

<https://doi.org/10.1038/s42003-025-08123-7>

# Parallel and convergent dynamics in the evolution of primary breast and lung adenocarcinomas



Robert A. Gatenby<sup>1,2</sup> , Jamie K. Teer<sup>3</sup> , Kenneth Y. Tsai<sup>1,4</sup>  & Joel S. Brown<sup>1,2</sup>

Cancer development requires an evolutionary transformation from mammalian cells fully regulated by and integrated into multicellular tissue to cancer cells that, as single cell protists, are individually subject to Darwinian selection. Through genetic and epigenetic mechanisms of inheritance, the evolving cancer phenotype must acquire independence from host controls, downregulate differentiated functions that benefit the host but not individual cells, and generate phenotypic traits that increase fitness in the context of the selection forces within the local microenvironment. Here, we investigate this evolutionary transition in breast (BRCA) and lung (LUAD, without *EGFR*, *KRAS* or *BRAF* driver mutations) adenocarcinomas using bulk mutation and expression data from the TCGA database. We define evolution selection for genes and molecular pathways based on 1) changes in gene expression compared to normal tissue, and 2) significantly larger or smaller observed mutation rates compared to those expected based on the gene size. We find BRCA and LUAD disable different genes and gene pathways associated with tissue-specific signaling and differentiated functions but promote common molecular pathways associated with cell cycle, cell-cell interactions, cytoskeleton, voltage gated ion channels, and microenvironmental niche construction. Thus, tissue-specific parallel evolution in early cancer development is followed by convergence to a common cancer phenotype.

The multistep transition from normal mammalian cells to a malignant phenotype during carcinogenesis represents a fundamental evolutionary transformation. Somatic cells are not directly subject to natural selection as their death, survival, phenotype, movement, and proliferation are normally controlled by the host tissue<sup>1,2</sup>. That is, mammalian cells within differentiated tissue contribute to and are defined by the fitness of the evolving multicellular organism. In contrast, survival and proliferation of individual cancer cells are governed by Darwinian dynamics determined by interactions of their heritable phenotype with local environmental selection forces. This transition from one evolutionary state to another is our working definition of carcinogenesis<sup>1</sup>.

Carcinogenesis is often framed as a series of specific genetic events. Here we present an alternative model in which carcinogenesis is viewed as a sequence of phenotypic adaptations that give rise to a cancer cell that is evolutionarily distinct from its mammalian cell progenitors. In the context of the “evolutionary transformation” model, several key adaptations are necessary and, therefore, expected in all cancers: 1) Independence from controlling host signals; 2) Acquisition of alternative signaling pathways to optimize fitness (proliferation); 3) Maintenance of molecular machinery for

proliferation; 4) Down regulation of whole-organism differentiated functions which benefit the host but do not contribute to fitness (proliferation) of an individual cell; 5) Promotion of phenotypic plasticity to rapidly detect and adapt to spatially and temporally variable environmental selection forces; 6) Niche construction to generate an extracellular environment that optimizes cancer cell survival and proliferation.

Thus, we view accumulating molecular changes in each cancer cell as the legacy of this evolutionary “march to independence”. Although each cancer must pass through predictable evolutionary changes, their specific evolutionary adaptations and corresponding molecular dynamics are subject to differences in the cell of origin and the selection pressures imposed by local dynamics. Here we use bulk data from patient cohorts in the TCGA database to investigate similarities and differences in the molecular dynamics driving key evolutionary steps in adenocarcinomas of the lung (without *EGFR*, *KRAS*, or *BRAF* driver mutations) (LUAD) and breast (BRCA). LUAD without common drivers typically evolves in a smoking-induced mutagenic<sup>3</sup> and inflammatory environment<sup>4</sup> while BRCA emerges from a plastic tissue that expands and regresses in response to systemic hormone levels<sup>5</sup>. We choose these as oncogenic extremes with the intent of

<sup>1</sup>Cancer Biology and Evolution Program, Tampa, FL, USA. <sup>2</sup>Integrated Mathematical Oncology Department, Tampa, FL, USA. <sup>3</sup>Biostatistics and Bioinformatics Department, Tampa, FL, USA. <sup>4</sup>Pathology Department Moffitt Cancer Center, Tampa, FL, USA. ✉ e-mail: [Robert.Gatenby@moffitt.org](mailto:Robert.Gatenby@moffitt.org)

exploring divergent, parallel, and convergent evolutionary pathways during carcinogenesis.

The large TCGA data sets for LUAD and BRCA can be analyzed through multiple bioinformatic methods. Here, we use evidence of evolutionary selection to reveal key genetic and molecular pathways. While carcinogenesis is often viewed as a sequence of genetic mutations<sup>6</sup>, we note Darwinian dynamics only require a “mechanism of inheritance” for selected phenotypes, which can, therefore, be genetic or epigenetic. Thus, we also view expression data as a reflection of Darwinian cost/benefit optimization. That is, the change in expression of each gene in a cancer cell compared to its normal progenitor is determined by the fitness benefit of the encoded protein compared to the cost of synthesizing it.

Our present work builds upon and extends techniques we have developed for studying under- and over-mutated genes, and gene expression arising through non-mutational gene-regulation including changes in expression, the spliceosome, or copy number<sup>7</sup>. Previously we have compared gene mutations across three subtypes of lung cancer (KRAS mutated, EGFR, and non-driver mutation), focusing primarily on those genes in common that are rarely mutated<sup>8</sup>, or associated specifically with immune evasion and activation<sup>9</sup>. Lastly, we have examined in detail the mutational and expression pathways to carcinogenesis in BRCA and LUAD cancers to identify key selection pressures opportunities and avoids hazards for each cancer type<sup>10</sup>. We take advantage of evolutionary principles to examine parallel and convergent evolution, and co-adaptations in LUAD with no known driver mutations (high mutation rates) with breast cancer (BCA) typified by fewer mutations. This allows us to integrate using the TCGA data sets both mutational changes (under- or over-mutated) and comparisons of gene expression between normal and lung and breast cells with that seen in cancerous lung and breast.

Our approach complements other approaches to using mutations and gene expression to compare cancers and infer evolutionary processes such as positive and negative selection<sup>11</sup>. Here we go beyond standard differential gene expression approaches such as edgeR<sup>12</sup> and DESeq2<sup>13</sup> by considering gene expression and mutations together as evolutionary driving forces in cancer progression as well as overall rates of expression change across groups of genes. Similar to our interests, the Curated Cancer Cell Atlas provides single cell RNAseq from diverse cancers and studies that are grouped into meta-programs that identify possible co-adapted syndromes within both normal and cancer cells<sup>14</sup>. For instance, we use under-mutated and over-mutated genes as evidence for stabilizing or directional selection, in a manner similar to MutSigCV and complementary to approaches using the ratio of synonymous to non-synonymous mutations (e.g., dNdScv).

Eco-evolutionary mathematical models<sup>15</sup> of carcinogenesis and clinical observations<sup>16</sup> demonstrate normal cells can evolve to a malignant phenotype through a wide range of genetic/epigenetic changes. Molecular data that traces the evolutionary arc of a single cancer is, therefore, difficult to generalize. However, we view each cancer within a TCGA cohort as a “replicate experiment” in carcinogenesis. Thus, *patterns* of genetic and transcriptional changes can provide insights into parallel and convergent adaptive strategies resulting from different or common evolutionary selection forces.

To illustrate, we have previously noted the analogy of cancer evolution to that of cavefish<sup>17</sup>. Cavefish converge on a common phenotype (without eyes or skin pigment) through diverse genetic pathways<sup>18,19</sup> -- the cave morph has evolved from > 85 different fish species. Thus, molecular sampling from a single cave population may be difficult to generalize but adding data from multiple caves reveals a general pattern of genetic and epigenetic selection on genes associated with, for example, eyes, skin pigment, and adaptation to hypoxia<sup>20</sup>. Furthermore, specific attributes of the cave system may result in repeatable and predictable divergent trait evolution. For instance, cave fish in a sulfidic stream have larger heads, longer gill filaments, and fewer and larger offsprings than those inhabiting a stream without hydrogen sulfide<sup>21</sup>.

Hutchinson described Darwinian dynamics as “The Ecological Theater and Evolutionary Play”<sup>22</sup>. Within that analogy, we find the evolutionary plays for breast and lung cancers have different first acts but both their later acts and associated ecological stages tend to converge (Fig. 1).

## Methods

### Gene list acquisition

We divided the TCGA lung adenocarcinoma cohort based on known driver mutations in KRAS (G12, G13, Q61, A146), BRAF (V600, N581, G464, G466, G469, G596, D594), and EGFR (L858, S768, L861, G719, T790, indels in exons 18–21). The 313 non-EGFR/KRAS/BRAF LUAD patients were selected as the study group because of a high prevalence (90%) of smoking (and associated mutagenic, inflammatory environment) that results in the highest mutation rate among the different lung cancer cohorts<sup>8</sup>. Somatic mutations along with gene expression in normal tissue and cancers were downloaded from the TCGA PanCancer Atlas<sup>23</sup>. We similarly obtained the entire breast cancer cohort (727 patients) (BRCA), including mutation and expression data from the TCGA PanCancer Atlas. RNAseq gene-level expression was downloaded from <https://api.gdc.cancer.gov/data/3586c0da-64d0-4b74-a449-5ff4d9136611> (EBPlusPlusAdjustPANCAN\_IlluminaHiSeq\_RNASeqV2.geneExp.tsv). Mean expression was calculated across all samples in each organ and disease status group, and then log2 transformed. In contrast to LUAD, BRCA has minimal exposure to environmental toxins and, therefore, evolves into a cancer with low mutational burden. However, BRCA arises in a normal tissue background that undergoes (or has undergone) cycles of expansion and contraction in response to systemic hormone changes.

Tumor and normal sequence alignment files used to calculate individual base coverage were downloaded from Genome Data Commons. A base was considered sufficiently covered if coverage depth was  $\geq 14$  in tumor samples and  $\geq 8$  in normal samples as has been previously described: <https://www.synapse.org/#!Synapse:syn1695394>. The fraction of each gene’s protein coding bases (using the longest RefSeq transcript) covered by sufficient sequence data was calculated for each sample using the Negative Storage Model<sup>24</sup>. Gene-level depth of coverage was then determined by calculating the number of bases sufficiently covered by sequencing for each of the RefSeq coding genes (with 25 base-pair flanking regions). This procedure measures the fraction of each gene (longest transcript) covered by sequencing data. To address sequencing artifacts that falsely decrease mutation rates, we excluded genes with a low average depth of coverage frequency ( $< 50\%$ ) and those with errors in the RefSeq gene model.

