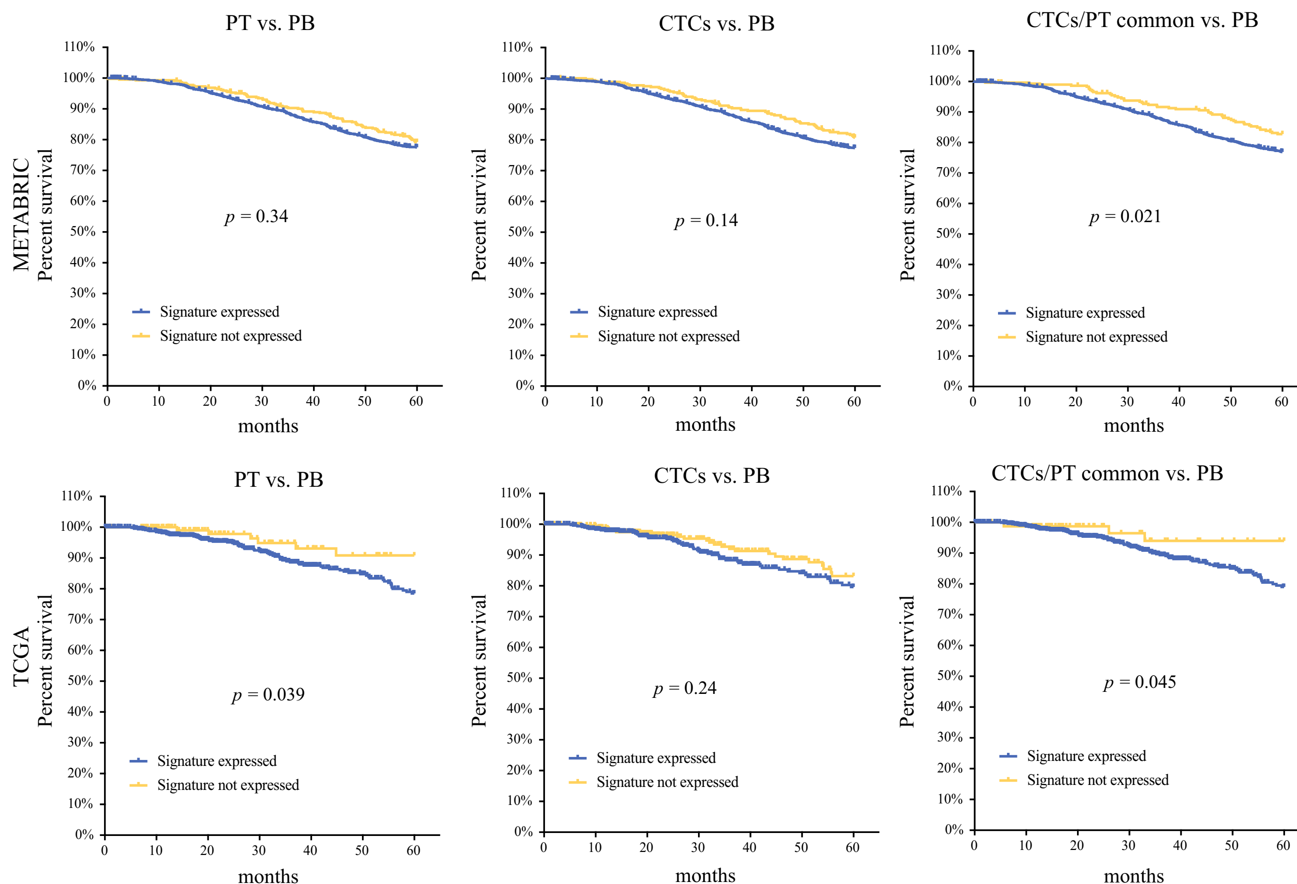

FIG. 4 Overall survival analysis based on differential gene expression in RNA-seq of CTCs, PTs, and PB in the METABRIC and TCGA breast cancer data sets. Differential gene expression results of the top 75 upregulated genes for each of the following groups (CTCs vs. PB, CTCs and PT common genes vs. PB, and PTs vs. PB) were used in cBioPortal to generate overall survival tables for the METABRIC (n = 2509) and TCGA (n = 817) breast cancer datasets. The tables were used to create Kaplan–Meier survival curves in GraphPad Prism. *CTCs* circulating tumor cells, *PTs* primary tumors, *PB* peripheral blood, *METABRIC* Molecular Taxonomy of Breast Cancer International Consortium, *TCGA* The Cancer Genome Atlas

characterizing potential precursors of metastatic disease at the whole transcriptome level, with comparison to matched PTs. This study is one of the first to perform whole transcriptome profiling of CTCs in non-metastatic breast cancer patients, although others have evaluated multigene panels.[30] We showed that RNA-seq of CTCs is feasible and that such profiling could be used to better understand what tumor characteristics in CTCs may allow for metastasis so that CTCs could be treated with molecularly-directed therapies.

