

SCIENTIFIC DATA

OPEN Data Descriptor: Transcriptomic profiling of human breast and melanoma cells selected by migration through narrow constraints

Received: 02 August 2017

Accepted: 05 October 2017

Published: 14 November 2017

Dominika A. Rudzka¹, William Clark¹, Ann Hedley¹, Gabriela Kalna¹ & Michael F. Olson^{1,2}

The metastatic spread of cancer cells is a step-wise process that starts with dissociation from primary tumours and local invasion of adjacent tissues. The ability to invade local tissues is the product of several processes, including degradation of extracellular matrices (ECM) and movement of tumour cells through physically-restricting gaps. To identify properties contributing to tumour cells squeezing through narrow gaps, invasive MDA-MB-231 human breast cancer and MDA-MB-435 human melanoma cells were subjected to three successive rounds of selection using cell culture inserts with highly constraining 3 µm pores. For comparison purposes, flow cytometry was also employed to enrich for small diameter MDA-MB-231 cells. RNA-Sequencing (RNA-seq) using the Illumina NextSeq 500 platform was undertaken to characterize how gene expression differed between parental, invasive pore selected or small diameter cells. Gene expression results obtained by RNA-seq were validated by comparing with RT-qPCR. Transcriptomic data generated could be used to determine how alterations that enable cell passage through narrow spaces contribute to local invasion and metastasis.

Design Type(s)	cell type comparison design
Measurement Type(s)	transcription profiling assay
Technology Type(s)	RNA sequencing
Factor Type(s)	cell line • protocol
Sample Characteristic(s)	Homo sapiens • breast cancer cell line • melanoma cell line

¹Cancer Research UK Beatson Institute, Garscube Estate, Switchback Road, Glasgow G61 1BD, UK. ²Institute of Cancer Sciences, University of Glasgow, Glasgow G12 8QQ, UK. Correspondence and requests for materials should be addressed to M.F.O. (email: m.olson@beatson.gla.ac.uk).

Background & Summary

The metastatic spread of cancer cells from primary tumours to distant sites is the most serious and deadliest aspect of the disease, with estimates of up to 90% of cancer related deaths being directly associated with metastasis^{1,2}. In addition, it has become apparent that the same processes that contribute to cancer metastasis also facilitate primary tumour growth^{3,4}. As a result, characterizing the properties of metastatic cells that enable their dissemination may identify actionable targets for chemotherapy that could reduce cancer spread and possibly tumour growth and progression.

The metastatic ability of cancer cells results from several changes in their normal behaviours⁵. The strength of cell-cell adhesions is often lessened, allowing individual or groups of cells to separate and move away from the primary tumour⁶. The movement away from the tumour into adjacent tissue may be promoted by changes in migratory behaviour, increased extracellular matrix (ECM) degrading activity, and the ability to squeeze between cells and ECM protein fibres⁷. In the next stage, locally invasive cells may spread further by moving through surfaces surrounding body cavities, via lymphatic or blood vessels, or through canalicular spaces. Ultimately, tumour cells may move to a secondary site, the location of which may be influenced by variables including the route taken, intrinsic properties of the target tissue, and accessibility to factors, such as tumour cell-generated exosomes, that condition a pre-metastatic niche. Common to several stages of the metastatic process is the ability of tumour cells to squeeze through narrow spaces. As a result, it can be predicted that changes in the deformability of tumour cells that enabled their movement through physically constraining conditions in their three dimensional environment would likely promote cancer spread⁸.

To select for cells that were better able to move through narrow gaps, MDA-MB-231 D3H2LN human breast cancer cells expressing firefly luciferase (Luc)⁹ (abbreviated MDA-MB-231) and MDA-MB-435 human melanoma cells (which were mistakenly used in the past as a breast cancer model until its cancer type was corrected)¹⁰ were subjected to three rounds of enrichment using tissue culture inserts with 3 µm pores (Fig. 1a). By plating cells on microporous membranes in serum-free medium in the inserts, and then transferring the inserts to tissue culture dishes containing medium supplemented with 10% fetal bovine serum (FBS), a chemotactic gradient was created to attract cells to move through the restricting narrow pores. Several independent 'Selected' populations of cells were established from both MDA-MB-231 and MDA-MB-435 cells in this manner. Given the possibility that selection through narrow pores would enrich for small cells, several independent 'FlowSorted' populations of small diameter MDA-MB-231 cells were also selected by three consecutive rounds of flow cytometry (Fig. 1b).

