# Transcriptomic Profiling of Sequential Tumors from Breast Cancer Patients Provides a Global View of Metastatic Expression Changes Following Endocrine Therapy.

AUTHOR(S)

Jean McBryan, Ailís Fagan, Damian McCartan, Fiona T. Bane, Damir Varešlija, Sinéad Cocchiglia, Christopher Byrne, Jarlath Bolger, Marie McIlroy, Lance Hudson, Paul Tibbitts, Peadar Ó Gaora, Arnold DK Hill, Leonie S. Young

**Transcriptomic profiling of sequential tumours from breast cancer patients provides a global view of metastatic expression changes following endocrine therapy**

Jean McBryan*[1], Ailís Fagan*[1], Damian McCartan[1], Fiona T. Bane[1], Damir Varešlija[1], Sinéad Cocchiglia[1], Christopher Byrne[1], Jarlath Bolger[1], Marie McIlroy[1], Lance Hudson[1], Paul Tibbitts[1], Peadar Ó Gaora[2], Arnold D. Hill[1] and Leonie S. Young[1].

*These authors contributed equally to this work.

[1]Endocrine Oncology Research Group, Department of Surgery, Royal College of Surgeons in Ireland, Dublin 2, Ireland.

[2]UCD School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Dublin 4, Ireland.

**Running title:** Transcriptomic profiles of metastatic breast cancer

**Key words:** Breast neoplasms; Neoplasm metastasis; Transcriptome; Tamoxifen; Drug resistance.

**Financial Support:** We kindly acknowledge the funding support from Science Foundation Ireland (08-IN1-B1853 and 12/1A/1294), the Health Research Board of Ireland (HRB/POR/2012/101) and Breast Cancer Campaign (2013MaySP022). This material is also based upon works supported by the Irish Cancer Society Collaborative Cancer Research Centre grant, CCRC13GAL.

**Corresponding author:** Leonie Young, Endocrine Oncology Research Group, Department of Surgery, Royal College of Surgeons in Ireland, Dublin 2, Ireland. Tel: +353 1 4028576. lyoung@rcsi.ie

**Conflict of interest:** The authors declare there is no conflict of interest.

**Text word count:** 4,028

**Total number of figures and tables:** 4

**TRANSLATIONAL RELEVANCE:**

To date, next generation sequencing has been completed for a substantial number of primary breast tumours and provides an in depth understanding of the biology and heterogeneity of this disease. Here primary tumours together with matched nodal and liver metastatic patient tumours have been sequenced, offering the first in depth view of the transcriptomic changes occurring during disease progression. Patient heterogeneity was strongly evident in both primary and metastatic tumours. In addition, common adaptations in cell-cell communication were detected as metastatic endocrine-resistant disease took hold. These findings emphasise the need for ongoing personalised medicine throughout disease management. In particular, it is important to re-evaluate the biology of metastatic disease, following treatment intervention, in order to monitor adaptations and identify the optimal treatment strategy.

**ABSTRACT**

**Purpose** Disease recurrence is a common problem in breast cancer and yet the mechanisms enabling tumour cells to evade therapy and colonise distant organs remain unclear. We sought to characterise global expression changes occurring with metastatic disease progression in the endocrine resistant setting.

**Experimental Design** Here, for the first time, RNAsequencing has been performed on matched primary, nodal and liver metastatic tumours from tamoxifen-treated patients following disease progression. Expression of genes commonly elevated in the metastases of sequenced patients was subsequently examined in an extended matched patient cohort with metastatic disease from multiple sites. The impact of tamoxifen treatment on endocrine resistant tumours *in vivo* was investigated in a xenograft model.

**Results** The extent of patient heterogeneity at the gene level was striking. Less than 3% of the genes differentially expressed between sequential tumours were common to all patients. Larger divergence was observed between primary and liver tumours than between primary and nodal tumours, reflecting both the latency to disease progression and the genetic impact of intervening therapy. Furthermore, an endocrine-resistant *in vivo* mouse model demonstrated that tamoxifen treatment has the potential to drive disease progression and establish distant metastatic disease. Common functional pathways altered during metastatic, endocrine-resistant progression included extracellular matrix receptor interactions and focal adhesions.

**Conclusions** This novel global analysis highlights the influence of primary tumour biology in determining the transcriptomic profile of metastatic tumours, as well as the need for adaptations in cell-cell communications to facilitate successful tumour cell colonisation of distant host organs.

**INTRODUCTION**

The vast majority of breast cancers are steroid receptor positive and depend on estrogen for growth. The use of estrogen receptor modulators, such as tamoxifen and estrogen depletion strategies, including aromatase inhibitors, have been widely successful. Disease recurrence however is common, with between 10 and 25% of patients developing extensive local or distant metastases. Most patients with metastases will succumb to their disease, as targeted therapies which provide significant clinical benefit remain elusive.

Metastatic breast cancer secondary to endocrine treatment is heterogeneous, ranging from discreet lesions to diffuse multi-organ disease. The dominant site of metastatic involvement can have a significant bearing on clinical outcome. Metastatic disease of the liver has been reported to be a predictor of poor outcome, with a median survival of 18 months (1), whereas metastatic burden confined to the bone is thought to be more indolent (2).

Colonisation of the host organ is the ultimate step of a progressive disease path. Most breast cancer cells that enter the circulation and infiltrate distant organs die due to the hostile nature of the host microenvironment (3). For metastatic growth to be established, disseminated breast cancer cells need to survive a period of latency, prolonged exposure to therapy and subsequently re-initiate growth when appropriate. Evidence of the impact of steroid treatment on tumour progression is emerging. Steroid receptor switching and in particular loss of the progesterone receptor (PgR), an estrogen receptor (ER) target, is observed in approximately 30% of patients (4, 5). Furthermore, gain of function mutations in the *ESR1* gene have been identified in metastatic lesions from endocrine treated patients (6, 7).

The mechanisms required to enable cancer cells to overcome hostile forces and re-initiate growth at distant organs in endocrine treated patients have not been fully elucidated. Lessons from *in vitro* studies and murine models have provided important information regarding discrete tools employed by breast cancer cells to prosper at distant metastatic sites. For example, the CXCR4/SDF1 pairing

has been identified as a mediator of metastatic cell survival in the bone (8, 9); transcriptional inhibitors of differentiation, Id1 and Id3, have been associated with lung metastasis (10, 11); and serpins, as metastatic functionaries in the brain (12). However, to date, no global overview of gene expression alterations which are essential for successful colonisation of cells at metastatic sites in breast cancer patients has been undertaken.

To clearly define mediators of colonisation in endocrine treated patients we undertook gene expression analysis from sequential tumour samples. RNAseq analysis was performed on matched primary, nodal and distant metastatic tumours from breast cancer patients, all of whom had developed a liver metastasis following tamoxifen treatment. At a functional level, studies in xenograft models of endocrine resistant tumours were undertaken to examine the role of tamoxifen in the development of metastatic disease secondary to therapy resistance. Over represented pathways identified by the RNAseq analysis were further studied at the transcript and protein level in an expanded cohort of patients with metastases of multiple distant organs.

## MATERIALS AND METHODS

### Gene expression and bioinformatic analysis

Following ethical approval, formalin fixed paraffin embedded (FFPE) tumour samples were macrodissected to select regions with >70% tumour cells. RNA isolation was performed with High Pure FFPE RNA Micro Kit (Roche) and subjected to duplex specific thermostable nuclease (DSN) treatment. RNA Sequencing was performed by BGI (Hong Kong) using the Illumina HiSeq 2000 with >40million reads. Insufficient RNA was obtained from the nodal sample of patient one for sequencing. Reads were aligned to the hg19 genome using two different types of analysis: one for the differential expression and the other for use with the TCGA data. For the differential expression analysis, the reads were aligned using TopHat and genes were quantified using HTseq. For usage with TCGA data, the data was aligned using MapSplice and the genes were quantified using RSEM.

Full details of bioinformatics analysis are provided in Supplementary Methods. Data from this study has been deposited in the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE58708.

cDNA was synthesised as previously described (13). Gene expression was assessed using Taqman technology (Applied Biosystems) on the ABI PRISM 7500 platform with the following probes: EPPK1 Hs01104050_s1; COL4A2 Hs01098873_m1; COL4A1 Hs00266237_m1; SPDEF Hs01026050_m1; KRT19 Hs00761767_s1; AHDC1 Hs00210424_m1; NCOR2 Hs00196955_m1; AKAP8L Hs00205106_m1. EIF2B Hs00426752_ml was used as an internal control. The comparative $C_T$ ($\Delta\Delta C_T$) method was applied to analyse relative mRNA expression levels.

**Immunohistochemistry**

5µm tumour sections were deparaffinised and rehydrated as previously described (4). Antigen retrieval was performed with sodium citrate at pH6 or EDTA at pH9 (Ki67). Dako Envision+ HRP kit was used as per manufacturer's instructions. Primary antibodies incubated for 1 hour at room temperature were: Ki67 (1:200, MIB-1, M7240, Dako); KRT19 (1:200, HPA002465, Sigma); LMNB2 (1:75, MAB3536, Merck Millipore); KIF12 (1:100, NBPI-86029, Novus); EVPL (1:30, Ab49397, Abcam); MCM4 (1:50, HPA004873, Sigma); p-mTOR (1:100, Ser2448, #2976, Cell Signalling); and active β-catenin (1:150, 05-665, Merck Millipore). The following primary antibodies were incubated overnight at 4°C: p-p70S6K (1:100, Thr389, #9205, Cell Signalling); p-HER2 (1:400, Tyr1248, 06-229, Merck Millipore); p-STAT3 (1:100, Tyr705, #9131, Cell Signalling); and p-ERK 1/2 (1:400, Thr202/Tyr204, #4370, Cell Signalling). ER (1:60, RM9101S, Thermo Fisher Scientific), PgR (1:60, H3569, DAKO) and HER2 (neat, 4B5, 790-2991, Ventana) were stained on an automated program. H-score was assigned by two independent assessors on a scale of 0-300 with 300 representing the highest intensity staining in 100% of cells.

**Xenograft Studies**

All mouse experiments were performed in accordance with the European Communities Council Directive (86/609/EEC) and were reviewed and approved by Research Ethics Committee under license from the Department of Health. Ten 6-week old female BALB/c SCID mice (Harlan) were implanted with 17-B-estradiol pellets (0.36 mg/pellet, 60-day release) alone or in combination with 4-hydroxytamoxifen pellets (5 mg/pellet, 60-day release). All pellets were obtained from Innovative Research of America and were readministered at week 8. Luciferase-tagged LY2 cells ($1 \times 10^6$ cells) were mixed with 50% Matrigel (BD biosciences) and implanted into the $4^{th}$ inguinal mammary gland by injection. Tumours were imaged using an IVIS whole body imaging system (Xenogen Corp) to detect luciferase activity. Five weeks after cell implantation, primary tumours were surgically removed, formalin fixed and paraffin embedded. The mice were monitored by IVIS once per week for the next 10 weeks to detect metastasis of the labelled cells. The fold increase in metastatic flux was compared by t-test between each group of animals.

**RESULTS**

**Gene expression profiles of breast cancer patients reflect the histopathology of the original primary tumour**

RNA was extracted from matched breast cancer primary, node and liver metastatic tumour tissue from three tamoxifen treated luminal breast cancer patients. The patients represented the spectrum of the luminal classification, from luminal A (patient one) to luminal B2 (patient three) (Table 1). Despite the diverse pathology of each of the primary breast tumours (Table 1, Fig. 1A and Supplementary Fig. S1), positive steroid receptor status qualified each of the patients for adjuvant endocrine treatment.

Correspondence analysis of the sequenced RNA from the individual tissue samples revealed that each of the sequential patient samples clustered together, rather than with the metastatic site (Fig 1B). The relative divergence in gene expression from the primary to the metastatic tumours was reflective of the disease latency period. Patient one with a recorded metastasis-free survival period of 11 years had the greatest difference in gene expression profile between the primary and metastatic tumours (Fig. 1B). Patient two and patient three had shorter disease-free periods and correspondingly substantially less alterations in gene expression, suggesting that some of the genes necessary for successful progression are present in the primary tumour.

RNAseq data from the breast cancer sequential samples was analysed with breast cancer samples from The Cancer Genome Atlas by correspondence analysis of almost 16,000 genes (Fig. 1C). Nodal and metastatic samples clustered in amongst the primary tumours. PAM50 genes were used to classify the breast cancer sequential samples and the TCGA samples into the major breast cancer subtypes. Tumours from patients one and two clustered on the luminal end of the horizontal axis. The tumours from patient three straddled the luminal/HER2 border. Interestingly, while the primary tumour from patient one was luminal A, the metastatic tumour from this patient was classified as luminal B

**Metastatic gene expression alterations are heterogeneous in endocrine treated breast cancer patients**

To define the changes involved in disease progression, differentially expressed genes (DEGs) were identified between matched primary and liver metastatic tumours from each patient. This analysis identified 1,268 upregulated and 2,606 downregulated genes between matched primary and metastatic tumours. Remarkably, less than 3% of DEGs were common to all three patients highlighting the extent of patient heterogeneity (Fig. 2A). Only 31 upregulated and 69 downregulated genes were shared by the three patients (Supplementary Tables S1 and S2). The individuality of each patient is evident in the heat maps of DEGs where each patient elevated their

own unique set of genes in the liver metastatic sample (Fig. 2B). Of note, little difference in gene expression was detected between the primary tumour and the nodal metastases, both of which were treatment naïve and were resected at the time of initial surgery (Fig. 2B).

Analysis of enriched KEGG pathway terms was performed for the DEGs of each patient and highlighted a number of important pathways that were unique to each patient (Supplementary Fig. S2 and S3). The luminal A tumour from patient one was highly endocrine dependent with strong expression of both ER and PgR. The large divergence from primary to metastasis on correspondence analysis as well as the large number of differentially expressed genes is indicative of the multiple adaptations that were required for successful tumour progression and colonisation. Following endocrine treatment, steroid receptor status remained strong, however alterations in growth factor signalling networks, including MAPkinase, Jak-STAT and ErbB2 signalling pathways were observed (Supplementary Fig. S2). As protein function cannot always be directly inferred from gene expression analysis, expression of phosphoproteins from these pathways was examined and confirmed in matched primary and metastatic tumours from a luminal A patient (Supplementary Fig. S4). In contrast to patient one, the luminal B1 primary tumour from patient two displayed limited endocrine dependence, with low steroid receptor expression and no HER2. Limited divergence on correspondence analysis and relatively low number of gene alterations suggest that the genes required for metastatic progression may have been present in the primary tumour. Pathways that were altered in the liver metastatic tumour include DNA replication and cell cycle (Supplementary Fig. S2), also confirmed at the protein level (MCM4) (Supplementary Fig. S4). The luminal B2 tumour from patient three was strongly positive for ER, PgR and HER2. From correspondence analysis, genes from the primary, node and liver metastasis clustered with the HER2 subtype using the PAM50 classification. Alterations in the WNT signalling pathway observed in the metastasis, at both the gene and phosphoprotein level (Supplementary Fig. S2 and S4), signify tumour adaptation to endocrine therapy and activation of the epithelial to mesenchymal transition network.