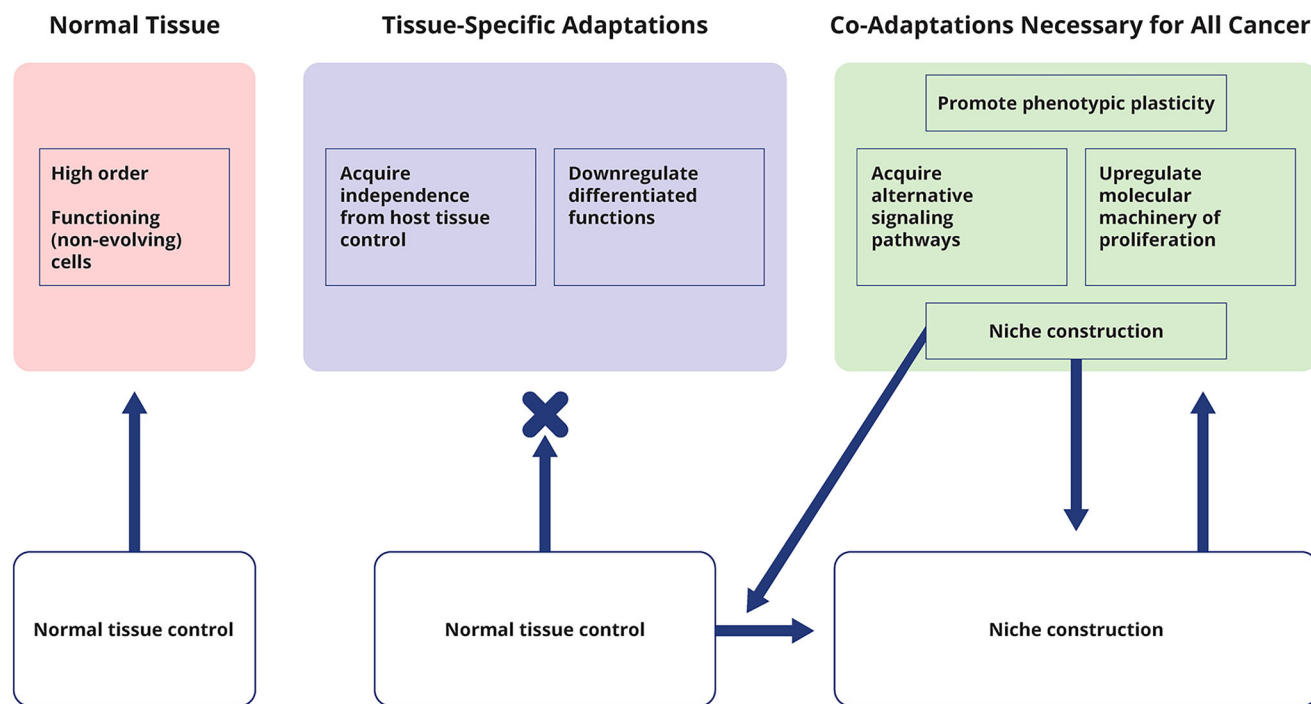
TCGA data provides whole exome sequencing with paired tumor/normal analysis to exclude germline mutations and focus on somatic mutations, which were downloaded from <https://api.gdc.cancer.gov/data/1c8cf5f-e52d-41ba-94da-f15ea1337efc> (mc3.v0.2.8.PUBLIC.maf.gz). The identified mutations include non-synonymous, synonymous, intronic, UnTranslated region, and intergenic. For our analysis, we limited the mutations to protein altering mutations: non-synonymous and truncating (stop-gain, frameshifting indel, splicing).

To minimize potential artifacts related to expression, and to focus on genes that are functional, we examined average differential gene expression between tumor samples and normal adjacent tissue. Genes were included in the analysis of mutations only if their average expression (log base 2) was  $\geq 1.5$  in either cancer or normal tissue. In each gene, the average expression in normal tissue and tumor samples across the cohort was analyzed separately to supplement the genetic data.

When gene expression increased  $\geq 4$ -fold in cancer cells, we considered this to indicate greater utility to the cancer cells than to the functioning of normal cells. Conversely a  $\geq 75\%$  decrease in expression indicated a gene function that benefited normal cells but not cancer cells.

### Observed mutational frequency as a function of its fitness effects

Similar to Cannataro et al.<sup>25</sup>, we characterize the fitness benefit of a gene mutation based on “evolutionary triage”<sup>15</sup>, which links the fitness effect of each genetic change to cellular proliferation. Thus, genetic mutations or



**Fig. 1 | The evolutionary play and ecological stages in cancer development.** The cell population in normal tissue has limited opportunities to evolve as the fates of individual cells are determined by tissue control. Initially, cancer evolution occurs on the ecological “stage” of normal tissue. Transition to somatic evolution requires severing tissue controls so that the cell’s fate is determined by its own fitness in the

context of its environment. Cancer cells initially increase their fitness by ceasing differentiated tissue functions that consume resources while not increasing the probability of survival and proliferation. Subsequently, similar co-adapted traits show broad convergence among different cancer types including transition to a highly modified and engineered ecological stage.

non-mutational changes in expression or splicing that increase fitness (and, therefore proliferation) will be observed more frequently than expected by chance alone. Conversely, mutations that decrease fitness are eliminated by more fit cancer cell lineages. Mutations in these genes, therefore, are observed less frequently than expected by chance alone. We note this could be interpreted as an intrinsically lower mutation rate in this gene.

We identify mutations as “over-mutated” based on their observed frequency of non-synonymous mutations in each gene against that gene’s size (number of base pairs) in each cohort. Assuming the probability of mutation was approximately equal for every base pair (see below) in an expressed gene, the background mutation rate was determined by regressing the mutational frequency of each gene against gene size. The distance of each gene to the regression line was then determined, and this standardized residual was compared across all genes. Negative and positive residual values indicate under-mutated and over-mutated genes, respectively. We scored a gene as over-mutated and under positive selection if it was  $\geq 2$  standard deviations above the neutral line. Similarly, a gene was labelled “under-mutated” (i.e., conserved) if it was  $\geq 2$  standard deviations below the neutral line. In other words, our primary metric for natural selection was based on whether mutations to a gene were less (stabilizing selection) or more (directional selection) frequent than expected by chance.

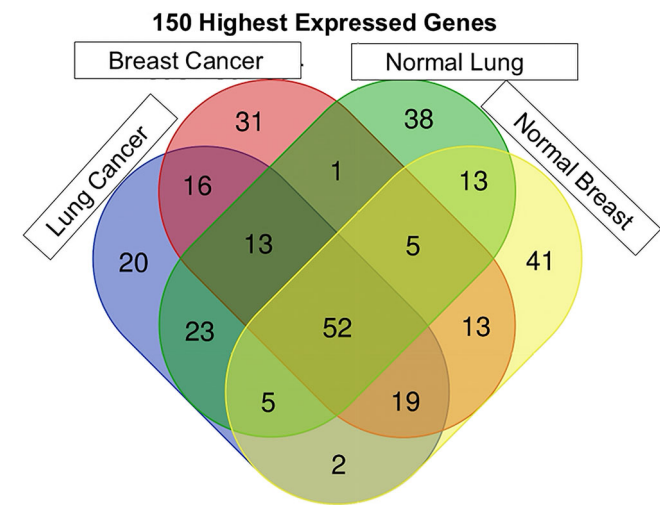
Our approach is similar to prior studies identifying “essential genes” in bacterial<sup>26</sup> and human genomes<sup>27</sup>. We note, however, that evolutionary selection on genes will be context dependent. That is, genes necessary for cancer cell proliferation in the host may differ significantly from genes required for cancer cells competing in a culture dish.

We recognize that our assumption of a roughly equal probability for mutations in all base pairs differs from prior studies that find variation in mutation rates related to gene expression, differences in DNA repair, and chromosomal location<sup>28</sup>. However, as discussed in prior publications<sup>8,9</sup>, these studies typically are carried out in populations over time, during which

competition among cancer cells will occur. Thus, variation in the *observed* frequency of mutations in a gene that decreases fitness and eliminated by natural selection could be interpreted as the result of intra-genomic variation in mutation rates. For example, Monroe et al.<sup>29</sup> observed that in *Arabidopsis thaliana* “genes subject to stronger purifying selection have a lower mutation rate”. We hypothesize important genes are *observed* to be mutated less often<sup>30</sup> because such mutations reduce fitness and are not transmitted across generations. Thus, we use apparent differences in the number of mutations observed in a cohort compared to the number expected based on the gene size as an estimate of its contributions to fitness. A genetic mutation observed significantly more frequently than expected indicates that mutation increases fitness while a gene that has significantly fewer mutations than expected must encode a protein whose normal function is essential for optimal fitness of the cancer cell.

### Identifying pathways and functions under selection

The collected TCGA data includes the expression of each gene in the cancer and its corresponding normal tissue. These data are used to estimate the average transcriptomic state of the normal and cancer cells. To determine coordinated functions of related groups of genes, curated lists of selected or conserved genes were subjected to pathway analysis using DAVID (Database for Annotation and Integrated Discovery), available at the website <https://david.ncifcrf.gov> and the ConsensusPathDB functional annotation tool at website [http://cpdb.molgen.mpg.de/CPDB/fct\\_annot](http://cpdb.molgen.mpg.de/CPDB/fct_annot). Both Gene Ontology (GO)- DIRECT and -FAT were used to identify significant ontologies, including Biological Process (BP), Cell Compartment (CC), and Molecular Function (MF). We then performed functional annotation clustering. We selected clusters based on high numbers of queried genes or based on biological significance. Statistical *p* values are from bioinformatic tools and represent the probability that the observed cluster could be the result of chance alone. Venn diagrams were constructed using the Ghent University VIB/UGent Center website: <http://bioinformatics.psb.ugent.be/webtools/Venn/>.



**Fig. 2 |** Overlap and differences in the 150 most highly expressed genes in normal breast and lung tissue and breast and lung cancers as a metric of convergence. In normal breast and lung tissue, 75 of the 150 highest expressing genes are common. This increases to 100 of the 150 highest expressing genes in breast and lung cancers. See Supplemental Figs. 1, 2, 3, and 4 for specific genes and functional pathways.

Comparing expression changes

To investigate non-mutational evolutionary strategies, we used the average expression for each gene in each TCGA cohort. Similarities and differences in the transcriptional states of normal tissue and cancer tissue in each cohort were measured by comparing the average expression of genes under investigation in normal breast and normal lung tissue. The average expression of each gene was then plotted for lung cancer and breast cancers in the same TCGA cohort. Variations in the linear regression from each gene set in the normal tissue and cancers were used as a metric of divergence or convergence.

Statistics and reproducibility

We used chi-square tests of heterogeneity when comparing counts in response two categorical factors such as where one factor might be cancer versus normal or lung versus breast, and the other factor might be up-regulated versus downregulated expression or in common or exclusive to one group or the other. We used chi-square tests of homogeneity when there is a single factor (cancer versus normal or lung versus breast) and there is an expectation that counts should be equal between the categories of the factor.