Parental, Selected and FlowSorted populations of MDA-MB-231 and MDA-MB-435 cells were fixed and stained with phalloidin to reveal filamentous actin (F-actin) structures and cell morphology. Selected and FlowSorted MDA-MB-231 cells, as well as Selected MDA-MB-435 cells, were notably smaller than their originating parental cells (Fig. 1c). In addition, Selected cells from both tumour cell lines were marked by fewer cytoplasmic and cortical distinct F-actin fibres, more actin-rich protrusive regions (indicated by white arrows), and more irregular circumferential membranes compared to their respective parents (Fig. 1c). When invasion into three dimensional fibroblast-conditioned dense collagen matrices was examined, Selected MDA-MB-231 cells were markedly more invasive than Parental or FlowSorted cells (Fig. 1d). In addition, pore-selection changed relatively non-invasive Parental MDA-MB-435 cells into highly invasive Selected cells (Fig. 1d).

To determine how gene expression differs between Parental, Selected and FlowSorted populations of MDA-MB-231 and MDA-MB-435, total RNA was extracted and enriched for polyA⁺ mRNAs, and then subjected to RNA-Sequencing (Fig. 2 and Table 1). The study has been described at the NCBI Bioproject (PRJNA327913), with descriptions of the MDA-MB-231 cells (SAMN07311741) and MDA-MB-435 cells (SAMN07311743). Primary data are available at the NCBI Sequence Read Archive (Table 2 (available online only) and Data Citation 1). The transcriptomic data generated by this study may reveal important modifiers of the physical properties that enable tumour cells to move through narrow spaces, as well as regulators of cell size, that contribute to the metastatic spread of cancer.

Methods

Cell culture

Human MDA-MB-231 D3H2LN Luc⁹ (abbreviated MDA-MB-231) breast cancer cells (Caliper LifeScience, Hopkinton MA USA) and MDA-MB-435 melanoma cells (ATCC, Teddington UK) were grown in Hyclone MEM/EBSS media, supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 10 U/ml penicillin and 10 µg ml⁻¹ streptomycin, 1% MEM/NEAA (non-essential amino acid), 1% Sodium Pyruvate (all Gibco, Fisher Scientific, Loughborough, UK) at 37 °C with 5% CO₂ in a humidified incubator. Cell line identities were validated by the Cancer Research UK Beatson Institute Molecular Services using the GenePrint 10 system STR multiplex assay (Promega, Southampton UK) that amplifies 9 tetranucleotide repeat loci and Amelogenin gender determining marker.

Cell line selection

Independent MDA-MB-231 or MDA-MB-435 pore-selected (Selected) cell populations were established by seeding 1 × 10⁶ cells in 10 ml serum-free medium on 3 µm pore membranes in 7.5 cm cell culture inserts (Corning, Fisher Scientific, Loughborough, UK). Inserts were placed in 10 cm dishes containing 10 ml serum-containing medium, and left for four days in standard tissue culture conditions to allow cells

