JNCI J Natl Cancer Inst (2018) 0(0): pky005

doi: 10.1093/jncics/pky005
Brief Communication

BRIEF COMMUNICATION

Major Impact of Sampling Methodology on Gene Expression in Estrogen Receptor–Positive Breast Cancer

Qiong Gao, Elena López-Knowles, Maggie Chon U Cheang, James Morden†, Ricardo Ribas, Kally Sidhu, David Evans, Vera Martins, Andrew Dodson, Anthony Skene, Chris Holcombe, Elizabeth Mallon, Abigail Evans, Judith M. Bliss, John Robertson, Ian Smith, Lesley-Ann Martin, Mitch Dowsett; on behalf of the POETIC Trial Management Group and Trialists

†Deceased.
See the Notes section for the full list of authors' affiliations.
Correspondence to: Mitch Dowsett, PhD, FMedSci, Biochemical Endocrinology, Translational Research at Breast Cancer Now Research Centre, Ralph Lauren Centre for Breast Cancer Research, The Royal Marsden, Fulham Road, London SW3 6JJ, UK (e-mail: mitch.dowsett@icr.ac.uk).

Abstract

To investigate the impact of sampling methodology on gene expression data from primary estrogen receptor–positive (ER+) breast cancer biopsies, global gene expression was measured in core-cut biopsies at baseline and surgery from patients randomly assigned to receive either two weeks of presurgical aromatase inhibitor (AI; n = 157) or no presurgical treatment (n = 56). Those genes most markedly altered in the AI group (eg, *FOS*, *DUSP1*, *RGS1*, *FOSB*) were similarly altered in the no treatment group; some widely investigated genes that were apparently unaffected in the AI group (eg, *MYC*) were counter-altered in the control group, masking actual AI-dependent changes. In the absence of a control group, these artefactual changes would likely lead to the most affected genes being the erroneous focus of research. The findings are likely relevant to all archival collections of ER+ breast cancer.

Analysis of gene expression in biopsies taken before and after treatment of primary breast cancer (BC) is frequently undertaken to study mechanisms of response and resistance. We and others have identified artefactual changes in gene expression that can result from the study procedures (1,2). Importantly, however, the degree of impact of those changes has not been evaluated in the context of a specific therapy. The current study reveals the potential for profound errors in data interpretation that could occur if such artefacts are not identified or are ignored.

The PeriOperative Endocrine Therapy-Individualising Care (POETIC; CRUK/07/015) (3) trial randomly assigned 4486 postmenopausal women with primary estrogen receptor–positive (ER+) BC 2:1 to receive perioperative aromatase inhibitor (AI; two weeks presurgery + two weeks postsurgery, termed AI-treated) or no perioperative treatment (termed control). Core-cut biopsy samples in RNA later were analyzed from both the baseline and surgical sample from 213 patients (157 were all good-quality available AI-treated and 56 were randomly chosen controls). High-quality genome-wide expression data (GSE105777) were analyzed to identify statistically significant altered gene expression and were compared between the AI-treated and control groups. Classical clinical factors were well balanced between the two groups (Supplementary Table 1).

A total of 3269 genes (n = 1504 upregulated, n = 1765 downregulated) from treated tumors and 110 genes (n = 70 upregulated, n = 40 downregulated) from control tumors were

Received: December 19, 2017; **Accepted:** February 16, 2018

differentially expressed between baseline and surgical samples, with *P* values of less than .001 by paired *t* test on the normalized expression data (Supplementary Tables 2 and 3 and Supplementary Figure 1, available online). In the control group, 25 of the top 30 differentially expressed genes ranked by an absolute fold-change (FC) were upregulated at surgery (Table 1), as previously reported (1). Many of these were early-response or stress-associated genes (*FOS*, *RGS1*, *DUSP1*, *FOSB*, *JUN*, and *MYC*), as well as *CD69* (FC = 1.39, *P* = 1.48E-4), an early T-cell activation antigen associated with immune response (4), and *TOB1* (FC = 1.28, *P* = 5.2E-06), an important mediator involved in T-cell activation and a critical determinant of estrogen-independent ER+ BC survival (5). The five downregulated genes included those encoding hemoglobin (*HBA* and *HBB*).

