

Early molecular changes and a pivotal transition define triple-negative breast cancer progression in a murine model: a stage-resolved transcriptomic study

Irene Tadeo-Cervera

`irene.tadeo@uchceu.es`

Universidad Cardenal Herrera-CEU, CEU Universities

Cristian Herrera-Olivares

Universidad Cardenal Herrera-CEU, CEU Universities

Celeste Moya-Valera

INCLIVA Biomedical Research Institute

Deborah La Fauci

Polo Universitario Annunziata

Raquel Bosch-Romeu

Universidad Cardenal Herrera-CEU, CEU Universities

Luis Álvarez

Universidad Cardenal Herrera-CEU, CEU Universities

Aida Artigot

Universitat de València

Marina Pascual

Universidad Cardenal Herrera-CEU, CEU Universities

Antonella Locascio

Universidad Cardenal Herrera-CEU, CEU Universities

Joaquín Ortega

Universidad Cardenal Herrera-CEU, CEU Universities

Ana Reula

Universidad Cardenal Herrera-CEU, CEU Universities

Ana Bárbara García-García

INCLIVA Biomedical Research Institute

Joan Climent

Universidad Cardenal Herrera-CEU, CEU Universities

Research Article

Keywords: Triple-negative breast cancer, Orthotopic 4T1 model, Stage-resolved transcriptomics, Tumour development, Extracellular matrix.

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11048428/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Early molecular changes and a pivotal transition define triple-negative breast cancer progression in a murine model: a stage-resolved transcriptomic study

Irene Tadeo-Cervera^{1,2*†}, Cristian Herrera-Olivares^{1†}, Celeste Moya-Valera³, Deborah La Fauci⁴, Raquel Bosch-Romeu^{5,6}, Luis Álvarez¹, Aida Artigot⁷, Marina Pascual⁸, Antonella Locascio⁹, Joaquín Ortega¹⁰, Ana Reula¹¹, Ana Bárbara García-García^{3,12}, Joan Climent^{1,2,6,13*}

¹Grupo de investigación GenosysCEU, Departamento de Producción y Sanidad Animal, Salud Pública Veterinaria y Ciencia y Tecnología de los Alimentos, Facultad de Veterinaria, Universidad Cardenal Herrera-CEU, CEU Universities, 46115 Alfara del Patriarca, Spain.

²ECMOR Chair@CEU-UCH, Universidad Cardenal Herrera-CEU, CEU Universities, 46115 Alfara del Patriarca, Spain

³Genomics and Diabetes Unit, INCLIVA Biomedical Research Institute, 46010 Valencia, Spain.

⁴Department of Veterinary Sciences, Veterinary Physiology Unit, Polo Universitario Annunziata, 98168 Messina, Italy

⁵Departamento de Matemáticas, Física y Ciencias Tecnológicas, Universidad Cardenal Herrera-CEU, CEU Universities, 46115 Alfara del Patriarca, Spain;

⁶ESI International Chair@CEU-UCH, Universidad Cardenal Herrera-CEU, CEU Universities, 46115 Alfara del Patriarca, Spain

⁷Unidad Central de Investigación de Medicina (UCIM), Universitat de València, 46010 Valencia, Spain

⁸Departamento de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Cardenal Herrera-CEU, CEU Universities, 46115 Alfara del Patriarca, Spain.

⁹GenosysCEU, Departamento de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Cardenal Herrera-CEU, CEU Universities, 46115 Alfara del Patriarca, Spain.

¹⁰Grupo de Investigación Pathovet, Departamento de Producción y Sanidad Animal, Salud Pública Veterinaria y Ciencia y Tecnología de los Alimentos, Facultad de Veterinaria, Universidad Cardenal Herrera-CEU, CEU Universities, 46115 Alfara del Patriarca, Spain.

¹¹GenosysCEU, Departamento de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Cardenal Herrera-CEU, CEU Universities, 12006 Castellón de la Plana, Spain

¹²CIBER de Diabetes y Enfermedades Metabólicas Asociadas, Instituto de Salud Carlos III, Madrid, Spain

¹³Biomedical Research Institute, Universidad Cardenal Herrera-CEU, CEU Universities, 46115 Alfara del Patriarca, Spain

* Corresponding authors: irene.tadeo@uchceu.es;

joan.climentbataller@uchceu.es

† Co-first authors

Abstract

Background: Triple-negative breast cancer (TNBC) is an aggressive breast cancer subtype associated with rapid tumour growth, early metastatic dissemination and limited therapeutic options. Transcriptomic studies have provided extensive insight into TNBC biology, but how the mammary tissue transcriptome changes throughout tumour growth remains poorly understood.

Methods: We conducted stage-resolved RNA sequencing of mammary tissue collected at successive stages of orthotopic 4T1 tumour growth in immunocompetent BALB/c mice, alongside histopathological evaluation. Parental 4T1 cell line samples were included to differentiate transcriptional features associated with the in vivo tumour context.

Results: Histopathological evaluation confirmed progressive local invasion across successive stages, beginning with replacement of the cutaneous muscle layer and culminating in complete dermis replacement by the Late stage. Substantial transcriptomic variations from Control mammary tissue were already evident at the Early stages of tumour growth (G1-G2, days 3-7 post-inoculation). An individual analysis of the G1 stage, which lacked the macroscopically detectable features of a tumour, confirmed that the principal transcriptomic alterations were already present at this point. Across successive stages, the rate of gene-expression change was unevenly distributed, with the most pronounced shift occurring during the Medium 1 to Medium 2 transition, after which comparatively few additional changes were observed despite continued tumour growth. Functional enrichment analyses

showed a progressive increase in the representation of extracellular matrix organisation and inflammatory pathways, alongside a decline in epithelial differentiation- and metabolism-related programmes. Comparison with the parental 4T1 cell line revealed that a substantial component of the in vivo transcriptomic landscape, particularly extracellular matrix and immune-related programmes, was not represented in cultured tumour cells, highlighting the contribution of the tissue microenvironment to tumour biology.

Conclusions: This stage-by-stage approach shows that extensive molecular alterations in mammary tissue are already observed within days of tumour cell inoculation, preceding the appearance of a macroscopic tumour mass. Transcriptomic change was not constant but centred on a pivotal transition between intermediate stages, after which most transcriptional programmes were already established. This study offers a stage-resolved molecular framework for interpreting tissue-level changes during TNBC progression, and highlights that a substantial part of this landscape reflects TME-associated programmes absent from cultured tumour cells alone.

Keywords

Triple-negative breast cancer; Orthotopic 4T1 model; Stage-resolved transcriptomics; Tumour development; Extracellular matrix.

Background

Triple-negative breast cancer (TNBC) tends to grow rapidly and is more likely to metastasise at an early stage, often resulting in poorer patient outcomes [1]. The

main challenge in TNBC is the absence of estrogen receptors (ER), progesterone receptors (PR), and human epidermal growth factor receptor 2 (HER2), which limits the availability of targeted treatment options compared to other breast cancer subtypes [2,3]. Therefore, improving our understanding of the biological processes underlying the development and progression of this aggressive cancer type is crucial for identifying new therapeutic strategies and improving patient outcomes.

Large-scale transcriptomic studies have already advanced our understanding of TNBC biology, identifying molecular subtypes [4,5] and revealing the importance of processes such as the Extracellular Matrix (ECM) organisation [6,7], the immune infiltration [8] and the metabolic adaptations, including shifts in oxidative phosphorylation and glycolytic metabolism, during tumour progression [9,10].

However, these data have been obtained mostly from tumours sampled at a single time point of tumour development, thus providing limited information about when specific gene-regulatory changes first occur, how the transcriptional alterations evolve across successive stages of tumour development, and which biological transitions accompany the progression from early stages of tumour establishment to advanced disease [11]. Characterising how the mammary tissue transcriptome evolves through successive stages is essential for understanding the biological events that underlie tumour establishment and progression. Depicting this evolution is further complicated by the fact that TNBC is a highly heterogeneous disease and that tumour development is accompanied not only by gradual tumour growth but also by continuous changes in tissue composition, resulting from the

110 evolving balance between normal mammary tissue, tumour cells and the inclusion
111 of other components of the tissue microenvironment (TME) [12–14].

112 Addressing this gap in temporal resolution requires experimental models that allow
113 mammary tissue to be sampled across successive stages of tumour development
114 under controlled conditions. Orthotopic implantation of 4T1 mammary carcinoma
115 cells into immunocompetent BALB/c mice provides an established experimental
116 model that reproduces key features of human TNBC, including rapid local tumour
117 growth, spontaneous metastasis and interaction with an intact immune system [15–
118 17], while providing access to stages of tumour development that are difficult or
119 impossible to sample serially in human patients. This model therefore offers an
120 opportunity to characterise the evolution of mammary tissue transcriptomic
121 changes throughout tumour establishment and progression.

122 In the present study, we performed a stage-resolved transcriptomic
123 characterisation of mammary tissue during orthotopic 4T1 tumour growth by
124 integrating RNA sequencing with histopathological evaluation. The parental 4T1 cell
125 line was included as a reference to distinguish transcriptional programmes
126 associated with the in vivo tissue context from those intrinsic to tumour cells
127 maintained in culture. Using this approach, we aimed to determine when early
128 transcriptional alterations become detectable in mammary tissue during tumour
129 development, identify the major biological transitions accompanying disease
130 progression, and establish a stage-resolved molecular framework for interpreting
131 transcriptomic changes throughout orthotopic TNBC progression.

Methods

In order to characterise molecular changes across successive stages of mammary tumour development, we designed a stage-resolved study using the orthotopic 4T1 model in immunocompetent BALB/c mice, combining bulk RNA sequencing with histopathological evaluation. The parental 4T1 cell line was included as a reference for tumour cell-intrinsic transcriptional programmes. The experimental design, sample selection strategy and analytical procedures are described below.

Animal model and experimental design

Female BALB/cJrj mice (5 weeks old; Janvier Labs, France) were housed under controlled environmental conditions with ad libitum access to food and water. Following a one-week acclimatisation period, animals were assigned to control or experimental groups. All procedures were approved by the Committee for Animal Experimentation and Welfare of the Universitat de València (CEBA; protocol 2022 VSC PEA 0100) and conducted in accordance with Directive 2010/63/EU, Spanish legislation governing the protection of animals used for scientific purposes, and institutional guidelines. All animal procedures were performed at the animal facility of the Universitat de València.

Control animals received an orthotopic injection of 100 µL phosphate-buffered saline (PBS) into the left fourth mammary gland, whereas experimental animals received 10^4 4T1 mammary carcinoma cells (ATCC CRL-2539; LGC Standards S.L.U., Spain) suspended in 100 µL PBS at the same anatomical site. Cells were maintained in RPMI-1640 supplemented with 10% fetal bovine serum and 1%

penicillin/streptomycin under standard culture conditions, routinely tested negative for mycoplasma contamination by PCR, and used within 5 passages after thawing.

To characterise transcriptomic changes across successive stages of tumour growth (hereafter denoted G[n], with G referring to 'Group' and n indicating the group number), mammary tissue was collected at predefined times following tumour cell inoculation. Samples were obtained from Control and G1 animals on day 3, G2 on day 7, G3 on day 10, G4 on day 15, and G5 on day 17 post-inoculation (n = 5 per group; **Figure 1A**). To increase the number of biological replicates available at the final sampling point, transcriptomic data from two independent experimental cohorts, G6 and G7, were incorporated into the study in addition to the main experimental cohort. These cohorts corresponded to time points equivalent to G5 (G6: n = 5, day 17; G7: n = 6, day 16) and were collected under the same anaesthetic protocol, by mastectomy rather than by euthanasia and full dissection (see **Sample collection** below).

The samples included in the transcriptomic analyses were selected from a larger experimental cohort comprising nine tumour-bearing animals per experimental group and five control animals. To maximise the transcriptional signal and following a similar volume-based selection approach previously reported for transcriptomic analyses in murine tumours [18], a predefined, volume-enriched sampling strategy was established before transcriptomic analysis: for groups with measurable tumours (G2 to G7), the five animals with the largest tumour volumes were selected to maximise tumour representation within the tissue analysed by RNA sequencing. For G1, and in the event of tied tumour volumes, animals with body weight closest

to the group mean were selected instead. This strategy was designed for stage characterisation and was not intended to provide a random representation of the full tumour-volume distribution at each time point; accordingly, the groups defined in this study should be interpreted as representing successive stages of tumour development rather than strictly chronological time points.

The experimental groups were classified into four successive stages: Early (G1–G2), Medium 1 (G3), Medium 2 (G4) and Late (G5–G7). The grouping of G5, G6, and G7 into a single Late stage group was subsequently supported by the histopathological findings and by the absence of statistically significant global transcriptomic separation among G5, G6, and G7 (see **Statistical analysis below**).

Animal welfare monitoring

Animals were monitored daily throughout the study for general health, body weight, tumour growth and clinical signs of discomfort. Humane endpoints were established before the start of the experiment in accordance with the approved animal welfare protocol. Body weight and tumour dimensions (length and width) were measured every 3–5 days using a digital calliper. Tumour volume was calculated using the modified ellipsoidal formula: $V = (\pi/6) \times \text{length} \times \text{width}^2$. No animals included in the present study reached predefined humane endpoints before the planned experimental termination, nor did they exceed the maximum tumour burden or body weight loss thresholds established by the approved animal experimentation protocols. Individual body weight and tumour volume measurements for the subset of samples included in the transcriptomic analyses,

as well as descriptive statistics of tumour volume for the complete experimental cohort, are provided in **Supplementary Table S1**

Sample collection

Animals from Control to G5 were euthanised by cervical dislocation under anaesthesia with 5% isoflurane, and mammary tissues were collected immediately after euthanasia. In G6 and G7, mammary tumours were collected by mastectomy under the same anaesthetic protocol once animals had reached the predefined Late stage of tumour development. Following tissue collection, samples were immediately placed in cold PBS and dissected. A small fragment intended for RNA sequencing was rapidly processed and stored at -80 °C, whereas the remaining, adjacent tissue was allocated for histopathological analysis.