For the genes of a gene family we compared, using least square regression, each gene’s expression in normal lung with its expression in normal breast, or each gene’s expression in lung cancer with its expression in breast cancer. Both the slope (a slope closer to 1 indicates a greater convergence of genes expression) and the R-squared (closer to 1 indicates a tighter correlation in gene expression with less residual variation around the fitted line) indicate the degree of correspondence in gene expression between normal tissues and cancer. While much variation in these values when comparing across gene families, overall, across all gene families, lung and breast cancer generally showed higher slopes and greater R-squared values than the corresponding normal lung and breast.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Initial and final cell states

Prior mathematical models show the evolutionary arc of each cancer is strongly influenced by the initial state of the normal progenitor cells and

**Table 1 |** Differences in genetic and epigenetic mechanisms of inheritance LUAD and BRCA

	Breast Cancers N = 727	Lung Cancers N = 313
Average mutations per tumor	34	273
Genes mutated in >10% of cohort	2	166
Evolutionary Selected mutations	93	343
Evolutionarily conserved genes	130	260
Genes with ≥ 4-fold increase expression	963	558
Genes with ≥ 75% decreased expression	822	914
Genes expressed in cancer but not in normal tissue	273	111
Genes expressed in normal lung or breast but not in corresponding cancer	108	32

LUADs, which typically arise in a mutagenic environment, have a larger mutational profile while BRCA, which arise from tissue that is epigenetically plastic (i.e. expands and contracts in response to systemic hormones), evolves largely through epigenetic inheritance.

the final state of the optimally fit cancer cells<sup>15</sup>. As a general estimate of differences and similarities in initial and final transcriptional states, we identified the 150 most highly expressed genes in each group. As evidence for convergent evolution, in normal lung and normal breast tissue, fewer highly expressed genes were common, 75, than for LUAD and BRCA with 100 of their respective 150 highly expressed genes in common ( $\chi^2 = 8.57$ ,  $df = 1$ ,  $p < 0.005$ , test for heterogeneity) (Fig. 2, Supplemental Fig. 1).

Functional annotations of the 150 highest expressing genes in lung and breast cancer show common (Supplemental Fig. 2) genes associated with cytoplasmic translation, signaling, antigen presentation and processing, unfolded protein response, glycolysis, lysosomes, serine endopeptidase inhibitors, spliceosome, the melanosome, and the extracellular matrix including both collagen production and degradation.

Of the 150 genes most highly expressed in normal lungs, LUAD maintained or increased expression of 96. This was a significantly greater proportion than for BRCA which maintained or increased expression of 78 of the 150 genes most highly expressed in normal breast tissue ( $\chi^2 = 4.43$ ,  $df = 1$ ,  $p < 0.05$ , test for heterogeneity) (Supplemental Figs. 3 and 4). This may reflect variations in tissue background during somatic evolution. The stress of inflamed, smoking-related environment in “normal” lung tissue may require upregulation of genes that also contribute to fitness in the cancer environment.

Mutations and expression changes in transition from normal to malignant states

LUADs without known driver mutations, unlike BRCA, typically arise in a smoking induced mutagenic and inflamed ecosystem. Consistent with this, BRCA average 34 mutations/tumor compared to 273 mutations/tumor in LUADs. As an alternative metric of mutational frequency, we find only 2 gene mutations (*TP53* and *PIK3CA*) occurred in >10% of the breast cancer cohort compared to 231 in lung cancer (Table 1, Supplemental Fig. 5). LUAD have significantly more evolution-selected mutations than BRCA (321 versus 93 (Supplemental Fig. 6), respectively;  $\chi^2 = 366.4$ ,  $df = 1$ ,  $p < 0.001$ , test for homogeneity). Of these, sixteen are common to both cancers (Supplemental Fig. 6).

While mutations are extensively studied in cancers, the conceptual model of evolutionary triage expects that conserved genes reveal valuable insights into critical pathways that must be preserved for optimal fitness<sup>15</sup>. We identified significantly more conserved genes in LUAD than in BRCA (260 versus 130, respectively,  $\chi^2 = 43.3$ ,  $df = 1$ ,  $p < 0.001$ , test for homogeneity) with only one (*HTT* – the Huntingtin gene) conserved in both (Supplemental Fig. 7).



As a measure of transcriptional changes, breast cancers have 963 genes with  $\geq 4$ -fold increased expression versus normal breast tissue compared to 558 in lung cancer with 246 common to both cancers. However, there are more genes with  $>75\%$  decreased expression (compared to normal tissue) in LUAD than BRCA (914 versus 822, respectively,  $\chi^2 = 13.3$ ,  $df = 1$ ,  $p < 0.001$ , test for homogeneity) with 255 of these genes in common (Supplemental Fig. 8).

### Evolutionary steps in the transition to a malignant phenotype

**General patterns of mutations and expression change.** LUADs have far more mutations than BRCA but the latter exhibits more genes with large expression changes (Table 1). However, there is some overlap of mutational selection in LUAD and expression changes in BRCA as 79 of 341 genes with evolution selected mutations in LUAD have  $>2$ -fold expression and 126 have  $>50\%$  decreased expression in BRCA. (Supplemental Fig. 9).

Although parallel changes in expression are common in BRCA and LUAD, 60 genes show  $>4$ -fold increased expression in BRCA and  $>75\%$  decreased expression in LUAD (Supplemental Fig. 8), and 12 genes have the opposite divergence. However, this apparently divergent expression changes generated an overall pattern of convergence so that there was greater similarity of expression in the cancers compared to normal tissue (Supplemental Fig. 10). This group of genes is enriched ( $p < 0.05$ ) for 7 genes associated with collagen-containing extracellular matrix and 5 membrane transporters (*SLC29A4*, *SLC7A10*, *SLC5A8*, *SLC6A4*, *SLC1A*) which have diverse functions including the uptake of amino acids, monocarboxylates, and neurotransmitters<sup>31–33</sup>.

**Abrogating response to normal tissue controls.** To evolve, cancer cells must become independent of local tissue controls that govern their phenotypic properties, position, and proliferation/death. Compared to normal breast and lung, respectively, we expect BRCA and LUAD to significantly downregulate the expression of genes associated with local signaling pathways that govern cellular proliferation, death, phenotype, function and location. As noted above, we see  $\geq 75\%$  decreased expression (compared to normal tissue) in 822 genes in breast cancer and 914 genes in lung cancer with 255 common to both. The functional annotations (Supplemental Fig. 11) of the genes with decreased expression exhibit both convergent and parallel evolution (see below). For example (Supplemental Fig. 11), BRCA and LUAD downregulate 152 and 153 genes, respectively associated with signal transduction. Of these, 59 are common to both while 93 and 92, respectively, are specific for each tumor type. BRCA and LUAD downregulated 5 and 8 genes, respectively, in the HIPPO pathway with no gene in common. Conversely, they down regulate the same genes in the VEGFA-VEGFR signaling pathway. For adrenergic signaling, peptide-ligand receptors, calcium signaling, and G alpha (i) signaling, roughly half of the downregulated genes in each tumor type are unique while the other half are common to both. BRCA downregulated 9 genes in the RAF/MAPK cascade in common with LUAD which additionally downregulated 14 other genes in the pathway. The opposite pattern with 6 to 14 genes downregulated genes in common to both cancers and BRCA only decreasing expression in an additional 8 to 19 genes involved in chemokine, cytokine, and PPAR signaling as well as the neuroactive ligand-receptor signaling pathway. LUADs uniquely downregulate genes in the Netrin-UNC5B, corticotropin releasing hormone, ERBB, AGE-RAGE, PI3/AKT, and secretin receptor signaling pathways. Similarly, BRCA uniquely down regulates genes in the insulin, glucagon, apelin, AMPK, oxytocin, cGMP-PKG, nuclear receptor, retinoic acid, estrogen, adipocytokine, and NRIH2- and NRIH3-mediated pathways.

A critical factor in normal tissue communication is cell-cell interactions. We expect this interaction will normally be tissue specific and, for example, find relatively low correlation with genes related to cell-cell interactions such as the Cadherin, Protocadherin, and Claudin gene families in normal breast and lung (Fig. 3). However, both families show strong convergence in BRCA and LUAD suggesting selection for a common cancer phenotype that governs a cancer cell's interactions with both host cells and other cancer cells.

**Downregulation of differentiated functions.** As independence from host control evolves, the specialized functions of mammalian cells, like eyes in a cavefish, impose an ongoing cost while usually providing no evolutionary fitness benefit. These functions are generally tissue specific. Thus, BRCA turned off 108 genes highly expressed in normal breast compared to 32 genes expressed in normal lung but turned off in LUAD with only 7 in common (Fig. 4). Similarly, BRCA and LUAD downregulated 44 and 33 developmental genes, respectively, with no overlap (Supplemental Fig. 11) and each tumor downregulated multiple genes associated with cell-cell adhesion/communication, circadian entrainment, vascular smooth muscle, and prostaglandin synthesis with no genes in common. LUAD downregulated 18, 7, and 8 genes related to complement, surfactant, and oxygen exchange, respectively with 4, 3, and 4 in common with BRCA. BRCA downregulated 22 and 31 genes, respectively; associated with adipogenesis and muscle proteins with 8 and 15 in common with LUAD. LUADs uniquely downregulated genes associated with cilia, ECM-receptor interactions, and clathrin-mediated endocytosis. BRCA uniquely downregulated genes associated with steroid metabolism, androgen/estrogen biosynthesis, and triacylglycerol synthesis.

**Maximizing cancer cell fitness.** In cavefish, specific phenotypic properties (loss of eyes and skin pigment with increased tactile sensory function) maximize fitness. Thus, although multiple, diverse fish species enter and live in caves, they evolutionarily converge on an optimal phenotype for their common environment. We hypothesized that this might be observed in cancers. That is, once cells are independent of host tissue, they will tend to converge on a common phenotype that is optimally fit for a single cell protist living in situ.

