Relationship between stemness and transcriptionally-inferred PI3K activity in human breast cancer

Ralitsa R. Madsen[1,*], Oscar M. Rueda[3,4,5], Xavier Robin[6], Carlos Caldas[3,4,5], Robert K. Semple[2,a], Bart Vanhaesebroeck[1,a,*]

[1]University College London Cancer Institute, Paul O'Gorman Building, University College London, London, UK.
[2]Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK.
[3]Cancer Research UK Cambridge Institute and Department of Oncology, Li Ka Shing Centre, University of Cambridge, Cambridge, UK
[4]Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust, Cambridge, UK
[5]NIHR Cambridge Biomedical Research Centre and Cambridge Experimental Cancer Medicine Centre, Cambridge University Hospital NHS Foundation Trust, Cambridge, UK
[6]SIB Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50–70, CH-4056 Basel, Switzerland

[a]These authors contributed equally to this work.

*Corresponding authors: Ralitsa R. Madsen (R.R.M.), Bart Vanhaesebroeck (B.V.)

Email: r.madsen@ucl.ac.uk (R.R.M.); bart.vanh@ucl.ac.uk (B.V.)
0000-0001-8844-5167 (R.R.M.)
0000-0002-7074-3673 (B.V.)

Keywords
PI3K, breast cancer, systems medicine, stemness, *PIK3CA*

Author Contributions
Overall conceptualisation and study design by R.R.M., under supervision from B.V., R.K.S. and O.R. O.R and X.R. reviewed the bioinformatic code. R.R.M., B.V. and R.K.S. wrote the manuscript. All authors reviewed and edited the final version.

This PDF file includes:
- Main Text
- Figure 1
- Figure 2
- Figure 3
- Figure 4

ABSTRACT

The development of reliable, prognostically informative molecular tests to direct targeted cancer therapy is a major challenge. In 2019, the PI3Kα inhibitor alpelisib was approved for the treatment of advanced breast cancer, in combination with the oestrogen receptor degrader fulvestrant, with some evidence for improved therapeutic response in patients classified as having tumours which were positive for mutation in *PIK3CA*, the gene encoding PI3Kα. Using human pluripotent stem cells, we recently demonstrated that the *PIK3CA*^{H1047R} oncogenic hotspot variant shows marked *PIK3CA* allele dose-dependent activation of PI3K signalling and induction of self-sustained stemness. Together with recent discoveries of multi-copy and double-*cis PIK3CA* mutations in human cancers, this calls for a re-evaluation of the *PIK3CA* genotype-phenotype relationship and the current use of binary stratification by *PIK3CA* mutation status. Using computational analyses, we thus investigated the relationship between *PIK3CA* mutational status, PI3K activity/signalling strength and stemness. Stemness and PI3K activity scores were calculated using open-source methods and well-established transcriptional signatures. We report that a high PI3K pathway activity score, but not the presence of *PIK3CA* mutation *per se*, predicts increased breast cancer dedifferentiation and higher stemness, correlating with reduced overall survival. Our data (1) corroborate reports that the presence of a *PIK3CA* mutation per se does not predict high PI3K pathway activation or poor prognosis; (2) suggest that stratification of breast cancer for PI3K-based therapy might benefit from the use of a PI3K pathway activity score rather than binary *PIK3CA* mutation status alone; (3) suggest that combination of PI3K pathway inhibitors with differentiation-promoting treatments warrants evaluation in aggressive breast cancers with high PI3K activity and stemness scores.

INTRODUCTION

Activating mutations in *PIK3CA* are among the most common somatic point mutations in cancer, together with inactivation or loss of the tumour suppressor *PTEN*, a negative regulator of PI3K [1–3]. However, experimental evidence suggests that heterozygous expression of a strongly activating *PIK3CA* mutation is not sufficient on its own to transform cells in vitro or to induce tumourigenesis in vivo (reviewed in Ref. [4]). This is further supported by observations in people with disorders in the *PIK3CA*-related overgrowth spectrum (PROS) which often carry the same *PIK3CA* mutations as found in cancer but feature benign tissue overgrowth without excess risk of adult malignancy [4].

We and others have recently shown that many *PIK3CA*-associated cancers harbour multiple independent mutations activating the PI3K pathway, including multiple *PIK3CA* mutations in cis or trans [3,5–8]. We further reported that human induced pluripotent stem cells (iPSCs) with two endogenous alleles of the strongly activating cancer "hotspot" mutation *PIK3CA*^{H1047R} exhibit pronounced phenotypic differences to isogenic cells heterozygous for the same *PIK3CA* variant [6]. These differences include partial loss of epithelial morphology, widespread transcriptional reprogramming and self-sustained stemness in vitro and in vivo [6]. Collectively, these genetic and cellular observations suggested that major cellular programmes implicated in cancer maintenance and progression may be exquisitely sensitive to the strength of the pathological PI3K signal.

As well as being a model for normal early development, PSCs share key characteristics with cancer cells, including developmental plasticity, the capacity for indefinite self-renewal, rapid proliferation and high glycolytic flux [9]. Indeed, oncogenesis commonly features aberrant reactivation of primitive embryonic and tissue repair pathways [10–12], while accumulating evidence suggests that tumour heterogeneity and therapeutic resistance is determined in part by cancer stem cells [13]. Seminal genome-wide studies have demonstrated a striking enrichment for embryonic gene signatures in human cancers, indicative of tumour dedifferentiation and aggressive disease [12,14]. *PIK3CA*^{H1047R} (over)expression or overexpression of wild-type *PIK3CA* have previously been linked to dedifferentiation and stemness in mouse cancer models [15–21], particularly of the breast, yet gene dose-dependent regulation by PI3K signalling has not been addressed. Importantly, no systematic profiling has previously investigated whether the link between PI3K and stemness extends to human cancer; and if it does exist, whether it may be used to guide patient stratification both for prognostic and therapeutic purposes.