In addition to mammary tissue, three vials of the parental 4T1 cell line, all derived from the same original stock, were independently thawed, cultured, and pelleted for RNA sequencing.

Histopathological analyses

Mammary tissue samples were fixed in 10% neutral-buffered formalin, paraffin-embedded, sectioned and stained with haematoxylin and eosin (H&E) using standard histopathological procedures. Histopathological processing and evaluation were performed at Universidad Cardenal Herrera-CEU, CEU Universities (Alfara del Patriarca, Spain). All sections were evaluated by the same board-certified pathologist, who assessed the presence of tumour tissue and its histopathological characteristics at each developmental stage, including the extent of necrosis,

mitotic activity in areas of highest density (hotspots), nuclear morphology, and other notable histological features. This evaluation was descriptive and qualitative in nature. Representative images were acquired using a NanoZoomer slide scanner (Hamamatsu Photonics, Hamamatsu, Japan).

RNA seq analysis

Total RNA was extracted from mammary tissue and cell line samples by Seqplexing (Paterna, Valencia, Spain) using the MagMAX™ mirVana™ Total RNA Isolation Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. RNA quantity and integrity were assessed using a QIAGEN automated electrophoresis system before library preparation, and only samples meeting quality criteria were sequenced.

Strand-specific 3'mRNA libraries were prepared by Seqplexing (Paterna, Valencia, Spain) using an oligo(dT)-primed protocol incorporating unique molecular identifiers (UMIs): the first 10 nucleotides of Read 1 contained six random nucleotides followed by the fixed sequence TATA. Libraries were sequenced on an Illumina NovaSeq X Series platform using paired-end sequencing.

Raw reads were processed using a custom pipeline. Read quality was assessed with FastQC before and after preprocessing. Adapters were removed by automatic paired-end detection with fastp, applying its default quality filtering (Phred ≥ 15) and discarding reads shorter than 25 nucleotides. UMIs were extracted from Read 1, retaining the random sequence for downstream deduplication and discarding the fixed TATA sequence. Reads were trimmed to remove poly(A) tails and adapter

sequences, discarding those shorter than 20 nucleotides, and aligned to the *Mus musculus* reference genome (GRCm39). Alignments were sorted and indexed, and gene-level counts were assigned against the GENCODE M38 annotation. UMI-based deduplication was then performed, collapsing reads that shared the same UMI and gene assignment to generate raw gene-level count matrices for downstream analysis.

Statistical analysis

Tumour volume distribution

Tumour volume did not meet normality assumptions (Kolmogorov-Smirnov test with Lilliefors correction, $p < 0.001$); group differences were therefore assessed with the Kruskal-Wallis test, with FDR-adjusted post hoc pairwise comparisons.

Differential expression analysis

Differential expression was assessed using DESeq2 (v1.46.0), modelling counts with disease stage as the explanatory variable and Control as the reference level. Genes with fewer than ten counts in every sample were excluded before model fitting. Six pairwise contrasts were extracted: Early vs Control, Medium 1 vs Early, Medium 2 vs Medium 1, Late vs Medium 2, Cell line vs Late, and Late vs Control. P-values were corrected for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) procedure, and genes with an FDR-adjusted p-value below 0.05 were considered differentially expressed.

Assessment of group structure

Global sample relationships were examined using principal component analysis (PCA) of the 1,000 most variable genes, and PERMANOVA (vegan R package, v2.7-3) was used to test for group separation based on Euclidean distances. G5, G6 and G7 were defined as replicates of the Late biological stage based on tumour burden, collection time and histopathological composition. PERMANOVA did not detect statistically significant global transcriptomic separation among the three cohorts ($R^2 = 0.251$, $p = 0.086$). G3 (Medium 1) and G4 (Medium 2) likewise showed no significant separation ($R^2 = 0.280$, $p = 0.077$), but samples were consistently closer to members of their own group than to the other; the two stages were therefore retained as separate levels in the differential expression model. PCA and PERMANOVA were performed on variance-stabilised expression values obtained with the variance-stabilizing transformation (VST) implemented in DESeq2.

Temporal trend analysis

Temporal expression dynamics across G1–G4 and the pooled Late stage were modelled using maSigPro (R package, v1.78.0), fitting a polynomial regression to each gene and retaining genes with a significant temporal profile ($FDR < 0.05$) whose model explained at least 60% of the observed variance ($R^2 \geq 0.6$). Retained genes were grouped by hierarchical clustering; a three-cluster solution was adopted based on silhouette width and interpretability of the resulting trajectories.

Functional enrichment

Functional interpretation of the up- and down-regulated genes from each contrast, and of each temporal cluster, was performed by over-representation analysis against Gene Ontology Biological Process, Kyoto Encyclopedia of Genes and

287 Genomes (KEGG) and Reactome, with terms considered significant at an adjusted
288 p-value and q-value below 0.05.

289 **Software**

290 All analyses were performed in R (v4.4.1), except the tumour volume distribution
291 analysis assessed in IBM SPSS Statistics (v29.0.2.0).

292 **Results**

293 Macroscopic and histopathological progression define successive
294 stages of tumour development.

295 Tumour growth was monitored throughout the study to characterise its
296 macroscopic progression (**Figure 1A and 1B**). No measurable tumours were
297 detected in Control or G1 animals. Tumour volume increased gradually from G2
298 onwards, reaching its highest values in the Late stage.

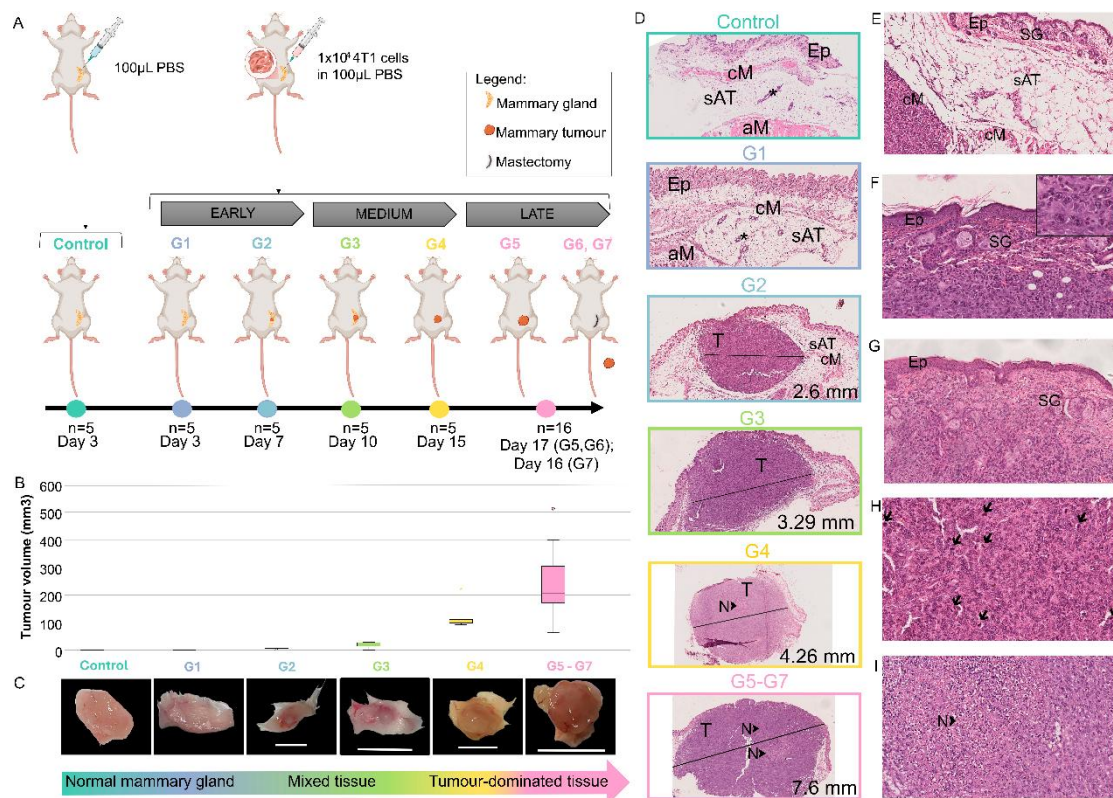


Figure 1. Experimental design and macroscopic/histopathological progression of mammary tumour growth. (A) Female BALB/cJRj mice received an orthotopic injection of PBS (Control) or 10⁴ 4T1 cells into the left fourth mammary gland. Mammary tissue was collected across successive stages of tumour growth. Early, Medium 1 and Medium 2, and Late denote the developmental stages used throughout the study. G6 and G7 samples were obtained by mastectomy after tumours reached the predefined Late stage. (B) Tumour volume at tissue collection across experimental groups. (C) Representative macroscopic images of mammary glands/tumours for each stage, acquired before tissue sectioning for frozen and histopathological processing (white scale bars, 10 mm). Samples illustrate the progressively increasing proportion of tumour tissue relative to the surrounding host tissue with advancing stage. (D) Representative histological images (H&E) of mammary tissue at each developmental stage, selected independently as the most

representative sections for each group; these correspond to different animals from those shown in (C), and tumour diameters measured on the histological sections therefore differ from the macroscopic measurements in (C). The largest tumour diameter is indicated by a black line, with the corresponding value (mm) shown in the bottom right corner of each panel. Ep, epidermis; cM, cutaneous muscle; aM, abdominal muscle; sAT, subcutaneous adipose tissue (fat pad); asterisks (*) indicate mammary tissue, shown in the Control and G1 panels. T, tumour tissue, shown from G2 onwards. N (arrowhead) indicates necrosis, shown in the Medium 2 (G4) and Late-stage (G5-G7) panels. **(E-G)** High-magnification images illustrating the sequential local invasion of the skin: **(E)** infiltration of the cutaneous muscle layer; **(F)** infiltration, with partial replacement of the dermis. Inset: detail of nuclear atypia; **(G)** complete replacement of the dermis, with sebaceous glands (SG) completely surrounded by tumour cells. **(H-I)** High-magnification images from Medium 2 and Late-stage tumours showing marked nuclear atypia, with **(H)** detailing a mitotic hotspot (arrows) and **(I)** an area of necrosis (N).

Descriptive tumour volume statistics for the groups with detectable tumour mass (G2 to G5-G7) are summarised in **Table 1**. Overall, tumour volume differed significantly among experimental groups (Kruskal-Wallis test, $H(5) = 35.51$, $p < 0.001$). Pairwise comparisons among tumour-bearing groups showed that the Late stage (G5–G7) exhibited significantly larger tumour volumes than G1 (FDR-adjusted $p < 0.001$), G2 (FDR-adjusted $p = 0.010$) and G3 (FDR-adjusted $p = 0.020$). In addition, G4 tumours were significantly larger than those in G1 (FDR-adjusted $p = 0.017$). No other pairwise comparisons reached statistical significance.

Table 1. Tumour volume at tissue collection across successive stages of tumour development.

STAGE	GROUP	n	Mean (mm ³)	SD	Minimum	Maximum
Early	G2	5	6.22	3.45	0.52	9.20
Medium 1	G3	5	18.06	11.98	0.52	30.15
Medium 2	G4	5	133.03	55.94	96.62	231.84
Late	G5-G7	16	255.21	112.52	66.35	536.78

SD: standard deviation. Tumour volume was calculated using the modified ellipsoidal formula: $V = (\pi/6) \times \text{length} \times \text{width}^2$. Only experimental groups in which a measurable tumour mass was detected are shown. Values correspond to the animals included in the transcriptomic analyses.

The relative contribution of host and tumour tissue to the section used for RNA sequencing was assessed macroscopically at the time of dissection (**Figure 1C**), and tissue was sampled representatively with respect to this visible composition. From G2 to G3, the dissected mass comprised an increasing volume of tumour tissue alongside a correspondingly decreasing, though still substantially present, volume of host mammary stroma. From G4 onwards, host tissue was essentially absent from the dissected sample beyond the overlying skin, such that the fragment analysed by RNA sequencing was almost exclusively tumour tissue (**Figure 1B**).

Histopathological evaluation (**Figure 1D-I**) revealed the expected features of the normal mammary gland in Control animals, comprising the overlying skin (with epidermis, dermis, sebaceous glands and the underlying cutaneous muscle layer), beneath which lay the mammary fat pad itself (composed predominantly of adipose

tissue interspersed with sparse mammary tubuloepithelial structures), resting on the underlying abdominal muscular wall. G1 samples were histologically indistinguishable from Control, despite the presence of inoculated tumour cells. Tumour tissue first became histologically evident in G2, presenting as discrete foci of reduced size, consistent with the localised nature of the orthotopic inoculation, and already displaying histological features associated with malignancy (nuclear atypia and high mitotic activity). Local invasion followed a broadly consistent pattern across successive stages: the underlying cutaneous muscle layer began to be infiltrated and partially replaced as early as G2, whereas the overlying dermis was progressively infiltrated and replaced throughout the Medium 1 stage (G3) and the Medium 2 stage (G4), and was consistently and extensively replaced by the Late stage (G5-G7). Nuclear atypia was present from the earliest tumour foci onwards (**Figure 1F**). Necrosis was generally absent in G2 and G3, appeared inconsistently from the Medium 2 stage (G4) onwards, and became a consistent, though generally limited (<50%), feature of the tumour mass (mostly in the central areas) from the Late stage onwards. Mitotic activity was already high from the earliest tumour foci (G2) onwards, with considerable heterogeneity both within and across groups; within individual tumours, mitotic activity also tended to be higher towards the periphery of the mass than in central areas. Histopathological evaluation of the Late stage (G5, G6 and G7) showed a consistent pattern with tumour tissue almost completely replacing the hypodermis, muscular layer and dermis, elevating the epidermis and leaving only a residual rim of covering epithelium together with foci of necrosis within the tumour mass, without evidence of systematic differences among these three groups.