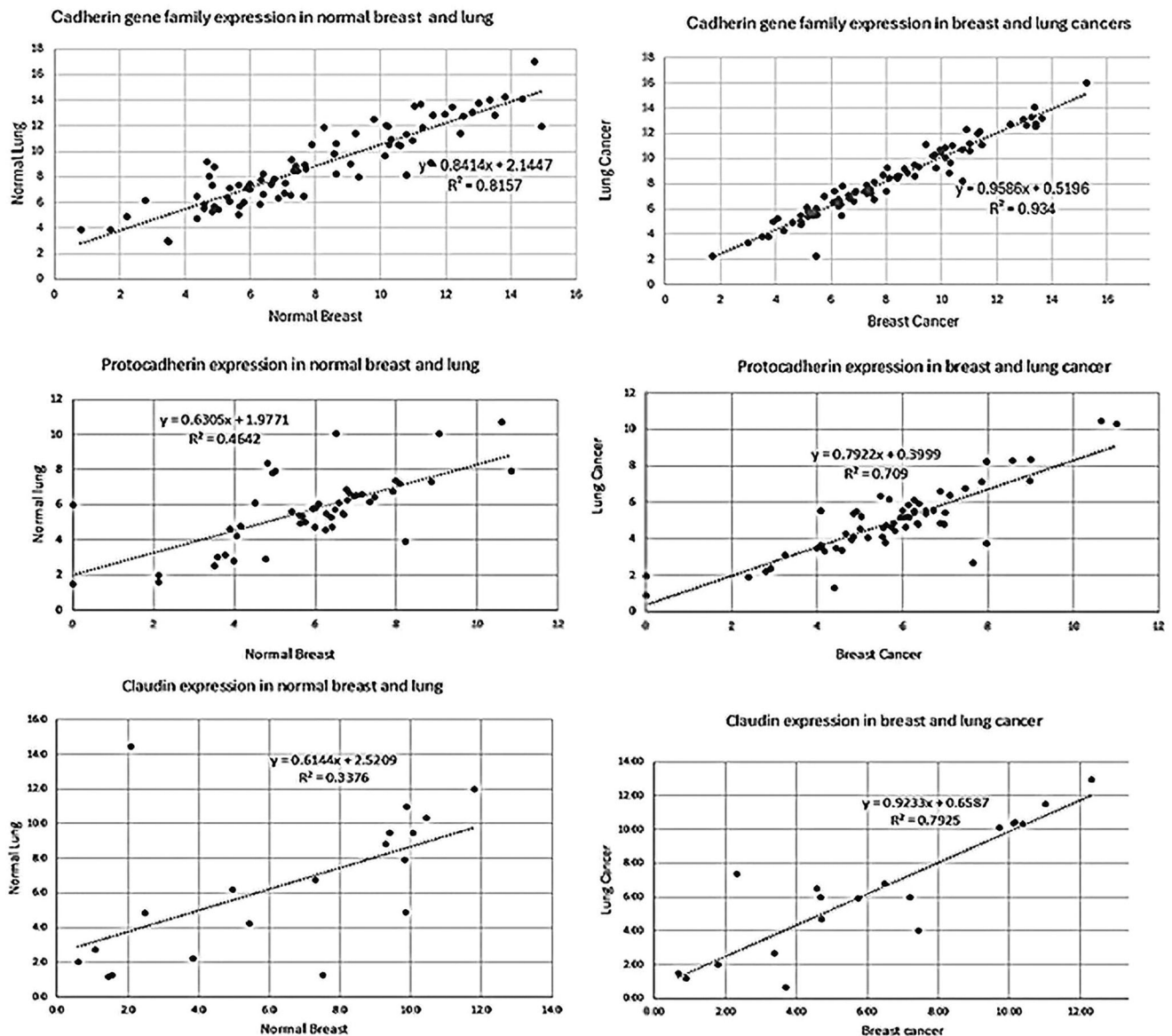
### Promoting phenotypic plasticity – accessing restricted regions of the genome

Phenotypic variability within individual cancer populations is well recognized. Here, we propose the *ability* to rapidly change phenotypic properties is a trait explicitly selected during cancer development. That is, the cancer environment, due to variations in blood flow and host response<sup>34</sup>, can experience sudden, potentially lethal changes in local environmental conditions. This selects for cells that can rapidly detect changes in environmental conditions and alter their phenotype to survive and proliferate. In other words, phenotypic plasticity is, itself, a selected trait.

Normal mammalian cells have epigenetic restrictions on regions of the genome<sup>35</sup> to maintain stable differentiated functions<sup>36</sup>. We identified 273 and 111 genes (27 in common) with high expression ( $> 1.5$ ) in BRCA and LUAD, respectively, but no or minimal expression ( $< 0.4$ ) in their corresponding normal tissue (Supplemental Fig. 12). The smaller number in LUAD may reflect prior “unlocking” of restricted genes due to smoking induced inflammation. LUAD (Supplemental Fig. 13) eliminated expression of 32 genes expressed in normal lung compared to 108 in BRCA with 7 genes in common ( $\chi^2 = 89.7$ ,  $df = 1$ ,  $p < 0.001$ , test for homogeneity).

The 27 activated genes common to both LUAD and BRCA include telomerase reverse transcriptase, (*TERT*), *LIN28B*, otherwise expressed only in embryonic stem cells<sup>37</sup>, *CDH18*, expressed on the embryonic subplate<sup>38</sup>, and *CSAG1*, which maintains integrity of the mitotic centrosome with defective p53<sup>39</sup>. LUAD and BRCA each activate 10 members of the MAGE gene family (with 7 in common (Supplemental Fig. 14)), which are typically described as “cancer/testes antigens”<sup>40</sup>. This strong evolutionary selection suggests MAGE genes contribute to cancer cell fitness although their precise function is currently unclear<sup>41</sup>. *PLAC1*, also described as a cancer testes gene<sup>42</sup>, has increased expression in both cancers, while BRCA uniquely activate *CTAG2*, *CT83*, *PAGE2*, *PAGE2B*, and *PAGE5*.

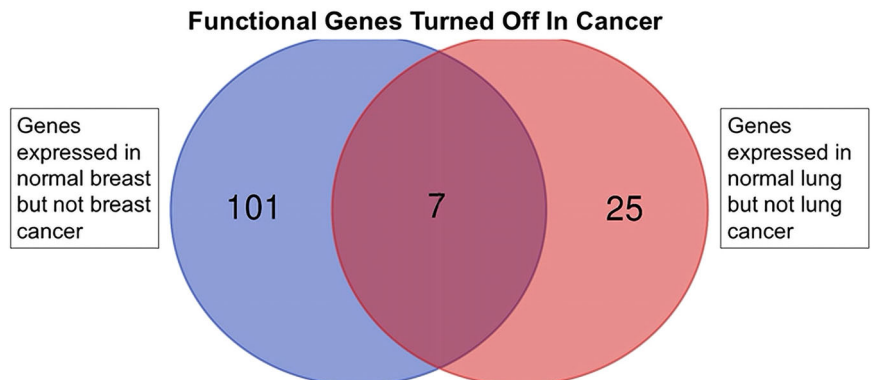
In BRCA, access to restricted elements of the genome may be promoted by frequent mutations in 7 genes associated with pluripotency (*TP53*, *AKT1*, *KRAS*, *TBX3*, *PIK3CA*, *PIK3R1* and *GATA3*). Of these, 4 (*AKT1*, *KRAS*, *TBX3* *GATA3*) have  $\geq 8$ -fold increased expression in BRCA. *PIK3CA* is the most highly mutated gene in BRCA but its expression, and that of *PIK3R1*,



**Fig. 3 | Rewiring genes associated with cell-cell interaction.** Left figures plot average expression of each member of the protocadherin (top panel), cadherin (middle panel) and claudin (lower panel) family of genes in normal breast (x-axis) and normal lung (y-axis) from their TCGA cohorts compared to their expression in

breast (x-axis) and lung (y-axis) in the right panels. There is pattern of convergence as the expression of members of each gene family become more similar in the cancers compared to the normal tissue.

**Fig. 4 | Genes expressed in normal tissue but not in the cancers arising from that tissue.** Early evolutionary selection will decrease differentiated function that benefits the host. The associated molecular machinery incurs a cost to the nascent, evolving cancer cell but typically provides no fitness benefit. The involved genes are highly tissue-specific and show minimum overlap. See Supplemental Figs. 12 and 13.



**Table 2 | Transcriptional selection (defined as >4-fold increased or 75% decreased expression) is extensively observed in both tumor types**

Functional annotations of genes with greater than 4-fold increased expression in BRCA and LUAD		
Pathway name	Number of genes	p-value
Cell Cycle	67	8.9e-42
Cell Cycle, Mitotic	58	1.9e-36
Cell Cycle Checkpoints	39	4.7e-30
Mitotic Prometaphase	27	1.3e-19
Mitotic Anaphase	27	1.9e-19
Mitotic Metaphase and Anaphase	27	2.2e-19
PLK1 signaling	16	3.7e-19
Separation of Sister Chromatids	23	3.7e-19
M Phase	35	5.0e-19
Amplification of signal from unattached kinetochores via MAD2 inhibitory signal	19	5.8e-17
Mitotic Spindle Checkpoint	20	1.0e-16
EML4 and NUDC in mitotic spindle formation	19	1.3e-16
Polo-like kinase mediated events	9	1.9e-13
FOXN1 transcription factor network	12	3.2e-13
APC/C-mediated degradation of cell cycle proteins	12	6.0e-13
Regulation of mitotic cell cycle	12	6.0e-13
Aurora B and Aurora A Signaling	18	7.0e-12
Kinesins	10	3.1e-10
Meiosis	15	3.5e-10
E2F transcription factor network	12	5.1e-10
TP53 Regulation of Cell Cycle Genes	10	1.8e-09
Signaling by Rho GTPases	31	7.8e-09
Progesterone Signaling	11	1.5e-07
ATM Signaling	6	1.7e-07
DNA Replication	10	1.9e-07
Collagen formation	10	5.7e-07
Golgi-to-ER retrograde transport	11	6.9e-07
ATR signaling pathway	7	7.0e-07
Activation of NIMA Kinases NEK9, NEK6, NEK7	4	1.3e-06
Activation of Matrix Metalloproteinases	6	2.6e-06
Chromosome Maintenance	11	2.7e-06
Extracellular matrix organization	16	2.9e-06
DNA Repair	17	3.0e-06
Matrix Metalloproteinases	6	3.9e-06
Activation of ATR in response to replication stress	6	1.2e-05
Cellular senescence	10	7.1e-05
p53 signaling pathway	7	7.2e-05
p73 transcription factor network	7	0.0001
Regulation of PLK1 Activity at G2/M Transition	7	0.0002
Homologous recombination	5	0.0002
Nucleosome assembly	5	0.0003
uPAR-mediated signaling	5	0.0001
Estrogen responsive protein efp controls cell cycle	3	0.001
Targets of C-MYC transcriptional activation	6	0.001

**Table 2 (continued) | Transcriptional selection (defined as >4-fold increased or 75% decreased expression) is extensively observed in both tumor types**

Functional annotations of genes with greater than 4-fold increased expression in BRCA and LUAD		
Pathway name	Number of genes	p-value
Signal Transduction	51	0.001
Generic Transcription Pathway	29	0.002
Syndecan-1-mediated signaling events	4	0.003
Telomere Maintenance	6	0.003
MicroRNAs in cancer	11	0.004
ECM-receptor interaction	5	0.008
RNA Polymerase II Transcription	29	0.009

Patterns in functions of genes upregulated demonstrate common pathways that are broadly necessary for optimizing cancer cell fitness.

are ~50% decreased. Likely related, expression of *GATA4*, which can induce pluripotency<sup>43</sup>, is increased >32-fold in BRCA. Interestingly, *TBX3* shows 90% decreased expression in LUAD. However, LUAD have strong genetic and epigenetic selection on *HOX* genes, which regulate pluripotency in stem cells<sup>44</sup>. LUAD also exhibit 8-fold increased expression of *HMGB3*, a multifunctional gene that can promote stemness<sup>45</sup>.

### Optimizing the molecular machinery of proliferation

Among the upregulated genes in BRCA and LUAD, approximately half are associated with the molecular machinery of proliferation (Table 2). This includes large, identical gene sets associated with cell cycle, mitosis, nucleosome assembly and DNA synthesis (all  $p < 0.001$ ).