Of note, the mTOR pathway which is a known predictor of endocrine resistance (14), was found to be elevated in the liver metastases of patients one and three at the transcript level in comparison to the primary tumour (Supplementary Fig. S2). Expression levels of phospho-mTOR and phospho-p70S6 kinase in matched primary and metastatic tumours from endocrine treated patients confirmed activation of the mTOR pathway (Supplementary Fig. S5).

**Cell to cell communication is a common pathway in liver metastasis**

Although at the gene level, patient heterogeneity was very apparent in the endocrine resistant metastatic tumours, a number of functional pathways were common to all three patients (Fig. 2C). Alterations in several cancer pathways including small cell lung carcinoma, renal cell carcinoma and pathways in cancer were observed. Of interest, pathways important in cell to cell interaction were represented in gene sets that were both enhanced and suppressed in the metastatic liver. Alterations in genes important in ECM receptor interactions and focal adhesion indicate the importance of communication between the tumour and the host liver for successful colonisation.

**Endocrine treatment contributes to the development of metastasis secondary to endocrine resistance**

Of the sequential tumour samples sequenced here, primary and nodal tumours were biopsied at the same time. Liver tumours were biopsied following a latency period and also following treatment intervention. To determine the impact of endocrine treatment on the development of distant metastasis in endocrine resistant tumours we employed a xenograft model. Endocrine resistant cells were implanted into the mammary fat pad of immune-compromised mice in the presence of estrogen (Fig. 3A). Once established, the primary tumours were resected. Local disease recurrence occurred in animals treated and untreated with the SERM tamoxifen. Distant metastasis however was only established in the tamoxifen treated group (Fig. 3A). Metastatic disease was observed in

the bone, lung and liver (Fig. 3B). Expression of the proliferation marker, Ki67 and keratin 19, a marker of disseminated cancer cells, which was elevated in all of the patient liver metastases, was found to be enhanced in the xenograft liver metastatic tissue compared with the primary (Fig. 3B). These data demonstrate that tamoxifen has the potential to contribute to the development of metastatic disease progression in endocrine resistant tumours. Analysis of the patient RNAseq data revealed that the majority of the identified upregulated DEGs were in fact elevated in metastatic tumours relative to both matched primary and nodal tumours (Fig. 3C). These data suggest that for the majority of these genes, expression did not increase when the cells first moved away from the primary site to the node but did increase following treatment with tamoxifen, the latency period and intravasation of the distant organ. Remarkably, of the common DEGs displaying this pattern of expression (Supplementary Table S3), almost half have defined roles in cell to cell communication and cell structure (Supplementary Fig. S6).

**Cellular communication and cell structure proteins are evident at other metastatic sites**

An extended cohort of matched primary and metastatic patient samples were employed to examine the RNA and protein expression of shared genes from the RNAseq data. The patient cohort included a range of metastatic sites including the liver, lung, bone and peritoneum of endocrine treated patients (Supplementary Table S4). Real-time PCR confirmed the elevated expression of genes in liver metastases relative to matched primary tumours (Fig. 3D). Increases in gene expression were also detected at other metastatic sites (Fig. 3D).

At the protein level, immunohistochemical analysis of matched primary and metastatic tumours revealed strong keratin 19 protein expression, consistent with that observed in the metastatic tumours from the tamoxifen treated xenografts (Fig. 3E). Protein expression of the cellular communication proteins, lamin B2, Kif12 and envoplakin were confirmed in the metastatic tumours

of the endocrine treated patients. Greater expression of these proteins was found in the metastatic tumour cells in comparison to the surrounding normal host tissue (Fig. 3E). Furthermore analysis of unmatched metastatic brain datasets revealed enhanced expression of our defined 31 gene set in the metastatic tumours compared with normal brain tissue (Supplementary Fig. S7). Taken together these data support a role for cellular communication in breast cancer colonisation of distant organs following treatment resistance.

**DISCUSSION**

The ability of selected tumours to adapt and evade endocrine therapy is well established. Enhanced expression of plasticity networks (4, 15, 16), as well as mediators of EMT (17, 18), in the primary tumour has been associated with treatment failure. Information regarding changes which occur with tumour progression on treatment however is not readily available. Here we describe the first global transcriptomic analysis of sequential primary and metastatic tumour samples from endocrine resistant breast cancer patients.

Several initiating events of tumour adaptability have been proposed, including the degree of intra-tumour heterogeneity (19, 20) as well as the presence of stem cell populations (21). Though the source of the treatment adaptability is not addressed in this study, the resultant mediators which enable the successful colonisation of the tumour at a distant site are elucidated.

Here global mapping of gene expression alterations which occur in metastatic breast cancers revealed a significant degree of patient heterogeneity. The relative divergence in gene expression reflected the histology of the original tumour and the disease latency period. While disease progression in some cancers is rapid (22, 23), in ER positive breast cancer, metastasis may occur decades after resection of the primary tumour (24). The greatest degree of divergence in gene expression was observed in the ER positive PgR positive luminal A patient (patient one) with a metastasis-free survival period of >10 years. Differences in gene expression suggest that the primary

tumour underwent significant alterations over an extended period of time to successfully develop liver metastasis. In contrast, the low ER positive, PgR/HER2 negative tumour (patient two) displayed relatively little alteration in gene expression between the primary and the metastatic tumours. These data suggest that the primary tumour from patient two may harbour metastatic progression genes (8) which could have altered functions at the primary and metastatic sites. The extent of divergence in gene expression between primary and metastatic tumours would therefore appear to be inversely related to the aggressiveness of the primary tumour.

Signal pathways enriched at the metastatic site were also patient specific and closely related to the biology of the primary tumour. The ER/PgR/HER2 positive patient (patient three) displayed elevations in plasticity networks, including the WNT signalling pathways, whereas the ER/PgR positive, HER2 negative patient (patient one) had enhanced growth factor signalling. Increases in de-differentiation signals and in particular induction of steroid receptor/growth factor receptor cross talk have been well described as mechanisms of tumour adaptability to endocrine treatment (16, 25, 26). Furthermore elevations in key components of the mTOR pathway were observed in both patient one and patient three in metastatic liver tissue. Activation of the mTOR pathway has previously been associated in endocrine resistance (14, 27) and elevated expression of p-mTOR, p-4EBP1 and p-p70S6K has been reported in metastatic tumours in comparison to matched primary tissue (28).

The impact of sequential endocrine therapy in the treatment of endocrine resistant disease in patients however remains to be fully clarified. Though several clinical studies suggest that some metastatic patients respond to continued endocrine treatment, a subset of patients receives little or no benefit (29-31). Here using endocrine resistant xenografts to model sequential treatment strategies we found that the development of distant metastasis was dependent on continued tamoxifen treatment. Though the molecular mechanisms driving this continued resistance have not been described to date, *ex vivo* studies demonstrating gain of function mutations in breast metastatic patient tumours following endocrine treatment suggest that ER may remain an important

player (7). In this study the majority of DEGs were elevated in the metastatic tumours relative to both the primary and the node, with little alteration observed between the primary and the matched nodal tissue. The gene expression changes that occur during the latency period, under the pressure of treatment are therefore those that are important for the establishment of distant metastatic disease.

Though several elegant studies have employed *in vitro* and *in vivo* models to determine key players in breast cancer colonisation of distant organs, including the bone, lung and brain (8-12), there is little information regarding signalling networks important in metastatic liver disease. In this study common pathways elevated in liver metastatic tumours included pathways in cancer, ECM receptor interactions and focal adhesions. Extracellular matrix proteins have previously been shown to be important in colon cancer liver metastasis (32), moreover enhanced Claudin 2-extracellular matrix interactions have been demonstrated in liver-aggressive breast cancer *in vivo* models (33). The ability of breast tumour cells to communicate with and impact host cells would therefore appear to be essential to its ability to colonise and survive at a distant site.

Elucidation of differential gene expression in metastatic tumours is a key step to understanding the complex mechanisms controlling tumour adaptability. Here we define gene expression alterations which occur in endocrine treated metastatic liver tumours. These global studies build on model systems and clinical observations, which along with further metastatic sequencing studies will enable rational personalised treatment strategies to be developed for endocrine resistant breast cancer patients.

**AUTHORS' CONTRIBUTION**

Conception and design: D.M., M.M., L.Y.

Development of methodology: J.M., A.F., D.M., C.B.

Acquisition of data: J.M., A.F., D.M., F.B., D.V., S.C., C.B., J.B., L.H., P.T.

Analysis and interpretation of data: J.M., A.F., P.O., L.Y.

Writing, review, and/or revision of the manuscript: J.M., A.F., L.Y.

Study supervision: P.O., A.H., L.Y.

**ACKNOWLEDGEMENTS**

The authors thank Robert Clarke, Georgetown University, for his gift of LY2 cells.

**REFERENCES**

1. Nicolini A, Giardino R, Carpi A, Ferrari P, Anselmi L, Colosimo S, et al. Metastatic breast cancer: an updating. Biomed Pharmacother. 2006;60:548-56.
2. Sherry MM, Greco FA, Johnson DH, Hainsworth JD. Metastatic breast cancer confined to the skeletal system. An indolent disease. Am J Med. 1986;81:381-6.
3. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am J Pathol. 1998;153:865-73.
4. McBryan J, Theissen SM, Byrne C, Hughes E, Cocchiglia S, Sande S, et al. Metastatic progression with resistance to aromatase inhibitors is driven by the steroid receptor coactivator SRC-1. Cancer Res. 2012;72:548-59.
5. Broom RJ, Tang PA, Simmons C, Bordeleau L, Mulligan AM, O'Malley FP, et al. Changes in estrogen receptor, progesterone receptor and Her-2/neu status with time: discordance rates between primary and metastatic breast cancer. Anticancer Res. 2009;29:1557-62.
6. Robinson DR, Wu YM, Vats P, Su F, Lonigro RJ, Cao X, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat Genet. 2013;45:1446-51.
7. Jeselsohn R, Yelensky R, Buchwalter G, Frampton G, Meric-Bernstam F, Gonzalez-Angulo AM, et al. Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. Clin Cancer Res. 2014;20:1757-67.
8. Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009;9:274-84.
9. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell. 2003;3:537-49.
10. Gupta GP, Perk J, Acharyya S, de Candia P, Mittal V, Todorova-Manova K, et al. ID genes mediate tumor reinitiation during breast cancer lung metastasis. Proc Natl Acad Sci U S A. 2007;104:19506-11.
11. Swarbrick A, Roy E, Allen T, Bishop JM. Id1 cooperates with oncogenic Ras to induce metastatic mammary carcinoma by subversion of the cellular senescence response. Proc Natl Acad Sci U S A. 2008;105:5402-7.
12. Valiente M, Obenauf AC, Jin X, Chen Q, Zhang XH, Lee DJ, et al. Serpins promote cancer cell survival and vascular co-option in brain metastasis. Cell. 2014;156:1002-16.

13. O'Hara J, Vareslija D, McBryan J, Bane F, Tibbitts P, Byrne C, et al. AIB1:ERalpha transcriptional activity is selectively enhanced in aromatase inhibitor-resistant breast cancer cells. Clin Cancer Res. 2012;18:3305-15.
14. Beelen K, Opdam M, Severson TM, Koornstra RH, Vincent AD, Wesseling J, et al. Phosphorylated p-70S6K predicts tamoxifen resistance in postmenopausal breast cancer patients randomized between adjuvant tamoxifen versus no systemic treatment. Breast Cancer Res. 2014;16:R6.
15. Pathiraja TN, Nayak SR, Xi Y, Jiang S, Garee JP, Edwards DP, et al. Epigenetic reprogramming of HOXC10 in endocrine-resistant breast cancer. Sci Transl Med. 2014;6:229ra41.
16. McIlroy M, McCartan D, Early S, P OG, Pennington S, Hill AD, et al. Interaction of developmental transcription factor HOXC11 with steroid receptor coactivator SRC-1 mediates resistance to endocrine therapy in breast cancer [corrected]. Cancer Res. 2010;70:1585-94.
17. Loh YN, Hedditch EL, Baker LA, Jary E, Ward RL, Ford CE. The Wnt signalling pathway is upregulated in an in vitro model of acquired tamoxifen resistant breast cancer. BMC Cancer. 2013;13:174.
18. Manavalan TT, Teng Y, Litchfield LM, Muluhngwi P, Al-Rayyan N, Klinge CM. Reduced expression of miR-200 family members contributes to antiestrogen resistance in LY2 human breast cancer cells. PLoS One. 2013;8:e62334.
19. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499:214-8.
20. Kleppe M, Levine RL. Tumor heterogeneity confounds and illuminates: assessing the implications. Nat Med. 2014;20:342-4.
21. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature. 2013;501:328-37.
22. Janne PA, Freidlin B, Saxman S, Johnson DH, Livingston RB, Shepherd FA, et al. Twenty-five years of clinical research for patients with limited-stage small cell lung carcinoma in North America. Cancer. 2002;95:1528-38.
23. Briele HA, Das Gupta TK. Natural history of cutaneous malignant melanoma. World J Surg. 1979;3:255-70.
24. Lee YT. Patterns of metastasis and natural courses of breast carcinoma. Cancer Metastasis Rev. 1985;4:153-72.
25. Arpino G, Wiechmann L, Osborne CK, Schiff R. Crosstalk between the estrogen receptor and the HER tyrosine kinase receptor family: molecular mechanism and clinical implications for endocrine therapy resistance. Endocr Rev. 2008;29:217-33.
26. Montemurro F, Di Cosimo S, Arpino G. Human epidermal growth factor receptor 2 (HER2)-positive and hormone receptor-positive breast cancer: new insights into molecular interactions and clinical implications. Ann Oncol. 2013;24:2715-24.
27. Tokunaga E, Hisamatsu Y, Tanaka K, Yamashita N, Saeki H, Oki E, et al. Molecular mechanisms regulating the hormone sensitivity of breast cancer. Cancer Sci. 2014;105:1377-83.
28. Beelen K, Hoefnagel LD, Opdam M, Wesseling J, Sanders J, Vincent AD, et al. PI3K/AKT/mTOR pathway activation in primary and corresponding metastatic breast tumors after adjuvant endocrine therapy. Int J Cancer. 2014;135:1257-63.
29. Barrios C, Forbes JF, Jonat W, Conte P, Gradishar W, Buzdar A, et al. The sequential use of endocrine treatment for advanced breast cancer: where are we? Ann Oncol. 2012;23:1378-86.
30. Goss PE, Ingle JN, Pater JL, Martino S, Robert NJ, Muss HB, et al. Late extended adjuvant treatment with letrozole improves outcome in women with early-stage breast cancer who complete 5 years of tamoxifen. J Clin Oncol. 2008;26:1948-55.
31. Mauri D, Pavlidis N, Polyzos NP, Ioannidis JP. Survival with aromatase inhibitors and inactivators versus standard hormonal therapy in advanced breast cancer: meta-analysis. J Natl Cancer Inst. 2006;98:1285-91.