In the AI-treated cohort, in contrast to the mainly upregulated genes in the control group, 22 of the top 30 ranked genes (Table 1) were downregulated after two weeks of AI treatment. Most of these corresponded to proliferation, estrogen-responsive, or cycling genes. However, 11 of the 30 top-ranked differentially expressed genes in the AI-treated group (eight upregulated and three downregulated) were among the 12 most altered genes in the control group, including *FOS*, *DUSP1*, and *RGS1*. These latter three genes were the most affected in both cohorts. A scatterplot of the changes in individual gene expression between the POETIC AI-treated and control cohorts (Figure 1) shows the near identical quantitative levels of change for the genes most affected by each treatment. In essence, the changes of greatest magnitude in the AI-treated tumors are completely artefactual. In addition, there are widely investigated genes, such as *MYC*, that are statistically significantly altered in the control samples, but apparently unaffected in the treated group. Correction for the degree of change in *MYC* expression within the control group reveals the highly statistically significant suppression of *MYC* as a result of AI therapy (*P* = .0003).

We examined gene expression data from an independent set of core cuts taken at baseline and two weeks after starting AI treatment in the FAIMoS study (6). The same markedly affected genes identified in the POETIC AI-treated arm were not observed in the top-ranked 30 genes identified in FAIMoS (Supplementary Table 4). This discrepancy is likely to be explained by the way the tissue core was taken and manipulated after therapy. The key difference between the POETIC and FAIMoS studies was that for the former the post-therapy core was from an excised tumor while in FAIMoS the core was taken from the breast with the tumor in situ. Thus, in FAIMoS, the processes of sampling for the baseline and on-treatment core biopsies were identical, while in POETIC, the second core biopsy was subject to a variable degree of ischemic conditions prior to

Table 1. Top ranked genes in control and AI-treated samples. Top 30 regulated genes in control tumors and in AI-treated tumors.