While the “mutator phenotype” to generate new mutations and accelerate evolution is often posited as a necessary trait of cancers, this likely has limits as excessive numbers of mutations risks an “error catastrophe”<sup>46</sup> in which the progressively degraded cellular information is insufficient to maintain optimal fitness. We find that both cancers upregulate the same 17 genes associated with DNA repair ( $p = 3.0e-6$ ). Interestingly, both tumors upregulate 15 genes associated with meiosis ( $p = 3.5e-10$ ) which may participate in repair of DNA double strand breaks<sup>47</sup> but has also been posited to drive oncogenesis<sup>48</sup>.

Optimal fitness requires upregulation of the molecular machinery for genetic transcription. Consistent with this, both cancers upregulate 29 genes associated with RNA Polymerase II and 12 genes in the E2F transcription network (both ( $p < 0.01$ )). We find convergence of expression in both gene families as well as the APOBEC, E2F, and DDX gene families (Supplemental Fig. 15).

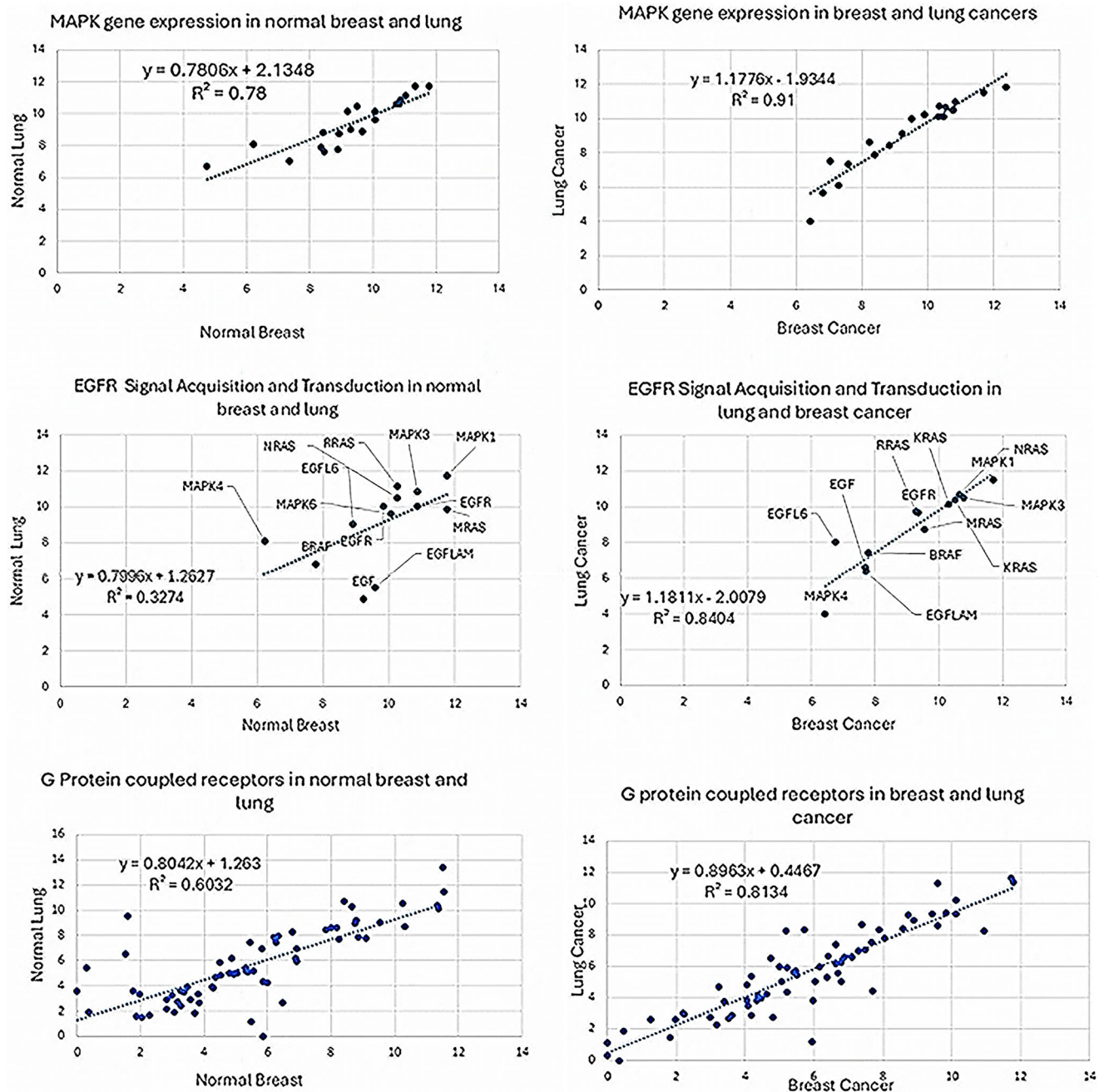
*TP53* is the most mutated gene in LUAD and BRCA and both cancers have ≥4-fold increase in 7 genes associated with p53 regulation of cell cycle ( $p = 7.2e-5$ ) (Table 2).

Cancer cells are assumed to be immortal and, in a seeming paradox, both cancers upregulate 10 genes related to cellular senescence ( $p = 7.1e-5$ ) and 6 related to telomere maintenances ( $p < 0.01$ ). This observation warrants further investigation (Table 2).

### Alternative signaling pathways

While loss of tissue control is necessary for cancer initiation, alternative signaling pathways, not governed by host tissue signals, are necessary for responding to environmental opportunities and threats, as well as maintaining pro-proliferative signaling. Thus, both cancers upregulate the same 51 genes associated with signal transduction ( $p < 0.01$ ), 29 common genes in the “generic” transcription pathway ( $p < 0.01$ ), and 11 microRNAs associated with cancer ( $p < 0.01$ ). Both tumors upregulate 6 genes encoding targets to C-MYC activation, 7 members of the p53 signaling pathway, 12 members of the FOXN1 transcription network, and 7 members of the p73 signaling pathway ( $p < 0.01$  for all) (Table 2).





**Fig. 5 | Promoting alternative signaling pathways.** In the left panels, expression of each gene in normal breast (x-axis) and normal lung (y-axis) is plotted. In the right panels, the expression of each gene in breast cancer (x-axis) and lung cancer (y-axis) is plotted. Top panels demonstrate convergence of expression in MAPK pathway genes in breast and lung cancers compared to their normal tissue counterparts.

Middle panels demonstrate changes in specific components of EGFR, RAS, and RAF signaling pathways. Note, for example, convergence on expression of *EGF*, *NRAS*, *KRAS*, and *RRAS*. Lower panels show convergence with elements of the G Coupled Protein signaling families are more similar in breast and lung cancers compared to normal breast and lung.

Other pathways with genes upregulated in both cancers include *PLK1*, *Aurora A* and *B*, *Rho GTPase*, progesterone, estrogen, *ATM*, *ATR*, *NIMA* kinases, and *Syndecan-1* ( $p < 0.01$  for all).

We find that expression of genes associated with the MAPK pathway and EGFR transduction are similar in normal breast and lung cells and further converge in breast and lung cancer. Expression of members of the Rho and G protein coupled receptors (Fig. 5) and C-C Motif chemokine ligands (Supplemental Fig. 16) converge in breast and lung cancers, suggesting an important and common role in optimizing fitness.

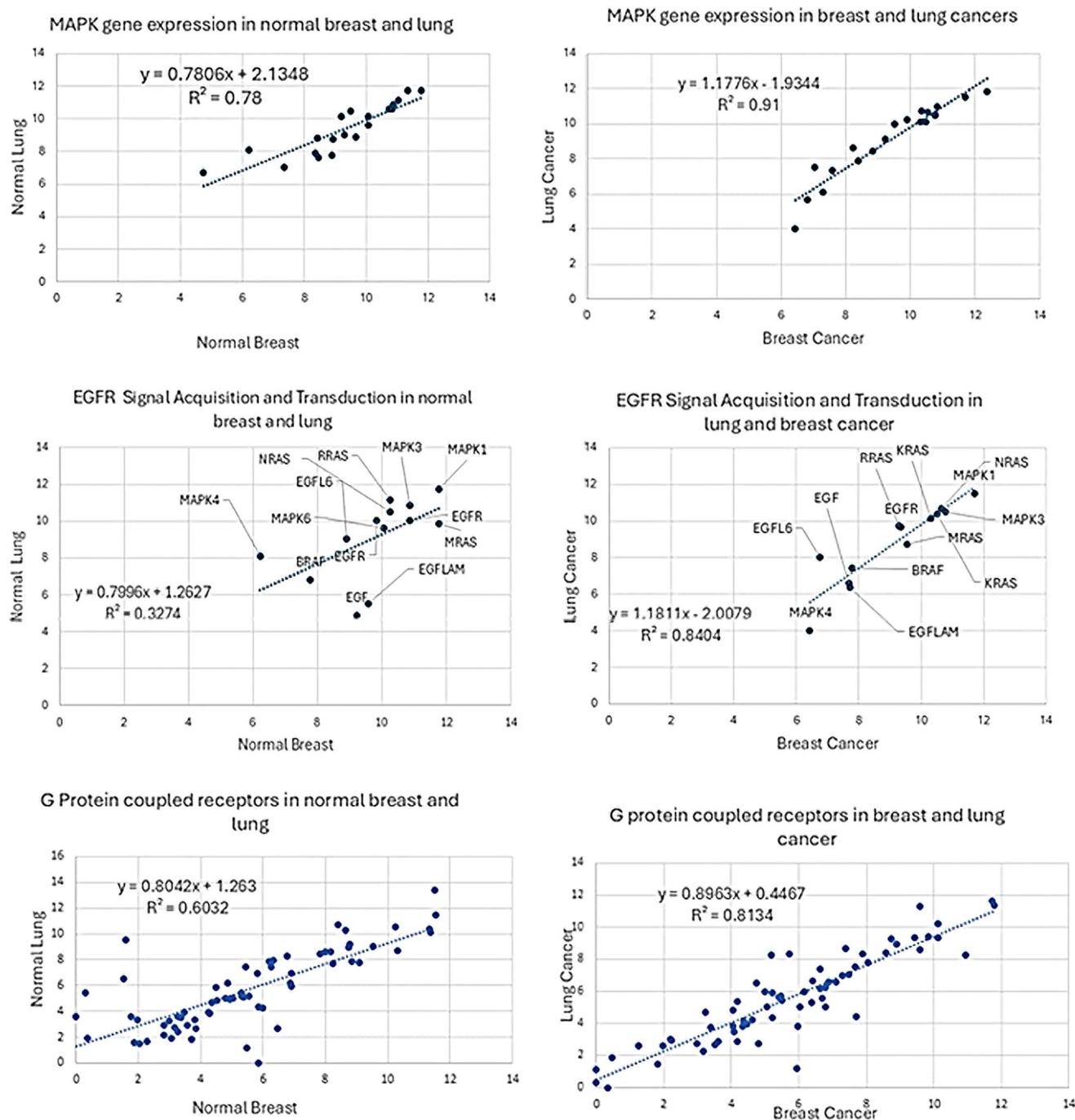
Although there is broad convergent evolution in many signaling pathways, tumor-specific changes as a legacy of the normal tissue of origin persist. For example, genes from the BMP pathway show no convergence

(Supplemental Fig. 16). LUADs increase the expression of *ESR1* nearly 8-fold but expression is still much less than *BRCA* which increases expression by  $> 3$ -fold. Interestingly, *BRCAs* decrease expression *ESR2*  $> 90\%$  while it is highly expressed in normal lung with increased expression in LUAD. *PGR* is highly expressed in both normal breast and lung with  $\sim 50\%$  increased expression in LUAD and *BRCA*.

### Cellular infrastructure

In the often-harsh intratumoral environment, cancer cells must maintain the necessary infrastructure for survival and proliferation. Both cancers upregulate 10, 11, and 11 genes associated with kinesins, chromosome maintenance, and Golgi-to-ER retrograde transport (all  $p < 0.01$ ),





**Fig. 6 | Upregulating alternative signaling pathways.** Convergence of some potassium and sodium ion channels suggests a significant signaling role in breast and lung cancers.

respectively, with relative convergence in the expression patterns of genes encoding members of the actin and myosin families (Supplemental Fig. 17).

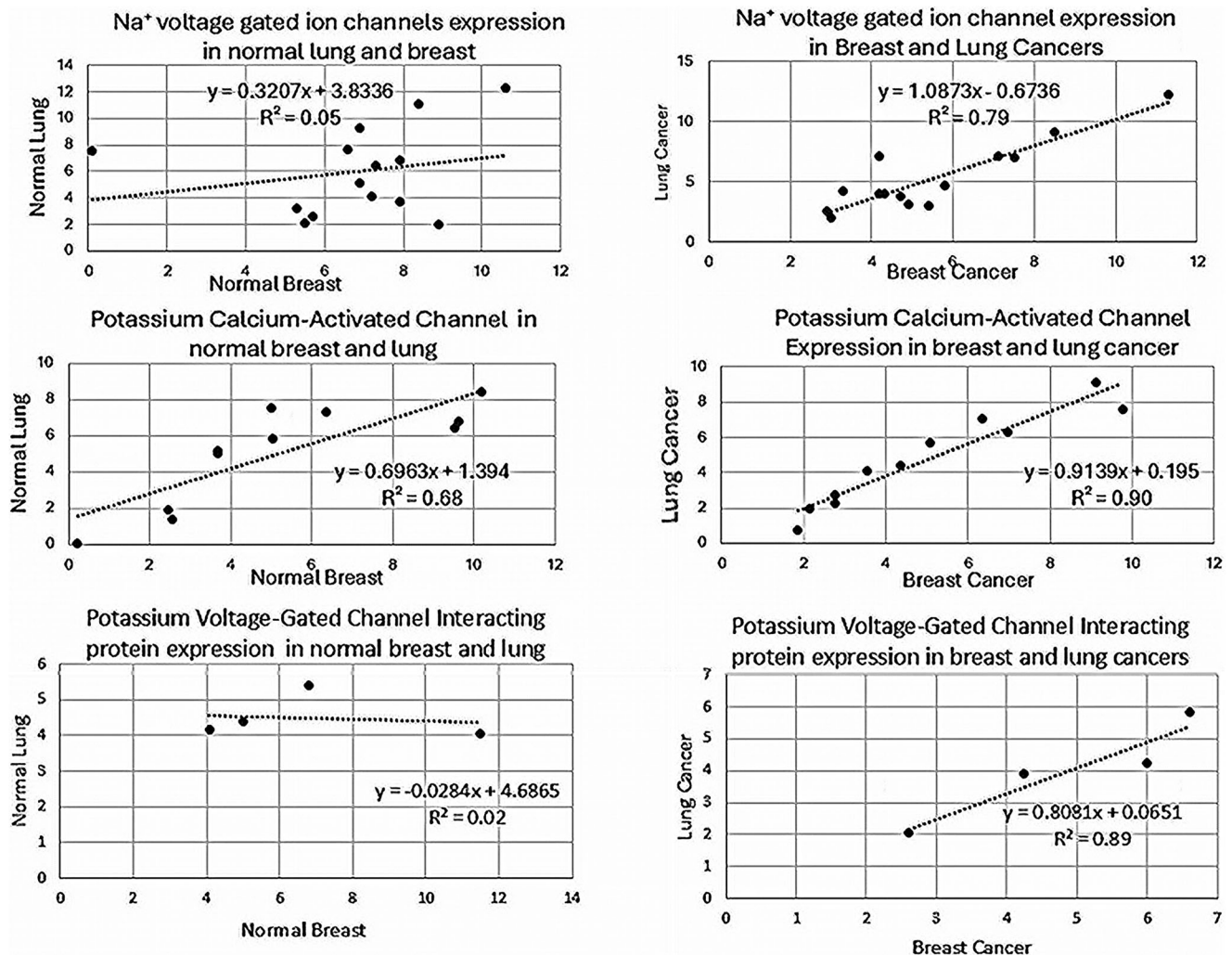
A necessary function of cancer cells is foraging as they transport diverse molecules in and out of the cell. In Supplemental Fig. 18, we demonstrate convergence in the ABCA and ABCB families. The former is associated with lipid transport<sup>49</sup> and that latter with diverse development functions<sup>50</sup>. Both have prognostic significance in LUAD<sup>51,52</sup>. In contrast, we show moderate convergence within the SCLC gene family and no convergence in TMEM genes (Supplemental Fig. 19).

### Ion channels and neural receptors

Differentiated mammalian cells typically maintain a high transmembrane potential, but the potential is small or absent in cancer cells<sup>53–55</sup>. Both cancers

demonstrate strong selection on channels associated with membrane ion fluxes including ion channels and neural receptors (Supplemental Figs. 20 and 21). BRCA increase ( $\geq 4$ -fold) expression of 15 genes associated with voltage gated ion channels. LUADs show increased expression of 7 genes with 3 in common. BRCA decreased ( $\geq 75\%$  decrease) expression of 25 genes associated with voltage gated ion channels, and 22 in LUAD with 5 in common (Supplemental Fig. 20, Table 2). The pattern demonstrates tissue-specific variations in the voltage gated ion channels primarily responsible for the transmembrane potential in normal breast and lung cells but general convergence of expression in LUAD and BRCA particularly in Na<sup>+</sup> voltage gated ion channels. (Fig. 6, Supplemental Fig. 21).

Among ion channels associated with neural receptors we see strong selection and expression convergence in the RYR (Supplemental Fig. 22)



**Fig. 7 | Niche construction – remodeling the extracellular matrix.** The tumor microenvironment is subject to complex dynamics that mix host response with cancer cell niche construction. In the left panels, expression of each gene in normal breast (x-axis) and normal lung (y-axis) is plotted. In the right panels, the expression

of each gene in breast cancer (x-axis) and lung cancer (y-axis) is plotted. Common patterns of increased and decreased collagen genes and metalloproteinase genes result in a convergence of expression patterns (see Supplementary Fig. 19).

among other neuroreceptors we see convergence in the 5-HT family which promotes cyclic AMP (Supplemental Fig. 22) but not cholinergic, adrenergic, or glutamate receptors (Supplemental Fig. 23).

**Niche engineering and host response.** Recent evolutionary theory recognizes the role of ecological engineering as adaptations for organisms to enhance their fitness by causing favorable changes to their environment<sup>56,57</sup>. Because these changes can persist over multiple generations, they represent a form of “ecological inheritance.”

Three competing dynamics govern the extracellular tumor environment: 1) a host response of deploying components of the extracellular matrix that can inhibit tumor growth. 2) tumor cell signaling to promote angiogenesis, blunt the immune response, and degrade host extracellular matrix to enable invasion. 3) tumor cell cooption of host mesenchymal or immune cells to produce an environment that maximizes their own fitness.

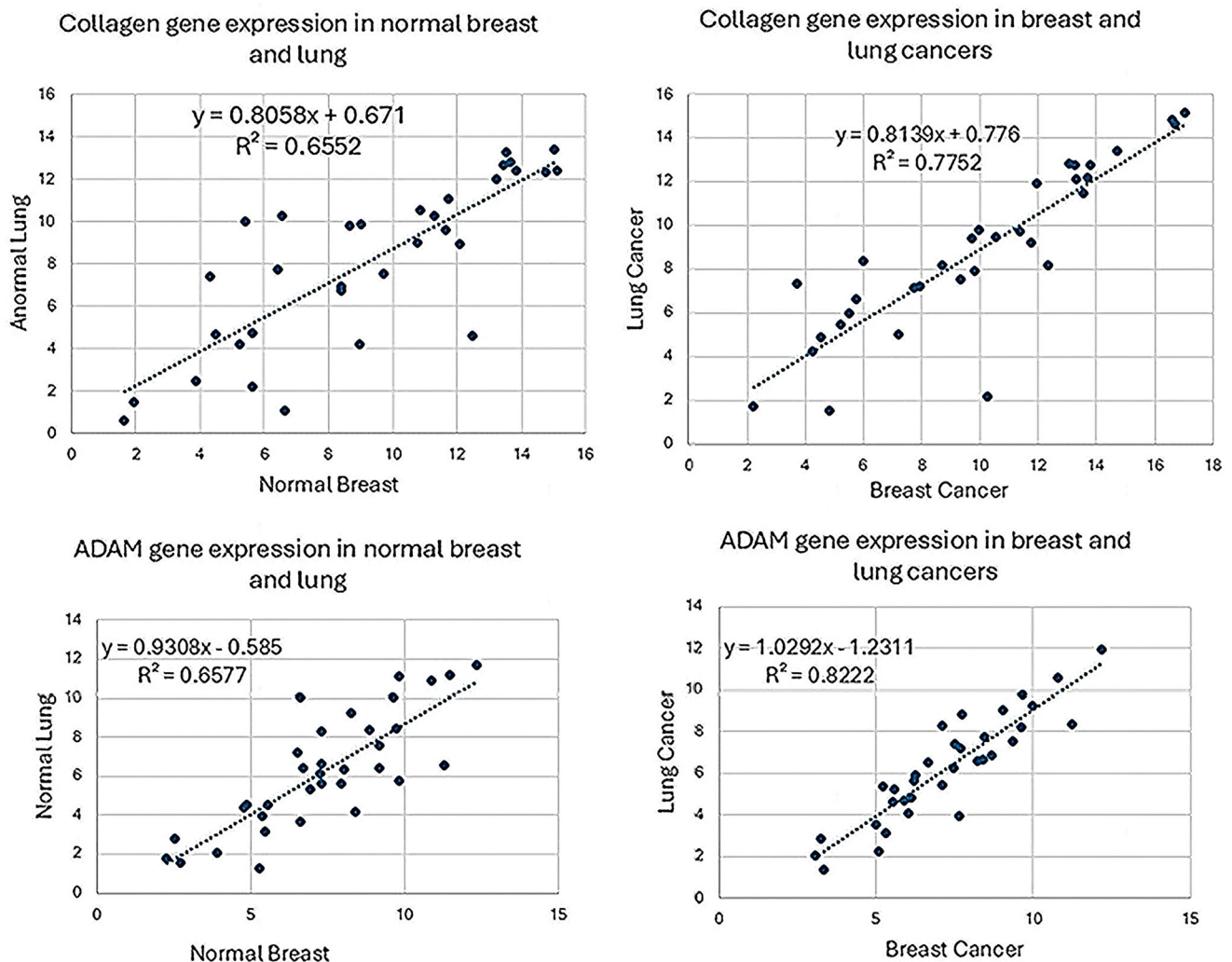
We find that 19 and 12 genes related to ECM are among the highest 150 expressed in BRCA and LUAD, respectively. These include 8 collagen genes in BRCA and 6 in LUAD in LUAD with 5 of these in common ( $p < 0.001$  for all) (Fig. 1, Supplemental Fig. 1).

Mutation and expression data in LUAD and BRCA demonstrate extensive genetic and epigenetic selection on elements of the ECM (Supplemental Fig. 24). Genes with  $\geq 4$ -fold increased expression in both LUAD and BRCA are highly enriched ( $p < 0.001$ ) for collagen degradation and

components of the extracellular matrix including 4 collagen genes and 6 matrix metalloproteinases. In LUAD, we find high mutation rates in most collagen genes and some members of the ADAMTS, Mucin, and Fraser gene families. In BRCA, there is strong evolutionary selection applied to members of the Fibrillin and Mucin gene families.

The most highly expressed collagen genes in normal breast and lung tissue, *COL1A1*, *COL1A2*, *COL3A1*, *COL6A6* and *COL6A3*, have a further 2- to 6-fold increased expression in both cancers. *COL10A1* and *COL11A1*, which are produced by acid-adapted cancer cells<sup>58</sup>, are moderately expressed in both normal tissues but have  $>50$ -fold increased expression in both cancers. Conversely, *COL4A3* is highly expressed in both normal tissues with  $\sim 90\%$  decreased expression in both cancers. *COL2A1* is highly expressed in normal breast tissue and further increased by  $\sim 10$ -fold in breast cancer but minimally expressed in normal lungs and increased about 2-fold in lung cancer.

Divergent patterns of expression changes are observed in *COL16A1*, *COL17A1*, *COL19A1*, *COL3A1*, *COL4A3BP*, *COL4A4*, *COL4A6*, *COL6A1* and *COL6A6*. However, as shown in Fig. 7, the net result is that expression of collagen genes is more similar in BRCA and LUAD than in normal breast and lung. An even greater convergence is seen in expression of members of the ADAM gene family (Fig. 5). However, little or no convergence is seen in other gene families associated with the extracellular space such as MMP, carbonic anhydrase, or mucin (Supplemental Fig. 25)



**Fig. 8 | Niche construction - angiogenesis.** In the left panels, expression of each gene in normal breast (x-axis) and normal lung (y-axis) is plotted. In the right panels, the expression of each gene in breast cancer (x-axis) and lung cancer (y-axis) is plotted. Both cancers increase the expression of some VEGF family members that promote

angiogenesis while other members of the family as well as angiopoietin genes that promote lymphangiogenesis or vascular maturation are downregulated. The overall selection forces result in relative convergence of expression in LUAD and BRCA.

A key component of cancer niche construction is angiogenesis (Fig. 8, Supplemental Fig. 24). In both cancers, *HIF1A* is conserved (1 observed mutation in 1040 samples) with about 20% increased expression. *EPAS1* (*HIF2A*) is highly expressed in both normal breast and normal lung (among the 150 highest expressed genes in normal lung) but decreases 75% and 90% in expression in BRCA and LUAD, respectively. Expression of *VEGFA* increased by 10 to 20% in both cancers while *VEGFB* and *VEGFC* decreased 10–50%. Interestingly, we find broadly decreased expression of angiopoietin genes as well as decreased expression of *TEK* (cell surface receptor for *ANGPT1*, *ANGPT2*, and *ANGPT4*) in both. *ANPTL4* is increased 3-fold in expression in LUAD, and expression of *ANGPTL6* increased by a similar amount in BRCA. The net effect is relative convergence of the VEGF and ANGPT gene families (Fig. 8)

**Tumor to host signaling pathways.** Cancer cells can co-opt fibroblasts and other host cells. As potential tumor-to-host signaling pathways, we find both cancers significantly increase expression of some ligands while decreasing expression of their associated receptors. For example, both cancers show a marked increased expression of Neuromedin U (*NMU*) but marked decreased expression of its receptor (*NMUR1*) (Supplemental Fig. 21) suggesting that cancer cells use the NMU protein to manipulate non-cancer cells within the tumor. For example, NMU

binding to *NMUR1* in lymphoid cells, causes release of IL-25 which strongly amplifies allergic inflammation<sup>59</sup>. *NMU* may also promote cancer cachexia<sup>60</sup>. Similarly, *EDN2* (Supplemental Fig. 21) is upregulated in both cancers while its receptors are down regulated or not expressed. BRCA highly upregulates the expression of *UCN3*, *CGA*, and *CPLX3* (neither is expressed in normal breast) which produce signaling proteins.

A more general tumor to host signaling is through production of acid<sup>58</sup>. Both cancers upregulate genes associated with glycolysis which will produce lactic acid as a byproduct (Supplemental Fig. 24 and 25). Similarly, both cancers highly upregulate *CA9* expression (by 32 to 64-fold) while generally decreasing expression of other members of the Carbonic Anhydrase gene family. *CA9* buffers cancer cells against acidic conditions thus allowing cancer cells to adapt to conditions that are lethal or induce dysfunction in local host cells.

## Discussion

We examined patterns of gene mutations and transcriptional variations in two cancer populations with different cells of origin, tissue environments, and mutation rates. LUAD develops in smoking-related inflamed and mutagenic environments and has direct contact with the external environment. Conversely, BRCA emerges in tissue sheltered from environmental pathogens and cyclically respond to changing systemic hormone levels.



When carcinogenesis is viewed as an evolutionary transition, several predictable steps must be achieved by all cancers: 1) Independence from controlling host signals; 2) Acquisition of alternative signaling to optimize Darwinian fitness (proliferation); 3) Maintenance of molecular machinery for proliferation; 4) Down regulation of differentiated, whole-organism functions; 5) Promotion of phenotypic plasticity to adapt to temporally and spatially heterogeneous environments; 6) Niche construction to generate an extracellular environment that optimizes cancer cell survival and proliferation. In general, evolutionary dynamics 1 and 4 represent alterations that will be dependent on the structure and function of the normal tissue of origin while dynamics 2, 3, 5, and 6 represent adaptations broadly necessary for optimal fitness of cancer cells, as free-living protists in situ.

We investigated each of these dynamics by identifying genes and gene families under evolutionary selection. That is, we assume natural selection produces consistent patterns of co-adaptations among the genetic and epigenetic properties of cancer populations. We use directional (positive) and stabilizing (negative) selection on suites of genes as evidence of adaptive strategies across each patient cohort.

Because all replicating populations are subject to random gene mutations, we identify evolution favored mutations as well as gene conservation based on the observed frequency of the mutation in the cohort compared to the expected frequency based on the gene size. As discussed above, this assumption differs from models in which the mutation rate is assumed to vary across the genome. We note our assumption that the observed frequency of a mutation is due to a fitness gain or loss (with a corresponding difference in proliferation) could also be interpreted as a difference in the mutation rate. We minimize the potential error through application of stringent criteria for evidence of evolutionary selection (at least 2 standard deviations from the neutral line).

We find downregulated canonical signaling pathways related to differentiated functions and interactions with local tissue through decreased expression and/or increased mutations. This represents parallel evolution as the general evolutionary trends (independence from host control and loss of differentiated functions) are identical but the specific molecular pathways under selection are tissue specific.

In contrast, adaptations that increase the fitness of a cancer cell within host tissue, including new signaling pathways, upregulating molecular machinery of gene transcription and cell cycle, niche construction, and phenotypic plasticity, demonstrate significantly more overlap in BRCA and LUAD.

Perhaps the most predictable commonality is upregulation of the molecular machinery of proliferation. We find LUAD and BRCA upregulate a common set of core genes associated with cell cycle, mitosis, and DNA repair. However, we also find common selection on signaling pathways that promote proliferation. TP53, often described as a tumor suppressor, is the most mutated gene in both cancers and both upregulate elements of the p53 signaling pathway consistent with a primary role in promoting tumor fitness and proliferation<sup>61</sup>. Strong genetic/epigenetic selection on voltage gated ion channels suggests a reduction of the transmembrane gradient is a common strategy in LUAD and BRCA, to increase proliferative signals<sup>62</sup> while decreasing differentiated function<sup>63</sup> and increasing pluripotency<sup>54</sup>.

Recent evolutionary theory has recognized the critical role of ecological engineering as adaptations for optimizing an organism's fitness by causing favorable changes in its environment<sup>56,57</sup>. We find, despite marked differences in the ecology of their normal tissue, both cancers engineer broadly similar environments though genetic/epigenetic selection of specific collagen genes which promote survival in an acidic environment and inhibit immune response. Because this niche construction can persist over multiple generations, it represents a form of ecological inheritance in which descendants inherit the environment sculpted by their forebears.

In conclusion, BRCA and LUAD, which arise from different cellular progeny, evolve in different tissues, and rely on different mechanisms of inheritance, exhibit parallel evolutionary strategies in gaining independence from host control and downregulating the differentiated function of their

specific tissue type, and convergent evolution towards strategies that optimize their fitness once they have become free-living protists in situ.

## Data availability

All data was obtained from the publicly available TCGA website (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>). Data selected for the analysis is included in Supplementary Materials.

## Code availability

The Negative Storage Model software used to summarize the mutation data is under Intellectual Property protection; interested parties should contact the Moffitt Office of Innovation for licensing options: <https://www.moffitt.org/research-science/academic-and-industry-partnerships/office-of-innovation/available-technologies/>. Details of the R code used to determine over- and under-mutation are provided in Supplemental Software.