Based on the extremely heterogeneous tumor biology of CTCs, it seems unlikely that single marker-based assays or limited panel multimarker assays could adequately characterize the tumor biology of CTCs to identify opportunities for targeted therapies, and that entire pathways should be examined instead, until suitable targets are known a priori. This information could complement DNA sequencing to predict therapeutic targets.[31] We have also recently published findings showing that multiple target enrichment in NanoString PAM50 resulted in false positive detection of gene transcripts in spiked CTC mimics, and noted potential confounding issues related to the overlap of genes of interest for signal in enriched CTCs and PB.[32] Our xCell analysis indicated that all sample types profiled contained mixed populations of tumor and immune cells as its default is to show a heterogeneous population. Any nucleated cells that are EpCAM high, CD45 low are isolated with our gating strategy, but we have previously shown that this method achieves highly enriched CTCs and that healthy controls do not have CTCs detected via IE/FACS.[8,14] xCell phenotyping indicated that CTC lysates contained erythrocytes, and, indeed, we found high expression of β-globin (HBB) in CTCs relative to PTs. Zheng et al. previously reported high HBB expression in CTCs; HBB expression was shown to decrease reactive oxygen species in CTCs, protecting them from apoptosis.[11]

The xCell analysis suggested that CTCs express immune markers, perhaps as a strategy to evade immune surveillance.

The limitation of perhaps an insufficient number of replicates for centroid-based intrinsic subtype classification for CTCs, with their inherently lower sequencing coverage, led us to conclude that we should focus our attention on highly expressed genes in CTCs, differentially expressed genes in CTCs versus PTs and CTCs versus PB, as well as pathway analysis, to better understand the tumor biology of CTCs.

CTCs highly expressed *HAND2*, which is reported to decrease *ESR1* transcriptional activity.[33] CTCs also strongly expressed *WNT8A,*[34] *PIK3CB,*[35] and *ESR2*,[36] in addition to the *OR52H1, CATSPER4*, and *CLRN1* genes, which were not previously associated with breast cancer. EGFR was expressed in half of the CTCs (Fig. 2) and AR was expressed in 37.5% of CTCs, while *PDL1* was not expressed in any of the CTCs, in contrast to a previously published report.[37] We interpret the fact that a 75-gene signature of the most upregulated genes in common between CTCs and PTs was prognostic for survival in two independent datasets, such that CTCs as a liquid biopsy may add value to determining prognosis based on PT features.

CONCLUSIONS

It is feasible to use RNA-seq of CTCs in non-metastatic breast cancer patients as a liquid biopsy of the status of tumor biology, as has been previously demonstrated in metastatic malignancies.[38,39] This approach is hypothesis-generating for potentially targeting tumor biology based on CTCs to eradicate the minimal residual disease responsible for subsequent recurrence.

ACKNOWLEDGEMENT This project was supported by a Society of Surgical Oncology Clinical Investigator Award, a California Breast Cancer Research Program IDEA award, and a STOP Cancer Marni Levine Memorial Seed Grant (JL). The project was also supported in part by grant UL1TR001855 from the National Center for Advancing Translational Science (NCATS) and grant P30CA014089 from the National Cancer Institute of the US National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. Magnetic cell separators were kindly made by the laboratory of Maciej Zborowski (Cleveland Clinic) based on the design provided by BD Biosciences.