32. Van den Eynden GG, Majeed AW, Illemann M, Vermeulen PB, Bird NC, Hoyer-Hansen G, et al. The multifaceted role of the microenvironment in liver metastasis: biology and clinical implications. Cancer Res. 2013;73:2031-43.
33. Tabaries S, Dong Z, Annis MG, Omeroglu A, Pepin F, Ouellet V, et al. Claudin-2 is selectively enriched in and promotes the formation of breast cancer liver metastases through engagement of integrin complexes. Oncogene. 2011;30:1318-28.

**FIGURE LEGENDS**

**Figure 1.** Molecular profiling of matched tumour samples from three heterogeneous breast cancer patients. **A,** Immunohistochemical staining of classic biomarkers in primary breast tumour tissue (top) and H&E staining of matched liver metastatic tumour tissue (bottom) showing strong heterogeneity between the 3 patients. ER, estrogen receptor; PgR progesterone receptor; H&E, haematoxylin and eosin staining. **B,** Correspondence analysis of RNAseq data from primary (P), nodal (N) and metastatic liver (M) tumours of the three patients. Patient heterogeneity is evident and tumours do not cluster based on site of origin. Patient 1 shows the largest divergence between primary and metastatic tumours. **C,** Correspondence analysis of RNAseq data from the primary, nodal and metastatic samples from the three patients (n=8) and breast cancer patient samples from The Cancer Genome Atlas (n=840). Each dot or shape represents an individual tumour, coloured based on PAM50 profiling.

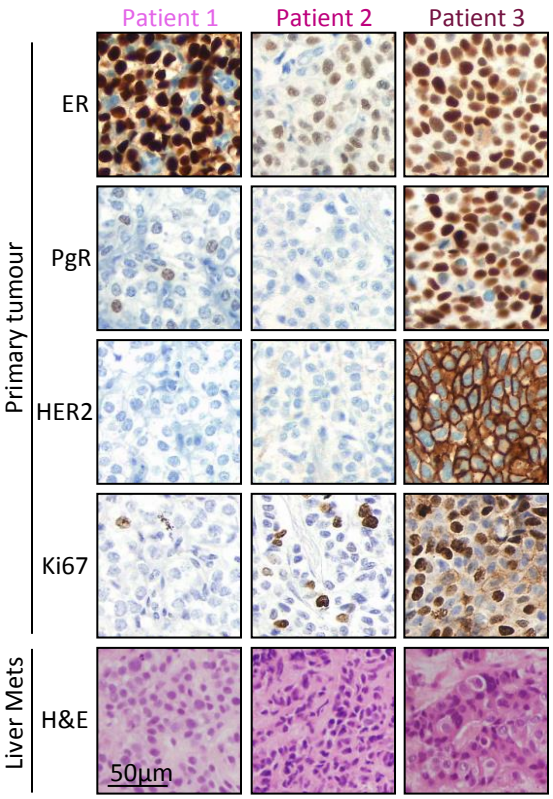
**Figure 2.** Matched primary and metastatic tumours display altered expression of genes involved in cell-cell interactions and cancer pathways. **A,** Venn diagrams showing DEGs for each individual patient between their primary and metastatic tumours. Left panel: genes upregulated in metastatic tumour relative to matched primary tumour. For all of these genes, expression in a normal liver sample was less than the primary tumour. Right panel: genes downregulated in metastatic tumour relative to matched primary tumour. DEGs exhibited a fold change >1.5 and >50 counts per million. **B,** Heat maps displaying the 1268 upregulated (left panel) and 2606 downregulated (right panel) DEGs from part A. Each patient displayed a distinct pattern of gene expression as highlighted. **C,** Network maps showing enriched KEGG pathways from the upregulated (left panel) and downregulated (right panel) DEGs. Details of pathways unique to each patient are provided in Supplementary Figs S2 and S3. Enriched KEGG pathways common to all 3 patients include a number

of cancer pathways and also cell interaction pathways such as focal adhesions and ECM receptor interactions.

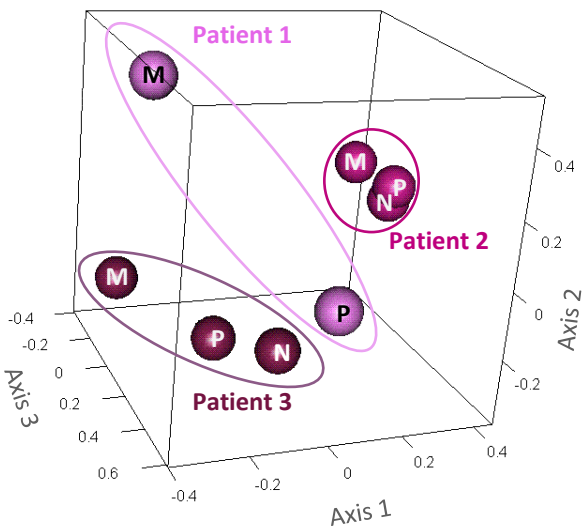
**Figure 3.** Tamoxifen contributes to the development of metastasis secondary to endocrine resistance. **A**, Representative *in vivo* bioluminescence images of mice following orthotopic injection of luciferase expressing endocrine resistant breast cancer cells. Five mice were treated with estrogen pellets (E2 only) and five mice had estrogen and 4-hydroxytamoxifen pellets (E2+4-OHT). Primary tumours were resected at week 5. Local recurrences developed in both groups but distant metastases were detectable only in tamoxifen treated mice. Graph displays the fold increase in metastatic flux with each line representing an individual mouse. *p-value<0.05 by unpaired t-test at week 15. **B,** *Ex vivo* bioluminescence imaging of organs from the mouse xenograft model. All estrogen plus tamoxifen treated mice displayed bioluminescence in at least one organ of their bone, lung and livers (n=5). No bioluminescence was detected in the same organs from estrogen-only treated animals (n=5). Scatter plots show total flux from each *ex vivo* organ. Representative images of immunohistochemical staining of matched primary and metastatic tumours from estrogen plus tamoxifen treated mouse xenografts are shown. The metastatic marker Keratin 19 was more strongly expressed in metastatic tumours than matched primary tumours (n=3). Ki67 staining is also shown to help identify tumour cells. **C,** Heat map listing the common 31 DEGs which are upregulated in liver metastatic tumours relative to matched primary tumours for all three patients. The majority of these genes are more strongly expressed in the liver metastasis, sampled after tamoxifen treatment, than in the nodal tumour, sampled before tamoxifen. Those genes with a fold change >1.5 fold between liver and node are highlighted in bold (n=21 genes). **D,** Heatmap of TaqMan gene expression fold changes of genes up regulated in metastasis from the RNA-seq data analysis in an extended cohort of matched primary breast cancer and metastases samples. To the left of the TaqMan expression is the original gene expression fold changes for the RNA-seq data. Metastatic sites of the extended cohort include liver, peritoneum (Perit.), contralateral breast (“Contra.”), bone and axilla (“Ax.”). The liver metastases patients in the extended cohort exhibited a similar gene

expression pattern to the RNA-seq data. **E,** Immunohistochemical staining of four DEGs in matched tumours of 7 independent endocrine-treated patients. H-scores for each sample are reported as a heat map. Details of patient samples and treatments are provided in Supplementary Table S4.

A

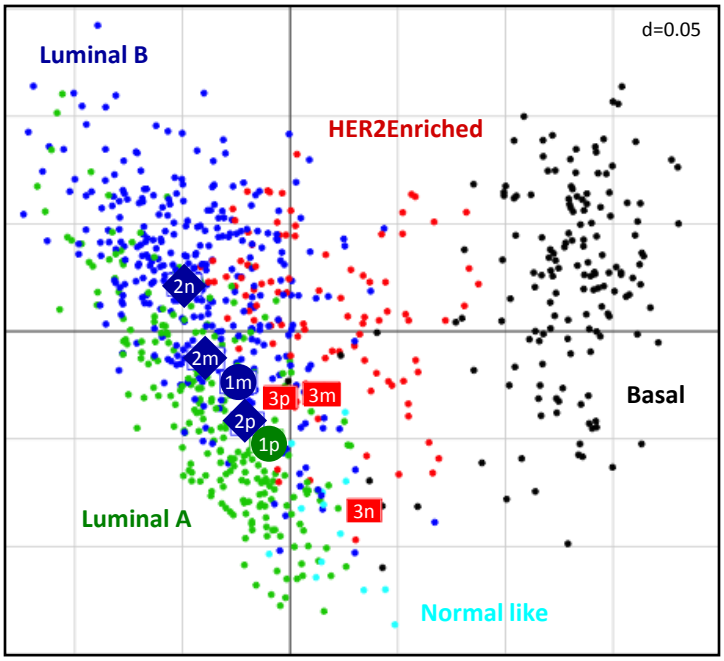



B




Correspondence analysis of matched tumours from 3 patients

C

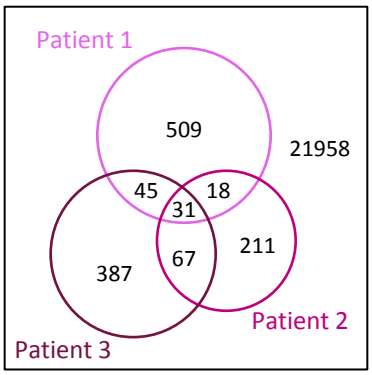



Correspondence analysis of tumours from 848 patient samples

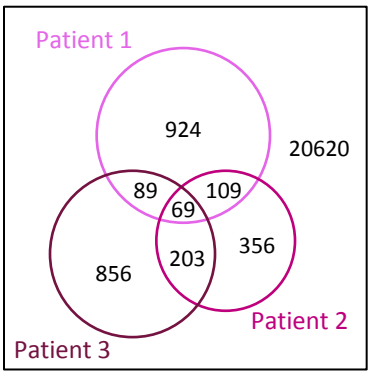
Figure 1

A

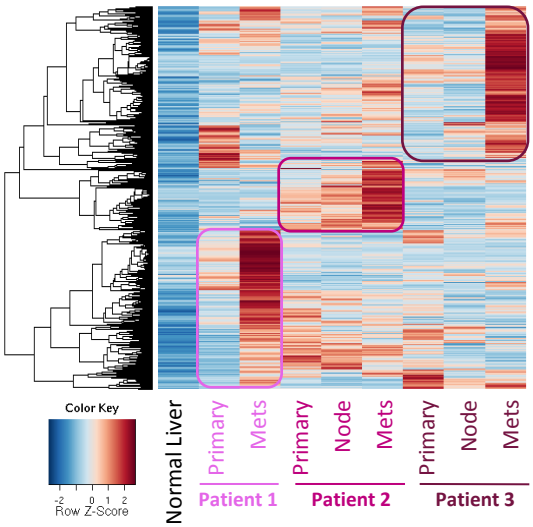
Genes Upregulated in Metastasis

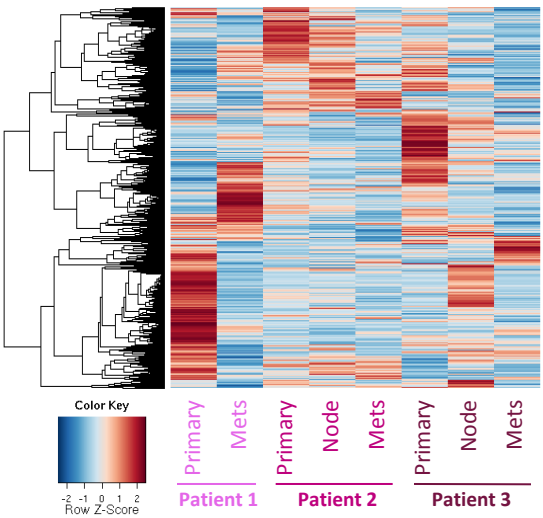



Genes Downregulated in Metastasis

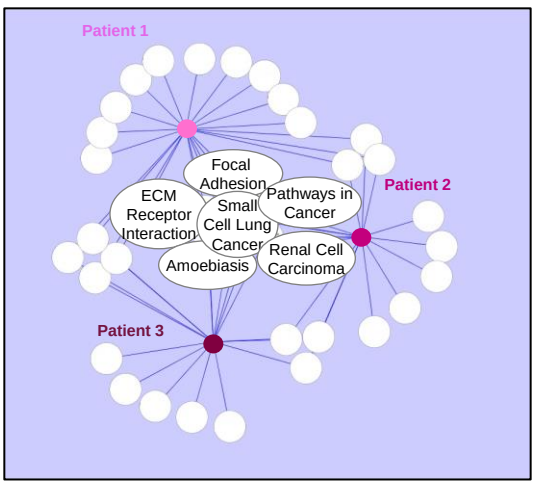



B




C



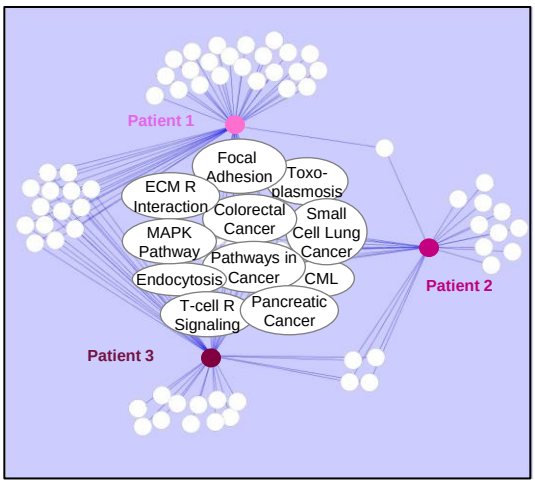




Figure 2

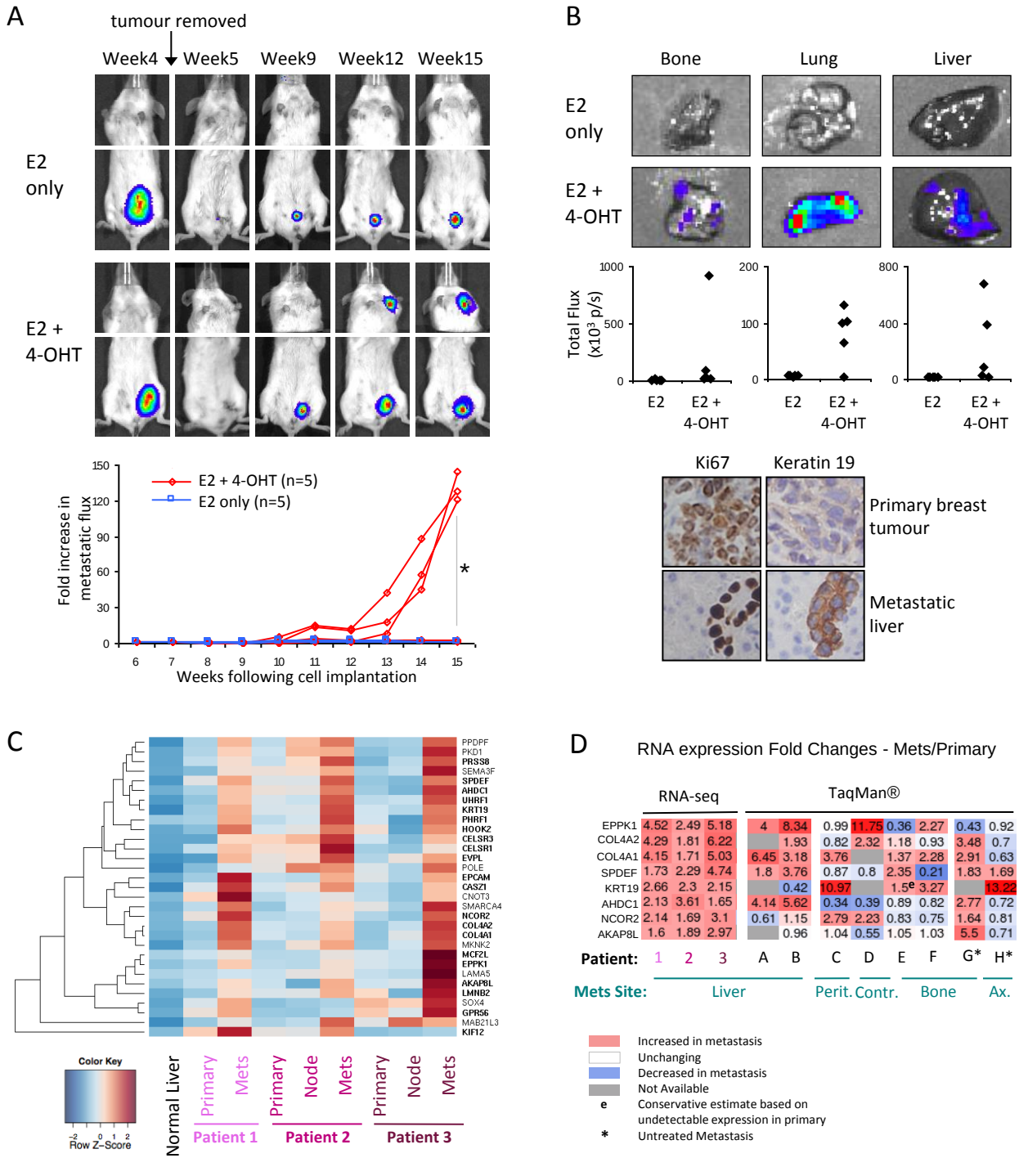

A
tumour removed
Week4
Week5
Week9
Week12
Week15
E2 only
E2 + 4-OHT
Fold increase in metastatic flux
E2 + 4-OHT (n=5)
E2 only (n=5)
*
Weeks following cell implantation
B
Bone
Lung
Liver
E2 only
E2 + 4-OHT
Total Flux (x10³ p/s)
E2
E2 + 4-OHT
Ki67
Keratin 19
Primary breast tumour
Metastatic liver
C
Color Key
Row Z-Score
Normal Liver
Primary
Mets
Node
Patient 1
Patient 2
Patient 3
D
RNA expression Fold Changes - Mets/Primary
RNA-seq
TaqMan®
Patient:
Mets Site:
Liver
Perit.
Contr.
Bone
Ax.
Increased in metastasis
Unchanging
Decreased in metastasis
Not Available
e Conservative estimate based on undetectable expression in primary
* Untreated Metastasis
E
KRT19
LMNB2
KIF12
EVPL
PT
MRT
MNT
Patient A liver
Patient B liver
Patient C peritoneum
Patient D contralateral
Patient I contralateral
Patient J lung
Patient K presacral
H score
0
1-49
50-149
>150
Not available
PT: Primary tumour tissue
MRT: Matched recurrent tumour tissue
MNT: Matched normal tissue at recurrence site

Figure 3

**Table 1: Tumour and Disease Characteristics for sequenced patient samples**

| | Patient 1 | Patient 2 | Patient 3 |
|---|---|---|---|
| Age at Diagnosis | 34 yr | 61 yr | 37 yr |
| Time to Distant Metastasis | 11 yr | 5 yr | 6 yr |
| Time to Death | 12 yr | 5 yr | 9 yr |
| **Primary Breast Pathology** | | | |
| ER | + | + | + |
| PgR | + | - | + |
| HER2 | - | - | + |
| P53 | - | - | - |
| Ki67 | 10% | 16% | 38% |
| Tumour Size | 50mm | 26mm | NA |
| Tumour Grade | 2 | 2 | 3 |
| Lymph Node Status | + | + | + |
| Lobular/Ductal | Lobular | Ductal | Ductal |
| OncotypeDX Risk | Low | High | High |
| Luminal Classification | A | B1 | B2 |
| **Surgery, Adjuvant Treatment** | | | |
| Surgery | WLE +ANC<br>Mastectomy[a] | WLE+ANC | Mastectomy+ANC |
| Chemotherapy | CMF x6<br>AC x4[a]<br>taxane x6[a] | AC x4<br>taxane x4 | CMF x6 |
| Radiotherapy | yes | yes | yes |
| Endocrine Therapy | 5 yr Tamoxifen<br>5yr Arimidex[a]<br>4yr Zoladex[a] | 5 yr Tamoxifen | 5 yr Tamoxifen |
| Herceptin | no | no | no |
| **Metastatic Tumour Pathology** | | | |
| Sites of Metastasis | Liver, bone | Liver, lung | Liver, lung, brain[b] |
| Biopsy | Liver | Liver | Liver |
| ER (liver biopsy) | + | NA | + |
| PgR (liver biopsy) | + | NA | + |
| HER2 (liver biopsy) | - | NA | + |
| p53 (liver biopsy) | NA | NA | - |
| Post-Metastatic Therapy | Fulvestrant | Arimidex | Herceptin<br>Letrozole<br>Fulvestrant |

[a]Treatment given 5 years following initial diagnosis

[b]Brain metastasis diagnosed 8 years following initial diagnosis

Abbreviations: yr, years; WLE, wide local excision; ANC, axilliary node clearance; CMF, cyclophosphamide methotrexate fluoracil; AC, adriamycin cyclophosphamide; NA, not available.

**Supplementary Table S1: List of 31 genes which are upregulated in metastatic tumours relative to matched primary tumours for all three patients**

| CommonToAll3Patients_UpInMetsvsPrimary (31 genes) | | Counts per Million Expression Values (genes highlighted in yellow are also down regulated in the nodes): | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PRIMARY | | | | NODE | | | METASTASIS | | | |
| GeneName | Description | Patient 1 | Patient 2 | Patient 3 | Mean Primary | Patient 2 | Patient 3 | Mean Node | Patient 1 | Patient 2 | Patient 3 | Mean Mets |
| AHDC1 | AT hook, DNA binding motif, containing 1 | 32.1980372 | 53.17648715 | 29.3071694 | 38.22723125 | 50.7415783 | 20.70643695 | 35.72400763 | 68.47514983 | 87.92471755 | 105.8556137 | 87.41849369 |
| AKAP8L | A kinase (PRKA) anchor protein 8-like | 34.81412772 | 37.31367064 | 38.58014097 | 36.90264644 | 14.47436048 | 41.89441895 | 28.18438971 | 55.67815461 | 70.44336722 | 114.4688653 | 80.19679571 |
| CASZ1 | castor zinc finger 1 | 107.4609491 | 105.2714186 | 88.03598933 | 100.256119 | 125.5528796 | 58.74849554 | 92.15068759 | 254.1438348 | 200.452817 | 163.6517801 | 206.0828106 |
| CELSR1 | cadherin, EGF LAG seven-pass G-type receptor 1 | 324.5964625 | 562.5892081 | 210.7597612 | 365.9818106 | 584.6665833 | 146.2291788 | 365.447881 | 707.9881475 | 1352.927023 | 534.8954065 | 865.2701925 |
| CELSR3 | cadherin, EGF LAG seven-pass G-type receptor 3 | 1.006188662 | 80.21537892 | 46.82278236 | 42.68144998 | 95.95362563 | 12.19914115 | 54.07638339 | 77.79226038 | 151.1165617 | 93.37264038 | 107.4271541 |
| CNOT3 | CCR4-NOT transcription complex, subunit 3 | 93.77678334 | 48.48974591 | 50.82962193 | 64.36538372 | 59.52377454 | 29.53476278 | 44.52926866 | 200.2617497 | 73.0331969 | 105.1066353 | 126.1338606 |
| COL4A1 | collagen, type IV, alpha 1 | 185.3399516 | 318.8786636 | 164.3949034 | 222.8711729 | 266.0680195 | 356.5038486 | 311.2859341 | 769.9525453 | 544.1232152 | 827.1218117 | 713.7325241 |
| COL4A2 | collagen, type IV, alpha 2 | 158.5753332 | 256.328694 | 123.4106586 | 179.4382286 | 234.5171664 | 232.4257419 | 233.4714541 | 680.2613245 | 464.7449356 | 768.077348 | 637.694536 |
| EPCAM | epithelial cell adhesion molecule | 21.33119964 | 32.44667012 | 42.58698053 | 32.12161677 | 28.62345442 | 26.32446248 | 27.47395845 | 119.2141133 | 74.58709471 | 75.64681828 | 89.81600876 |
| EPPK1 | epiplakin 1 | 28.37452028 | 39.65704126 | 61.70532932 | 43.24563029 | 35.12878498 | 50.24119974 | 42.68499236 | 128.3067152 | 98.80200219 | 319.5641168 | 182.2242781 |
| EVPL | envoplakin | 4.82970558 | 51.01337581 | 30.90990523 | 28.9176622 | 57.57217538 | 51.20428983 | 54.3882326 | 74.31237571 | 108.5138635 | 82.26279414 | 88.36301111 |
| GPR56 | G protein-coupled receptor 56 | 94.1792588 | 36.77289281 | 123.5251398 | 84.82576379 | 34.47825192 | 78.97338745 | 56.72581968 | 145.5938841 | 55.55184658 | 231.3094955 | 144.1517421 |
| HOOK2 | hook microtubule-tethering protein 2 | 13.88540354 | 26.85863249 | 34.11537688 | 24.95313764 | 30.41242033 | 8.346780787 | 19.37960056 | 56.01491765 | 54.90438916 | 51.30502032 | 54.07477571 |
| KIF12 | kinesin family member 12 | 205.8662003 | 140.2417186 | 16.82872618 | 120.9788817 | 128.3176451 | 30.97939792 | 79.64852152 | 430.7199178 | 300.2907511 | 53.30229605 | 261.437655 |
| KRT19 | keratin 19 | 138.6527977 | 218.2939862 | 204.9212235 | 187.2893358 | 220.6933389 | 168.8617959 | 194.7775674 | 369.3167917 | 502.1679744 | 441.2731066 | 437.5859576 |
| LAMA5 | laminin, alpha 5 | 76.87281381 | 103.8293444 | 95.8207062 | 92.17428814 | 178.4086904 | 68.21888143 | 123.3137859 | 116.6322634 | 228.4229775 | 823.5017495 | 389.5189968 |
| LMNB2 | lamin B2 | 28.17328255 | 42.90170827 | 54.4930181 | 41.85600297 | 44.23624775 | 21.187982 | 32.71211487 | 53.20855905 | 66.55862271 | 100.2382757 | 73.33515249 |
| MAB21L3 | mab-21-like 3 (C. elegans) | 150.7270616 | 182.0618712 | 291.8124016 | 208.2004448 | 214.3506417 | 575.6068443 | 394.978743 | 301.2906592 | 596.3082832 | 467.1128614 | 454.9039346 |
| MCF2L | MCF.2 cell line derived transforming sequence-like | 33.00298813 | 21.27059486 | 38.92358436 | 31.06572245 | 21.79285735 | 30.65836789 | 26.22561262 | 55.45364593 | 51.53761058 | 202.9731461 | 103.3214675 |
| MKNK2 | MAP kinase interacting serine/threonine kinase 2 | 195.603076 | 199.1865027 | 177.1023088 | 190.6306292 | 278.9160474 | 235.3150122 | 257.1155298 | 408.6058121 | 363.0941208 | 445.2676581 | 405.6558636 |
| NCOR2 | nuclear receptor corepressor 2 | 235.0456715 | 151.9585717 | 145.5055168 | 177.5032534 | 166.2111956 | 101.9270346 | 134.0691151 | 502.1136806 | 256.7816125 | 451.7588042 | 403.5513658 |
| PHRF1 | PHD and ring finger domains 1 | 52.92552364 | 57.50270983 | 62.7356595 | 57.72129766 | 63.10170635 | 37.23948351 | 50.17059493 | 89.57896649 | 112.139625 | 101.7362325 | 101.151608 |
| PKD1 | polycystic kidney disease 1 (autosomal dominant) | 86.93470043 | 105.9924557 | 62.96462176 | 85.29725931 | 178.8965902 | 78.17081237 | 128.5337013 | 159.2889141 | 186.2087538 | 273.1274561 | 206.2083747 |
| POLE | polymerase (DNA directed), epsilon, catalytic subunit | 68.21959131 | 119.3316423 | 111.3901399 | 99.64712453 | 185.8898205 | 87.32016823 | 136.6049944 | 108.662205 | 179.4751966 | 199.3530838 | 162.4968285 |
| PPDPF | pancreatic progenitor cell differentiation and proliferation | 98.80772665 | 110.679197 | 90.09664968 | 99.8611911 | 194.3467502 | 94.70385893 | 144.5253046 | 198.9146975 | 243.5734812 | 259.6458449 | 234.0446745 |
| PRSS8 | protease, serine, 8 | 69.02454224 | 71.20241499 | 30.68094297 | 56.96930007 | 113.6806514 | 64.20600605 | 88.94332872 | 107.3151528 | 197.2155299 | 192.737108 | 165.7559302 |
| SEMA3F | sema domain, immunoglobulin domain (Ig), short basic | 27.97204481 | 108.1555671 | 51.17306532 | 62.43355907 | 117.746483 | 70.94763669 | 94.34705983 | 113.6013961 | 163.6772356 | 266.136991 | 181.1385409 |
| SMARCA4 | SWI/SNF related, matrix associated, actin dependent | 116.1141716 | 91.93223202 | 144.81863 | 117.6216779 | 130.9197773 | 102.2480646 | 116.583921 | 238.9894983 | 165.1016419 | 280.8668996 | 228.3193466 |
| SOX4 | SRY (sex determining region Y)-box 4 | 156.5629559 | 56.0606356 | 276.0140056 | 162.879199 | 74.64866807 | 235.4755272 | 155.0620976 | 295.9024507 | 93.8813258 | 447.7642528 | 279.1826764 |
| SPDEF | SAM pointed domain containing ETS transcription factor | 91.96564374 | 89.40860212 | 39.26702775 | 73.5470912 | 87.33406265 | 27.6085826 | 57.47132263 | 159.2889141 | 205.1145104 | 186.2459618 | 183.5497955 |
| UHRF1 | ubiquitin-like with PHD and ring finger domains 1 | 16.30025633 | 39.29652271 | 29.65061279 | 28.41579728 | 33.82771887 | 32.74506309 | 33.28639098 | 64.43399344 | 97.76607032 | 100.6127649 | 87.60427622 |