Symbol	Rank in control†	FC (S/B) ‡	Parametric p-value	FDR	Rank in AI-treated†	Symbol	Rank in AI-treated†	FC (S/B) ‡	Parametric p-value	FDR	Rank in control†
FOS*	1	3.87	$< 1x10^{-07}$	$< 1x10^{-07}$	1	FOS*	1	4.14	$< 1x10^{-07}$	$< 1x10^{-07}$	1
RGS1*	2	3.22	$< 1x10^{-07}$	$< 1x10^{-07}$	3	DUSP1*	2	3.36	$< 1x10^{-07}$	$< 1x10^{-07}$	3
DUSP1*	3	3.06	$< 1x10^{-07}$	$< 1x10^{-07}$	2	RGS1*	3	3.33	$< 1x10^{-07}$	$< 1x10^{-07}$	2
HBA2*	4	-2.94	$9.00x10^{-07}$	$9.10x10^{-04}$	7	TFF1	4	-2.94	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
HBB*	5	-2.86	$3.40x10^{-06}$	$2.06x10^{-03}$	7	FOSB*	5	2.65	$< 1x10^{-07}$	$< 1x10^{-07}$	7
HBA1*	6	-2.50	$6.30x10^{-06}$	$3.21x10^{-03}$	12	TOP2A	6	-2.56	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
FOSB*	7	2.30	$9.00x10^{-07}$	$9.10x10^{-04}$	5	UBE2C	7	-2.44	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
RNY5	8	-2.13	$5.04x10^{-05}$	$1.39x10^{-02}$	83	HBB*	7	-2.44	$< 1x10^{-07}$	$< 1x10^{-07}$	5
CYR61*	9	2.05	$1.70x10^{-06}$	$1.43x10^{-03}$	10	HBA2*	7	-2.44	$< 1x10^{-07}$	$< 1x10^{-07}$	4
EGR1*	10	1.98	$1.00x10^{-06}$	$9.47x10^{-04}$	10	EGR1*	10	2.36	$< 1x10^{-07}$	$< 1x10^{-07}$	10
ZFP36*	11	1.97	$< 1x10^{-07}$	$< 1x10^{-07}$	18	CYR61*	10	2.36	$< 1x10^{-07}$	$< 1x10^{-07}$	9
SNORD3D*	12	1.86	$4.00x10^{-07}$	$4.66x10^{-04}$	30	CDC20	12	-2.13	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
TRK1	13	-1.72	$< 1x10^{-07}$	$< 1x10^{-07}$	49	NUSAP1	12	-2.13	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
JUN	14	1.67	$< 1x10^{-07}$	$< 1x10^{-07}$	65	HBA1*	12	-2.13	$< 1x10^{-07}$	$< 1x10^{-07}$	6
SGK1	15	1.62	$< 1x10^{-07}$	$1.68x10^{-04}$	32	SUSD3	15	-2.04	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
SNORD3A	15	1.62	$1.61x10^{-05}$	$6.10x10^{-03}$	79	FGFR3	16	-2.00	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
SNORD3C	17	1.61	$1.30x10^{-05}$	$5.05x10^{-03}$	97	NEK2	17	-1.96	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
LOC100132564	19	1.53	$4.12x10^{-04}$	$7.70x10^{-02}$	NA	ZFP36*	18	1.94	$< 1x10^{-07}$	$< 1x10^{-07}$	11
RASD1	20	1.52	$1.22x10^{-05}$	$4.87x10^{-03}$	39	UHRF1	19	-1.92	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
RGS2	21	1.51	$9.74x10^{-05}$	$2.38x10^{-02}$	33	PRC1	19	-1.92	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
SNORD13	22	1.49	$5.52x10^{-05}$	$1.49x10^{-02}$	80	ASPM	19	-1.92	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
CCL3L3	23	1.48	$6.80x10^{-06}$	$3.21x10^{-03}$	80	AGR2	19	-1.92	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
KLF6	24	1.47	$5.29x10^{-04}$	$9.02x10^{-02}$	99	PDZK1	23	-1.85	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
APOLD1	25	1.45	$2.00x10^{-06}$	$1.60x10^{-03}$	71	PTTG1	24	-1.82	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
ATF3	26	1.43	$3.71x10^{-05}$	$1.12x10^{-02}$	138	ADCY1	24	-1.82	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
SPRY1	26	1.43	$9.59x10^{-05}$	$2.38x10^{-02}$	36	CDCA5	26	-1.79	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
MYC	28	1.41	$2.28x10^{-05}$	$8.23x10^{-03}$	NA	CCNB2	26	-1.79	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
CEBPD	29	1.39	$4.60x10^{-06}$	$2.68x10^{-03}$	226	KIAA0101	26	-1.79	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
BTG2	29	1.39	$9.20x10^{-06}$	$3.77x10^{-03}$	138	STC2	26	-1.79	$< 1x10^{-07}$	$< 1x10^{-07}$	NA
CD69	29	1.39	$1.48x10^{-04}$	$3.31x10^{-02}$	65	SNORD3D*	30	1.77	$< 1x10^{-07}$	$< 1x10^{-07}$	12

*Highlights the 11 genes regulated in AI-treated group that are among the 12 most regulated genes in control. AI = aromatase inhibitor; B = baseline; FC = fold-change; FDR = false discovery rate; S = surgery.

†Rank: by an absolute fold-change.

‡FC (S/B): fold-change of individual genes at surgery compared with baseline.