Recently, the PI3Kα-specific inhibitor alpelisib (Piqray/NVP-BYL719; Novartis) received FDA approval for the treatment of advanced hormone-receptor (HR)-positive, HER2-negative breast cancers, following a randomised phase III trial evaluating alpelisib with fulvestrant *versus* fulvestrant alone [22]. The trial concluded that a clinically-relevant benefit of the combination therapy was more likely in patients with *PIK3CA*-mutant tumours [22]. The FDA-approval of alpelisib was accompanied by approval of the companion diagnostic therascreen® *PIK3CA* test (QIAGEN) which detects 11 *PIK3CA* "hotspot" mutations. Despite these advances, a substantial proportion of patients with *PIK3CA*-mutant tumours failed to improve on the combination therapy [22], highlighting the need for further refinements of current patient stratification strategies.

Mutation-centric approaches for stratification of patients with cancer are relatively easy to implement, using well-established technical protocols and analytical pipelines. Such "hard-wired" genomic information nevertheless has limited predictive value for cellular behaviour, which is governed by additional layers of biological complexity and buffering, including transcriptional and translational regulation. Gene expression signatures are used to classify functionally distinct cell types, and "functional genomic tests" are also gaining traction in clinical oncology [23]. This is exemplified by the FDA-approved MammaPrint® 70-gene signature test which is used to aid treatment decisions in early-stage breast cancers. Moreover, computational initiatives like the Broad Institute's Molecular Signatures Database Hallmark Gene Set Collection (mSigDb) have provided the community with "refined" gene sets that can be used to evaluate the activity of major biological pathways based on transcriptional data [24].

In this study, we computed PI3K "activity" scores for breast cancers based on well-established and publicly-available transcriptional signatures. The scores represent inferred PI3K activity, equivalent to a 'record' of past PI3K pathway activation. A similar approach was used to compute transcriptional stemness scores. We next used these scores to evaluate the relationship between PI3K signalling strength, stemness and dedifferentiation in two large breast cancer cohorts (METABRIC and TCGA breast carcinoma). We demonstrate that a high PI3K pathway activity score correlates strongly with cancer dedifferentiation/grade, stemness gene expression and reduced survival. In line with previous studies [25,26], binary classification of samples according to *PIK3CA* genotype resulted in a negative relationship between *PIK3CA* mutant status, PI3K signalling strength and stemness. Further stratification that also took into account *PIK3CA* mutant allele dosage revealed the expected positive correlation with PI3K signalling strength and stemness. Nevertheless, given the multitude of known genetic and non-genetic causes of PI3K hyperactivation in cancer, including complex signalling feedback loops, we conclude that *PIK3CA* mutational status alone is not sufficient to predict PI3K pathway activation and/or stemness. The implications of our findings in the context of cancer therapy are discussed.

RESULTS

Transcriptional PI3K pathway activity in breast cancer is associated with increased stemness and tumour dedifferentiation

In order to determine whether dose-dependent PI3K activation – irrespective of its genetic basis – was linked to stemness in human cancer, we implemented openly available tools to calculate phenotypic scores for PI3K signalling strength and stemness (see Supplementary Data for annotated source code to reproduce all of the following steps).

For the PI3K activity score, we used Gene Set Variation Analysis (GSVA) [27] and the "HALLMARK_PI3K_AKT_MTOR_SIGNALING" gene set from the Broad Institute's Molecular Signature Database (MSigDB). This gene set consists of 105 genes upregulated upon PI3K pathway activation across multiple studies [24] (Supplementary Table 1). To compute a stemness score, we used the PLURINET gene signature (n = 299 genes), developed based on machine learning methods to facilitate robust classification of human pluripotent stem cells [28] (Supplementary Table 2). Of note, only 4 genes were shared between the PI3K activity and stemness gene lists, thus precluding a confounding effect on the relationship between stemness and PI3K activity scores reported below.

We next used breast cancer transcriptomic data to assess correlations among PI3K scores, stemness scores and clinical characteristics/outcomes. We used the METABRIC breast cancer dataset [29] due to its large sample size and high-quality information on cancer grade, a surrogate measure of dedifferentiation and stemness. The PI3K activity score in METABRIC breast tumours correlated significantly with the stemness score (**Fig. 1A**; Spearman's Rho = 0.49, $p < 2.2e-16$) as well as tumour grade and thus dedifferentiation (**Fig. 1B**).

Upon stratification of METABRIC breast cancers into those with "high" and "low" PI3K pathway activity, we found that around 90% of ER-negative tumours but only up to 40% of ER-positive tumours had "high" PI3K activity score (**Fig. 2A**). PI3K activity and stemness scores were highest in the more aggressive PAM50 breast cancer subtypes (**Fig. 2B**), including Basal, HER2 and Luminal B [30]. This contrasts with the known enrichment of *PIK3CA* mutations in the ER-positive tumours, in particular the Luminal A subtype (**Fig. S1A**) (demonstrated previously [26,31]). Several of these findings, including the strong relationship between inferred PI3K activity and stemness scores, were reproduced using available TCGA breast cancer transcriptomic data **(Fig. S1B, S1C)**.