Received: 31 October 2024; Accepted: 23 April 2025;

Published online: 21 May 2025

## References

1. Brown J. S. et al. Updating the definition of cancer. *Mol. Cancer Res.* 2023. Epub 2023/07/06. <https://doi.org/10.1158/1541-7786.MCR-23-0411> (2023).
2. Gatenby, R. A. & Brown, J. Mutations, evolution and the central role of a self-defined fitness function in the initiation and progression of cancer. *Biochim Biophys. Acta Rev. Cancer* 1867, 162–166 (2017).
3. Wang, X. et al. Association between Smoking History and Tumor Mutation Burden in Advanced Non-Small Cell Lung Cancer. *Cancer Res.* **81**, 2566–2573 (2021).
4. Lugade, A. A. et al. Cigarette smoke exposure exacerbates lung inflammation and compromises immunity to bacterial infection. *J. Immunol.* **192**, 5226–5235 (2014).
5. Ramakrishnan, R., Khan, S. A. & Badve, S. Morphological changes in breast tissue with menstrual cycle. *Mod. Pathol.* **15**, 1348–1356 (2002).
6. Vineis, P., Schatzkin, A. & Potter, J. D. Models of carcinogenesis: an overview. *Carcinogenesis* **31**, 1703–1709 (2010).
7. Group, P. T. C. et al. Genomic basis for RNA alterations in cancer. *Nature* **578**, 129–136 (2020).
8. Freischel A. R., et al. Evolutionary Analysis of TCGA Data Using Over- and Under- Mutated Genes Identify Key Molecular Pathways and Cellular Functions in Lung Cancer Subtypes. *Cancers (Basel)* **15**. <https://doi.org/10.3390/cancers15010018> (2022).
9. Luddy, K. A., Teer, J. K., Freischel, A., O'Farrelly, C. & Gatenby, R. Evolutionary selection identifies critical immune-relevant genes in lung cancer subtypes. *Front. Genet.* **13**, 921447 (2022).
10. Gatenby, R. A. et al. Lung adenocarcinomas without driver genes converge to common adaptive strategies through diverse genetic, epigenetic, and niche construction evolutionary pathways. *Med. Oncol.* **41**, 135 (2024).
11. Heumos, L. et al. Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550–572 (2023).
12. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
13. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
14. Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
15. Gatenby, R. A., Cunningham, J. J. & Brown, J. S. Evolutionary triage governs fitness in driver and passenger mutations and suggests targeting never mutations. *Nat. Commun.* **5**, 5499 (2014).
16. D'Entro, S. C. et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* **184**, 2239–54 e39 (2021).



17. Gatenby, R. A., Gillies, R. J. & Brown, J. S. Of cancer and cave fish. *Nat. Rev. Cancer* **11**, 237–238 (2011).
18. Strickler, A. G. & Jeffery, W. R. Differentially expressed genes identified by cross-species microarray in the blind cavefish *Astyanax*. *Integr. Zool.* **4**, 99–109 (2009).
19. Casane, D. & Retaux, S. Evolutionary Genetics of the Cavefish *Astyanax mexicanus*. *Adv. Genet.* **95**, 117–159 (2016).
20. van der Weele C. M., Jeffery W. R. Cavefish cope with environmental hypoxia by developing more erythrocytes and overexpression of hypoxia-inducible genes. *Elife* **11**. <https://doi.org/10.7554/eLife.69109> (2022).
21. Riesch, R., Plath, M. & Schlupp, I. Toxic hydrogen sulphide and dark caves: pronounced male life-history divergence among locally adapted *Poecilia mexicana* (Poeciliidae). *J. Evol. Biol.* **24**, 596–606 (2011).
22. Hutchinson G. E. The ecological theater and the evolutionary play. New Haven,: Yale University Press; 1965. xiii, 139 p. p.
23. Ellrott, K. et al. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell Syst.* **6**, 271–81.e7 (2018).
24. Gonzalez-Calderon G., Liu R., Carvajal R., Teer J. K. A negative storage model for precise but compact storage of genetic variation data. Database (Oxford). 2020;. <https://doi.org/10.1093/database/baz158> (2020).
25. Cannataro, V. L., Gaffney, S. G. & Townsend, J. P. Effect Sizes of Somatic Mutations in Cancer. *J. Natl Cancer Inst.* **110**, 1171–1177 (2018).
26. Glass, J. I. et al. Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. USA* **103**, 425–430 (2006).
27. Chen, H. et al. New insights on human essential genes based on integrated analysis and the construction of the HEGIAP web-based platform. *Brief. Bioinform.* **21**, 1397–1410 (2020).
28. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
29. Monroe, J. G. et al. Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* **602**, 101–105 (2022).
30. Zhang, J. Important genomic regions mutate less often than do other regions. *Nature* **602**, 38–39 (2022).
31. Zengin, T. & Onal-Suzek, T. Analysis of genomic and transcriptomic variations as prognostic signature for lung adenocarcinoma. *BMC Bioinforma.* **21**, 368 (2020).
32. Li, P., Li, Y., Bai, S., Zhang, Y. & Zhao, L. miR-4732-3p prevents lung cancer progression via inhibition of the TBX15/TNFSF11 axis. *Epigenomics* **15**, 195–207 (2023).
33. Zhang, R. & Ma, A. High expression of MYEOV reflects poor prognosis in non-small cell lung cancer. *Gene* **770**, 145337 (2021).
34. Gillies, R. J., Brown, J. S., Anderson, A. R. A. & Gatenby, R. A. Eco-evolutionary causes and consequences of temporal changes in intratumoural blood flow. *Nat. Rev. Cancer* **18**, 576–585 (2018).
35. Consortium GT, et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
36. Patel T., Hobert O. Coordinated control of terminal differentiation and restriction of cellular plasticity. *Elife* **6**. <https://doi.org/10.7554/eLife.24100> (2017).
37. Gewalt, T., Noh, K. W. & Meder, L. The role of LIN28B in tumor progression and metastasis in solid tumor entities. *Oncol. Res.* **31**, 101–115 (2023).
38. Oeschger, F. M. et al. Gene expression analysis of the embryonic subplate. *Cereb. Cortex.* **22**, 1343–1359 (2012).
39. Sapkota H., Wren J. D., Gorbisky G. J. CSAG1 maintains the integrity of the mitotic centrosome in cells with defective p53. *J. Cell Sci.* **133**. <https://doi.org/10.1242/jcs.239723> (2020).
40. Schooten, E., Di Maggio, A., van Bergen En Henegouwen, P. M. P. & Kijanka, M. M. MAGE-A antigens as targets for cancer immunotherapy. *Cancer Treat. Rev.* **67**, 54–62 (2018).
41. Lee, A. K. & Potts, P. R. A Comprehensive Guide to the MAGE Family of Ubiquitin Ligases. *J. Mol. Biol.* **429**, 1114–1142 (2017).
42. Mahmoudian, J. et al. PLAC1: biology and potential application in cancer immunotherapy. *Cancer Immunol. Immunother.* **68**, 1039–1058 (2019).
43. Zhu, L. et al. Transcription Factor GATA4 Regulates Cell Type-Specific Splicing Through Direct Interaction With RNA in Human Induced Pluripotent Stem Cell-Derived Cardiac Progenitors. *Circulation* **146**, 770–787 (2022).
44. Steens, J. & Klein, D. HOX genes in stem cells: Maintaining cellular identity and regulation of differentiation. *Front Cell Dev. Biol.* **10**, 1002909 (2022).
45. Ma, H. et al. HMGB3 promotes the malignant phenotypes and stemness of epithelial ovarian cancer through the MAPK/ERK signaling pathway. *Cell Commun. Signal.* **21**, 144 (2023).
46. Eigen, M. Error catastrophe and antiviral strategy. *Proc. Natl. Acad. Sci. USA* **99**, 13374–13376 (2002).
47. Lingg, L., Rottenberg, S. & Francica, P. Meiotic Genes and DNA Double Strand Break Repair in Cancer. *Front Genet.* **13**, 831620 (2022).
48. McFarlane, R. J. & Wakeman, J. A. Meiosis-like Functions in Oncogenesis: A New View of Cancer. *Cancer Res.* **77**, 5712–5716 (2017).
49. Albrecht, C. & Viturro, E. The ABCA subfamily-gene and protein structures, functions and associated hereditary diseases. *Pflug. Arch.* **453**, 581–589 (2007).
50. Yang, Y. et al. Genome-wide identification and expression profiles of ABCB gene family in Chinese hickory (*Carya cathayensis* Sarg.) during grafting. *Plant Physiol. Biochem.* **168**, 477–487 (2021).
51. Yang, Y. et al. Comprehensive Analysis of ABCA Family Members in Lung Adenocarcinoma with Prognostic Values. *Mol. Biotechnol.* **64**, 1441–1453 (2022).
52. Xiao, H. N., Zhao, Z. Y., Li, J. P. & Li, A. Y. Comprehensive pan-cancer analysis: essential role of ABCB family genes in cancer. *Transl. Cancer Res.* **13**, 1642–1664 (2024).
53. Yang, M. & Brackenbury, W. J. Membrane potential and cancer progression. *Front Physiol.* **4**, 185 (2013).
54. Sempou, E. et al. Membrane potential drives the exit from pluripotency and cell fate commitment via calcium and mTOR. *Nat. Commun.* **13**, 6681 (2022).
55. Bonzanni M., et al. Defined extracellular ionic solutions to study and manipulate the cellular resting membrane potential. *Biol. Open* ;9. <https://doi.org/10.1242/bio.048553> (2020).
56. Odling-Smee F. J., Laland K. N., Feldman M. W. Niche construction : the neglected process in evolution. Princeton: Princeton University Press; 2003. xii, 472 p. p.
57. Laland, K., Matthews, B. & Feldman, M. W. An introduction to niche construction theory. *Evol. Ecol.* **30**, 191–202 (2016).
58. Damaghi, M. et al. Collagen production and niche engineering: A novel strategy for cancer cells to survive acidosis in DCIS and evolve. *Evol. Appl.* **13**, 2689–2703 (2020).
59. Wallrapp, A. et al. The neuropeptide NMU amplifies ILC2-driven allergic lung inflammation. *Nature* **549**, 351–356 (2017).
60. Wu, Y. et al. Neuromedin U is regulated by the metastasis suppressor RhoGDI2 and is a novel promoter of tumor formation, lung metastasis and cancer cachexia. *Oncogene* **26**, 765–773 (2007).
61. Walerych, D., Napoli, M., Collavin, L. & Del Sal, G. The rebel angel: mutant p53 as the driving oncogene in breast cancer. *Carcinogenesis* **33**, 2007–2017 (2012).
62. Sundelacruz, S., Levin, M. & Kaplan, D. L. Role of membrane potential in the regulation of cell proliferation and differentiation. *Stem Cell Rev. Rep.* **5**, 231–246 (2009).

63. Sundelacruz, S., Levin, M. & Kaplan, D. L. Membrane potential controls adipogenic and osteogenic differentiation of mesenchymal stem cells. *PLoS One* **3**, e3737 (2008).

### Author contributions

Theoretical development: R.A.G., J.S.B. Data acquisition and modeling, J.K.T. Data analysis: J.K.T., R.A.G., K.Y.T., J.S.B. Manuscript preparation and editing: R.A.G., J.S.B., K.Y.T., J.K.T.

### Competing interests

A patent has been issued to Moffitt Cancer Center on behalf of JKT for the Negative Storage Model used to analyze data in this manuscript. The other authors have no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-08123-7>.

**Correspondence** and requests for materials should be addressed to Robert A. Gatenby.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Christina Karlsson Rosenthal.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025