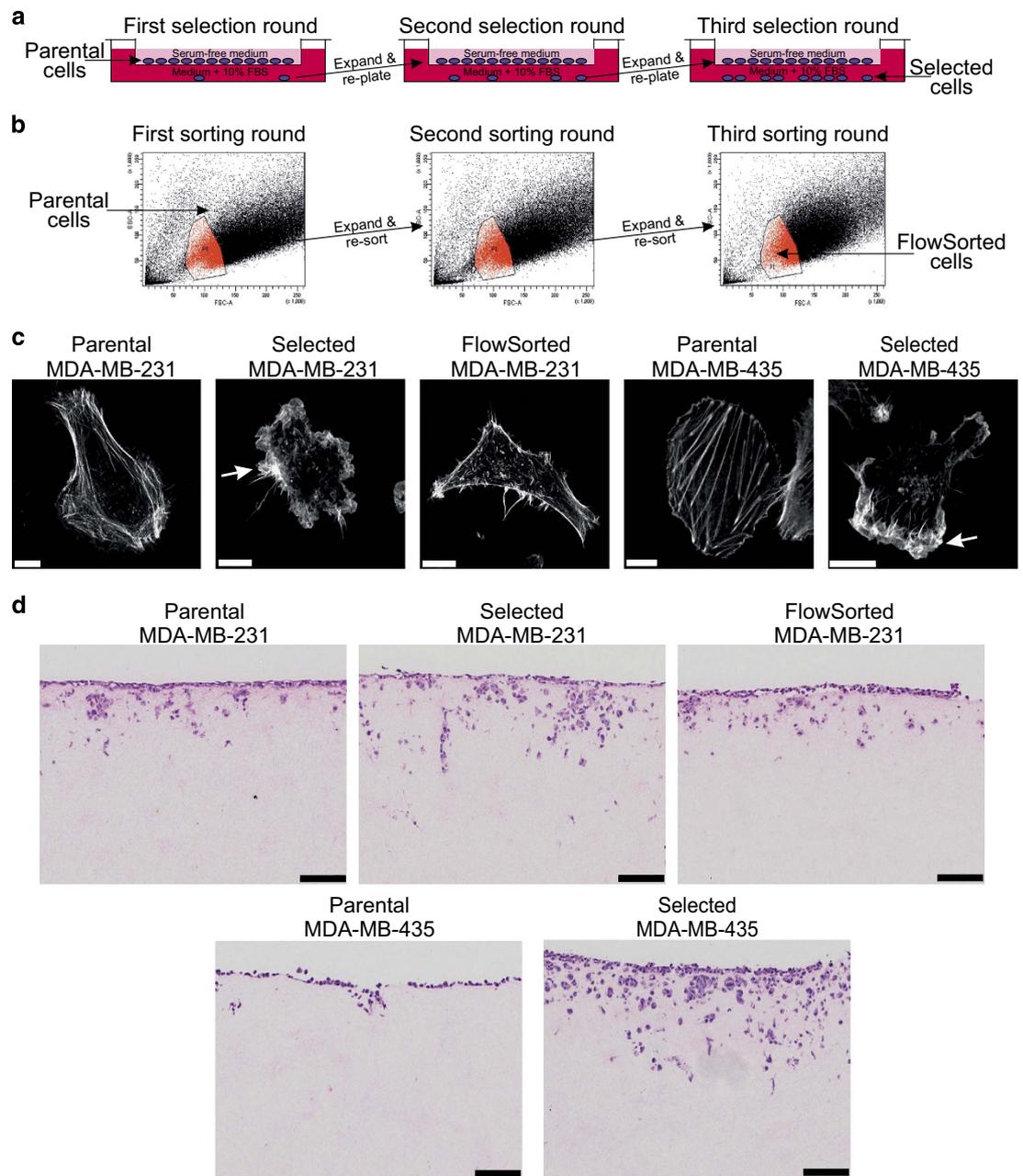


Figure 1. Selection strategies to enrich for invasive or small diameter cells. **(a)** MDA-MB-231 or MDA-MB-435 cells were plated in serum-free medium into tissue culture inserts with 3 μm pores, and placed in tissue culture dishes containing medium with 10% fetal bovine serum (FBS). Cells able to migrate through pores and attach to the underlying dishes were expanded, re-plated and the selection repeated twice more. At the end, several independent Selected cell populations were established. **(b)** MDA-MB-231 cells were sorted by flow cytometry (side scatter and forward scatter) and gated in the P1 region (red) for small diameter cells. After two additional rounds of expansion and sorting, several independent FlowSorted populations were established. **(c)** Representative Parental, Selected and FlowSorted MDA-MB-231 cells, as well as Parental and Selected MDA-MB-435 cells, were fixed, stained with Alexa Fluor488-conjugated phalloidin to enable visualization of filamentous actin structures. Actin-rich protrusive regions have been indicated for Selected cells with white arrows. Scale bars = 10 μm . **(d)** H&E-stained sections of cell invasion into collagen matrix after 8 days. Scale bar = 100 μm .

to migrate through the pores. The inserts were then removed, media was changed and plates were placed back in the incubator to expand the selected cell population. The selection process was repeated twice more as described above.