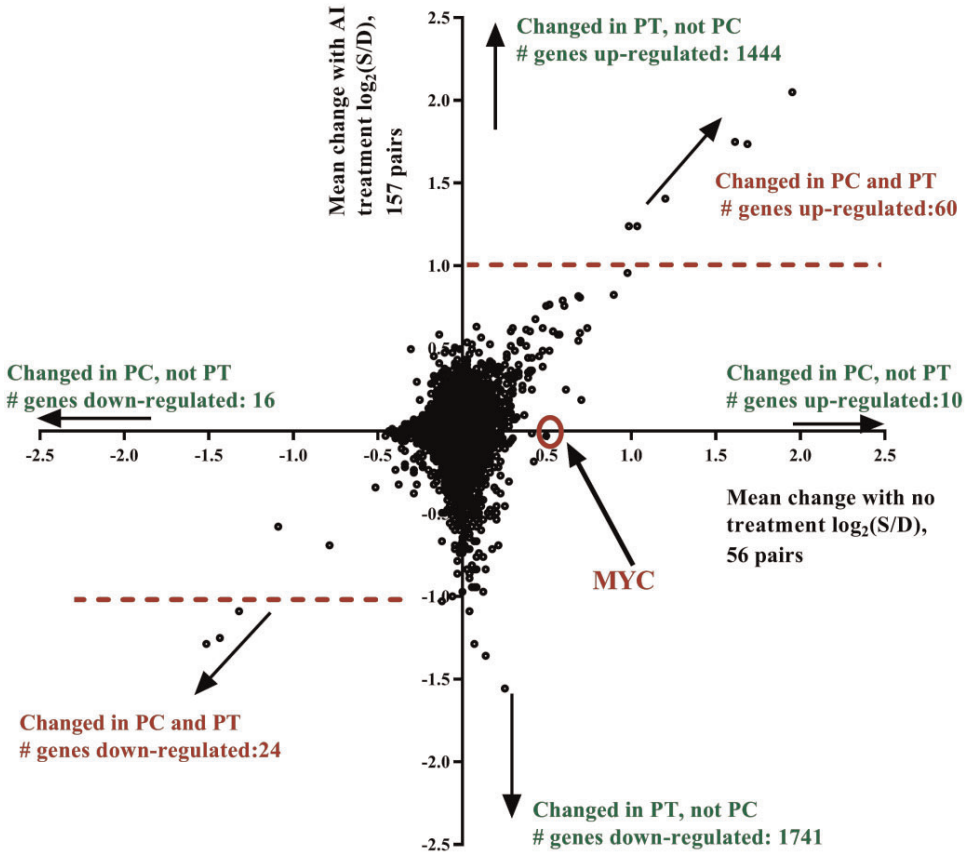

Figure 1. Whole-genome gene expression changes between baseline and surgery in PeriOperative Endocrine Therapy-Individualising Care (POETIC): aromatase inhibitor (AI)–treated (n = 157) vs control (n = 56). The largest changes (above or below the **dotted line**) in gene expression seen with AI treatment are artefactual. PC = POETIC control; PT = POETIC treated.

placement in RNA later. In many cases, the time of ischemia will have been variably extended by the time taken to x-ray the biopsy sample. Our earlier report described the changes that result from that delay in genes such as *RGS1* and *DUSP1*, which are among the most affected genes in the AI-treated and control arms in POETIC. Of note, in support of this explanation, the decrease seen in MYC expression in the POETIC AI-treated group *after correction* for the artefactual increase in the controls concurs with the statistically signifcant decrease seen in FAIMoS.

We conclude that the majority of the most upregulated genes (eg, *FOS*, *DUSP1*, *RGS1*, *FOSB*) and a small number of the most downregulated genes are identified as a result of preanalytical sample processing. In addition, the true effect of AI treatment on other genes can be hidden by counteractive artefactual change. In the absence of a control group, investigators are likely to focus on the most extensive gene changes, yet these will include many ascribed wrongly to the effect of experimental intervention; some genes that would be the focus will be wrongly ignored because they are apparently unaffected by therapy. It is notable that our observations have been made in the context of withdrawal of estrogen stimulation, the strongest transcriptional drive for ER+ BC. The artefacts are likely to be pronounced relative to true effects in the context of less impactful therapy. Future presurgical studies should ensure that core cuts taken at surgery are either taken in an identical fashion to those at baseline or that a control group of patients is included to identify any process-related changes.

It should also be recognized that the majority of tissue-related studies in BC occur in archival excision specimens that will have been subject to similar or perhaps greater ischemic conditions before fixation than the core-cut samples in POETIC. Investigators should establish that the collection process does not affect expression of the genes of interest prior to assuming that the observed expression reflects the true expression in the tumor in situ.

FUNDING

This work was supported by the Mary-Jean Mitchell Green Foundation, Breast Cancer Now, working in partnership with Walk the Walk and the National Institutes of Health Research Royal Marsden/Institute of Cancer Research Biomedical Research Centre. The POETIC trial (C1491/A8671/CRUK/07/015, C1491/A15955, C406/A8962), from which samples were obtained for this study, was supported by Cancer Research UK, as is Institute of Cancer Research- Clinical Trials and Statistical Unit through its core program grant.

Notes

Affiliations of authors: Breast Cancer Now Research Centre (QG, ELK, RR, LAM, MD) and Clinical Trials and Statistics Unit (MCUC, JM, JMB), The Institute of Cancer Research, London, UK; Ralph Lauren Centre for Breast Cancer Research (ELK, KS, DE, VM, AD, MD) and Breast Unit, Royal Marsden Hospital, London, UK (IS); Royal Bournemouth Hospital, Bournemouth, UK (AS); Royal Liverpool University Hospital, Liverpool, UK (CH); Queen

Elizabeth University Hospital Glasgow, Govan, UK (EM); Poole General Hospital, Dorset, UK (AE); University of Nottingham, Nottingham, UK (JR).

We thank all participating patients and staff at POETIC centers and ICR-CTSU who contributed to trial management and data and sample collection, and the independent data monitoring committee and trial steering committee, for their oversight of the trial.

QG analyzed the data and drafted the manuscript. ELK extracted RNA and drafted the manuscript. MC generated subtypes. JM provided data and composed Supplementary Table 1. KS and DE recorded the samples for the study. VM and AD sectioned and reviewed the histopathology of the samples. AS, HC, EM, EA, and JR were involved in sample acquisition. MD, IS, and JB were involved in conception and design of POETIC. LAM, RR, and MD drafted the manuscript. All authors read and approved the final manuscript.

References

1. Lopez-Knowles E, Gao Q, Cheang MC, et al. Heterogeneity in global gene expression profiles between biopsy specimens taken peri-surgically from primary ER-positive breast carcinomas. *Breast Cancer Res*. 2016;18(1):39.
2. Pearce DA, Arthur LM, Turnbull AK, et al. Tumor sampling method can significantly influence gene expression profiles derived from neoadjuvant window studies. *Sci Rep*. 2016;6:29434.
3. Dowsett M, Smith I, Robertson J, et al. Endocrine therapy, new biologicals, and new study designs for presurgical studies in breast cancer. *J Natl Cancer Inst Monogr*. 2011;2011(43):120–123.
4. Shiow LR, Rosen DB, Brdickova N, et al. CD69 acts downstream of interferon-alpha/beta to inhibit S1P1 and lymphocyte egress from lymphoid organs. *Nature*. 2006;440(7083):540–544.
5. Zhang YW, Nasto RE, Varghese R, et al. Acquisition of estrogen independence induces TOB1-related mechanisms supporting breast cancer cell proliferation. *Oncogene*. 2016;35(13):1643–1656.
6. Smith IE, Walsh G, Skene A, et al. A phase II placebo-controlled trial of neoadjuvant anastrozole alone or with gefitinib in early breast cancer. *J Clin Oncol*. 2007;25(25):3816–3822.