Together, these macroscopic and histopathological findings confirm a pattern of progressive tumour establishment and growth, supporting the use of these groups as a staging framework for the subsequent transcriptomic analyses.

Global transcriptomic dynamics across tumour growth

To investigate the global transcriptomic dynamics associated with tumour growth, principal component analysis (PCA) was performed on all mammary tissue samples and the 4T1 cell line, revealing a stage-dependent organisation of the samples (**Figure 2A**). PC1 and PC2 explained 53% and 24% of the total transcriptomic variance, respectively. Samples from G1 and G2 displayed highly similar transcriptional profiles. They clustered separately from both Control and subsequent disease stages, supporting their classification as a single Early stage for downstream transcriptomic and differential expression analyses. However, G1 was additionally analysed separately to evaluate transcriptional alterations at day 3, before macroscopic tumour detection. G3 and G4 occupied distinct positions in the PCA, supporting their classification as two transcriptionally distinct intermediate stages (Medium 1 and Medium 2) (**Figure 2**). Likewise, Late-stage samples (G5-G7) formed an overlapping cluster, further supporting their classification as a single Late stage. The 4T1 cell line clustered close to the Late-stage samples while remaining clearly separated from all mammary tissue samples, highlighting persistent transcriptomic differences between cultured tumour cells and tumour tissue in vivo.

The expression heatmap of the 50 most variable genes, with samples grouped according to their assigned disease stage, further illustrated the global

transcriptomic changes accompanying tumour growth (**Figure 2C**). The heatmap revealed distinct blocks of genes with similar expression trajectories. *Spp1* and the epithelial keratin *Krt18* showed the opposite trend to most other genes in this set, remaining low in Control and Early samples and increasing markedly from the later stages onward and in the cell line. A second block, comprising skeletal muscle-associated transcripts (*Ttn*, *Pvalb*, *Myh1*, *Ckm*) showed highest expression in Control and Early samples and declined progressively thereafter. A third block, dominated by keratin and cornified envelope genes (*Krt5*, *Krt10*, *Loricrin*, *Flg*), showed a progressive overall decrease from Early/Medium 1 onward. Biological replicates within each stage displayed consistent expression profiles, and the 4T1 cell line formed a visually distinct block, consistent with the transcriptional differences from mammary tissue described above.

Differential expression analysis confirmed the extensive transcriptomic transitions suggested by PCA and heatmap analysis (**Figure 2B**). Comparison of the Early stage with Controls identified 3,306 differentially expressed genes (DEGs), indicating that widespread molecular alterations are already present at this early stage. Similarly, extensive transcriptomic changes were observed between the Early and Medium 1 stages (3,077 DEGs) and between Medium 1 and Medium 2 (2,795 DEGs). In contrast, only 628 DEGs differentiated Medium 2 from the Late stage, indicating that the most pronounced transcriptomic transitions occur during the early and intermediate phases of tumour development, whereas the Late stage is characterised by a comparatively more stable transcriptional profile. Consistent with this progressive divergence, comparison of the Late stage with Controls

identified the largest overall transcriptional difference (4,767 DEGs). Notably, a substantial number of genes (1,658 DEGs) also distinguished the Late stage from the 4T1 cell line, highlighting persistent transcriptional differences between in vivo tumour tissue and cultured tumour cells.

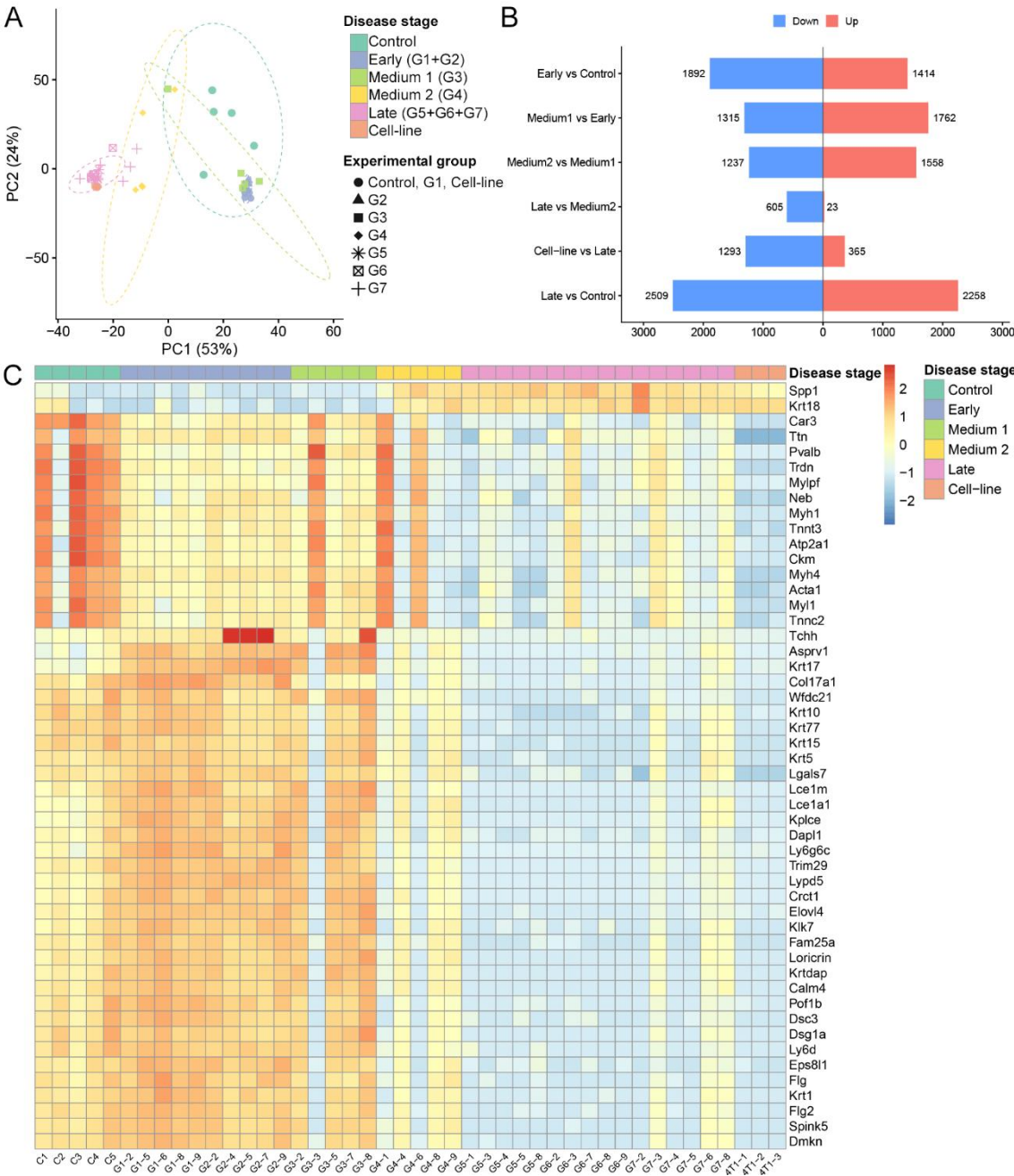


Figure 2. Global transcriptomic dynamics across successive stages of mammary tumour development. (A) Principal component analysis (PCA) of

431 mammary tissue transcriptomes across disease stages and the 4T1 cell line (PC1:
432 53%, PC2: 24% of variance). **(B)** Number of differentially expressed genes (DEGs;
433 $\text{FDR} < 0.05$, $|\log_2\text{FC}| \geq 1$) per pairwise comparison, shown separately for genes with
434 increased and decreased expression. **(C)** Heatmap of the 50 most variable genes
435 across samples, ordered by disease stage, showing stage-dependent
436 transcriptomic organisation and distinct clustering of the 4T1 cell line.

437 Given the extensive transcriptional divergence already observed at the Early stage,
438 we next characterised the earliest molecular alterations associated with tumour
439 development in greater detail.

440 Substantial transcriptomic alterations are detectable from the
441 earliest stages of tumour growth.

442 Functional enrichment analysis of the 3,306 DEGs identified between the Early
443 stage (G1-G2) and Control mammary tissue provided further insight into the
444 biological processes underlying this early divergence (**Figure 3**). Of these genes,
445 1,414 showed increased expression, and 1,892 showed decreased expression
446 (**Figure 2B**), at a stage when tumours, if present, remained small (G2) or
447 macroscopically undetectable (G1). Among the most differentially expressed
448 genes, those with increased expression included some keratin/cornified envelope-
449 associated genes (**Figure 3A**). These findings support a marked divergence in the
450 transcriptomic composition of mammary tissue during the earliest stage. Genes
451 with decreased expression in Early relative to Control were predominantly enriched
452 for biological processes related to cellular respiration and oxidative

phosphorylation, including multiple components of the mitochondrial electron transport chain, such as *Ndufa1*, *Ndufa5*, *Ndufa8*, *Ndufb1*, *Cox5a*, *Cox7a1*, *Cox7b* and *Cox7c* (**Figure 3B; Supplementary Table S3**). Conversely, genes with increased expression were primarily associated with cell cycle progression and chromosome segregation, consistent with increased representation of proliferative and cell-cycle-associated transcriptional programmes and were enriched for pathways related to skin/epidermis development and keratinocyte differentiation (**Figure 3B; Supplementary Table S3**). To determine whether these molecular alterations were already present at the earliest sampling point, G1 (day 3) samples were analysed separately. Of the DEGs identified in the Early vs Control comparison, 2,881 (87.1%) were also detected as DEGs in the same direction in the G1 vs Control comparison alone (highlighted in bold in **Supplementary Table S3**), indicating that a substantial proportion of the transcriptional changes captured in the pooled Early group were already established at this point. Functional enrichment analysis confirmed that the major biological patterns identified in the pooled Early stage were already evident at day 3 (**Supplementary Figure S2**): despite the absence of a macroscopically detectable tumour, genes with increased expression in G1 were similarly enriched for skin/epidermis development and keratinocyte differentiation. Genes with decreased expression remained predominantly associated with cellular respiration and oxidative phosphorylation, closely mirroring the enrichment profile observed in the combined Early group.

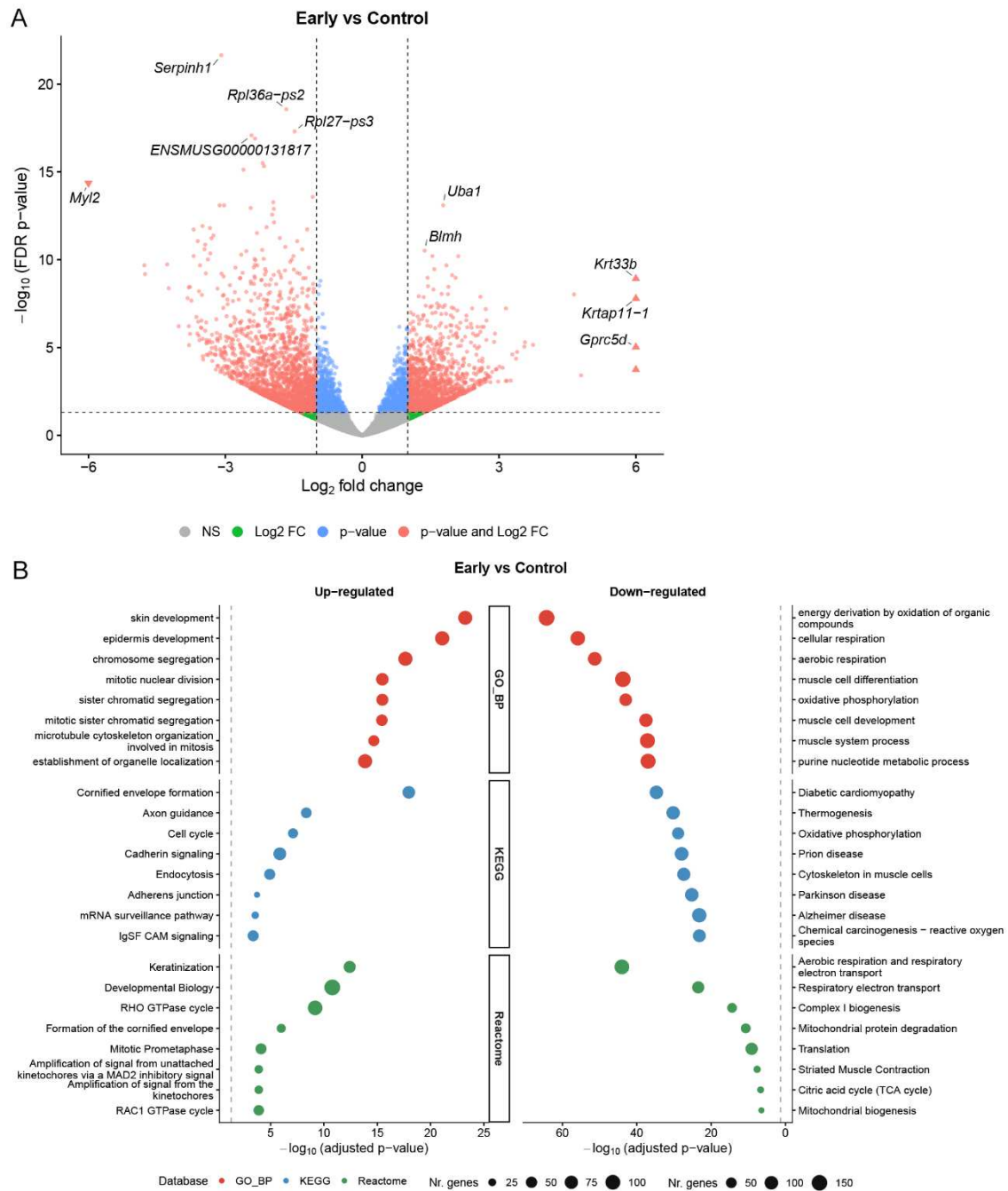


Figure 3. Early-stage transcriptomic alterations. (A) Volcano plot showing differentially expressed genes (DEGs) between Early stage (G1-G2) and Control mammary tissue ($\text{FDR} < 0.05$, $|\text{Log}_2 \text{FC}| \geq 1$). Dashed lines indicate significance and fold-change thresholds. Triangles indicate genes that fall outside the plotting limits. The five genes with the greatest increase and the five with the greatest decrease in expression, ranked by Euclidean distance from the origin in $\text{Log}_2 \text{FC}$ vs. $-\log_{10}(\text{FDR})$

space, are labelled. **(B)** Functional enrichment analysis of genes with increased and decreased expression in the Early-versus-Control comparison. Representative significantly enriched Gene Ontology Biological Process (GO_BP), KEGG and Reactome terms are shown. Dot size represents the number of genes associated with each term, and the x-axis indicates enrichment significance ($-\log_{10}$ adjusted p-value).