**Supplementary Table S2: List of 69 genes which are downregulated in metastatic liver tumours relative to matched primary tumours for all three patients**

| CommonToAll3Patients_DownInMetsvsPrimary (69 genes) | | Counts per Million Expression Values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PRIMARY | | | | NODE | | | METASTASIS | | | |
| GeneName | Description | Patient 1 | Patient 2 | Patient 3 | Mean Prima | Patient 2 | Patient 3 | Mean Node | Patient 1 | Patient 2 | Patient 3 | Mean Mets |
| ADCY7 | adenylate cyclase 7 | 111.4857038 | 52.81596859 | 61.13292367 | 75.14486535 | 53.99424357 | 51.84634989 | 52.92029673 | 48.38162225 | 29.78304129 | 40.694493 | 39.61971885 |
| AKAP12 | A kinase (PRKA) anchor protein 12 | 57.55399149 | 56.42115416 | 64.91080098 | 59.62864887 | 42.60991511 | 73.1948469 | 57.90238101 | 28.8493664 | 22.27253523 | 23.3431601 | 24.82168724 |
| ARHGAP30 | Rho GTPase activating protein 30 | 238.6679507 | 75.52863768 | 76.01547063 | 130.0706863 | 59.36114128 | 170.6274611 | 114.9943012 | 15.26659078 | 44.28608748 | 49.05808512 | 36.2035878 |
| ARHGAP31 | Rho GTPase activating protein 31 | 78.48271567 | 52.99622787 | 65.71216889 | 65.73037081 | 47.32627976 | 80.89956763 | 64.11292369 | 29.74740115 | 27.84066903 | 38.822047 | 32.13670573 |
| ARHGDIB | Rho GDP dissociation inhibitor (GDI) beta | 267.4449465 | 205.1350589 | 147.9096206 | 206.8298753 | 191.2567182 | 276.7278861 | 233.9923021 | 34.91110098 | 85.98234529 | 81.388986 | 67.42747742 |
| BTG1 | B-cell translocation gene 1, anti-proliferative | 162.1976124 | 65.43411808 | 61.93429158 | 96.52200735 | 82.78033126 | 99.03776434 | 90.9090478 | 38.50323998 | 26.67524568 | 34.45300635 | 33.21049734 |
| BTG2 | BTG family, member 2 | 421.995525 | 664.2554411 | 302.5736278 | 462.9415313 | 415.5279889 | 426.4883952 | 421.0081921 | 213.7322709 | 197.7334959 | 153.7902312 | 188.418666 |
| BTN3A1 | butyrophilin, subfamily 3, member A1 | 94.38049653 | 102.3872702 | 90.44009307 | 95.73595326 | 61.9632735 | 137.721883 | 99.84257824 | 31.76797935 | 49.59523832 | 45.18836339 | 42.18386035 |
| CCND2 | cyclin D2 | 236.6555734 | 173.0489073 | 101.773725 | 170.4927352 | 137.9130077 | 163.7253154 | 150.8191616 | 28.73711206 | 87.01827716 | 16.22786531 | 43.99441817 |
| CDR1 | cerebellar degeneration-related protein 1, 34kDa | 226.9961622 | 262.6377687 | 300.055043 | 263.229658 | 104.5731886 | 179.776817 | 142.1750028 | 66.34231729 | 61.7674378 | 20.09758703 | 49.40244738 |
| CELF2 | CUGBP, Elav-like family member 2 | 209.2872418 | 61.64867323 | 70.6348575 | 113.8569242 | 70.58283648 | 139.3270331 | 104.9549348 | 11.67445177 | 18.38779071 | 14.23058958 | 14.76427735 |
| CLIP3 | CAP-GLY domain containing linker protein 3 | 53.73047457 | 81.29693459 | 60.67499915 | 65.2341361 | 67.81807099 | 69.82403158 | 68.82105129 | 33.78855754 | 35.09219213 | 28.08668996 | 32.32247987 |
| CYR61 | cysteine-rich, angiogenic inducer, 61 | 169.240933 | 110.3186784 | 79.67886681 | 119.7461594 | 72.37180238 | 163.7253154 | 118.0485589 | 55.79040896 | 6.086099742 | 21.8452033 | 27.90723733 |
| DOCK10 | dedicator of cytokinesis 10 | 92.77059467 | 88.86782428 | 51.28754645 | 77.64198847 | 78.71449967 | 156.0205947 | 117.3675472 | 36.03364442 | 35.61015806 | 22.34452223 | 31.32944157 |
| EGR1 | early growth response 1 | 1065.553793 | 663.8949226 | 233.6559873 | 654.3682345 | 345.1077857 | 776.0900982 | 560.598942 | 43.77919415 | 34.05626026 | 63.41350443 | 47.08298628 |
| ELN | elastin | 99.00896438 | 76.42993407 | 72.00863106 | 82.48250984 | 24.06972304 | 48.7965646 | 36.43314382 | 23.34890355 | 38.07049626 | 31.95641168 | 31.1252705 |
| ETS1 | v-ets avian erythroblastosis virus E26 oncogene | 246.1137468 | 86.70471294 | 101.0868382 | 144.6350993 | 75.13656786 | 357.1459087 | 216.1412383 | 62.07665222 | 36.9050729 | 35.20198474 | 44.72790329 |
| EZH1 | enhancer of zeste 1 polycomb repressive complex 2 | 154.3493408 | 121.1342351 | 55.63782941 | 110.3738018 | 101.6457899 | 75.92360216 | 88.78469601 | 85.08879274 | 80.28472 | 33.82885768 | 66.40079014 |
| FKBP1A | FK506 binding protein 1A, 12kDa | 168.4359821 | 125.4604578 | 160.617026 | 151.5044886 | 116.1201503 | 204.0145842 | 160.0673673 | 80.48636463 | 65.0047249 | 105.9804434 | 83.82384432 |
| FLJ10038 | Uncharacterized Protein FLJ10038 | 78.88519113 | 57.68296911 | 81.39608376 | 72.654748 | 83.10559779 | 65.49012618 | 74.29786198 | 27.16555124 | 38.07049626 | 38.57238754 | 34.60281168 |
| FMNL3 | formin-like 3 | 224.9837849 | 238.6632847 | 215.2245253 | 226.2905316 | 185.0766542 | 223.5974161 | 204.3370351 | 88.23191437 | 88.96064942 | 104.9818056 | 94.05812312 |
| FOS | FBJ murine osteosarcoma viral oncogene homolog | 485.9891239 | 248.3972857 | 63.65150854 | 266.0126394 | 145.7194043 | 385.5570664 | 265.6382353 | 12.12346915 | 10.48881019 | 11.10984625 | 11.24070853 |
| FYB | FYN binding protein | 173.6681631 | 95.53741759 | 64.91080098 | 111.3721272 | 118.397016 | 156.8231698 | 137.6100929 | 27.95133165 | 47.3938831 | 30.45845488 | 35.26788988 |
| GOSR2 | golgi SNAP receptor complex member 2 | 57.35275376 | 111.0397155 | 54.37853697 | 74.25700209 | 96.27889216 | 71.26866672 | 83.77377944 | 34.79884663 | 66.81760568 | 35.82613341 | 45.81419524 |
| GPCPD1 | glycerophosphocholine phosphodiesterase GDE1 | 100.2163908 | 76.9707119 | 76.58787628 | 84.59165965 | 67.16753794 | 51.04377481 | 59.10565638 | 47.93260488 | 40.9193089 | 37.69857941 | 42.18349773 |
| IFI16 | interferon, gamma-inducible protein 16 | 64.39607439 | 71.20241499 | 95.47726281 | 77.02525073 | 49.11524566 | 82.98626283 | 66.05075424 | 31.655725 | 46.48744271 | 20.7217357 | 32.9549678 |
| IGLL5 | immunoglobulin lambda-like polypeptide 5 | 189.9684195 | 1005.846774 | 106.8108947 | 434.208696 | 577.8359862 | 224.8815362 | 401.3587612 | 8.419075798 | 250.9544957 | 12.23331385 | 90.53562846 |
| IL16 | interleukin 16 | 204.0550607 | 58.40400622 | 73.15344237 | 111.8708364 | 19.19072513 | 135.7957028 | 77.49321397 | 22.56312314 | 30.94846464 | 13.10712198 | 22.20623659 |
| INPP4A | inositol polyphosphate-4-phosphatase, type I, 107kDa | 104.2411454 | 58.04348767 | 84.94499881 | 82.4098773 | 72.69706891 | 64.20600605 | 68.45153748 | 32.89052279 | 25.25083935 | 48.18427699 | 35.44187971 |
| KLF12 | Kruppel-like factor 12 | 107.2597114 | 70.12085932 | 75.32858385 | 84.23638486 | 36.26721782 | 76.08411717 | 56.1756675 | 57.81098715 | 29.39456684 | 42.44210927 | 43.21588775 |
| KLF6 | Kruppel-like factor 6 | 161.1914237 | 272.552029 | 200.914384 | 211.5526122 | 142.6293723 | 359.5536339 | 251.0915031 | 93.05885116 | 83.39251561 | 91.87468359 | 89.44201679 |
| LOC1001308 | uncharacterized LOC100130899 | 58.15770469 | 72.64448922 | 86.54773463 | 72.44997618 | 82.12979821 | 49.75965469 | 65.94472645 | 19.98127323 | 45.45151084 | 55.67406098 | 40.36894835 |
| LOC728264 | uncharacterized LOC728264 | 249.736026 | 242.088211 | 278.8760338 | 256.9000903 | 210.1221768 | 266.4549251 | 238.288551 | 88.5686774 | 70.70235019 | 129.0739441 | 96.11499055 |
| MAML2 | mastermind-like 2 (Drosophila) | 172.259499 | 65.43411808 | 74.06929141 | 103.9209695 | 61.9632735 | 108.0266052 | 84.99493934 | 35.58462704 | 20.97762039 | 39.69585514 | 32.08603419 |
| MARS | methionyl-tRNA synthetase | 96.19163613 | 69.39982221 | 108.0701872 | 91.2205485 | 89.93619487 | 38.04205859 | 63.98912673 | 37.94196826 | 38.07049626 | 72.02675602 | 49.34640685 |
| MBNL1 | muscleblind-like splicing regulator 1 | 249.736026 | 195.5813171 | 176.1864598 | 207.1679343 | 146.0446709 | 272.3939807 | 209.2193258 | 129.9905303 | 110.3267443 | 85.75802666 | 108.6917671 |
| MCL1 | myeloid cell leukemia 1 | 297.2281309 | 394.7678198 | 480.1338612 | 390.7099373 | 430.6528825 | 497.7570619 | 464.2049722 | 196.2205933 | 224.797216 | 198.9785946 | 206.665468 |
| MFAP4 | microfibrillar-associated protein 4 | 152.739439 | 80.75615675 | 70.17693298 | 101.2241762 | 28.46082116 | 21.187982 | 24.82440158 | 11.00092571 | 40.66032593 | 21.72037357 | 24.46054174 |
| MIAT | myocardial infarction associated transcript (non-protein | 93.97802107 | 111.5804934 | 69.37556506 | 91.64469317 | 91.56252751 | 81.7021427 | 86.63233511 | 9.31711055 | 18.25829923 | 9.362229985 | 12.31254659 |
| MYCBP2 | MYC binding protein 2, E3 ubiquitin protein ligase | 215.5256115 | 172.868648 | 157.068111 | 181.8207902 | 124.9023466 | 296.1502029 | 210.5262748 | 100.8044009 | 100.096917 | 84.63455907 | 95.17862566 |
| NFIB | nuclear factor I/B | 91.76440601 | 124.91968 | 128.4478284 | 115.0439715 | 76.92553377 | 135.6351878 | 106.2803608 | 46.58555275 | 26.54575419 | 76.14613721 | 49.75914805 |
| NLRC5 | NLR family, CARD domain containing 5 | 310.1073458 | 135.0141996 | 70.52037637 | 171.8806406 | 136.7745748 | 155.5390497 | 146.1568123 | 18.74647544 | 89.47861535 | 40.19517407 | 49.47342162 |
| PCSK7 | proprotein convertase subtilisin/kexin type 7 | 136.4391826 | 52.99622787 | 72.69551785 | 87.37697611 | 58.06007517 | 77.3682373 | 67.71415623 | 18.07294938 | 30.94846464 | 36.57511181 | 28.53217528 |
| PLEKHG2 | pleckstrin homology domain containing, family G (with | 134.4268053 | 89.5888614 | 114.4811305 | 112.8322657 | 64.89067225 | 156.8231698 | 110.856921 | 45.12624628 | 58.78913368 | 70.65362895 | 58.18966964 |
| PLXDC1 | plexin domain containing 1 | 69.02454224 | 74.80760056 | 131.4243378 | 91.75216019 | 53.99424357 | 41.73390394 | 47.86407375 | 26.82878821 | 28.22914348 | 30.45845488 | 28.50546219 |
| PLXNC1 | plexin C1 | 77.47652701 | 81.29693459 | 93.30212133 | 84.02519431 | 66.84227141 | 79.29441748 | 73.06834444 | 27.39005993 | 40.66032593 | 31.33226302 | 33.12754963 |
| PRRX1 | paired related homeobox 1 | 56.74904056 | 87.60600933 | 162.9066487 | 102.4205662 | 83.91876411 | 97.27209917 | 90.59543164 | 7.072023671 | 4.143727484 | 7.364954255 | 6.19356847 |
| PTPRC | protein tyrosine phosphatase, receptor type, C | 160.7889483 | 92.29275057 | 85.2884422 | 112.790047 | 53.99424357 | 238.2042825 | 146.099263 | 24.69595568 | 40.78981742 | 14.35541931 | 26.6137308 |
| PTRF | polymerase I and transcript release factor | 142.2750769 | 111.2199748 | 161.9907996 | 138.4952838 | 93.67675994 | 169.0223109 | 131.3495354 | 41.75861596 | 66.29963974 | 76.27096695 | 61.44307421 |
| RPS12 | ribosomal protein S12 | 222.9714076 | 59.12504334 | 282.1959866 | 188.0974792 | 53.66897705 | 189.0866878 | 121.3778324 | 128.5312239 | 32.37287097 | 119.3372249 | 93.41377323 |
| SEC14L1 | SEC14-like 1 (S. cerevisiae) | 81.30004392 | 86.88497222 | 77.84716872 | 82.01072829 | 73.1849687 | 38.68411865 | 55.93454367 | 35.92139007 | 54.77489768 | 47.18563912 | 45.96064229 |
| SEPT6 | septin 6 | 289.581097 | 128.8853841 | 125.5858001 | 181.3507604 | 101.4831566 | 279.2961263 | 190.3896415 | 39.51352908 | 54.90438916 | 29.70947649 | 41.37579824 |
| SLC7A6 | solute carrier family 7 (amino acid transporter light | 165.2161784 | 101.3057145 | 105.4371212 | 123.986338 | 87.9845957 | 118.2995662 | 103.1420809 | 87.78289699 | 59.56608258 | 63.91282336 | 70.42060098 |
| SNED1 | sushi, nidogen and EGF-like domains 1 | 68.21959131 | 179.8987599 | 115.7404229 | 121.286258 | 107.8258539 | 62.44034089 | 85.13309739 | 25.93075346 | 114.5999632 | 69.90465056 | 70.14512241 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SPARCL1 | SPARC-like 1 (hevin) | 56.14532736 | 71.56293355 | 156.6101865 | 94.7728158 | 16.26332638 | 59.71158563 | 37.987456 | 9.092601862 | 33.27931135 | 3.744891994 | 15.3722684 |
| STARD9 | StAR-related lipid transfer (START) domain containing | 110.6807529 | 68.49852582 | 110.4742909 | 96.55118986 | 18.37755881 | 85.71501808 | 52.04628844 | 43.10566809 | 24.73287342 | 31.70675222 | 33.18176457 |
| STK4 | serine/threonine kinase 4 | 141.6713637 | 109.9581599 | 104.9791966 | 118.8695734 | 112.3795853 | 175.4429115 | 143.9112484 | 72.85306924 | 66.42913122 | 69.77982082 | 69.68734043 |
| TAGLN | transgelin | 194.798125 | 129.2459027 | 217.1707045 | 180.4049107 | 62.77643982 | 252.3296038 | 157.5530218 | 56.23942633 | 50.63117019 | 126.7021791 | 77.85759189 |
| TAGLN2 | transgelin 2 | 240.8815658 | 217.7532084 | 355.2349478 | 271.2899073 | 189.305119 | 162.4411953 | 175.8731572 | 143.1242886 | 136.225041 | 200.3517217 | 159.9003504 |
| TCF4 | transcription factor 4 | 135.8354694 | 98.42156604 | 170.4624033 | 134.9064796 | 83.43086432 | 134.9931277 | 109.211996 | 66.34231729 | 41.95524077 | 38.19789834 | 48.8318188 |
| TGFBR2 | transforming growth factor, beta receptor II | 149.720873 | 101.6662331 | 80.59471585 | 110.6606073 | 106.3621545 | 174.3193064 | 140.3407305 | 50.85121782 | 54.64540619 | 45.43802286 | 50.31154896 |
| TNC | tenascin C | 93.97802107 | 103.1083073 | 159.1287714 | 118.7383666 | 115.1443508 | 202.2489191 | 158.6966349 | 60.72960009 | 24.73287342 | 99.73895677 | 61.7338101 |
| TRAF1 | TNF receptor-associated factor 1 | 138.8540354 | 51.37389436 | 66.17009341 | 85.46600773 | 47.32627976 | 82.50471778 | 64.91549877 | 12.90924956 | 24.085416 | 34.07851715 | 23.6910609 |
| TRIM22 | tripartite motif containing 22 | 90.75821735 | 109.7779006 | 67.88731037 | 89.4744761 | 68.79387058 | 94.86437395 | 81.82912226 | 23.79792092 | 49.85422129 | 15.22922744 | 29.62712322 |
| TTN | titin | 663.2795663 | 320.6812564 | 389.9227304 | 457.9611843 | 185.8898205 | 794.0677799 | 489.9788002 | 70.04671064 | 63.32133561 | 169.8932668 | 101.0871043 |
| UTRN | utrophin | 442.5217737 | 376.3813734 | 586.7157937 | 468.5396469 | 251.4310258 | 541.096116 | 396.2635709 | 257.9604825 | 183.4894326 | 305.5831867 | 249.0110339 |
| VAMP1 | vesicle-associated membrane protein 1 (synaptobrevin | 188.5597553 | 171.0660553 | 87.23462142 | 148.9534773 | 119.8607154 | 95.18540398 | 107.5230597 | 48.38162225 | 61.50845484 | 53.30229605 | 54.39745771 |
| WIPF1 | WAS/WASL interacting protein family, member 1 | 256.3768712 | 123.2973465 | 102.2316495 | 160.6352891 | 117.9091162 | 133.3879776 | 125.6485469 | 32.32925107 | 53.47998284 | 47.56012832 | 44.45645408 |
| ZEB2 | zinc finger E-box binding homeobox 2 | 184.7362384 | 148.5336455 | 135.6601396 | 156.3100078 | 148.1589033 | 206.4223095 | 177.2906064 | 91.03827297 | 72.38573948 | 53.17746632 | 72.20049292 |