Independent small diameter MDA-MB-231 flow cytometry sorted (FlowSorted) populations were obtained by gating on cells with low forward scatter and side scatter parameters as indicated in the red region in Fig. 1b using a FACS Aria Fusion sorter (BD Biosciences, Oxford UK). FlowSorted cells were then grown using standard tissue culture conditions to expand the isolated sorted cell populations, followed by two additional rounds of sorting as described above.

Cells were stained and imaged as described in ref. 11. For filamentous actin imaging, 0.8×10^5 cells were seeded on 13 mm autoclaved cover slips and grown overnight. Cells were washed with phosphate buffered saline (PBS), then fixed with 4% paraformaldehyde/PBS solution, permeabilized with 0.5% (v/v) Triton X-100/PBS, blocked with 1% (w/v) bovine serine albumin (Sigma-Aldrich, Gillingham, UK)/PBS, and then incubated with Alexa Fluor488 phalloidin (Sigma-Aldrich, Gillingham, UK) (1:1,000 dilution) for 1 h at room temperature. Cover slips were washed twice with PBS and inverted on 7 μ l of ProLong Diamond Antifade Mountant with DAPI (Thermo Fisher Scientific, Renfrew UK) on a glass slide. Cells were imaged on a Zeiss 880 confocal microscope (Cambridge UK) using a 63X oil objective.

Collagen matrix invasion assays were performed as previously described¹¹. Primary human fibroblasts were mixed with rat tail collagen1 and placed in a cell culture incubator for a week to allow conditioning of the collagen. To remove fibroblasts from the collagen matrices, the disks were incubated with 5 μ g ml⁻¹ Puromycin for at least 24 h and then washed twice with medium. 2×10^5 cells were seeded on top of the disks and allowed to settle and grow over 2 days. Afterward, the collagen matrices were mounted onto grids to generate an air/liquid interface. After 8 days, the collagen disks were fixed in 4% paraformaldehyde overnight and processed using standard histological methods. H&E-stained sections were scanned and analyzed using Digital Slide Server (SlidePath, Leica, Milton Keynes UK) software.

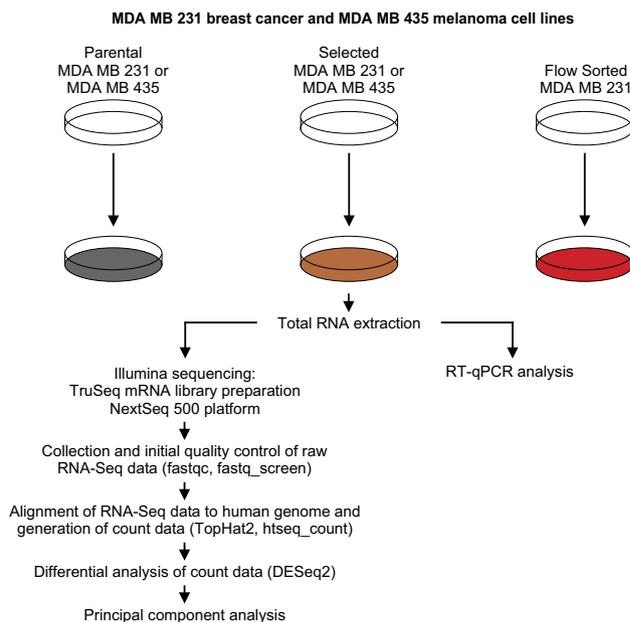


Figure 2. Overview and experimental design of the study.

Cell line	BioSample	Cell population	Enrichment method	Biological replicates	Protocol 1	Protocol 2
MDA-MB-231 D3H2LN Luc	SAMN07311741	Parental	None	4	RNA extraction	RNA-Seq
MDA-MB-231 D3H2LN Luc	SAMN07311741	Selected	Selection through 3 μ m pores	3	RNA extraction	RNA-Seq
MDA-MB-231 D3H2LN Luc	SAMN07311741	FlowSorted	Selection for small diameter by flow cytometry	3	RNA extraction	RNA-Seq
MDA-MB-435	SAMN07311743	Parental	None	4	RNA extraction	RNA-Seq
MDA-MB-435	SAMN07311743	Selected	Selection through 3 μ m pores	4	RNA extraction	RNA-Seq

Table 1. Experimental samples and design.