Temporal trajectory analysis defines opposing axes of ECM/inflammatory increase and epithelial/metabolic decline.

To investigate how gene expression evolved throughout mammary tumour development, temporal trend analysis was performed across the sequential experimental groups (G1, G2, G3, G4, and G5-G7). Among the clustering solutions evaluated, the three-cluster model provided the best representation of the data (average silhouette score = 0.345), identifying three major temporal programmes **(Figure 4A)**. Rather than representing stage-specific transcriptional signatures, these clusters describe coordinated patterns of gene expression shared by groups of genes across tumour growth. The full list of genes assigned to each cluster is provided in **Supplementary Table S2**.

Cluster 1 (n = 476) displayed a progressive increase in expression throughout tumour development **(Figure 4A)**. Functional enrichment analysis consistently associated this programme with extracellular matrix organisation, inflammatory response, and chemotaxis, as well as signalling pathways related to integrins, chemokines, tumour necrosis factor (TNF), and advanced glycation end-products–

receptor for advanced glycation end-products (AGE-RAGE) signalling. Reactome analysis further highlighted collagen and elastic fibre formation, as well as extracellular matrix organisation (**Figure 4B; Supplementary Figure S1**). Representative genes included multiple components involved in extracellular matrix organisation and remodelling (*Fln1*, *Col4a1*, *Col5a1*, *Col12a1*, *Hspg2*, *Lamb1*, *Lox*, *Loxl3*, *Plod2*, *Pcolce*), matrix-cell interactions (*Itga1*, *Itgb2*, *Itgb3*), inflammatory signalling (*Cxcl9*, *Cxcl10*, *Ccr1*, *Ccr2*, *Ccr5*) and transforming growth factor beta (TGF- β)-associated matrix regulation (*Tgfb1*, *Ltbp1-3*; **Supplementary Table S2**).

In contrast, Clusters 2 (n = 412) and 3 (n = 283) exhibited progressive decreases in expression during tumour development, although following distinct temporal trajectories (**Figure 4A**). Cluster 2 showed its most pronounced decline between G3 (Medium 1) and G4 (Medium 2) and was enriched for biological processes associated with epidermis development, keratinocyte differentiation, keratinisation and establishment of the skin barrier. Consistent with these enrichments, this cluster contained numerous members of the epidermal differentiation complex, including twelve genes of the late cornified envelope (Lce) family, together with *Loricrin*, *Filaggrin* (*Flg*, *Flg2*) and *Involucrin* (*Ivl*); alongside multiple keratin genes (*Krt1*, *Krt5*, *Krt10*, *Krt14*, *Krt17*; **Supplementary Table S2**). Cluster 3 displayed a more gradual decrease overall, spanning the entire course of tumour progression rather than collapsing at a single stage, and was enriched for biological processes related to skin development and fat cell differentiation (**Figure 4B; Supplementary Figure S1; Supplementary Table S2**). Representative genes included regulators of

526 adipocyte differentiation and lipid metabolism (*Cebpa*, *Dgat2*, *Scd1*, *Ppara*, *Acs1*)
527 alongside genes associated with epidermal adhesion and differentiation (*Col17a1*,
528 *Cdh3*, *Klf4*, *Scel*; **Supplementary Table S2**).

529 All three temporal programmes, despite their distinct overall shapes and
530 magnitudes, showed their largest overall change between G3 (Medium 1) and G4
531 (Medium 2) (**Figure 4A**). This local acceleration was considerably more pronounced
532 for Clusters 1 and 2 than for Cluster 3, which instead showed a comparatively
533 modest inflexion within its broader gradual decline. This transition coincided with
534 the shift, described above, towards an almost exclusively tumour-dominated
535 composition of the dissected sample from G4 onwards (**Figure 1**), the stage
536 separation identified by PCA and expression heatmap (**Figure 2A, C**), and the
537 extensive transcriptomic differences detected between these stages by pairwise
538 differential expression analysis (**Figure 2B**). Together, these complementary
539 analyses identify the transition from G3 (Medium 1) to G4 (Medium 2) as a point of
540 particularly pronounced transcriptional change during mammary tumour growth.

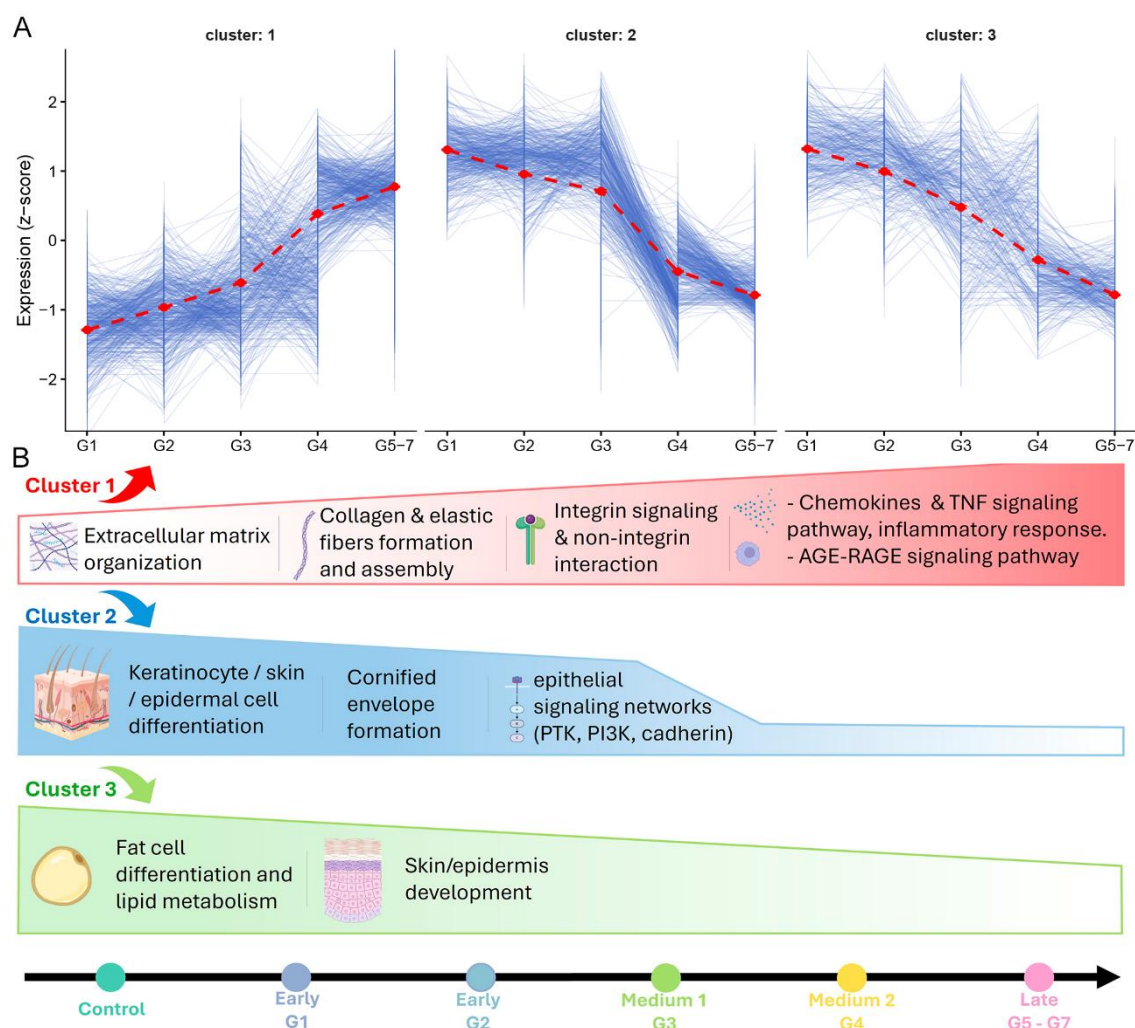


Figure 4. Temporal gene expression trajectories during mammary tumour growth. (A) Temporal tendency analysis identified three major clusters of genes with similar expression trajectories across successive stages of mammary tumour development (G1 to G5-G7). Blue lines represent individual gene expression trajectories (z-score), whereas the red dashed line indicates the average trajectory for each cluster. (B) Schematic summary of the major biological processes represented by each temporal programme, based on integrated Gene Ontology Biological Process (GO_BP), KEGG and Reactome enrichment analyses (Supplementary Figure S1). The width of each wedge reflects the relative expression trend of the corresponding programme across tumour progression. It

does not represent the number of genes or imply a temporal sequence among the annotated biological processes.

Pairwise comparisons support the temporal programmes and reveal their progressive establishment across disease stages.

To determine when the biological programmes identified by temporal trajectory analysis became detectable during tumour progression, differential expression and functional enrichment analyses were performed between consecutive disease stages (Medium 1 vs Early, Medium 2 vs Medium 1 and Late vs Medium 2), together with the cumulative comparison between Late-stage tumours and Control mammary tissue (**Supplementary Figure S3; Figure 5**). These pairwise analyses were consistent with the temporal patterns identified above and provided stage-specific resolution of the biological processes emerging throughout tumour progression.

The transition from Early to Medium 1 was characterised predominantly by enrichment of oxidative phosphorylation, mitochondrial respiration and aerobic respiration among genes with increased expression, indicating an increased representation of respiratory programmes following their marked reduction at the Early stage (**Supplementary Figure S3**). Conversely, genes with decreased expression during this transition were enriched for RNA processing, chromosome organisation and cell cycle-related processes, consistent with a reduction in the representation of cell-cycle and chromosome-organisation programmes that were prominent in the Early-versus-Control comparison.

The Medium 1-to-Medium 2 transition represented the most pronounced functional reorganisation observed between consecutive disease stages. Genes with increased expression at this transition were enriched for chromosome segregation, mitotic division and cell cycle regulation, whereas genes with decreased expression were predominantly associated with epidermis development, keratinocyte differentiation and keratinisation (**Supplementary Figure S3**). In addition, several genes implicated in extracellular matrix remodelling and inflammatory signalling showed significantly higher expression at this transition, including *Spp1* and *Ankrd1*, and remained elevated thereafter, consistent with the progressive establishment of the tumour-associated programme identified as Cluster 1 in the temporal tendency analysis.

Relatively few transcriptional changes distinguished Late-stage tumours from Medium 2, indicating that most of the major transcriptional programmes had already become detectable by the Medium 2 stage. Nevertheless, enrichment of processes related to positive regulation of chemotaxis and continued cell cycle activity suggests that inflammatory and proliferative transcriptional programmes remain represented during advanced disease (**Supplementary Figure S3**).

The cumulative Late-versus-Control comparison provided a cumulative view of the transcriptional changes associated with tumour progression (**Figure 5**). Genes with increased expression in Late-stage tumours were predominantly associated with cell cycle progression, DNA replication and chromosome segregation. In contrast, genes with decreased expression were enriched for oxidative phosphorylation, cellular respiration and muscle differentiation. Among the genes showing the most

significant increase in expression was *Spp1*, which first became detectable at higher levels during the Medium 1-to-Medium 2 transition and remained highly expressed thereafter, consistent with the progressive establishment of the extracellular matrix- and inflammatory- associated programme identified in Cluster 1.