A "high" PI3K activity score, but not mutant *PIK3CA* status, predicts reduced survival in breast cancer

Across all breast cancers, a "high" PI3K activity score was associated with reduced patient survival, with the median survival probability reduced by over 3 years in patients with a high PI3K pathway activity score (**Fig. 3A**). However, no such correlation was seen when patients were grouped by binary *PIK3CA* mutation status (**Fig. 1E**). The prognostic value of the PI3K score remained apparent when tested in ER-positive tumours only (**Fig. S1B**), suggesting it was not confounded by the higher PI3K activity score in the more aggressive ER-negative breast cancer subtypes (**Fig. 2**).

Stratification of breast cancers according to mutant *PIK3CA* allele dosage results in an unexpected biphasic relationship with PI3K activity and stemness scores

Next, using TGCA data and our previous allele copy number data for TCGA tumours [6], we tested whether breast cancers with multiple copies of *PIK3CA* "hotspot" mutations exhibit higher PI3K pathway activity and stemness scores than tumours with a single copy, as predicted by our iPSC findings [6,32]. In the presence of a single oncogenic *PIK3CA* missense variant, we found a paradoxical reduction in transcriptional PI3K pathway activity score, a relationship that was also reflected in lower stemness scores when only one copy of oncogenic *PIK3CA* was present (**Fig. 4A**). In contrast, in breast cancers with multiple oncogenic *PIK3CA* copies, both PI3K and stemness scores increased (**Fig. 4A**). This relationship was lost upon simple binary classification based on *PIK3CA* genotypes (i.e. wild-type *vs* mutant) (**Fig. S1E**). Nearly identical patterns were observed in the METABRIC cohort when we used publicly available allele copy number data from cBioPortal (**Fig. 4B**).

Taken together, these multiple analyses in large human breast cancer datasets provide strong evidence for an association between strong PI3K pathway activation, breast cancer stemness and clinical outcome (visually summarised in **Fig. 4C**), contrasting with the lack of prognostic value found on binary stratification based on *PIK3CA* mutation status. Our findings raise the possibility that stratification of cancers in general according to transcriptional indices of PI3K pathway activity may be of greater prognostic value and more useful in precision therapy than simple stratification by *PIK3CA* mutation status.

DISCUSSION

PIK3CA^{H1047R} is the most common activating *PIK3CA* mutation in human cancers and in PROS, a group of largely benign overgrowth disorders [4]. We recently found that *PIK3CA*-associated cancers often harbour multiple mutated *PIK3CA* copies, and demonstrated that homozygosity but not heterozygosity for *PIK3CA*^{H1047R} leads to self-

sustained stemness in iPSCs [6]. Here, we use computational analyses of large human breast cancer datasets to demonstrate a strong, positive relationship between the transcriptionally-inferred PI3K pathway activity, stemness gene expression and tumour grade. Importantly, we show that stratification of breast tumours according to single *vs* multiple copies of "hotspot" *PIK3CA* mutations results in subgroups with "low" and "high" PI3K activity scores, respectively. This is not observed upon binary classification into *PIK3CA* mutant and *PIK3CA* wild-type cancers, due to a paradoxical decrease in transcriptional PI3K pathway activity conferred by single *PIK3CA* mutations. This agrees with the negative relationship between *PIK3CA* mutant status and indices of transcriptomic PI3K pathway activity previously reported [25,26].

The apparent biphasic relationship between single *versus* multiple copies of *PIK3CA* mutation and stemness scores warrants further study, but is likely to reflect poorly understood regulatory aspects of intracellular signalling networks. We therefore caution against the use of a *PIK3CA*-mutant-centric approach to predict PI3K pathway activity, given that numerous alternative genetic changes – including *PIK3CA* amplification, *PTEN* and *INPP4B* loss – may converge on increased PI3K pathway activation [3,26,33,34]. These will be captured by the PI3K activity signature used in our study. Indeed, the aggressive basal breast cancer subtype, despite showing a relative lack of *PIK3CA* mutations, exhibits some of the highest PI3K activity and stemness scores (**Fig. S1A**). Although their mutual relationships has not been addressed previously, indices for PI3K activity and stemness have separately been associated with the basal breast cancer subtype in previous studies [11,12,26,33]. Moreover PI3K signalling was recently shown to promote stem cell-like traits in basal-like breast cancers [35], consistent with the notion that the aggressive nature of this cancer subtype is driven by a stem cell component [36].

In a separate study of iPSCs with heterozygous and homozygous *PIK3CA*^{H1047R} expression, BYL719 (alpelisib; Novartis) failed to reverse the increased stemness gene expression in homozygous *PIK3CA*^{H1047R} iPSCs [32]. This PI3Kα-selective inhibitor was recently approved for use in combination with anti-estrogen therapy in ER-positive breast cancers. In a randomised phase 3 trial that compared BYL719 with fulvestrant to fulvestrant alone, increased progression-free survival was seen in 26.6 % of patients with *PIK3CA*-mutant tumours vs 12.8 % of those without a *PIK3CA* variant [22]. This demonstrates the utility of *PIK3CA*-centric stratification, yet a substantial proportion of patients with *PIK3CA*-mutant tumours did not benefit from the BYL719 and fulvestrant combination [22]. Although recent studies demonstrated that double *PIK3CA* mutations in *cis* confer greater sensitivity to PI3Kα inhibition, these studies focused on inhibition of proliferation/growth [7,8]. Such parameters do not strictly correlate with dedifferentiation or 'stemness' which are more strongly linked with metastasis and ultimately death. This is exemplified in a recent pre-clinical study of ER-negative breast cancer models in which rapalogs effectively blunted primary tumour growth but failed to reduce the number of lung metastases [37]. Moreover, rapalog resistance was linked to stem cell-like features in both ER-positive and ER-negative breast cancer cell lines [37].