MDA-MB-231		MDA-MB-435	
Sample name	RINe	Sample name	RINe
Parental1	10.0	Parental1	10.0
Parental2	9.9	Parental2	10.0
Parental3	10.0	Parental3	10.0
Parental4	9.9	Parental4	10.0
Selected2-1	10.0	Selected1-1	10.0
Selected2-2	9.9	Selected1-2	10.0
Selected2-3	10.0	Selected1-3	10.0
Selected3-1	10.0	Selected1-4	9.6
Selected3-2	8.9	Selected2-1	10.0
Selected3-3	9.9	Selected2-2	10.0
FlowSorted1-1	10.0	Selected2-3	10.0
FlowSorted1-2	10.0	Selected2-4	10.0
FlowSorted1-3	10.0		
FlowSorted2-1	9.9		
FlowSorted2-2	10.0		
FlowSorted2-3	10.0		

Table 3. RNA integrity values for each sample used for RNA sequencing.

Primer target	Qiagen catalogue number
<i>LILRB1</i>	QT00024920
<i>TNFSF15</i>	QT00041965
<i>HBEGF</i>	QT00000455
<i>PTPRU</i>	QT00005677
<i>ENPP1</i>	QT00094787
<i>MXRA8</i>	QT00202790
<i>GAPDH</i>	QT00079247

Table 4. PCR primer catalogue numbers.

RNA isolation

1×10^6 cells were seeded into 6-well plates and allowed to settle and grow overnight. Cells were harvested with Trypsin and total RNA was extracted using the RNAeasy kit (Qiagen, Manchester UK) according to manufacturer's instructions. RNA was quantified using the Nanodrop spectrophotometer (Nanodrop, Thermo Fisher Scientific, Renfrew UK). The Agilent RNA ScreenTape assay and the Agilent 2,200 TapeStation system (both Agilent, Stockport UK) were used to determine the RNA integrity number equivalent (RINe; Table 3).

RNA-sequencing

Total RNA was used to generate an oligo dT-enriched library with the Illumina TruSeq RNA Library Preparation kit v2.0 (Illumina, Cambridge UK). Quality and quantity of the DNA library was assessed using the Agilent 2,100 Bioanalyzer (Agilent, Stockport UK) and the Qubit (Thermo Fisher Scientific, Renfrew UK), respectively. The library was run on the Illumina NextSeq 500 platform using the High Output 75 cycles kit (2×36 cycles, paired-end reads, single index) (both Illumina, Cambridge UK).

RNA-sequence analysis

RNA-Sequence analysis and alignment was carried out as described in reference¹². Quality control checks of raw RNA-Seq data files were done with fastqc v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and fastq_screen v0.4.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/). RNA-Seq reads were aligned to the human genome build GRCh38 with TopHat2.0.13¹³ and genome annotation using GRCh38.82.gtf. BAM files were further processed with HTseq0.6.1p1 (<http://www.huber.embl.de/users/anders/HTSeq/doc/count.html>). Differential analysis of count data was performed by the DESeq2 package (DESeq2)¹⁴. Regularized log transformation was used to transform the DESeq2 data for principal component analysis.

MDA-MB-231			MDA-MB-435		
Replicate	versus Replicate	Pearson's r	Replicate	versus Replicate	Pearson's r
Parental1	Parental2	0.9960374	Parental1	Parental2	0.9996197
Parental1	Parental3	0.9965765	Parental1	Parental3	0.9999907
Parental1	Parental4	0.9893210	Parental1	Parental4	0.9999899
Parental2	Parental3	0.9972366	Parental2	Parental3	0.9995527
Parental2	Parental4	0.9951007	Parental2	Parental4	0.9996528
Parental3	Parental4	0.9954966	Parental3	Parental4	0.9999925
Selected2-1	Selected2-2	0.9879274	Selected1-1	Selected1-2	0.9999951
Selected2-1	Selected2-3	0.9949912	Selected1-1	Selected1-3	0.9999855
Selected2-2	Selected2-3	0.9924361	Selected1-1	Selected1-4	0.9999968
Selected3-1	Selected3-2	0.9979480	Selected1-2	Selected1-3	0.9999798
Selected3-1	Selected3-3	0.9968270	Selected1-2	Selected1-4	0.9999967
Selected3-2	Selected3-3	0.9974801	Selected1-3	Selected1-4	0.9999865
FlowSorted1-1	FlowSorted1-2	0.9955907	Selected2-1	Selected2-2	0.9999973
FlowSorted1-1	FlowSorted1-3	0.9969312	Selected2-1	Selected2-3	0.9999610
FlowSorted1-2	FlowSorted1-3	0.9967215	Selected2-1	Selected2-4	0.9999835
FlowSorted2-1	FlowSorted2-2	0.9874292	Selected2-2	Selected2-3	0.9999461
FlowSorted2-1	FlowSorted2-3	0.9990905	Selected2-2	Selected2-4	0.9999739
FlowSorted2-2	FlowSorted2-3	0.9913348	Selected2-3	Selected2-4	0.9999938

Table 5. Pearson's correlation coefficient (r) values for biological replicates.

Quantitative PCR

Total RNA was used to prepare complementary DNA (cDNA) using a Quantitect Reverse Transcription kit (Qiagen, Manchester UK). To perform quantitative PCR, the DyNamo HS SYBR Green qPCR Kit (Thermo Fisher Scientific, Renfrew UK) was used in triplicate in a 20 μ l reaction mixture containing 10 μ l of master mix (master mix contains hot-start polymerase, SYBR green, PCR buffer, 5 mM MgCl₂, and dNTP mix), 6.35 μ l of nuclease-free water, 0.15 μ l of primer, and 0.4 μ l ROX passive reference dye. Primer catalogue numbers (Qiagen, Manchester UK) are in Table 4. Reaction mixtures were distributed into MicroAmp Fast Optical 96-well plates and 1.5 μ l of cDNA sample or standard added to each well. The plate was covered with optically transparent sealing film and run on an Applied 7,500 Fast Real-Time PCR system. A melting curve was performed to validate the presence of single PCR product. Data was analysed on Applied Biosystem 7,500 Software 2.0.5 and the expression level of genes of interests were calculated using Δ Ct method and normalized to GAPDH.

Data Records

A project overview has been submitted as the BioProject reference PRJNA327913, with descriptions of the BioSample MDA-MB-231 D3H2LN Luc cells (reference SAMN07311741) and of the BioSample MDA-MB-435 cells (reference SAMN07311743). Unprocessed RNA-Sequencing reads have been deposited as fastq files at the National Center for Biotechnology Information (NCBI) Sequence Reads Archive (SRA) with the reference SRP111915 (Data Citation 1).

The fastq files correspond to four independent biological replicates (BiolRep1–4) for the MDA-MB-435 melanoma Parental, Selected1 or Selected2 cell populations, or to four independent biological replicates (BiolRep1–4) for the MDA-MB-231 Parental and three independent biological replicates (BiolRep1–3) for the Selected2, Selected3, FlowSorted1 or FlowSorted2 cell populations, as indicated in Table 2 (available online only). For each sequencing reaction run on the Illumina NextSeq 500 instrument, four technical replicates were produced (TechRep1–4). Forward (R1) and reverse (R2) reads have been combined, with SRA accession numbers for the combined sequencing results also indicated in Table 2 (available online only). Please also see the associated Metadata Record.

Technical Validation

The quality of extracted RNA (RNA integrity number equivalent; RINe) of all samples was determined using the Agilent RNA ScreenTape assay and the Agilent 2,200 TapeStation system (Table 3). Following RNA-seq, correlation coefficients were calculated for all pairwise comparisons of biological replicates (Table 5), which were ≥ 0.9874292 for MDA-MB-231 cells and ≥ 0.9995527 for MDA-MB-435 cells. Principal component analysis revealed that the four biological replicates of Parental MDA-MB-231 cells (Fig. 3a; grey symbols) along with the three biological replicates of FlowSorted1 and FlowSorted2 populations (Fig. 3a; magenta and red symbols) clustered together, while the three biological replicates of

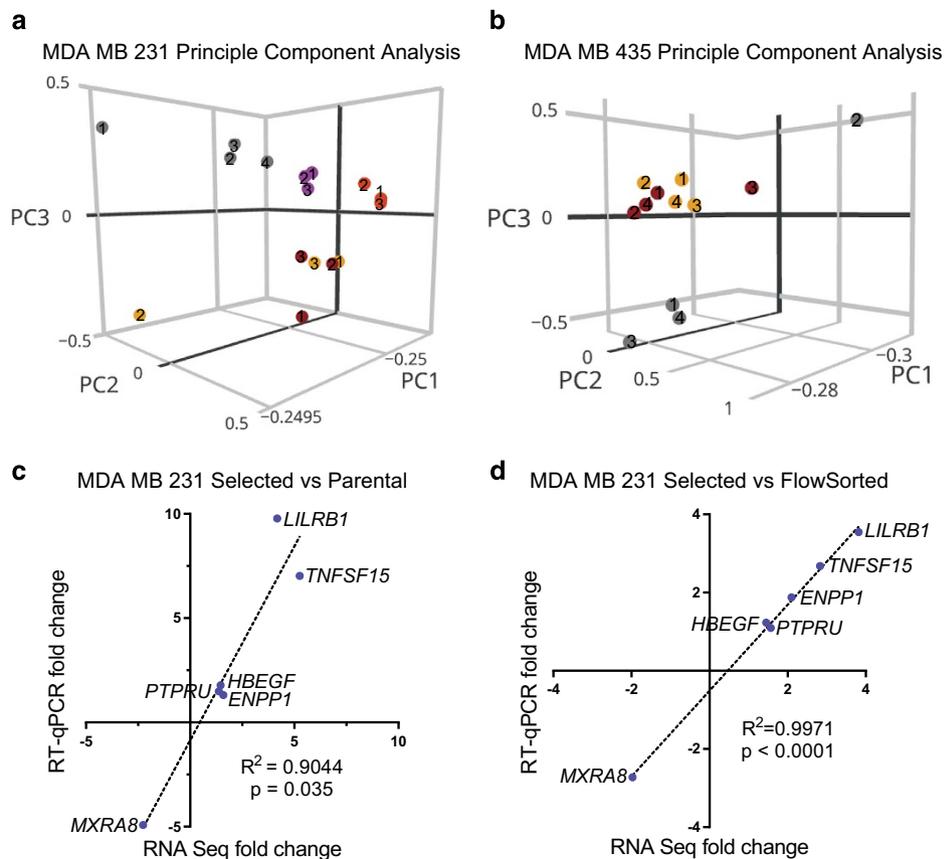


Figure 3. Quality control of RNA-Seq data. (a) Principal component plot of RNA-Seq (DESeq2) data from MDA-MB-231 Parental (grey symbols), Selected2 (orange symbol), Selected3 (brown symbol), FlowSorted1 (magenta symbol) and FlowSorted2 (red symbol) cell populations. Parental and FlowSorted populations clustered distinctly from Selected populations cluster. (b) Principal component plot of RNA-Seq (DESeq2) data from MDA-MB-435 Parental (grey symbols), Selected1 (brown symbol), and Selected2 (orange symbol) cell populations, with the number of the biological replicate indicated within each symbol. Parental cells clustered distinctly from the Selected populations cluster. (c) Linear regression plots of gene expression fold change (relative to *GAPDH*) for Selected MDA-MB-231 cells versus Parental or (d) versus FlowSorted cells obtained by RNA-Seq (DESeq2 data, relative to *GAPDH*) compared to RT-qPCR (relative to *GAPDH*). RNA transcripts expressed at relatively higher levels in Selected cells: *LILRB1*, *TNFSF15*, *HBEGF*, *PTPRU*, *ENPP1*. RNA transcript expressed at relatively lower levels in Selected cells: *MXRA8*. R^2 of Goodness of Fit with P value.

Selected2 and Selected3 populations (Fig. 3a; orange and brown symbols) formed a separate cluster. Similarly, three of the Parental MDA-MB-435 biological replicates (Fig. 3b; grey symbols) clustered together and were separate from the four biological replicates of Selected1 and Selected2 populations (Fig. 3b; orange and brown symbols). These results are consistent with the pore-selection procedure having enriched for independent cell populations with transcriptomic profiles distinct from the originating parental populations.