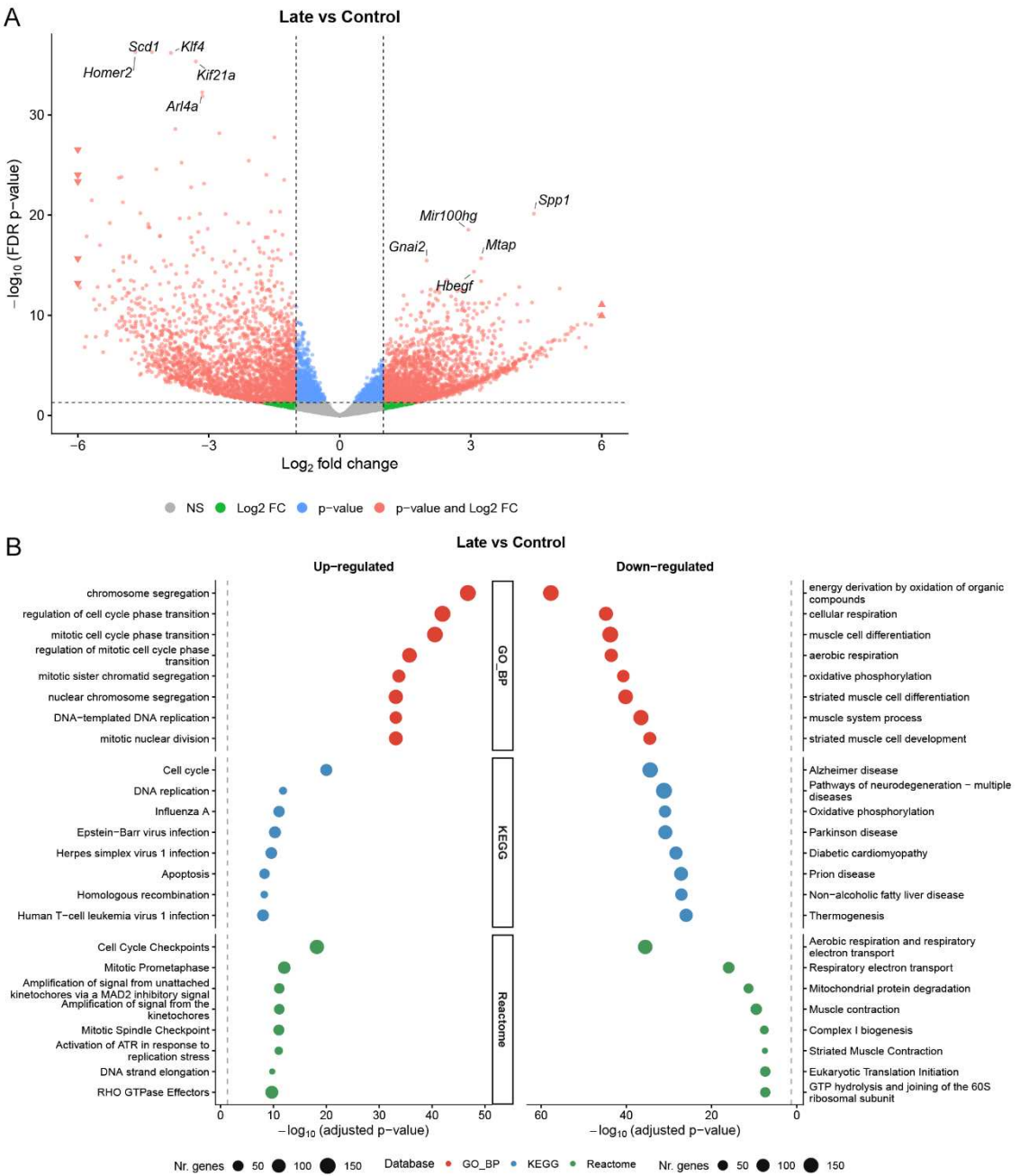


Figure 5. Differential gene expression and functional enrichment in Late-stage tumours relative to Control mammary tissue. (A)

Volcano plot showing differentially expressed genes (DEGs) between Late-stage tumours (G5-G7) and Control mammary tissue ($FDR < 0.05$, $|\log_2FC| \geq 1$). Dashed lines indicate significance and fold-change thresholds. Triangles denote genes that fall outside the plotting limits. The five genes with the greatest increase and the five with the greatest decrease in expression, ranked by Euclidean distance from the origin in $\log_2FC$ vs. $-\log_{10}(FDR)$ space, are labelled. **(B)** Functional enrichment analysis of genes with increased and decreased expression in the Late-versus-Control comparison. Representative significantly enriched Gene Ontology Biological Process (GO_BP), KEGG and Reactome terms are shown. Dot size represents the number of genes associated with each term, and the x-axis indicates enrichment significance ($-\log_{10}$ adjusted p-value). This enrichment profile is consistent with the cumulative establishment of the biological programmes identified by temporal trajectory analysis.

Comparison with the 4T1 cell line identifies tumour microenvironment-associated transcriptional programmes.

To determine which transcriptional programmes distinguish advanced tumours from tumour cells, differential expression analysis was performed comparing Late-stage tumours (G5-G7) with the parental 4T1 cell line (**Figure 6**). A total of 1,658 differentially expressed genes were identified (**Figure 2B**), indicating substantial transcriptional differences between advanced tumour tissue and the parental 4T1 cell line.

626 Genes with increased expression in the 4T1 cell line were predominantly enriched
627 for biological processes associated with ribosome biogenesis, cytoplasmic
628 translation, mitochondrial translation, rRNA processing and mitochondrial gene
629 expression (**Figure 6B**). Consistently, KEGG and Reactome analyses highlighted
630 ribosome, oxidative phosphorylation and translation-related pathways, coherent
631 with increased representation of protein synthesis, ribosome biogenesis and
632 mitochondrial translation programmes in cultured cells. Representative genes
633 included components of the cytoplasmic and mitochondrial translational
634 machinery (*Rpl26*, *Eif3g*, *Mrpl34*, *Mrpl40*) and rRNA processing (*Rrp9*), and
635 aminoacid biosynthesis (*Asns*) (**Figure 6A; Supplementary Table S3**)

636 In contrast, genes with increased expression in Late-stage tumours were strongly
637 enriched for biological processes related to chemotaxis, leukocyte migration,
638 myeloid leukocyte migration, and leukocyte cell-cell adhesion, as well as
639 extracellular matrix organisation, collagen formation, and cytokine-chemokine
640 signalling pathways (**Figure 6B**). Representative genes contributing to these
641 enrichments included *Col1a1*, *Apoe* and *Maf*, among the most differentially
642 expressed genes in this comparison (**Figure 6A**), together with the macrophage
643 markers *C1qa* and *Msr1*, both markedly higher in tumour tissue relative to the
644 isolated 4T1 cell line (**Supplementary Table S3**).

645 Interestingly, the extracellular matrix remodelling and immune-related programmes
646 that progressively emerged during tumour growth (**Figures 4 and 5**) were not
647 represented to the same extent in the transcriptional profile of cultured 4T1 cells
648 (**Figure 6B**). Conversely, translation- and ribosome-associated programmes were

649 comparatively enriched in vitro, indicating that these predominantly reflect tumour
650 cell-intrinsic functions rather than the complex biological processes operating
651 within the tumour tissue.

652 Together, these findings underscore that the transcriptomic landscape of advanced
653 tumours cannot be explained solely by tumour cell-intrinsic gene expression
654 programmes. Instead, a substantial proportion of the in vivo transcriptional profile
655 reflects biological processes associated with the TME, particularly extracellular
656 matrix remodelling and immune cell-associated processes, particularly leukocyte
657 migration and chemotaxis.

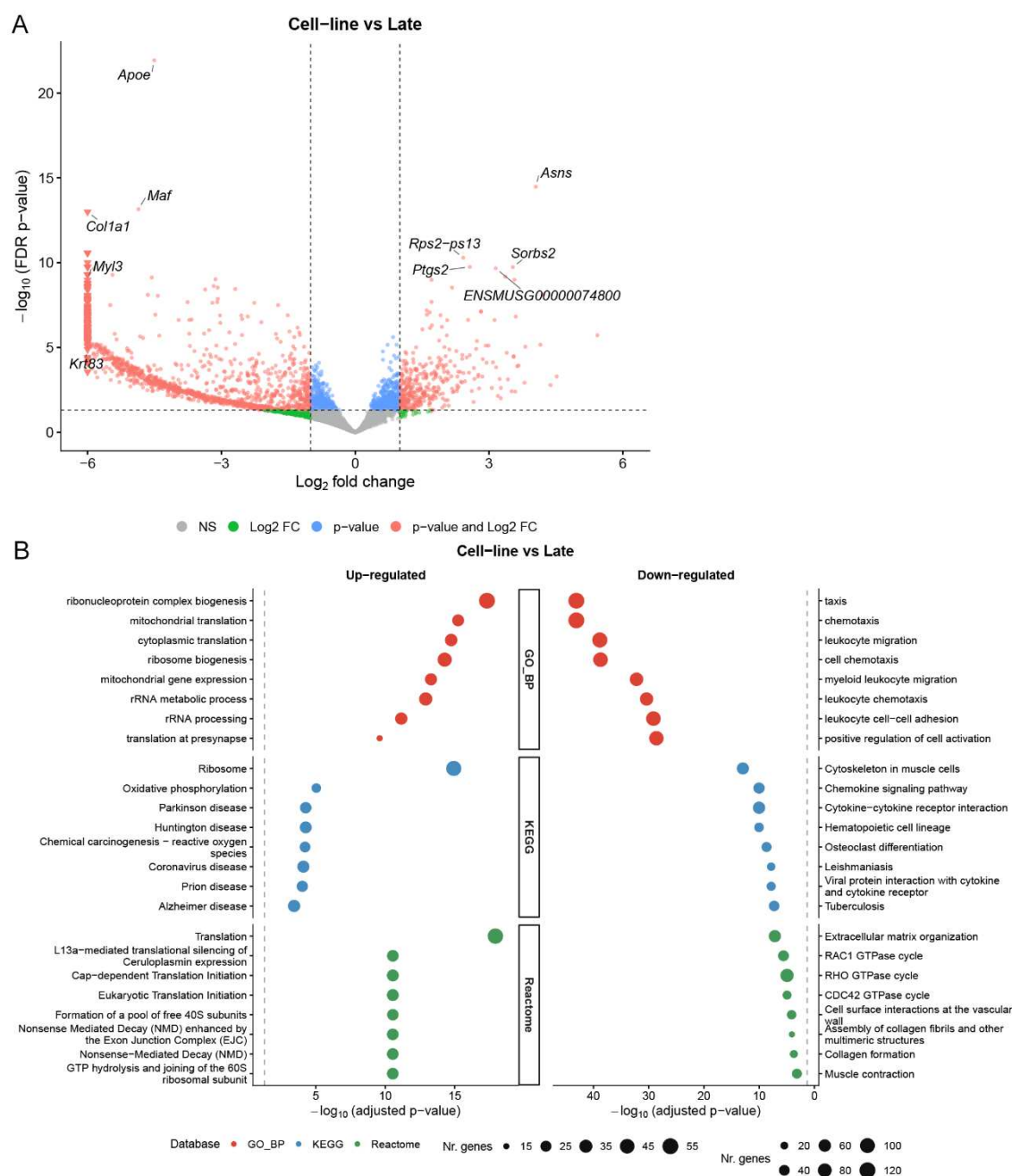


Figure 6. Transcriptomic comparison between the parental 4T1 cell line and Late-stage tumours. (A) Volcano plot showing differentially expressed genes (DEGs) between the parental 4T1 cell line and Late-stage tumours (G5-G7) (FDR < 0.05, $|\log_2 FC| \geq 1$). Dashed lines indicate significance and fold-change thresholds. Triangles denote genes that fall outside the plotting limits. The five genes with the greatest increase and the five with the greatest decrease in expression, ranked by

Euclidean distance from the origin in $\log_2\text{FC}$ vs. $-\log_{10}(\text{FDR})$ space, are labelled. **(B)** Functional enrichment analysis of genes with higher expression in the 4T1 cell line (left) and in Late-stage tumours (right). Representative significantly enriched Gene Ontology Biological Process (GO_BP), KEGG and Reactome terms are shown. Dot size represents the number of genes associated with each term, and the x-axis indicates enrichment significance ($-\log_{10}$ adjusted p-value). Translation- and ribosome-associated terms were enriched in cultured tumour cells, whereas extracellular matrix organisation and immune-related terms were enriched in tumours grown in vivo.

Discussion

Main findings

Most TNBC research focuses on established tumours, offering a static, late-stage view that can overlook the early events, temporal sequence and TME dynamics needed for early detection and targeted therapies [19]. By adopting a stage-resolved transcriptomic approach, this study identifies five principal findings: (1) extensive transcriptomic alterations were already present before a macroscopically detectable tumour formed; (2) tumour progression was mainly characterised by a limited number of coordinated biological programmes involving extracellular matrix remodelling, inflammation and progressive loss of epithelial differentiation; (3) transcriptional change varied markedly across stages, consistent with discrete transitions rather than uniform accumulation; (4) the transition between the Medium 1 and Medium 2 stages was the most pronounced of these, across

687 histopathological, transcriptomic and temporal analyses; and (5) many
688 transcriptional programmes prominent in vivo were not represented in the parental
689 4T1 cell line, highlighting the contribution of the TME. Together, these findings
690 support the view that TNBC progression should be understood as the evolution of a
691 tissue ecosystem, in which coordinated interactions between tumour cells and the
692 surrounding microenvironment are progressively established.

693 **Extensive transcriptomic alterations precede detectable tumour formation**

694 One of the most striking findings of this study is the identification of extensive
695 transcriptomic alterations before the development of a macroscopically detectable
696 tumour: 3,306 DEGs were already detected at the pooled Early stage, with 2,881 of
697 these genes already differentially expressed in G1 alone, when no macroscopic
698 tumour was detectable, and the mammary tissue remained histologically
699 indistinguishable from Control. Although early molecular alterations have been
700 described in experimental models and pre-neoplastic breast lesions [20],
701 transcriptomic studies of TNBC have largely focused on established tumours,
702 leaving the earliest stages of tumour onset comparatively unexplored [4,8,10,11].

703 Among these early changes, the transient activation of keratinocyte differentiation
704 and keratinisation programmes was particularly noteworthy, with genes such as
705 Krt23b, Krtap1-1 and Sprr2d among the most differentially expressed at this stage.
706 Because control animals underwent the same surgical manipulation and PBS
707 injection, these changes are unlikely to reflect a non-specific wound-healing
708 response; their reproducible detection and subsequent progressive decline instead
709 suggest an intrinsic component of the earliest tissue response to tumour cell

710 engraftment. Keratin-associated differentiation programmes are increasingly
711 recognised as extending beyond epidermal homeostasis into epithelial plasticity
712 and tissue remodelling [21], consistent with an adaptive epithelial response to
713 tumour cell presence rather than simple tissue injury.

714 In parallel, the early decrease of genes included in the oxidative phosphorylation
715 and other mitochondrial pathways is consistent with the metabolic adaptations that
716 accompany malignant transformation, broadly comparable to the shift towards
717 aerobic glycolysis described in breast cancer [22–25]. Although our data capture a
718 decrease in oxidative programme representation rather than a direct measurement
719 of glycolytic flux, its detection before macroscopic tumour formation suggests that
720 metabolic reorganisation is established remarkably early in TNBC progression.

721 Together, these findings indicate that the earliest stages of tumour development
722 involve coordinated biological responses (epithelial plasticity, metabolic
723 remodelling and tissue reorganisation) rather than isolated alterations. This is
724 conceptually consistent with field cancerization, whereby histologically normal
725 tissue adjacent to a tumour already shows molecular alterations before overt
726 malignancy [26–29]. Our findings extend this concept temporally, showing that
727 comparable tissue-wide reorganisation can be detected within days of tumour cell
728 engraftment, before any tumour mass is evident, raising the possibility that at least
729 part of what is captured as "field" alterations around human tumours reflects an
730 early, rapidly established response rather than a slowly accumulated one.