REFERENCES

1. Gupta GP, Massague J. Cancer metastasis: building a framework. *Cell.* 2006;127(4):679–695.
2. Cristofanilli M, Budd GT, Ellis MJ, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med.* 2004;351(8):781–791.
3. Lucci A, Hall CS, Lodhi AK, et al. Circulating tumour cells in non-metastatic breast cancer: a prospective study. *Lancet Oncol.* 2012;13(7):688–695.
4. Rack B, Schindlbeck C, Juckstock J, et al. Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. *J Natl Cancer Inst.* 2014;106(5): pii: dju066.
5. Zhang L, Riethdorf S, Wu G, et al. Meta-analysis of the prognostic value of circulating tumor cells in breast cancer. *Clin Cancer Res.* 2012;18(20):5701–5710.
6. Budd GT, Cristofanilli M, Ellis MJ, et al. Circulating tumor cells versus imaging: predicting overall survival in metastatic breast cancer. *Clin Cancer Res.* 2006;12(21):6403–6409.
7. Liu MC, Shields PG, Warren RD, et al. Circulating tumor cells: a useful predictor of treatment efficacy in metastatic breast cancer. *J Clin Oncol.* 2009;27(31):5153–5159.
8. Lang JE, Scott JH, Wolf DM, et al. Expression profiling of circulating tumor cells in metastatic breast cancer. *Breast Cancer Res Treat.* 2015;149(1):121–31.
9. Boral D, Vishnoi M, Liu HN, et al. Molecular characterization of breast cancer CTCs associated with brain metastasis. *Nat Commun.* 2017;8(1):196.
10. Yu M, Bardia A, Wittner BS, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science.* 2013;339(6119):580–584.
11. Zheng Y, Miyamoto DT, Wittner BS, et al. Expression of β-globin by cancer cells promotes cell survival during blood-borne dissemination. *Nat Commun.* 2017;8:14344.
12. Magbanua MJ, Park JW. Isolation of circulating tumor cells by immunomagnetic enrichment and fluorescence-activated cell sorting (IE/FACS) for molecular profiling. *Methods.* 2013;64(2):114–8.
13. Lang JE, Magbanua MJ, Scott JH, et al. A comparison of RNA amplification techniques at sub-nanogram input concentration. *BMC Genom.* 2009;10:326.
14. Ring A, Mineyev N, Zhu W, et al. EpCAM based capture detects and recovers circulating tumor cells from all subtypes of breast cancer except claudin-low. *Oncotarget.* 2015;6(42):44623–44634.
15. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst.* 2005;97(16):1180–1184.
16. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol.* 2010;28(16):2784–2795.
17. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol.* 2008;26(3):317–325.
18. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol.* 2009;27(8):1160–1167.
19. Gendoo DM, Ratanasirigulchai N, Schroder MS, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics.* 2016;32(7):1097–1099.
20. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 15 Oct 2017
21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–2120.
22. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.

23. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–140.
24. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):220.
25. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res.* 2002;30(1):42–46.
26. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6(269):pl1.
27. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery.* 2012;2(5):401–404.
28. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486(7403):346–352.
29. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61–70.
30. Powell AA, Talasaz AH, Zhang H, et al. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. *PloS One.* 2012;7(5):e33788.
31. Gulbahce N, Magbanua MJM, Chin R, et al. Quantitative Whole Genome Sequencing of Circulating Tumor Cells Enables Personalized Combination Therapy of Metastatic Cancer. *Cancer research.* 2017;77(16):4530–4541.
32. Porras TB, Kaur P, Ring A, Schechter N, Lang JE. Challenges in using liquid biopsies for gene expression profiling. *Oncotarget.* 2018;9(6):7036–7053.
33. Fukuda T, Shirane A, Wada-Hiraike O, et al. HAND2-mediated proteolysis negatively regulates the function of estrogen receptor alpha. *Mol Med Rep.* 2015;12(4):5538–5544.
34. Saitoh T, Katoh M. Molecular cloning and characterization of human WNT8A. *Int J Oncol.* 2001;19(1):123–127.
35. Nakanishi Y, Walter K, Spoerke JM, et al. Activating Mutations in PIK3CB Confer Resistance to PI3 K Inhibition and Define a Novel Oncogenic Role for p110beta. *Cancer Res.* 2016;76(5):1193–1203.
36. Elebro K, Borgquist S, Rosendahl AH, et al. High estrogen receptor beta expression is prognostic among adjuvant chemotherapy-treated patients-results from a population-based breast cancer cohort. *Clin Cancer Res*. 2017;23(3):766–777.
37. Mazel M, Jacot W, Pantel K, et al. Frequent expression of PD-L1 on circulating breast cancer cells. *Mol Oncol.* 2015;9(9):1773–1782.
38. Kalinich M, Bhan I, Kwan TT, et al. An RNA-based signature enables high specificity detection of circulating tumor cells in hepatocellular carcinoma. *Proc Natl Acad Sci U S A.* 2017;114(5):1123–1128.
39. Miyamoto DT, Zheng Y, Wittner BS, et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science.* 2015;349(6254):1351–1356.