Further technical validation of the RNA-Sequencing results was provided by quantitative reverse transcription PCR (RT-qPCR) analyses of differences in gene expression between MDA-MB-231 Parental versus Selected (Fig. 3c) or MDA-MB-231 Selected versus FlowSorted cells (Fig. 3d) identified by RNA-Seq. Genes were selected on the basis of fold-change, statistical significance and number of sequence reads. The relatively higher expression of the genes encoding leukocyte immunoglobulin-like receptor subfamily B member 1 (*LILRB1*), tumour necrosis factor ligand superfamily member 15 (*TNFSF15*), ectonucleotide pyrophosphatase/phosphodiesterase 1 (*ENPP1*), protein tyrosine phosphatase receptor type U (*PTPRU*), heparin-binding epidermal growth factor (*HBEGF*), as well as the relatively lower expression of matrix remodelling associated 8 (*MXRA8*), in Selected versus Parental (Fig. 3c) as well as in Selected versus FlowSorted cells (Fig. 3d) were comparable in both RNA-Seq and RT-qPCR assays. In both cases, the fold-changes determined by either method fell on a single fitted straight line with $R^2 > 0.90$ and $P < 0.05$ (Fig. 3c,d). Although the agreement between RNA-Seq and RT-qPCR was good for this

limited set of genes, it is formally possible that the analytical methods employed might have underestimated gene expression levels¹⁵.

References

1. Sporn, M. B. The war on cancer. *Lancet* **347**, 1377–1381 (1996).
2. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559–1564 (2011).
3. Bernards, R. & Weinberg, R. A. Metastasis genes: A progression puzzle. *Nature* **418**, 823–823 (2002).
4. Waclaw, B. *et al.* A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* **525**, 261–264 (2015).
5. Olson, M. F. & Sahai, E. The actin cytoskeleton in cancer cell motility. *Clin. Exp. Metastasis* **26**, 273–287 (2009).
6. Kawauchi, T. Cell adhesion and its endocytic regulation in cell migration during neural development and cancer metastasis. *Int. J. Mol. Sci.* **13**, 4564–4590 (2012).
7. Wolf, K. & Friedl, P. Extracellular matrix determinants of proteolytic and non-proteolytic cell migration. *Trends Cell Biol.* **21**, 736–744 (2011).
8. Weihs, D., Mason, T. G. & Teitell, M. A. Bio-microrheology: a frontier in microrheology. *Biophys. J.* **91**, 4296–4305 (2006).
9. Jenkins, D. E., Hornig, Y. S., Oei, Y., Dusich, J. & Purchio, T. Bioluminescent human breast cancer cell lines that permit rapid and sensitive in vivo detection of mammary tumors and multiple metastases in immune deficient mice. *Breast Cancer Res.* **7**, R444–R454 (2005).
10. Ellison, G. *et al.* Further evidence to support the melanocytic origin of MDA-MB-435. *Mol. Pathol.* **55**, 294–299 (2002).
11. Rath, N. *et al.* ROCK signaling promotes collagen remodeling to facilitate invasive pancreatic ductal adenocarcinoma tumor cell growth. *EMBO Mol. Med* **9**, 198–218 (2017).
12. Rath, N., Kalna, G., Clark, W. & Olson, M. F. ROCK signalling induced gene expression changes in mouse pancreatic ductal adenocarcinoma cells. *Sci. Data* **3**, 160101 (2016).
13. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
14. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
15. Robert, C. & Watson, M. Errors in RNA-Seq quantification affect genes of relevance to human disease. *Genome Biol.* **16**, 1–16 (2015).

Data Citations

1. *NCBI Sequence Read Archive* SRP111915 (2017).

Acknowledgements

Funding for this project was from Cancer Research UK (A18276).

Author Contributions

Experiments were designed by D.A.R. and M.F.O., and performed by D.A.R. Enrichment of total RNA for polyA⁺ fraction, library construction and sample analysis on the Illumina NextSeq 500 platform by W.C. Sequence assembly and statistical analysis were performed by A.H. and G.K. Manuscript was written and figures prepared by D.A.R. and M.F.O.

Additional Information

Table 2 is only available in the online version of this paper.

Competing interests: The authors declare no competing financial interests.

How to cite this article: Rudzka, D. A. *et al.* Transcriptomic profiling of human breast and melanoma cells selected by migration through narrow constraints. *Sci. Data* 4:170172 doi: 10.1038/sdata.2017.172 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.