731 **Tumour progression is organised into coordinated biological programmes**
732 **rather than isolated molecular changes**

733 Beyond the early transcriptomic alterations, a major contribution of this study is
734 evidence that mammary tumour progression is organised into a limited number of
735 coordinated transcriptional programmes. While pairwise differential expression
736 analyses define molecular differences between individual stages, temporal
737 trajectory analysis resolved the subset of genes with well-defined temporal profiles
738 into only three dominant programmes, suggesting that the apparent complexity of
739 tumour progression arises from coordinated changes in a small number of higher-
740 order transcriptional programmes rather than the independent accumulation of
741 numerous individual alterations.

742 Among these, genes associated with extracellular matrix organisation, collagen
743 remodelling and inflammatory processes (Cluster 1) showed a progressive increase
744 in representation, consistent with well-recognised hallmarks of TNBC progression
745 linked to a tumour-supportive microenvironment and metastatic dissemination
746 [30,31]. The two remaining programmes both showed a progressive decrease in
747 representation, but with distinct trajectories: Cluster 2 (keratinocyte differentiation,
748 keratinisation) decreased sharply, concentrated between the Medium 1 and
749 Medium 2 stages, while Cluster 3 (skin development, fat cell differentiation, lipid
750 metabolism) decreased gradually throughout progression. This divergence
751 indicates that the reduced representation of programmes associated with normal
752 mammary tissue identity is not a single, uniform process, but follows different
753 temporal patterns across distinct tissue compartments, consistent with the
754 progressive replacement of normal tissue by the tumour ecosystem.

755 This interpretation is reinforced by concordance with the pairwise comparisons: the
756 extracellular matrix and inflammatory gene set identified by temporal clustering
757 showed a progressive increase in representation across consecutive stages,
758 peaking at the Medium 1 to Medium 2 transition, while the genes comprising
759 Clusters 2 and 3 corresponded closely to those showing higher expression at the
760 Early-stage relative to Control, before decreasing thereafter. Individual genes within
761 each programme showed considerable heterogeneity around this average
762 trajectory, expected in a tissue undergoing continuous remodelling, and supporting
763 the interpretation of these programmes as dominant tissue-level trends identified
764 through pathway-level enrichment rather than direct evidence of coordinated
765 regulatory mechanisms.

766 **The transition between Medium 1 and Medium 2 represents a major biological**
767 **inflexion point during tumour progression**

768 One of the most consistent findings of this study was the identification of the
769 transition between the Medium 1 and Medium 2 stages as a pivotal event during
770 tumour progression. This transition was supported by histopathological
771 assessment, global transcriptomic separation between stages, temporal trajectory
772 analysis and pairwise differential expression analyses. Notably, this convergence
773 did not rest on the sheer number of differentially expressed genes: the Medium 2
774 versus Medium 1 comparison yielded fewer DEGs (2,795) than either the Early
775 versus Control (3,306) or Medium 1 versus Early (3,077) comparisons. What
776 distinguishes this transition instead is the coincidence of multiple complementary
777 lines of evidence at the same point in tumour development: the histological shift

from mixed to tumour-dominated tissue, which corresponds to the largest macroscopic change in tissue composition observed in this study, as samples from the Medium 2 stage onward consisted almost exclusively of tumour tissue with little remaining host stroma; the sharpest inflexion in the temporal trajectories of the coordinated gene expression programmes; and the first significantly higher expression of extracellular matrix and inflammatory genes such as *Spp1* and *Ankrd1*. This agreement among different complementary approaches suggests that this transition captures a genuine biological event in tumour progression, although the concurrent change in tissue composition is likely to contribute to its transcriptomic magnitude. Other murine tumour progression models have similarly reported discrete transcriptional transitions instead of a gradual, uniform accumulation of molecular changes [32,33], suggesting that punctuated transcriptional change, not a steady one, may be a recurring feature of tumour progression across tissue types.

At the level of individual programmes, all three temporal clusters showed their largest overall change between the Medium 1 and Medium 2 stages, despite otherwise distinct trajectories: an abrupt, near-complete change in representation for Clusters 1 and 2, and a comparatively modest but detectable local acceleration within the gradual decline of Cluster 3. This shared, if unequally weighted, inflexion across all three programmes indicates that the Medium 1 to Medium 2 transition reflects a coordinated, tissue-wide change, not one confined to a single biological process.

Relatively few additional biological processes emerged after the Medium 2 stage, despite continued tumour growth. Instead, the pairwise comparisons suggest that the programmes already showing increased or decreased representation during the Medium 1 to Medium 2 transition continued to show the same trend throughout the Late stage, particularly extracellular matrix and inflammatory genes. This suggests that the major transcriptional transitions underlying TNBC progression may occur before advanced tumour burden is reached: subsequent tumour growth appears to be characterised mainly by a continuation of already established patterns.

Together, these findings identify the transition from Medium 1 to Medium 2 as a critical stage in the evolution of the tumour ecosystem, and defining the mechanisms underlying it may provide important insights into the establishment of the TME.

The tumour microenvironment is a major determinant of the transcriptomic landscape of advanced TNBC

The comparison between late-stage tumours and the parental 4T1 cell line provides additional insight into the biological origin of the transcriptional programmes identified throughout tumour progression. Late-stage tumours, in which tumour tissue represents the predominant component of the sample (**Figure 1C**), displayed a transcriptomic profile relatively close to that of the parental tumour cells in principal component space (**Figure 2**), consistent with the increasing proportion of tumour tissue within the analysed material at this stage. Nevertheless, a substantial number of genes remained differentially expressed between the two conditions (1,658 DEGs), indicating that even at this advanced stage, mammary tumours in vivo

823 show a transcriptional component not represented in tumour cells cultured in vitro.
824 Functional enrichment analyses showed that these differences were predominantly
825 associated with extracellular matrix organisation, collagen formation, cell adhesion
826 and inflammatory signalling, while the cultured 4T1 cells were mainly enriched for
827 translation, ribosome and biosynthesis-related processes. Among the genes
828 showing the highest expression in tumour tissue relative to cultured 4T1 cells were
829 canonical markers of tumour-associated macrophages, suggesting that a
830 substantial part of this transcriptional divergence reflects the presence of infiltrating
831 myeloid cells, in addition to tumour cell-intrinsic transcriptional differences. These
832 findings are consistent with a growing body of work reframing the TME as an active
833 participant in breast cancer progression, not a passive scaffold [34,35]. This
834 observation provides support for the interpretation of the temporal trajectory
835 analysis. The extracellular matrix programme that showed increasing
836 representation across tumour development and became dominant during the
837 Medium 1 to Medium 2 transition, was largely absent from the isolated tumour cells,
838 suggesting that it arises predominantly from tissue-level processes associated with
839 stromal activation and extracellular matrix deposition, not from tumour cell-
840 autonomous transcriptional change alone. Notably, osteopontin (*Spp1*), one of the
841 earliest and most persistently highly expressed genes within this programme, has
842 been shown to mediate an MZF1 to TGF- β 1-dependent transformation of
843 mesenchymal stem cells into cancer-associated fibroblasts in breast cancer [36],
844 offering a plausible mechanistic link to the stromal reorganisation this programme
845 likely reflects. These findings are consistent with the recognised role of cancer-

846 associated fibroblasts, infiltrating immune populations and matrix remodelling in
847 shaping the aggressive phenotype of TNBC [34,35].

848 The progressive increase in representation of extracellular matrix-related pathways
849 throughout tumour development is also consistent with the histopathological
850 evolution observed in this model. Temporal trajectory analysis indicated that this
851 programme was already trending upward from the earliest stages examined,
852 although this trend remained modest and was not detected as a significant signal in
853 the Early versus Control comparison, where keratinisation and respiration related
854 processes dominated instead. A clear, statistically significant signal for the
855 extracellular matrix and inflammatory programme did not emerge until the Medium
856 1 to Medium 2 transition, after which it became increasingly dominant as tumour
857 tissue progressively replaced the normal mammary gland. This suggests that
858 extracellular matrix remodelling begins early, gains momentum progressively, and
859 becomes clearly detectable only once a critical mass of tumour tissue is reached,
860 consistent with the established contribution of extracellular matrix remodelling to
861 tumour invasion, immune regulation and metastatic dissemination in TNBC [8], and
862 offering a temporal framework for when these processes become established
863 during disease progression.

864 Together, these findings indicate that the extensive extracellular matrix-related
865 changes observed in advanced tumours reflect the progressive establishment of a
866 complex tumour ecosystem, not solely the intrinsic transcriptional programme of
867 the tumour cells themselves and underline the value of studying tumour
868 progression in its native tissue context.

869 **Study limitations**

870 As with any bulk transcriptomic study of whole tissue, the programmes identified
871 here reflect the combined contribution of multiple cellular populations and cannot
872 be assigned to specific cell types. Because bulk RNA sequencing cannot distinguish
873 transcriptional changes within individual cell populations from changes resulting
874 from altered tissue cellular composition, the observed differences should therefore
875 be interpreted as alterations in the overall transcriptomic landscape of mammary
876 tissue, potentially reflecting changes in gene expression, cellular composition, or
877 both. One practical consequence is that, because tumour tissue represents an
878 increasingly large fraction of the sample at the later stages (**Figure 1**), the higher
879 representation of cell cycle and chromosome segregation genes in the Medium 1 to
880 Medium 2 and Late versus Control comparisons likely reflects both the growing
881 proportion of proliferating tumour cells within the tissue and a genuine increase in
882 proliferative signalling, without the present data allowing these two contributions to
883 be separated. Single-cell and spatial transcriptomic approaches, applied within a
884 similar stage-resolved design, would be a natural next step to resolve the cellular
885 origin of these programmes in finer detail.

886 The sample selection strategy for tumour-bearing groups (G2 to G7), in which the
887 five animals with the largest tumour volumes within each group were selected,
888 follows directly from the aim of maximising tumour representation within the
889 analysed tissue, a necessary approach given that RNA sequencing was only feasible
890 on a subset of the full cohort. As a result, the transcriptomic stages defined here
891 reflect this volume-enriched selection rather than a strictly random sample of

892 animals at each time point, which is worth bearing in mind when relating them to
893 elapsed time after inoculation or to tumour volume distributions in other
894 experimental settings.

895 Similarly, the temporal clusters identified in this study capture dominant, average
896 trajectories rather than expression patterns shared identically by every constituent
897 gene; some heterogeneity among individual genes, including transient and non-
898 monotonic profiles, is expected in any tissue undergoing continuous biological
899 remodelling and does not detract from the coherence of the programmes as a
900 whole.

901 The key differentially expressed genes highlighted throughout this study, including
902 *Spp1* and components of the extracellular matrix and oxidative phosphorylation
903 programmes, were identified from RNA sequencing data alone. Their consistency
904 across several complementary analytical approaches, namely differential
905 expression, temporal trajectory analysis and comparison with the parental cell line,
906 already provides converging support for their biological relevance; independent
907 validation at the protein or transcript level, for example by qPCR or
908 immunohistochemistry, would nonetheless be a valuable addition in future work.

909 Finally, these findings were obtained using a single orthotopic murine model of
910 triple-negative breast cancer. Extending this stage-resolved framework to additional
911 experimental models and human datasets will help clarify how broadly the
912 transcriptional programmes identified here apply beyond this specific model.

Conclusions

This study provides a stage-resolved transcriptomic framework for mammary tumour progression in an orthotopic model of triple-negative breast cancer. By integrating histopathological characterisation with temporal transcriptomic analyses, our findings suggest that extensive molecular alterations may arise before the development of a macroscopically detectable tumour and subsequently evolve through a limited number of coordinated biological programmes, with discrete transitions instead of a continuous, uniform accumulation of change. The identification of a pivotal transition between the Medium 1 and Medium 2 stages, together with the progressive increase in representation of extracellular matrix and inflammatory programmes largely absent in isolated tumour cells, indicates that the biological complexity of advanced TNBC arises from dynamic interactions between malignant cells and the surrounding TME. Collectively, these findings support the view that tumour progression should be understood as the evolution of a complex tissue ecosystem, not simply the accumulation of molecular alterations intrinsic to the tumour cells alone. Notably, most major transcriptional programmes were already established by the Medium 2 stage, with the Late stage characterised predominantly by their continuation instead of the emergence of new processes, suggesting that the biological events most relevant for understanding TNBC establishment may occur earlier in disease progression than is typically captured by studies focused on advanced tumours alone. This stage-resolved framework provides a foundation for future studies to define the earliest biological events

935 underlying TNBC progression and the mechanisms governing the transitions
936 identified here, which may represent opportunities for therapeutic intervention.

937 **List of abbreviations**

938 AGE-RAGE: Advanced glycation end-products-receptor for advanced glycation end-
939 products

940 CEBA: Committee for Animal Experimentation and Welfare (Universitat de València)

941 DEG(s): Differentially expressed gene(s)

942 ECM: Extracellular matrix

943 ER: Oestrogen receptor

944 FC: Fold change

945 FDR: False discovery rate

946 GO_BP: Gene Ontology Biological Process

947 H&E: Haematoxylin and eosin

948 HER2: Human epidermal growth factor receptor 2

949 KEGG: Kyoto Encyclopedia of Genes and Genomes

950 PBS: Phosphate-buffered saline

951 PCA: Principal component analysis

952 PERMANOVA: Permutational multivariate analysis of variance

953 PR: Progesterone receptor

954 qPCR: Quantitative polymerase chain reaction

955 SD: Standard deviation

956 TGF- β : Transforming growth factor beta

957 TME: Tumour microenvironment

958 TNBC: Triple-negative breast cancer

959 TNF: tumour necrosis factor

960 UMI: Unique molecular identifier

961 VST: Variance-stabilizing transformation

962 **Declarations**

963 **Ethics approval and consent to participate**

964 All animal procedures were approved by the Committee for Animal Experimentation
965 and Welfare of the Universitat de València (CEBA; protocol 2022 VSC PEA 0100) and
966 were conducted in accordance with Directive 2010/63/EU, Spanish legislation
967 governing the protection of animals used for scientific purposes, and institutional
968 guidelines. No human participants, data or tissue were used in this study.