Our results raise the possibility that tumours with high PI3K activity score may respond well to combined inhibition of the PI3K pathway and reversal of the stemness phenotype. A recent RNA-interference screen in breast cancer cell lines revealed several potential pro-differentiation agents, including the BET bromodomain inhibitor JQ1 [38]. The bromodomain 9 (BRD9) subunit of the SWI-SNF chromatin-remodelling complex has also been implicated as a driver of the high MYC transcriptional signature in *PIK3CA*^{H1047R}/*KRAS*^{G12V} double knock-in breast epithelial cells relative to single-mutant counterparts [39]. Our data in human iPSCs suggest TGFβ pathway inhibition as yet another strategy in the context of strong PI3K pathway activation [32]. Assessing efficacy of such treatments will, however, require monitoring of phenotypes beyond bulk tumour growth, such as stemness and metastatic potential.

Given a previous study which found a PI3K/AKT/mTOR gene signature score to correlate negatively with pan-cancer survival [25], dual PI3K-stemness score assessment may also be prognostically useful beyond breast cancer. Our computational method to infer phenotypic scores is applicable to other cancer contexts given sufficient transcriptomic data (N>30 tumour samples). It is, however, important to emphasise that our approach is unlikely to report instantaneous PI3K pathway activity, which is better assessed, for example, by phosphoprotein-based analyses [25]. By focusing on the transcriptome, we instead infer PI3K activity over time, as recorded in wider gene expression changes. We also note that a limitation of the current approach is the use of bulk transcriptomic data which fail to report on the heterogeneity in single-cell gene expression within a tumour sample.

Finally, the proposed stratification according to PI3K pathway activity does not require expensive transcriptomic studies; instead, we suggest development of a simple qPCR "scorecard" that captures the hallmark PI3K gene signature and key stemness signature genes in one assay (see SI Appendix for a proposed outline of such as *"PI3K-stemness scorecard"*), akin to multigene panels such as Oncotype Dx® and MammaPrint® for use in breast cancer [23]. This proposal will require evaluation of sensitivity and specificity and long-term assessment but may ultimately help the clinical translation of PI3K-based therapies, an effort which has turned out to be very challenging to date.

MATERIALS AND METHODS

METABRIC and TCGA transcriptomic data access and pre-processing

Normalised microarray-based gene expression for METABRIC breast tumour samples were obtained from Curtis et al. [29], and clinical data from Rueda et al. [40]. The relevant METABRIC mutation data were downloaded from cBioPortal in January (mutation-only) and March (mutation and copy number) 2020 [41]. TCGA breast invasive carcinoma (BRCA) RNAseq, mutational and clinical data were retrieved from the GDC server (using the legacy database) using the TCGAbiolinks package [42]. This package was also used for subsequent quantile filtering (quantile value = 0.4) of lowly-expressed gene and removal of tumour samples with low purity (cpe = 0.6). The resulting raw RSEM counts were normalised with the TMM method [43] and log2-transformed using the voom() function in the limma package prior to downstream use in GSVA computations. The TCGA BRCA mutation data with available copy number estimates for individual mutations were obtained from Madsen et al. [6].

Gene signature analyses of METABRIC and TCGA breast cancer cohorts

For a detailed workflow of all computational steps, the reader is referred to the annotated RNotebook provided on the accompanying OSF project page (doi:10.17605/OSF.IO/G8RF3). All computational analyses were performed using the R software. Briefly, normalised microarray-based gene expression for METABRIC breast tumour samples were obtained from Curtis et al. [29], and clinical data from Rueda et al. [40]. The relevant METABRIC mutation data were downloaded from cBioPortal in January (mutation-only) and March (mutation and copy number) 2020 [41]. TCGA breast invasive carcinoma (BRCA) RNAseq, mutational and clinical data were retrieved from the GDC server (using the legacy database) using the TCGAbiolinks package [42]. This package was also used for subsequent quantile filtering (quantile value = 0.4) of lowly-expressed gene and removal of tumour samples with low purity (cpe = 0.6). The resulting raw RSEM counts were normalised with the TMM method [43] and log2-transformed using the voom() function in the limma package prior to downstream use in GSVA computations. The TCGA BRCA mutation data with available copy number estimates for individual mutations were obtained from Madsen et al. [6].

The “HALLMARK_PI3K_AKT_MTOR_SIGNALING” (PI3K pathway activity) and “MUELLER_PLURINET” (“stemness”) signatures were retrieved from The Molecular Signature Database (MSigDB). We note that the “HALLMARK_PI3K_AKT_MTOR_SIGNALING” gene set also includes mTORC1-dependent gene expression changes, in contrast to other studies which have sought to separate AKT- and mTORC1-driven gene expression changes [44,45]. Individual signature scores were computed with the GSVA package, using the default Gaussian kernel and selecting ESdiff enrichment values as output [27]. The correlation between PI3K pathway activity and “stemness” scores in the two breast cancer cohorts was computed using Spearman’s rank correlation. Binary classification of PI3K pathway activity scores into “low” and “high” used the 0.25 and 0.75 score quantiles, respectively. The scores were next assessed in the context of available clinical and/or genetic attributes (*PIK3CA* missense mutations only). *PIK3CA* “hotspot” mutations (C420R, E542K, E545K, H1047L, H1047R) were defined according to their prevalence in cancer and evidence for severe functional impact in developmental overgrowth disorders [4,46].

Statistical analyses

The lm() function in R was used to fit linear models to METABRIC tumour grade and gene signature score data; the F statistic of each model and the accompanying R2 values are reported within the respective plots. Of note, the normality assumption was violated for the stemness versus grade linear model; however, as the sample size is large and as the assumption terms in the specified models are formed as a sum of several other independent quantities, this violation is expected to have a minimal impact on the reported results. Tukey’s Honest Significant Differences method was used to test for statistically significant (adjusted p-value < 0.05) differences in score means across different tumour grades or stages. Wilcoxon pairwise-comparison with Benjamini-Hochberg correction was used to assess differences in PI3K and stemness scores across tumours stratified according to *PIK3CA* copy number status. The relationship between PI3K pathway activity score and survival in METABRIC was assessed using a non-parametric log-rank test. Differences in score distributions across tumour subtypes and/or genotypes were assessed using a Chi-squared goodness-of-fit test.

Data and materials availability

Raw data and bespoke RNotebooks containing guided scripts and plots are available via the Open Science Framework (doi: 10.17605/OSF.IO/G8RF3). Individual scripts include information on the name and version of

applied R packages. Further information requests should be directed to and will be fulfilled by the corresponding authors, Ralitsa R. Madsen (r.madsen@ucl.ac.uk) and Bart Vanhaesebroeck (bart.vanh@ucl.ac.uk).

Funding

Work in the laboratory of B.V. is supported by Cancer Research UK (C23338/A25722) and the UK NIHR University College London Hospitals Biomedical Research Centre. R.K.S. is supported by the Wellcome Trust (105371/Z/14/Z, 210752/Z/18/Z). O.M.R. and C.C. are supported by Cancer Research UK.

Acknowledgements

We are grateful to Dr Neil Vasan (Memorial Sloan Kettering Cancer Center, New York) for excellent feedback on the manuscript.

Competing interests

R.K.S. is a consultant for HotSpot Therapeutics (Boston, MA, USA); B.V. is a consultant for Karus Therapeutics (Oxford, UK), iOnctura (Geneva, Switzerland) and Venthera (Palo Alto, CA, USA) and has received speaker fees from Gilead Sciences (Foster City, US).

315 **References**

316 1 Chang, M. T. et al. (2015) Identifying recurrent mutations in cancer reveals widespread lineage
317 diversity and mutational specificity. Nat. Biotechnol. **34**, 155–163.
318 2 Campbell, P. J. et al. (2020) Pan-cancer analysis of whole genomes. Nature **578**, 82–93.
319 3 Sanchez-Vega, F. et al. (2018) Oncogenic Signaling Pathways in The Cancer Genome Atlas.
320 Cell **173**, 321-337.e10.
321 4 Madsen, R. R. et al. (2018) Cancer-Associated PIK3CA Mutations in Overgrowth Disorders.
322 Trends Mol. Med. **24**, 856–870.
323 5 Yuan, T. L. and Cantley, L. C. (2008) PI3K pathway alterations in cancer: variations on a
324 theme. Oncogene **27**, 5497–5510.
325 6 Madsen, R. R. et al. (2019) Oncogenic PIK3CA promotes cellular stemness in an allele dose-
326 dependent manner. Proc. Natl. Acad. Sci. **116**, 8380–8389.
327 7 Vasan, N. et al. (2019) Double PIK3CA mutations in cis increase oncogenicity and sensitivity
328 to PI3Kα inhibitors. Science. **366**, 714–723.
329 8 Saito, Y. et al. (2020) Landscape and function of multiple mutations within individual
330 oncogenes. Nature **582**, 95–99.
331 9 Intlekofer, A. M. and Finley, L. W. S. (2019) Metabolic signatures of cancer cells and stem
332 cells. Nat. Metab. **1**, 177–188.
333 10 Lineweaver, C. H. et al. (2014) Targeting cancer's weaknesses (not its strengths): Therapeutic
334 strategies suggested by the atavistic model. BioEssays **36**, 827–835.
335 11 Malta, T. M. et al. (2018) Machine Learning Identifies Stemness Features Associated with
336 Oncogenic Dedifferentiation. Cell **173**, 338-354.e15.
337 12 Ben-Porath, I. et al. (2008) An embryonic stem cell-like gene expression signature in poorly
338 differentiated aggressive human tumors. Nat. Genet. **40**, 499–507.
339 13 Shibue, T. and Weinberg, R. A. (2017) EMT, CSCs, and drug resistance: The mechanistic link
340 and clinical implications. Nat. Rev. Clin. Oncol. **14**, 611–629.
341 14 Malta, T. M. et al. (2018) Machine Learning Identifies Stemness Features Associated with
342 Oncogenic Dedifferentiation. Cell **173**, 338-354.e15.
343 15 Van Keymeulen, A. et al. (2015) Reactivation of multipotency by oncogenic PIK3CA induces
344 breast tumour heterogeneity. Nature **525**, 119–23.
345 16 Koren, S. et al. (2015) PIK3CA(H1047R) induces multipotency and multi-lineage mammary
346 tumours. Nature **525**, 114–8.
347 17 Hanker, A. B. et al. (2013) Mutant PIK3CA accelerates HER2-driven transgenic mammary
348 tumors and induces resistance to combinations of anti-HER2 therapies. Proc. Natl. Acad. Sci.
349 U. S. A. **110**, 14372–14377.
350 18 van Veen, J. E. et al. (2019) Mutationally-activated PI3'-kinase-α promotes de-differentiation of
351 lung tumors initiated by the BRAFV600E oncoprotein kinase. Elife **8**, 1–33.
352 19 Riemer, P. et al. (2017) Oncogenic β-catenin and PIK3CA instruct network states and cancer
353 phenotypes in intestinal organoids. J. Cell Biol. **216**, 1567–1577.
354 20 Du, L. et al. (2016) Overexpression of PIK3CA in murine head and neck epithelium drives
355 tumor invasion and metastasis through PDK1 and enhanced TGFβ signaling. Oncogene **35**,
356 4641–4652.
357 21 Meyer, D. S. et al. (2011) Luminal expression of PIK3CA mutant H1047R in the mammary
358 gland induces heterogeneous tumors. Cancer Res. **71**, 4344–4351.
359 22 André, F. et al. (2019) Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced
360 breast cancer. N. Engl. J. Med. **380**, 1929–1940.
361 23 Giuliano, A. E. et al. (2017) Breast Cancer-Major changes in the American Joint Committee on
362 Cancer eighth edition cancer staging manual. CA. Cancer J. Clin. **67**, 290–303.
363 24 Liberzon, A. et al. (2015) The Molecular Signatures Database Hallmark Gene Set Collection.
364 Cell Syst. **1**, 417–425.
365 25 Zhang, Y. et al. (2017) A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway
366 Alterations. Cancer Cell **31**, 820-832.e3.
367 26 Koboldt, D. C. et al. (2012) Comprehensive molecular portraits of human breast tumours.
368 Nature **490**, 61–70.
369 27 Hänzelmann, S. et al. (2013) GSVA: Gene set variation analysis for microarray and RNA-Seq
370 data. BMC Bioinformatics **14**.
371 28 Müller, F. J. et al. (2008) Regulatory networks define phenotypic classes of human stem cell
372 lines. Nature **455**, 401–405.
373 29 Curtis, C. et al. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours

374 reveals novel subgroups. Nature **486**, 346–352.
375 30 Bernard, P. S. et al. (2009) Supervised risk predictor of breast cancer based on intrinsic
376 subtypes. J. Clin. Oncol. **27**, 1160–1167.
377 31 Stemke-Hale, K. et al. (2008) An integrative genomic and proteomic analysis of PIK3CA,
378 PTEN, and AKT mutations in breast cancer. Cancer Res. **68**, 6084–6091.
379 32 Madsen, R. R. et al. (2019) High-dose oncogenic PIK3CA drives constitutive cellular stemness
380 through self-sustained TGFβ pathway activation. bioRxiv (preprint).
381 33 López-Knowles, E. et al. (2010) PI3K pathway activation in breast cancer is associated with
382 the basal-like phenotype and cancer-specific mortality. Int. J. Cancer **126**, 1121–1131.
383 34 Liu, H. et al. (2020) The INPP4B Tumor Suppressor Modulates EGFR Trafficking and
384 Promotes Triple Negative Breast Cancer. Cancer Discov. CD-19-1262.
385 35 Thomas, C. et al. (2017) Pentraxin-3 is a PI3K signaling target that promotes stem cell–like
386 traits in basal-like breast cancers. Sci. Signal. **10**, eaah4674.
387 36 Sridharan, S. et al. (2019) Novel and Alternative Targets Against Breast Cancer Stemness to
388 Combat Chemoresistance. Front. Oncol. **9**, 1–19.
389 37 Mateo, F. et al. (2017) Stem cell-like transcriptional reprogramming mediates metastatic
390 resistance to mTOR inhibition. Oncogene **36**, 2737–2749.
391 38 Arfaoui, A. et al. (2019) A genome-wide RNAi screen reveals essential therapeutic targets of
392 breast cancer stem cells. EMBO Mol. Med. **11**.
393 39 Bell, C. M. et al. (2019) PIK3CA Cooperates with KRAS to Promote MYC Activity and
394 Tumorigenesis via the Bromodomain Protein BRD9. Cancers (Basel). **11**, 1634.
395 40 Rueda, O. M. et al. (2019) Dynamics of breast-cancer relapse reveal late-recurring ER-positive
396 genomic subgroups. Nature **567**, 399–404.
397 41 Cerami, E. et al. (2012) The cBio Cancer Genomics Portal: An open platform for exploring
398 multidimensional cancer genomics data. Cancer Discov. **2**, 401–404.
399 42 Colaprico, A. et al. (2016) TCGAbiolinks: An R/Bioconductor package for integrative analysis
400 of TCGA data. Nucleic Acids Res. **44**, e71.
401 43 Robinson, M. D. and Oshlack, A. (2010) A scaling normalization method for differential
402 expression analysis of RNA-seq data. Genome Biol. **11**, R25.
403 44 Creighton, C. J. (2007) A gene transcription signature of the Akt/mTOR pathway in clinical
404 breast tumors. Oncogene **26**, 4648–4655.
405 45 Sonnenblick, A. et al. (2019) pAKT pathway activation is associated with PIK3CA mutations
406 and good prognosis in luminal breast cancer in contrast to p-mTOR pathway activation. Breast
407 Cancer **5**, 1–9.
408 46 Mirzaa, G. et al. (2016) PIK3CA-associated developmental disorders exhibit distinct classes of
409 mutations with variable expression and tissue distribution. JCI Insight **1**, 1–18.
410
411
412

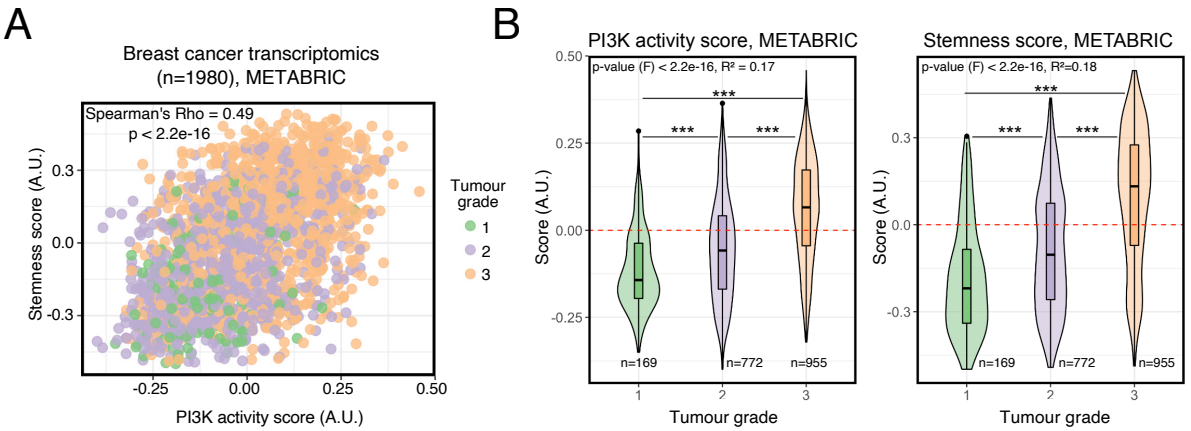

Fig. 1. Inferred PI3K pathway activation correlates with increased stemness and breast tumour dedifferentiation. (A) Correlation plot of PI3K activity and stemness scores, based on METABRIC breast cancer transcriptomes. The scores were determined using Gene Set Variation Analysis (GSVA) with mSigDb "HALLMARK_PI3K_AKT_MTOR_SIGNALING" and "MUELLER_PLURINET" gene signatures [24,27,28]. The gene lists used are included in Supplementary Tables 1 and 2. **(B)** PI3K activity and stemness score distributions across breast cancer grade (METABRIC). *** $p \leq 0.001$ according to Tukey's Honest Significant Differences method. The global p-value for each linear model is indicated within each plot. See also Fig. S1A, B.

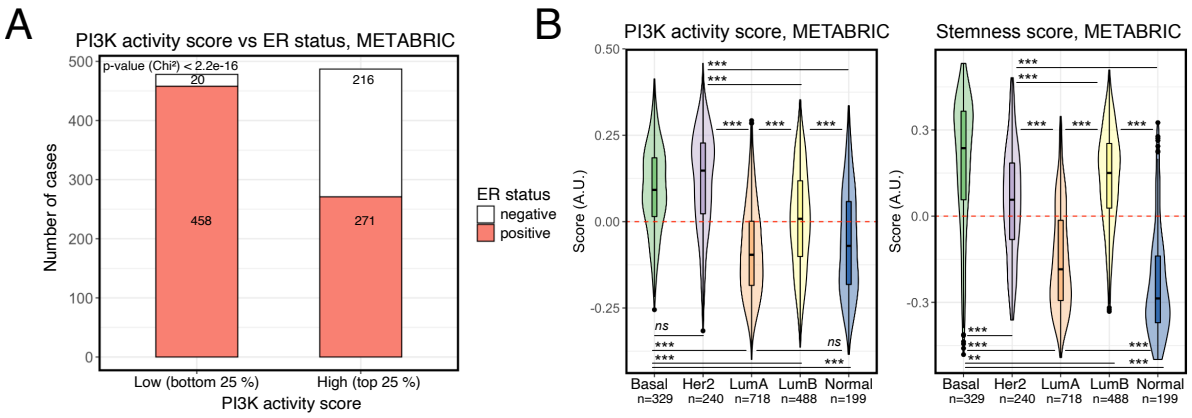

Fig. 2. High PI3K activity and stemness scores are enriched for in aggressive breast cancer subtypes. (A) PI3K activity score distribution in METABRIC breast tumours stratified according to ER status. **(B)** PI3K activity and stemness score distributions across METABRIC breast cancers stratified according to PAM50 subtype; ** $p \leq 0.01$, *** $p \leq 0.001$ according to Tukey's Honest Significant Differences method; *ns*: non-significant. See also Fig. S1C.

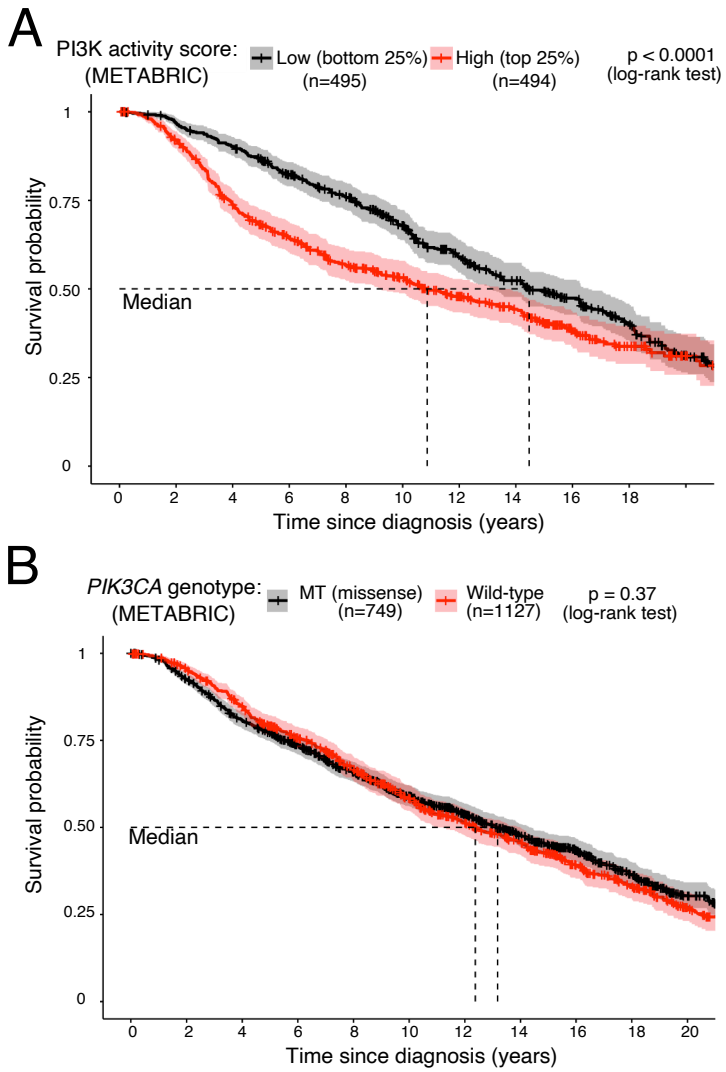

Fig. 3. PI3K activity score, but not *PIK3CA* genotype, is prognostic in breast cancer. (A) Kaplan-Meier plot of overall survival of patients stratified according to PI3K activity score (all METABRIC breast cancers). **(B)** Kaplan-Meier plot of overall survival of patients stratified according to the presence/absence of *PIK3CA* missense mutations (all METABRIC breast cancers). See also Fig. S1D.

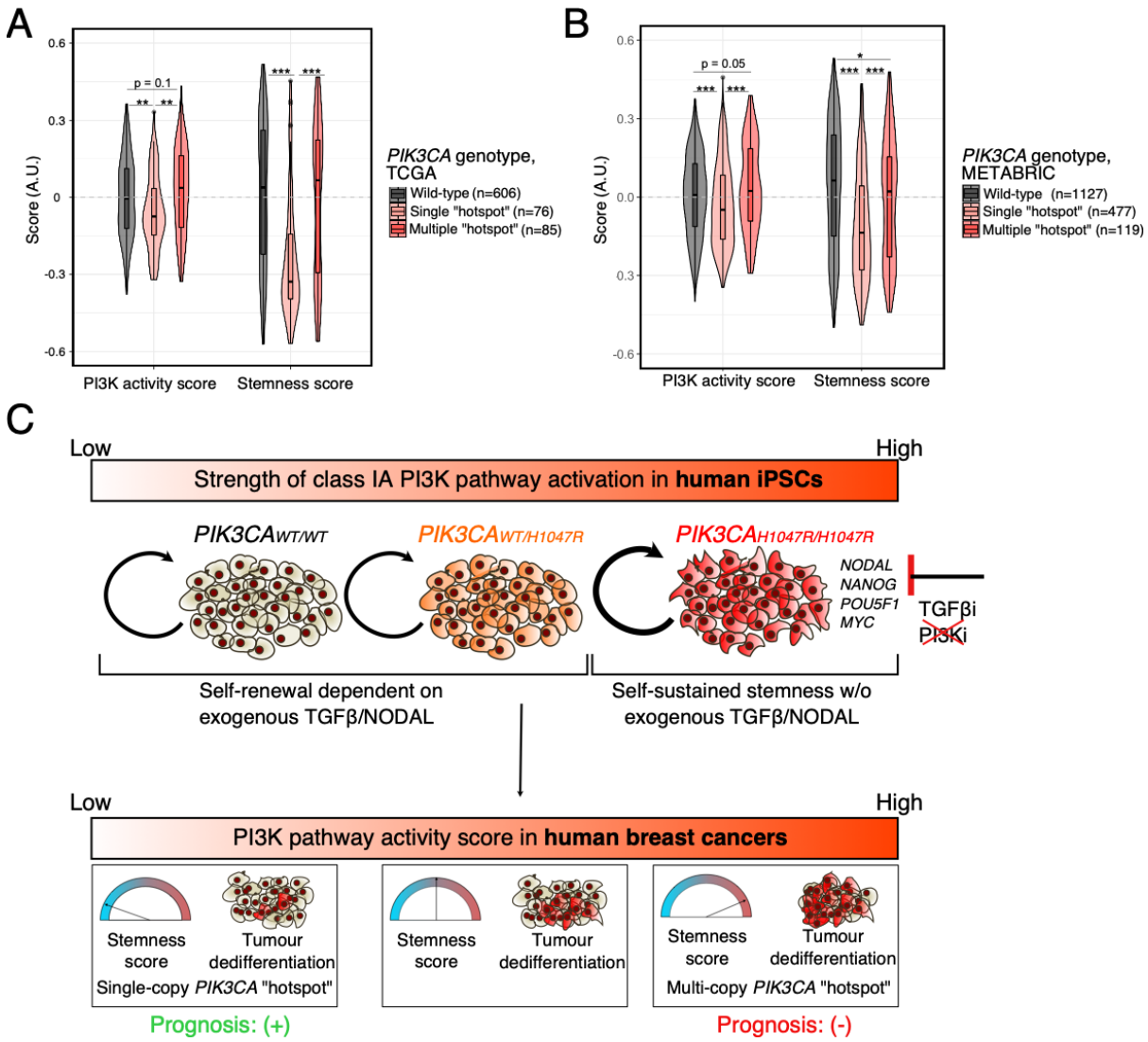

Fig. 4. A biphasic relationship between mutant *PIK3CA* allele dosage and PI3K activity / stemness scores **(A)** PI3K activity and stemness score distributions across TCGA breast cancers following stratification according to the presence or absence of single *vs* multiple copies of the *PIK3CA* "hotspot" alleles (C420R, E542K, E545K, H1047L, H1047R); * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ according to Wilcoxon pairwise-comparison with Benjamini-Hochberg correction. **(B)** As in (A) but performed using METABRIC breast cancer data. **(C)** Graphical summary of key conclusions. The current study builds on previous findings of increased stemness in human induced pluripotent stem cells (iPSCs) with homozygous but not heterozygous *PIK3CA*^{H1047R} expression [6]. We extend these findings to breast cancer and demonstrate that a high transcriptomic PI3K pathway activity score associates strongly with high stemness score, tumour dedifferentiation and reduced survival.