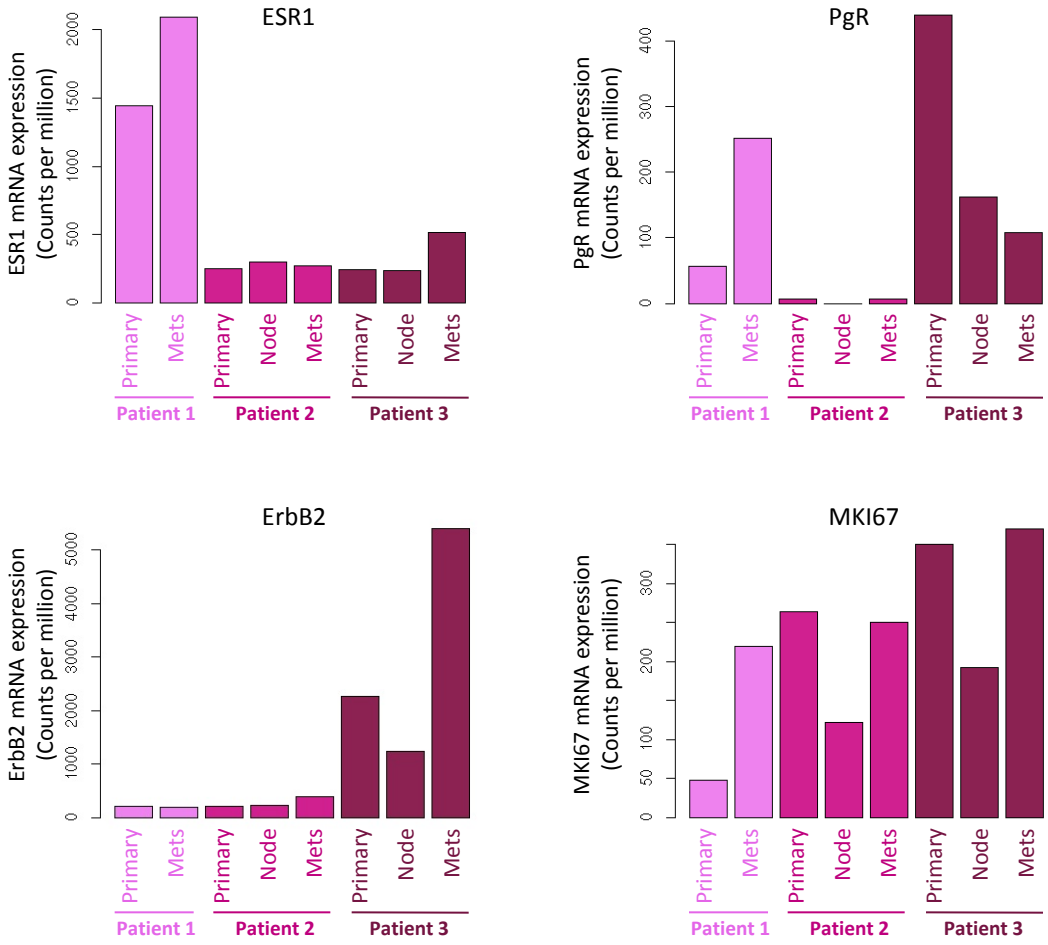
**Supplementary Table S3: List of 21 genes which are upregulated in the metastatic liver relative to both matched primary and matched node in all three patients (where available)**

| CommonToAll3Patients_UpInMetsvsPrimary&Node (21 Genes) | | Counts per Million Expression Values: | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PRIMARY | | | | NODE | | | METASTASIS | | | |
| GeneName | Description | Patient 1 | Patient 2 | Patient 3 | Mean Prima | Patient 2 | Patient 3 | Mean Node | Patient 1 | Patient 2 | Patient 3 | Mean Mets |
| AKAP8L | A kinase (PRKA) anchor protein 8-like | 34.81412772 | 37.31367064 | 38.58014097 | 36.90264644 | 14.47436048 | 41.89441895 | 28.18438971 | 55.67815461 | 70.44336722 | 114.4688653 | 80.19679571 |
| CASZ1 | castor zinc finger 1 | 107.4609491 | 105.2714186 | 88.03598933 | 100.256119 | 125.5528796 | 58.74849554 | 92.15068759 | 254.1438348 | 200.452817 | 163.6517801 | 206.0828106 |
| CELSR1 | cadherin, EGF LAG seven-pass G-type receptor 1 | 324.5964625 | 562.5892081 | 210.7597612 | 365.9818106 | 584.6665833 | 146.2291788 | 365.447881 | 707.9881475 | 1352.927023 | 534.8954065 | 865.2701925 |
| CELSR3 | cadherin, EGF LAG seven-pass G-type receptor 3 | 1.006188662 | 80.21537892 | 46.82278236 | 42.68144998 | 95.95362563 | 12.19914115 | 54.07638339 | 77.79226038 | 151.1165617 | 93.37264038 | 107.4271541 |
| COL4A1 | collagen, type IV, alpha 1 | 185.3399516 | 318.8786636 | 164.3949034 | 222.8711729 | 266.0680195 | 356.5038486 | 311.2859341 | 769.9525453 | 544.1232152 | 827.1218117 | 713.7325241 |
| COL4A2 | collagen, type IV, alpha 2 | 158.5753332 | 256.328694 | 123.4106586 | 179.4382286 | 234.5171664 | 232.4257419 | 233.4714541 | 680.2613245 | 464.7449356 | 768.077348 | 637.694536 |
| EPCAM | epithelial cell adhesion molecule | 21.33119964 | 32.44667012 | 42.58698053 | 32.12161677 | 28.62345442 | 26.32446248 | 27.47395845 | 119.2141133 | 74.58709471 | 75.64681828 | 89.81600876 |
| EPPK1 | epiplakin 1 | 28.37452028 | 39.65704126 | 61.70532932 | 43.24563029 | 35.12878498 | 50.24119974 | 42.68499236 | 128.3067152 | 98.80200219 | 319.5641168 | 182.2242781 |
| EVPL | envoplakin | 4.82970558 | 51.01337581 | 30.90990523 | 28.9176622 | 57.57217538 | 51.20428983 | 54.3882326 | 74.31237571 | 108.5138635 | 82.26279414 | 88.36301111 |
| GPR56 | G protein-coupled receptor 56 | 94.1792588 | 36.77289281 | 123.5251398 | 84.82576379 | 34.47825192 | 78.97338745 | 56.72581968 | 145.5938841 | 55.55184658 | 231.3094955 | 144.1517421 |
| HOOK2 | hook microtubule-tethering protein 2 | 13.88540354 | 26.85863249 | 34.11537688 | 24.95313764 | 30.41242033 | 8.346780787 | 19.37960056 | 56.01491765 | 54.90438916 | 51.30502032 | 54.07477571 |
| KIF12 | kinesin family member 12 | 205.8662003 | 140.2417186 | 16.82872618 | 120.9788817 | 128.3176451 | 30.97939792 | 79.64852152 | 430.7199178 | 300.2907511 | 53.30229605 | 261.437655 |
| KRT19 | keratin 19 | 138.6527977 | 218.2939862 | 204.9212235 | 187.2893358 | 220.6933389 | 168.8617959 | 194.7775674 | 369.3167917 | 502.1679744 | 441.2731066 | 437.5859576 |
| LMNB2 | lamin B2 | 28.17328255 | 42.90170827 | 54.4930181 | 41.85600297 | 44.23624775 | 21.187982 | 32.71211487 | 53.20855905 | 66.55862271 | 100.2382757 | 73.33515249 |
| MCF2L | MCF.2 cell line derived transforming sequence-like | 33.00298813 | 21.27059486 | 38.92358436 | 31.06572245 | 21.79285735 | 30.65836789 | 26.22561262 | 55.45364593 | 51.53761058 | 202.9731461 | 103.3214675 |
| NCOR2 | nuclear receptor corepressor 2 | 235.0456715 | 151.9585717 | 145.5055168 | 177.5032534 | 166.2111956 | 101.9270346 | 134.0691151 | 502.1136806 | 256.7816125 | 451.7588042 | 403.5513658 |
| PHRF1 | PHD and ring finger domains 1 | 52.92552364 | 57.50270983 | 62.7356595 | 57.72129766 | 63.10170635 | 37.23948351 | 50.17059493 | 89.57896649 | 112.139625 | 101.7362325 | 101.151608 |
| PRSS8 | protease, serine, 8 | 69.02454224 | 71.20241499 | 30.68094297 | 56.96930007 | 113.6806514 | 64.20600605 | 88.94332872 | 107.3151528 | 197.2155299 | 192.737108 | 165.7559302 |
| SPDEF | SAM pointed domain containing ETS transcription factor | 91.96564374 | 89.40860212 | 39.26702775 | 73.5470912 | 87.33406265 | 27.6085826 | 57.47132263 | 159.2889141 | 205.1145104 | 186.2459618 | 183.5497955 |
| UHRF1 | ubiquitin-like with PHD and ring finger domains 1 | 16.30025633 | 39.29652271 | 29.65061279 | 28.41579728 | 33.82771887 | 32.74506309 | 33.28639098 | 64.43399344 | 97.76607032 | 100.6127649 | 87.60427622 |

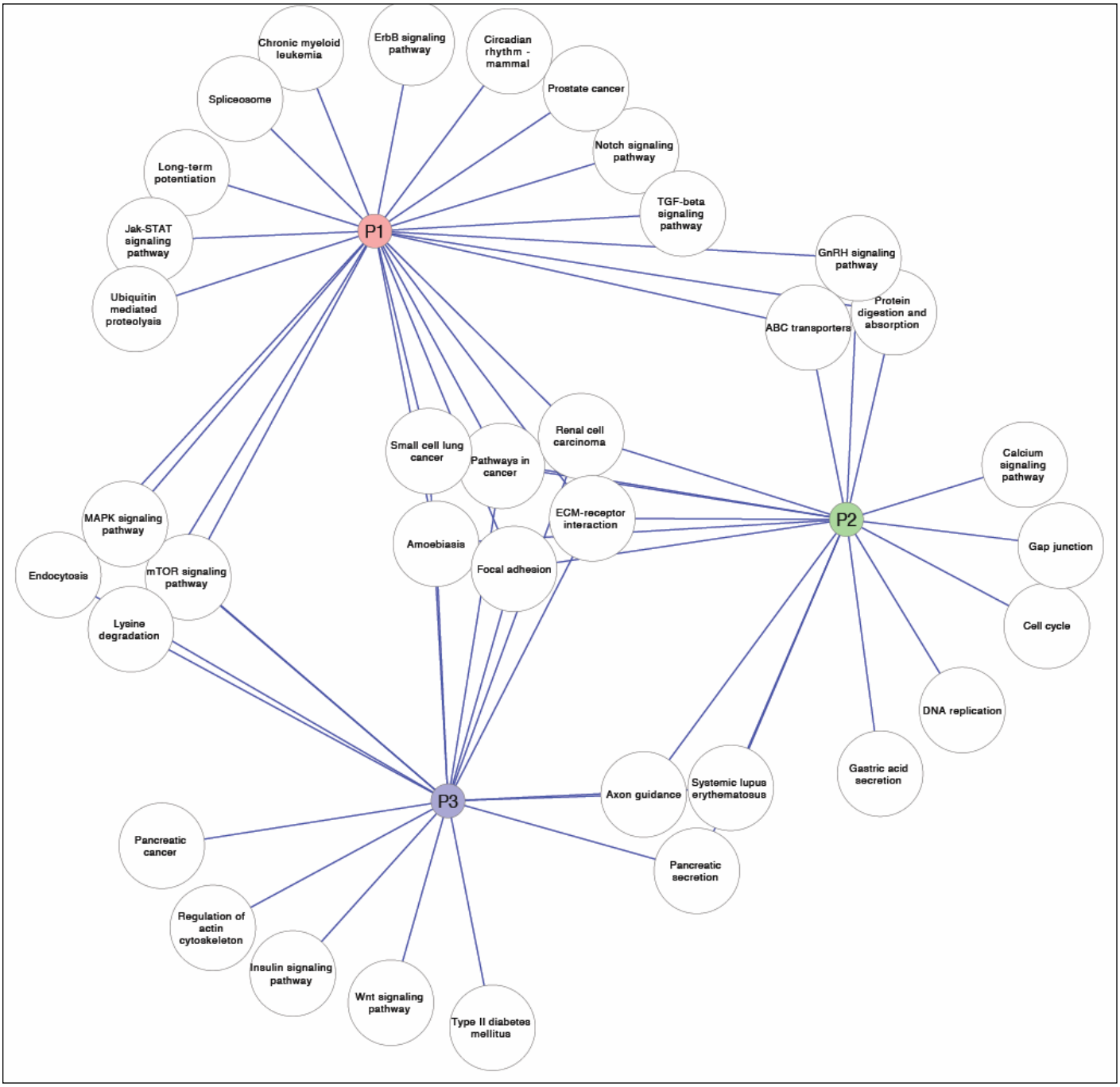
**Supplementary Table S4: Patient information for patients A-L**

| | Patient A | Patient B | Patient C | Patient D | Patient E | Patient F | Patient G | Patient H | Patient I | Patient J | Patient K | Patient L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Age at Diagnosis** | 39 | 44 | 62 | 70 | 70 | 55 | 71 | 45 | 50 | 72 | 36 | 48 |
| **Time to Recurrence** | 7yr | 5yr | 6yr | 2yr | 4yr | 4yr | 0yr | 5yr | 3yr | 3yr | 4yr | 6yr |
| **Endocrine Treatment** | Tamoxifen | Tamoxifen | Tamoxifen | AI | Tamoxifen | AI | none | none | AI | AI | Tamoxifen | Tamoxifen |
| **Primary Breast Pathology** | | | | | | | | | | | | |
| **ER** | + | + | + | + | + | + | + | + | + | + | + | + |
| **PgR** | + | - | - | - | - | - | - | - | + | + | + | + |
| **HER2** | - | - | - | + | + | + | + | NA | + | - | + | + |
| **Recurrent Tumour Pathology** | | | | | | | | | | | | |
| **Site of Recurrence** | liver | liver | peritoneum | contra breast | bone | bone | bone | axilla | contra breast | lung | presacral | contra breast |
| **ER** | + | + | - | + | + | + | NA | + | - | + | + | + |
| **PgR** | + | - | - | - | + | - | NA | - | - | - | + | - |
| **HER2** | - | - | - | + | + | + | NA | - | + | + | - | - |

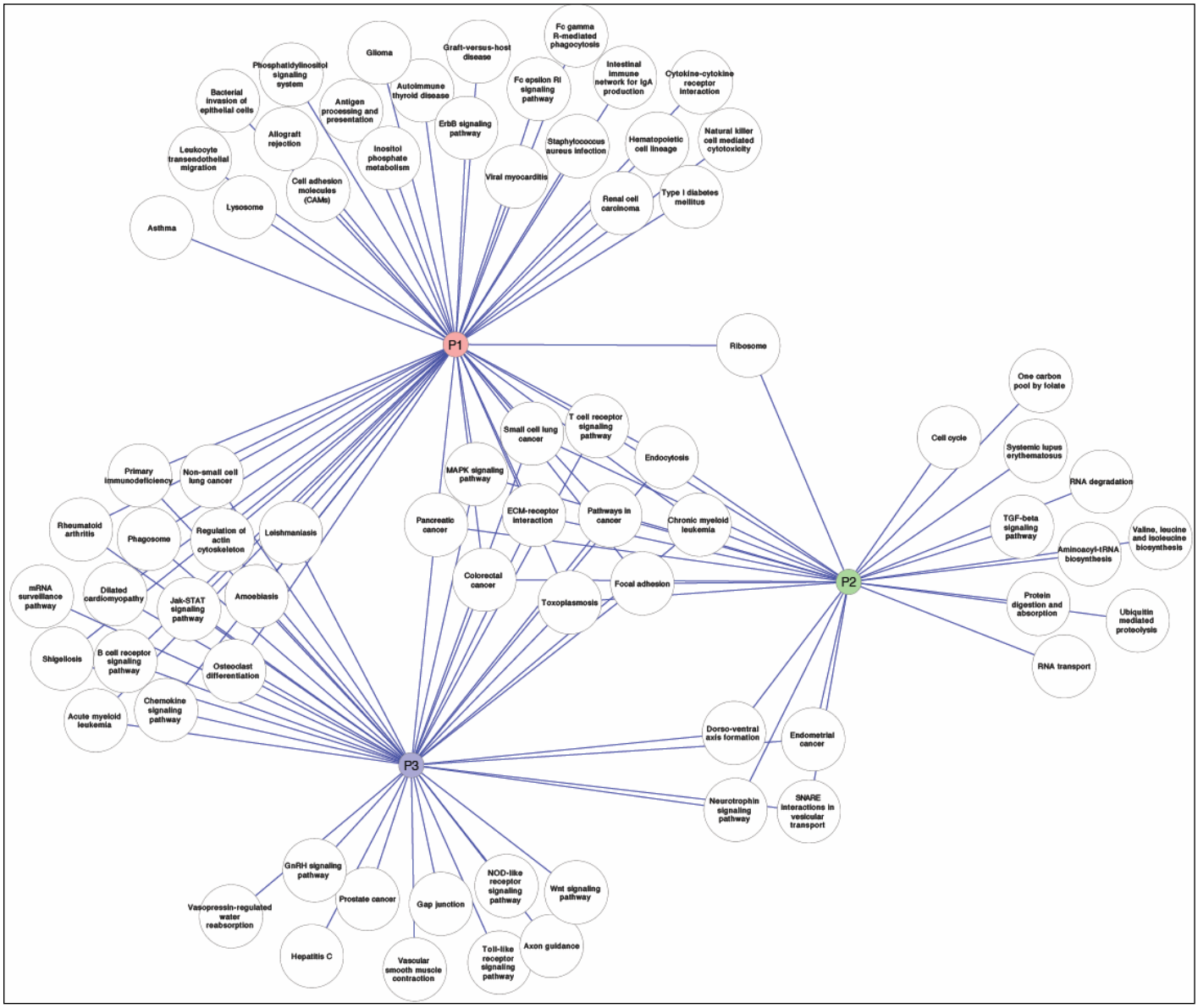
Abbreviations: AI, aromatase inhibitor; NA, not available; contra, contralatera



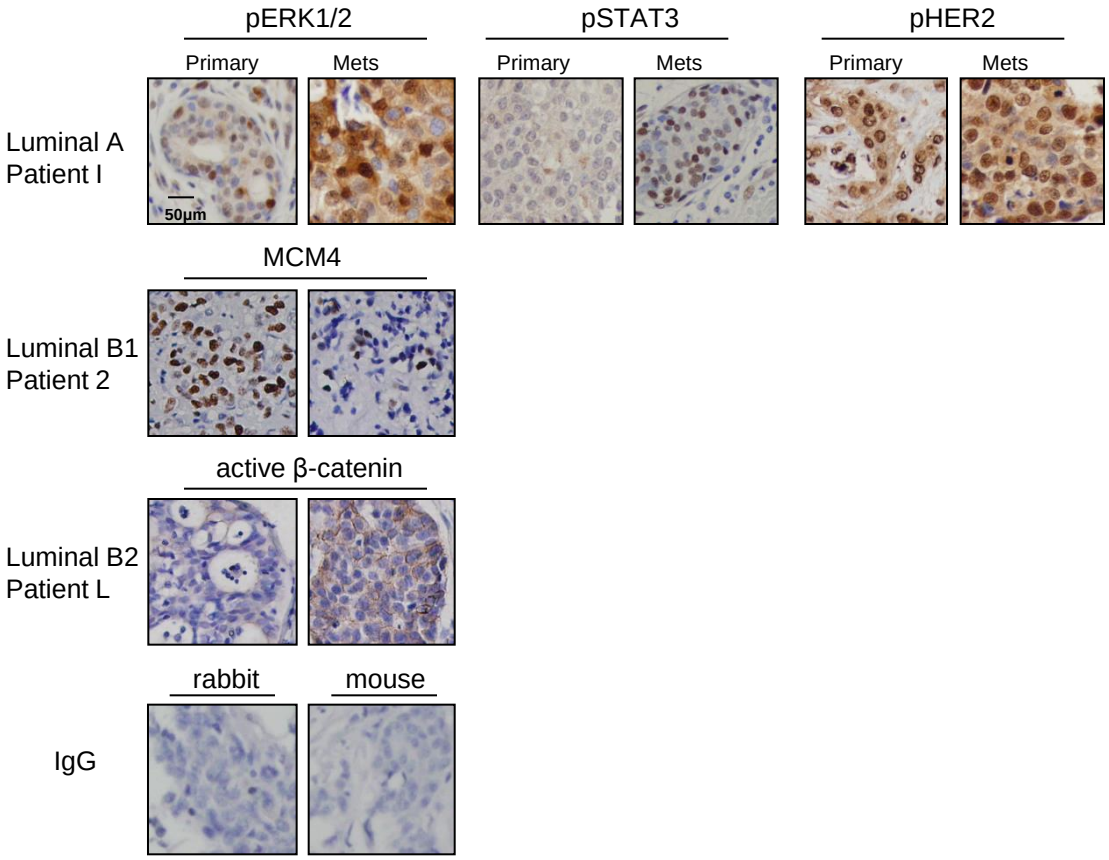

**Supplementary Figure S1: Gene expression profiles of classic biomarkers in matched tumour samples from 3 patients.** RNAseq data was mined for expression of the estrogen receptor (ESR1), the progesterone receptor (PgR), the human epidermal growth factor receptor 2 (ErbB2) and the proliferation marker Ki67 (MKI67). Gene expression in primary tumours largely reflected the protein expression observed by immunohistochemistry.