969 **Consent for publication**

970 Not applicable.

971 **Availability of data and materials**

972 The raw sequencing data generated in this study have been deposited in the
973 ArrayExpress database at BioStudies under the dataset title "Temporal

974 transcriptomic profiling of the response of mammary tissue following the orthotopic
975 inoculation of 4T1 cells in mice", accession number E-MTAB-17468.

976 **Competing interests**

977 The authors declare that they have no competing interests.

978 **Funding**

979 This work was supported by the Consellería de Innovación, Universidades, Ciencia
980 y Sociedad Digital, Generalitat Valenciana, under the "Convocatoria para la
981 realización de Proyectos I+D+i desarrollados por grupos de investigación
982 Emergentes" programme (grant number GV/2021/184); by Universidad CEU
983 Cardenal Herrera (FUSP-PPC-22-061, INDI 24/11, GIR 24/10) and by the ECMOR –
984 Chair of Modern Oncology Research, CEU UCH.

985 R. Bosch-Romeu received funding from the Spanish Ministry of Science, Innovation
986 and Universities through the *Ayudas para la Formación de Profesorado Universitario*
987 (*FPU 2022*) programme (grant FPU22/02810).

988 Cristian Herrera-Olivares is the beneficiary of a predoctoral scholarship by the
989 Universidad Cardenal Herrera-CEU under the “*Ayudas a la Formación de Jóvenes*
990 *Investigadores CEU*” programme (2024/25)

991 **Authors' contributions**

- 992 • Conceptualization and study design: I.T., J.C.
- 993 • Animal experiments and tissue collection: I.T., A.A., D.LF., M.P.
- 994 • Histopathological evaluation: J.O., C.H., I.T.

995 • RNA sequencing and bioinformatic analysis: L.A., C.M., R.B., AB.G.

996 • Statistical analysis: L.A., C.M., R.B., I.T., C.H.,

997 • Drafting of the manuscript: I.T., C.H., C.M., A.R., M.P., A.L.

998 • Critical revision of the manuscript: all authors

999 • Supervision and funding acquisition: I.T., J.C.

1000 I.T. and C.H. contributed equally to this work as co-first authors. I.T. and J.C. shared
1001 equal roles in study leadership, funding acquisition, and manuscript supervision as
1002 joint corresponding authors. All authors read and approved the final manuscript.

1003 **Acknowledgements**

1004 We thank Ana Díaz, responsible for the animal facility of the Universitat de València,
1005 for her assistance with animal procedures. We also thank the research technicians
1006 at CEU Cardenal Herrera University, Andrea Yuste, Rocío Rodrigo and Sacramento
1007 Olivares, for their support with logistics and material procurement. We further thank
1008 the Pathology team at Universidad Cardenal Herrera-CEU, CEU Universities as a
1009 whole, and technicians Elisabeth Blat, Virginia Alemany and Isela Giró, for their
1010 support with tissue processing and for providing access to facilities and equipment.

1011 **Declaration of AI use**

1012 AI-based tools were used solely to assist with language editing and improve the
1013 clarity and readability of the manuscript. The authors reviewed and verified all AI-
1014 assisted content and retain full responsibility for the scientific content,
1015 interpretation, and conclusions of the work.

1016 **References**

- 1017 1. Pajewska M, Partyka O, Czerw A, Deptała A, Sygit K, Gaska I, et al. Advanced and
1018 Metastatic Triple Negative Breast Cancer—Potential New Treatment. *Cancers (Basel)*.
1019 2025;17:1183. <https://doi.org/10.3390/cancers17071183>
- 1020 2. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer:
1021 challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol*.
1022 2016;13:674–90. <https://doi.org/10.1038/nrclinonc.2016.66>
- 1023 3. Irvin WJ, Carey LA. What is triple-negative breast cancer? *Eur J Cancer*.
1024 2008;44:2799–805. <https://doi.org/10.1016/j.ejca.2008.09.034>
- 1025 4. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al.
1026 Identification of human triple-negative breast cancer subtypes and preclinical models
1027 for selection of targeted therapies. *Journal of Clinical Investigation*. 2011;121:2750–67.
1028 <https://doi.org/10.1172/JCI45014>
- 1029 5. Chen Z, Liu Y, Lyu M, Chan CH, Sun M, Yang X, et al. Classifications of triple-negative
1030 breast cancer: insights and current therapeutic approaches. *Cell Biosci*. 2025;15:13.
1031 <https://doi.org/10.1186/s13578-025-01359-0>
- 1032 6. Hernandez K, Nguyen CH, Rijal G. Asporin increases the extracellular matrix cross-
1033 links and inhibits the cancer cell migration. *Tumor Biology*. 2025;47.
1034 <https://doi.org/10.1177/10104283241313441>
- 1035 7. Yu T, Zhang G, Sun W, Zhao W, Liu H, Guo Y, et al. Rotating magnetic field
1036 downregulating type XI collagen to suppress triple-negative breast cancer metastasis by
1037 inactivating the ITGB1/FAK/YAP signaling pathway. *Int J Biol Macromol*.
1038 2026;338:149677. <https://doi.org/10.1016/j.ijbiomac.2025.149677>
- 1039 8. Sabit H, Adel A, Abdelfattah MM, Ramadan RM, Nazih M, Abdel-Ghany S, et al. The
1040 role of tumor microenvironment and immune cell crosstalk in triple-negative breast
1041 cancer (TNBC): Emerging therapeutic opportunities. *Cancer Lett*. 2025;628:217865.
1042 <https://doi.org/10.1016/j.canlet.2025.217865>
- 1043 9. Gong Y, Ji P, Yang Y-S, Xie S, Yu T-J, Xiao Y, et al. Metabolic-Pathway-Based Subtyping
1044 of Triple-Negative Breast Cancer Reveals Potential Therapeutic Targets. *Cell Metab*.
1045 2021;33:51-64.e9. <https://doi.org/10.1016/j.cmet.2020.10.012>
- 1046 10. Wang Z, Jiang Q, Dong C. Metabolic reprogramming in triple-negative breast cancer.
1047 *Cancer Biol Med*. 2020;17:44–59. [https://doi.org/10.20892/j.issn.2095-](https://doi.org/10.20892/j.issn.2095-3941.2019.0210)
1048 3941.2019.0210

1049 11. Messal HA, Mourao L, Dörner L, Scheele CLGJ, van Rheenen J. Watching cancer
1050 begin: emerging tools to visualize the first steps of tumorigenesis. *Trends Cancer*.
1051 2026;12:652–66. <https://doi.org/10.1016/j.trecan.2026.02.007>

1052 12. Liu Y, Zou Y, Ye Y, Chen Y. Advances in the Understanding of the Pathogenesis of
1053 Triple-Negative Breast Cancer. *Cancer Med*. 2024;13.
1054 <https://doi.org/10.1002/cam4.70410>

1055 13. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov*. 2022;12:31–46.
1056 <https://doi.org/10.1158/2159-8290.CD-21-1059>

1057 14. Jovanović B, Temko D, Stevens LE, Seehawer M, Fassl A, Murphy K, et al.
1058 Heterogeneity and transcriptional drivers of triple-negative breast cancer. *Cell Rep*.
1059 2023;42:113564. <https://doi.org/10.1016/j.celrep.2023.113564>

1060 15. Schrörs B, Boegel S, Albrecht C, Bukur T, Bukur V, Holtsträter C, et al. Multi-Omics
1061 Characterization of the 4T1 Murine Mammary Gland Tumor Model. *Front Oncol*.
1062 2020;10. <https://doi.org/10.3389/fonc.2020.01195>

1063 16. Tao K, Fang M, Alroy J, Sahagian GG. Imagable 4T1 model for the study of late stage
1064 breast cancer. *BMC Cancer*. 2008;8:228. <https://doi.org/10.1186/1471-2407-8-228>

1065 17. Paschall A V., Liu K. An Orthotopic Mouse Model of Spontaneous Breast Cancer
1066 Metastasis. *Journal of Visualized Experiments*. 2016; <https://doi.org/10.3791/54040>

1067 18. Kato S, Fushimi K, Yabuki Y, Maru Y, Hasegawa S, Matsuura T, et al. Precision
1068 modeling of gall bladder cancer patients in mice based on orthotopic implantation of
1069 organoid-derived tumor buds. *Oncogenesis*. 2021;10:33.
1070 <https://doi.org/10.1038/s41389-021-00322-1>

1071 19. Liu Y, Zou Y, Ye Y, Chen Y. Advances in the Understanding of the Pathogenesis of
1072 Triple-Negative Breast Cancer. *Cancer Med*. 2024;13.
1073 <https://doi.org/10.1002/cam4.70410>

1074 20. Brunner AL, Li J, Guo X, Sweeney RT, Varma S, Zhu SX, et al. A shared transcriptional
1075 program in early breast neoplasias despite genetic and clinical distinctions. *Genome*
1076 *Biol*. 2014;15:R71. <https://doi.org/10.1186/gb-2014-15-5-r71>

1077 21. Li Y, Sun Y, Yu K, Li Z, Miao H, Xiao W. Keratin: A potential driver of tumor
1078 metastasis. *Int J Biol Macromol*. 2025;307:141752.
1079 <https://doi.org/10.1016/j.ijbiomac.2025.141752>

1080 22. Garg P, Singhal G, Horne D, Salgia R, Singhal SS. Metabolic reprogramming in breast
1081 cancer: Pathways driving progression, drug resistance, and emerging therapeutics.
1082 *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 2025;1880:189396.
1083 <https://doi.org/10.1016/j.bbcan.2025.189396>

1084 23. Gandhi N, Das G. Metabolic Reprogramming in Breast Cancer and Its Therapeutic
1085 Implications. *Cells*. 2019;8:89. <https://doi.org/10.3390/cells8020089>

1086 24. Liu S, Zhang X, Wang W, Li X, Sun X, Zhao Y, et al. Metabolic reprogramming and
1087 therapeutic resistance in primary and metastatic breast cancer. *Mol Cancer*.
1088 2024;23:261. <https://doi.org/10.1186/s12943-024-02165-x>

1089 25. Penkert J, Ripperger T, Schieck M, Schlegelberger B, Steinemann D, Illig T. On
1090 metabolic reprogramming and tumor biology: A comprehensive survey of metabolism
1091 in breast cancer. *Oncotarget*. 2016;7:67626–49.
1092 <https://doi.org/10.18632/oncotarget.11759>

1093 26. Heaphy CM, Bisoffi M, Fordyce CA, Haaland CM, Hines WC, Joste NE, et al.
1094 Telomere DNA content and allelic imbalance demonstrate field cancerization in
1095 histologically normal tissue adjacent to breast tumors. *Int J Cancer*. 2006;119:108–16.
1096 <https://doi.org/10.1002/ijc.21815>

1097 27. Trujillo KA, Heaphy CM, Mai M, Vargas KM, Jones AC, Vo P, et al. Markers of fibrosis
1098 and epithelial to mesenchymal transition demonstrate field cancerization in
1099 histologically normal tissue adjacent to breast tumors. *Int J Cancer*. 2011;129:1310–21.
1100 <https://doi.org/10.1002/ijc.25788>

1101 28. Kang T, Yau C, Wong CK, Sanborn JZ, Newton Y, Vaske C, et al. A risk-associated
1102 Active transcriptome phenotype expressed by histologically normal human breast
1103 tissue and linked to a pro-tumorigenic adipocyte population. *Breast Cancer Research*.
1104 2020;22:81. <https://doi.org/10.1186/s13058-020-01322-6>

1105 29. Abdalla M, Tran-Thanh D, Moreno J, Iakovlev V, Nair R, Kanwar N, et al. Mapping
1106 genomic and transcriptomic alterations spatially in epithelial cells adjacent to human
1107 breast carcinoma. *Nat Commun*. 2017;8:1245. <https://doi.org/10.1038/s41467-017-01357-y>

1109 30. Shukla M, Narayan P, Narayan M. Role of matrix metalloproteinases in triple
1110 negative breast cancer. *Therapeutics of Natural and Synthetic Compounds in Protease-*
1111 *Induced Cancer*. Elsevier; 2025. p. 3–18. [https://doi.org/10.1016/B978-0-443-26635-](https://doi.org/10.1016/B978-0-443-26635-5.00011-0)
1112 [5.00011-0](https://doi.org/10.1016/B978-0-443-26635-5.00011-0)

1113 31. Mohammad-Rezaei M, Shafaghat Z, Noroozbeygi M, Velu P, Safaee M, Heiranizadeh
1114 N, et al. Collagen remodeling in breast cancer progression: from molecular mechanisms
1115 to diagnostic and therapeutic opportunities. *Clinical and Translational Oncology*. 2026;
1116 <https://doi.org/10.1007/s12094-026-04450-1>

1117 32. Cai Y, Nogales-Cadenas R, Zhang Q, Lin J-R, Zhang W, O'Brien K, et al. Transcriptomic
1118 dynamics of breast cancer progression in the MMTV-PyMT mouse model. *BMC*
1119 *Genomics*. 2017;18:185. <https://doi.org/10.1186/s12864-017-3563-3>

33. Wang G, Luo X, Liang Y, Kaneko K, Li H, Fu X-D, et al. A tumorigenic index for quantitative analysis of liver cancer initiation and progression. *Proceedings of the National Academy of Sciences*. 2019;116:26873–80.
<https://doi.org/10.1073/pnas.1911193116>
34. Yuan Z, Li Y, Zhang S, Wang X, Dou H, Yu X, et al. Extracellular matrix remodeling in tumor progression and immune escape: from mechanisms to treatments. *Mol Cancer*. 2023;22:48. <https://doi.org/10.1186/s12943-023-01744-8>
35. Sun S, Zhao P, Wang C, Du J, Zhang T, He X, et al. Prominent protumoral cellular compartments of the tumor microenvironment in triple-negative breast cancer. *Front Cell Dev Biol*. 2026;13. <https://doi.org/10.3389/fcell.2025.1668583>
36. Weber CE, Kothari AN, Wai PY, Li NY, Driver J, Zapf MAC, et al. Osteopontin mediates an MZF1–TGF- β 1-dependent transformation of mesenchymal stem cells into cancer-associated fibroblasts in breast cancer. *Oncogene*. 2015;34:4821–33.
<https://doi.org/10.1038/onc.2014.410>

1.