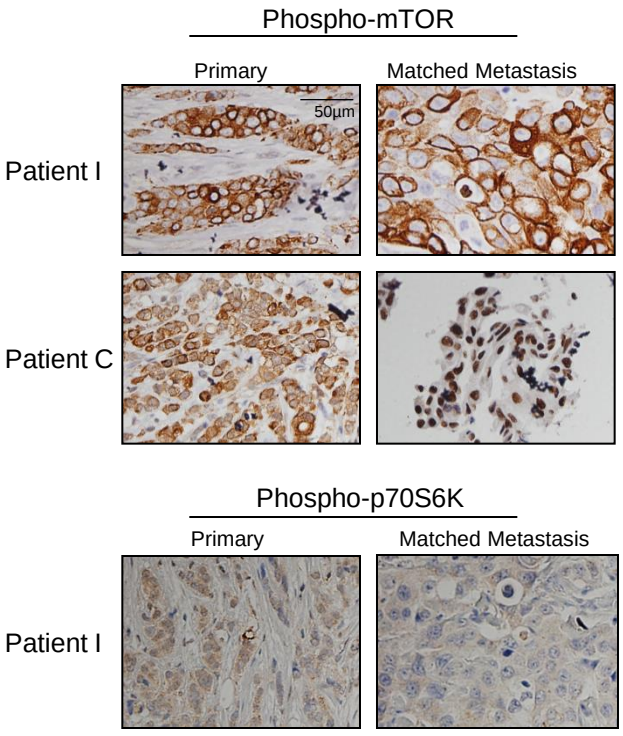

**Supplementary Figure S2: Network map showing the enriched KEGG pathways from upregulated differentially expressed genes (DEGs).** For each patient, relative gene expression was compared between the metastatic liver tumour and the primary breast tumour. In total, 1,268 genes were identified as being significantly upregulated in at least one patient (fold change >1.5 with liver metastatic tumour showing higher expression than primary tumour, >50 counts per million in the metastatic tumour, expression in a normal liver sample < expression in the primary tumour). Enriched KEGG pathway terms were identified for each patient using a hypergeometric test from the GOstats package in Bioconductor. The network map allows us to visualise the KEGG pathway terms which are common or unique.



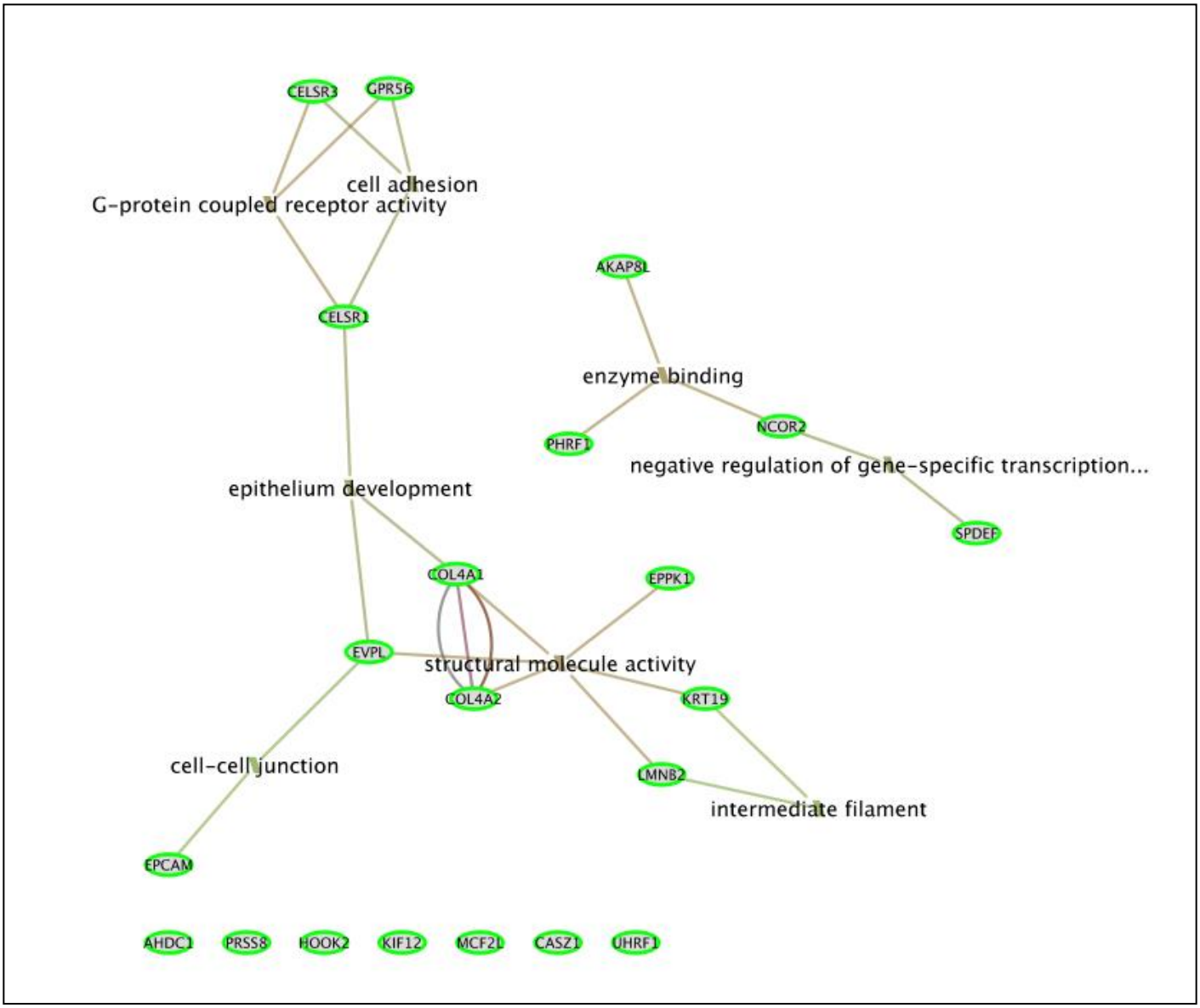

**Supplementary Figure S3: Network map showing the enriched KEGG pathways from downregulated differentially expressed genes (DEGs).** For each patient, relative gene expression was compared between the metastatic liver tumour and the primary breast tumour. In total, 2,606 genes were identified as being significantly downregulated in at least one patient (fold change >1.5 with liver metastatic tumour showing lower expression than primary tumour, >50 counts per million). Enriched KEGG pathway terms were identified for each patient using a hypergeometric test from the GOstats package in Bioconductor. The network map allows us to visualise the KEGG pathway terms which are common or unique.



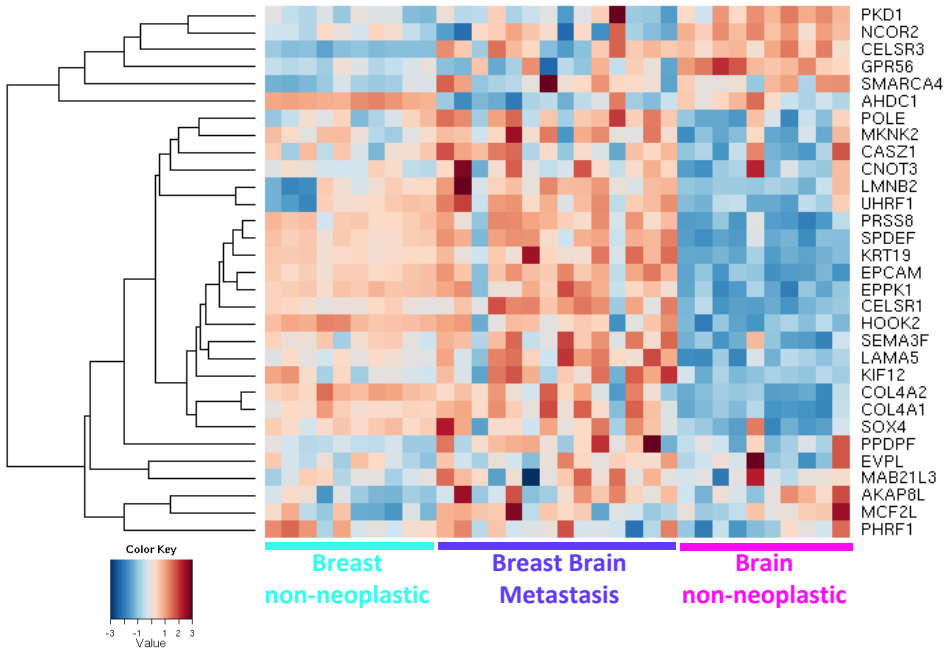

**Supplementary Figure S4: Protein expression in matched primary and metastatic tissue from endocrine treated breast cancer patients.** Immunohistochemistry analysis confirmed expression of phosphorylated growth factor signalling proteins, pERK1/2, pSTAT and pHER2 in both primary and metastatic tissue from a luminal A patient, Patient I (see Supplementary Table S4 for patient details). Protein expression of MCM4, a cell cycle protein, was confirmed in the matched primary and metastatic tumours from patient 2. Active β-catenin, a protein involved in the WNT signalling pathway, was detected in both primary and metastatic tumours from patient L, who presented with a luminal B2 primary tumour similar to patient 3 (see Supplementary Table S4 for patient details). Negative IgG controls for both rabbit and mouse antibodies are shown in the lower panel.




**Supplementary Figure S5: Phosphorylated mTOR signaling in endocrine resistant breast cancer tumours.** High levels of phosphorylated mTOR protein was detected by immunohistochemistry in primary tumour and matched metastatic tissue in patient I and patient C (see Supplementary Table S4 for patient details). Phosphorylated p-70S6K, a downstream target of mTOR, was also detected at the protein level in both the primary and matched tissue from patient I (Supplementary Table S4).




**Supplementary Figure S6: Interaction plot showing functional roles for 21 genes commonly upregulated in metastatic liver tumours of all three patients.** These 21 genes are significantly upregulated in the liver metastatic tumours relative to both primary and nodal tumours for all three patients, where available (fold change >1.5, sequence read >50 counts per million, expression in a normal liver sample < expression in the primary tumour). The 21 genes and their expression are listed in Supplementary Table S3. The plot was generated using the Exploratory Gene Analysis Network software tool. Each of the nodes in the network represents a gene while each of the edges represents an association between the genes. Associations between genes are either from PubMed or protein-protein interaction evidence. The 21 genes were ran through enrichment analysis of KEGG/GO terms using a hypergeometric test. Significant terms were selected and added to the plot above (e.g. cell-cell junction, intermediate filament etc.).




**Supplementary Figure S7: Genes upregulated in liver metastases are also strongly expressed in breast metastases to the brain.** Heat map of the common 31 DEGs using publicly available microarray data from an unmatched breast cancer brain metastasis study. The majority (18) of the 31 genes showed significant upregulation in the brain metastasis relative to non-neoplastic brain. Six of these genes also showed a significant upregulation relative to the non-neoplastic breast tissue. Genes were considered significant using a p-value cut off of <0.05 and a fold change of > 1.5.

**SUPPLEMENTARY METHODS**

**Data**

RNA-seq data from the 3 patient samples is described in the main body of the paper. The following publicly available data listed here was also used in the analysis. Normal human liver RNA-seq data was downloaded from the Illumina BodyMap 2.0 Project (E-MTAB-513 from Array Express http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/). Breast cancer samples (n=956) (1) in the form of level 3 RNA-seqV2 data were downloaded from The Cancer Genome Atlas (TCGA) data portal (http://tcga-portal.nci.nih.gov/). Gene expression microarray data from a breast brain metastasis study (2) were downloaded from NCBI Gene Expression Omnibus (GSE52604).

**Data Analysis**

The raw RNA-seq data was processed in two ways for two different types of analyses:
a) differential gene expression analysis;

b) combined analysis with the Breast Cancer Data from TCGA.

**a)** RNA-seq Differential gene expression analysis

Using default parameters, TopHat v1.4.1 (3) was used to align the reads from our patient samples to the hg19 genome. The reads were quantified into gene counts by using HT-Seq v0.5.3 (http://www-huber.embl.de/users/anders/HTSeq/). The counts were loaded into the software environment R and were processed by functions in the EdgeR package (4) from Bioconductor (5). The differential expression analysis was carried out on each patient individually using a negative binomial generalised linear model (6). A gene was considered to be differentially expressed if it exhibited a fold change of greater than 1.5 and if the gene had greater than 50 counts per million in the up regulated condition. When looking for differentially expressed genes between primary and metastasis, a normal RNA-seq liver sample from Illumina BodyMap 2.0 was used to reduce the selection of liver specific genes. Genes were filtered out if their expression was higher in the normal liver than in the primary breast cancer sample.

**b)** Combined analysis with TCGA

The RNA-seq data from our patient samples were processed in the same manner as the RNA-seqV2 (level 3) data from TCGA. Briefly, MapSplice v2.0 (7) was used to align the RNA-seq reads to the hg19 genome. RNA-seq by Expectation-Maximization (RSEM) v1.2.0 was used to calculate the expected counts of 20,532 human genes. The expected counts were upper quartile normalised and then filtered. They were filtered down to 15,964 genes based on two qualifying criteria: (i) if they were protein coding genes and (ii) if the mean normalised expected counts was greater than 10. The protein coding genes were retained to avoid influence from the expression differences in the non protein coding genes. Our RNA

extraction protocol produced reads from the pre-polyadenylated fraction that included a lot of non coding genes. The lowly expressed genes (mean < 10) were then removed to reduce the noise in the dataset. Out of the 956 TCGA breast cancer samples, 109 samples were removed because they were either from male patients, were normal breast samples or were outliers (n=2). The filtered normalised expected counts were then converted to a log2 scale. Each of the samples were classified by the PAM50 (8) classifier from the GeneFu package in Bioconductor. A correspondence analysis using the made4 package (9) in Bioconductor was carried out using the 15,964 human genes. Each of the samples was labeled by their PAM50 classification. OncotypeDX classifier was also carried out on our patient samples using the GeneFu package.

**KEGG pathway enrichment analysis**

Enriched KEGG (10) pathway terms were found using a hypergeometric test from the GOstats package in Bioconductor. The hypergeometric test was carried out on the genes found to be differentially expressed in the metastasis samples vs. primary samples on each patient separately.

**Gene Expression Microarray analysis:**

The breast cancer brain metastasis dataset contained processed gene expression data from 35 breast to brain metastases samples, 10 non-neoplastic brain samples and 10 non-neoplastic breast samples. The brain metastasis samples were classified as the PAM50 (8) subtypes by using the 'genefu' package in Bioconductor. 12 luminal B and 2 luminal A samples were classified from the brain metastasis group. These luminal samples were assessed for differential expression against the non-neoplastic brain and the non-neoplastic breast samples. All analyses were carried out in R using the Limma package. P-values were corrected for multiple testing using the Benjamini-Hochberg method.

**REFERENCES:**

1. Comprehensive molecular portraits of human breast tumours. Nature;490(7418):61-70.
2. Salhia B, Kiefer J, Ross JT*, et al.* Integrated genomic and epigenomic analysis of breast cancer brain metastasis. PLoS One;9(1):e85448.
3. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009;25(9):1105-11.
4. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics;26(1):139-40.
5. Gentleman RC, Carey VJ, Bates DM*, et al.* Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5(10):R80.
6. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res;40(10):4288-97.

7. Wang K, Singh D, Zeng Z*, et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res;38(18):e178.
8. Parker JS, Mullins M, Cheang MC*, et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27(8):1160-7.
9. Culhane AC, Thioulouse J, Perriere G, Higgins DG. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics 2005;21(11):2789-90.
10. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27-30.