Supplementary Material

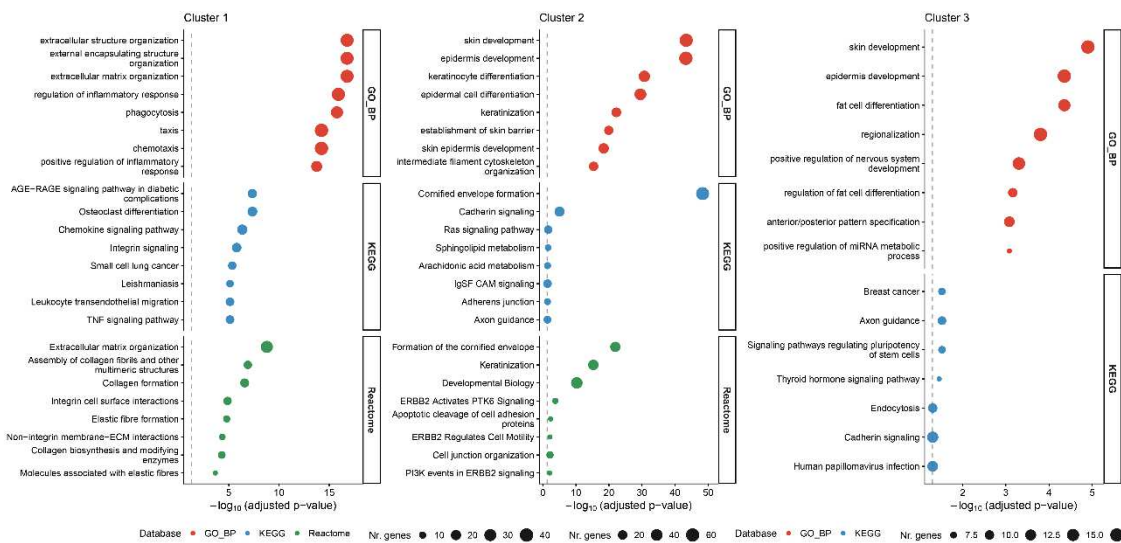
Supplementary Table S1. Individual body weight and tumour volume measurements for the subset of samples included in the transcriptomic analyses. Descriptive statistics of tumour volume for the complete experimental cohort are also included.

Supplementary Table S2. Gene membership of the three temporal expression clusters identified by temporal trend analysis (**Figure 4A**). No Reactome pathway reached significance for Cluster 3, likely reflecting its narrower *coverage of Mus musculus pathways compared with* Gene Ontology and KEGG.

Supplementary Table S3. Full differential expression results ($\log_2$ fold-change, adjusted p-value and functional annotation) for all pairwise comparisons between disease stages and the parental 4T1 cell line. Genes shown in bold within the Early

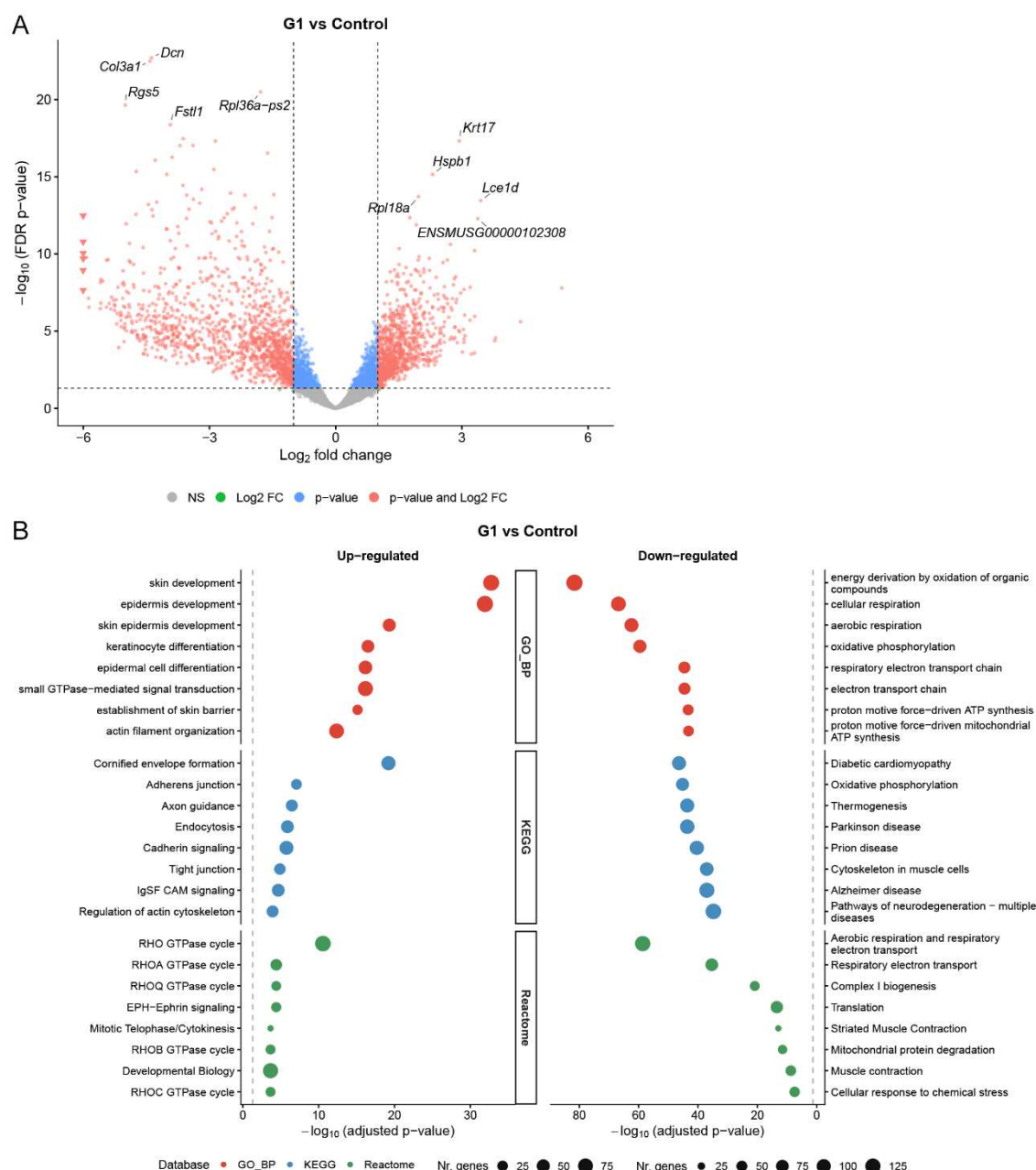
vs Control comparison were also identified as significant DEGs in the G1 vs Control comparison alone.

Supplementary Figure S1. Functional enrichment analysis of the three temporal gene expression clusters. Gene Ontology Biological Process (GO_BP), KEGG and Reactome terms significantly enriched among the genes composing each of the three temporal clusters identified in **Figure 4A**: Cluster 1, Cluster 2 and Cluster 3. No Reactome pathway reached significance for Cluster 3 (see **Supplementary Table S2**). Dot size represents the number of genes associated with each term, and the x-axis indicates enrichment significance ($-\log_{10}$ adjusted p-value).



Supplementary Figure S2. Differential expression and functional enrichment between G1 and Control mammary tissue. (A) Volcano plot showing differentially expressed genes (DEGs) between G1 (day 3 post-inoculation) and Control mammary tissue ($FDR < 0.05$, $|\log_2 FC| \geq 1$), examined separately from the pooled Early stage (G1-G2). Dashed lines indicate significance and fold-change thresholds.

1164 The five genes with the greatest increase and the five with the greatest decrease in
1165 expression, ranked by Euclidean distance from the origin in $\log_2\text{FC}$ vs. $-\log_{10}(\text{FDR})$
1166 space, are labelled. **(B)** Functional enrichment analysis of genes with increased and
1167 decreased expression in the G1-versus-Control comparison. Representative
1168 significantly enriched Gene Ontology Biological Process (GO_BP), KEGG and
1169 Reactome terms are shown. Dot size represents the number of genes associated
1170 with each term, and the x-axis indicates enrichment significance ($-\log_{10}$ adjusted p-
1171 value).



1172

1173 **Supplementary Figure S3. Differential expression and functional enrichment**

1174 **between consecutive disease stages. (A, C, E) Volcano plots showing**

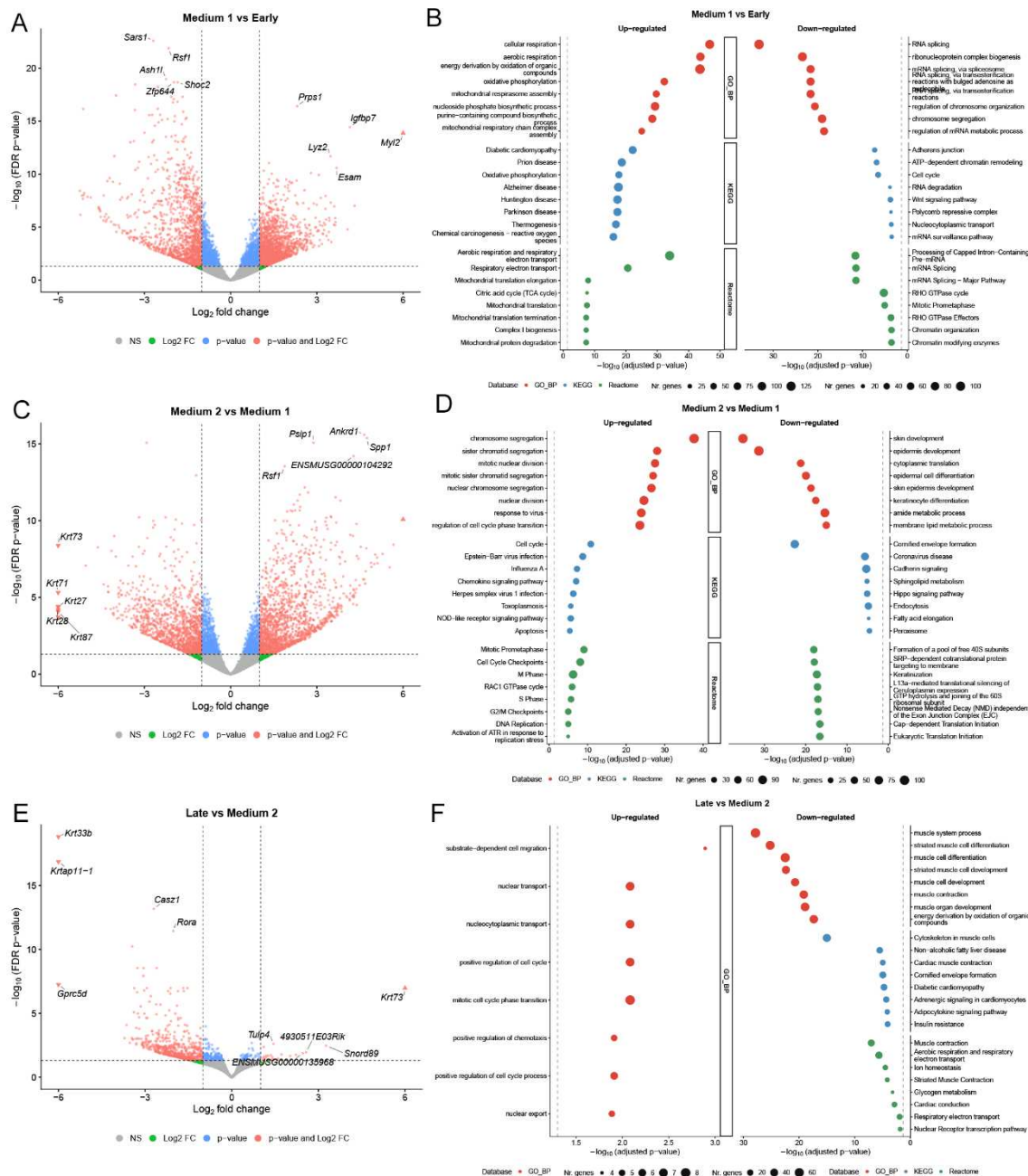
1175 differentially expressed genes (DEGs; $FDR < 0.05$, $|\log_2 FC| \geq 1$) between consecutive

1176 disease stages: **(A)** Medium 1 vs Early, **(C)** Medium 2 vs Medium 1, **(E)** Late vs

1177 Medium 2. Dashed lines indicate significance and fold-change thresholds. The five

1178 genes with the greatest increase and the five with the greatest decrease in

1179 expression, ranked by Euclidean distance from the origin in $\log_2\text{FC}$ vs. $-\log_{10}(\text{FDR})$
1180 space, are labelled. **(B, D, F)** Functional enrichment analysis of genes with
1181 increased and decreased expression for the corresponding comparison in **(A), (C)**
1182 and **(E)**, respectively. Representative significantly enriched Gene Ontology
1183 Biological Process (GO_BP), KEGG and Reactome terms are shown. Dot size
1184 represents the number of genes associated with each term, and the x-axis indicates
1185 enrichment significance ($-\log_{10}$ adjusted p-value).



1186

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfilescombinedcarteradedocs.pdf](#)