bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

High-dose oncogenic *PIK3CA* drives constitutive cellular stemness through self-sustained TGFβ pathway activation

Ralitsa R. Madsen[1,2,3,*], James Longden[4,5], Rachel G. Knox[2,3], Xavier Robin[4], Franziska Völlmy[4], Kenneth G. Macleod[6], Larissa Moniz[7], Neil O. Carragher[6], Nicholas McGranahan[7], Rune Linding[4,5], Bart Vanhaesebroeck[7], Robert K. Semple[1,2,3,*]

[1]Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK.
[2]Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge, Cambridge, UK.
[3]The National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge, UK.
[4]Biotech Research and Innovation Centre, University of Copenhagen, Copenhagen, Denmark.
[5]Humboldt-Universität zu Berlin, Berlin, Germany.
[6]Edinburgh Cancer Research UK Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh, UK.
[7]University College London Cancer Institute, Paul O'Gorman Building, University College London, London, UK.

*To whom correspondence should be addressed: rmadsen@ed.ac.uk or rsemple@ed.ac.uk

One-sentence summary: Biallelic genetic PI3Kα activation rewires signaling and induces constitutive stemness downstream from self-sustained TGFβ pathway activation.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Oncogenic *PIK3CA* mutations activate phosphoinositide 3-kinase-alpha (PI3Kα) and are among the commonest somatic mutations in cancer. We recently demonstrated that the "hotspot" variant *PIK3CA*^{H1047R} exerts striking allele dose-dependent effects on stemness in human pluripotent stem cells (hPSCs), and found multiple oncogenic *PIK3CA* copies in a substantial proportion of human cancers. This suggested that the consequences of oncogenic PI3K signaling may differ according to the strength of genetic *PIK3CA* activation. Here, to identify the stemness-promoting mechanism, we profiled isogenic wild-type, *PIK3CA*$^{WT/H1047R}$ and *PIK3CA*$^{H1047R/H1047R}$ iPSCs by high-depth transcriptomics, proteomics and reverse-phase protein arrays (RPPA). We report that the phenotypic switch in homozygous *PIK3CA*^{H1047R} hPSCs occurs downstream of signaling "rewiring" towards self-sustained TGFβ pathway activation and increased *NODAL* expression, which was no longer reversible by pharmacological PI3Kα inhibition. Gene expression analysis of *PIK3CA*-associated human breast cancers in The Cancer Genome Atlas revealed increased expression of *NODAL* according to tumor stage and *PIK3CA*^{H1047R} allele dosage. Together with the emerging link between NODAL re-expression and cancer aggressiveness, our data suggest that TGFβ pathway inhibitors warrant investigation in breast tumors stratified by *PIK3CA*^{H1047R} allele dosage.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Class IA phosphoinositide 3-kinases (PI3Ks) are evolutionarily conserved enzymes that catalyze the formation of the membrane-bound second lipid messenger phosphatidylinositol-3,4,5-trisphosphate (PIP_3). PI3Ks are activated downstream of many receptor tyrosine kinases, with the ensuing increase in PIP_3 and its byproduct $PI(3,4)P_2$ activating the serine/threonine kinases AKT and mTORC1. PI3K activation is best known for promoting cell survival, glucose uptake, anabolic metabolism, cell proliferation and cell migration (*1*). Among the class IA PI3K isoforms (PI3Kα, PI3Kβ, PI3Kδ), the ubiquitously-expressed PI3Kα, encoded by the *PIK3CA* gene in humans, is the main regulator of organismal growth, development and survival (*2*).

Activating mutations in *PIK3CA* are among the most common somatic point mutations in cancer, together with alterations in the tumor suppressors *PTEN* and *TP53* (cBioPortal, accessed August 2019) (*3*). Many *PIK3CA*-associated cancers harbor multiple cancer-promoting mutations within the pathway (*4*, *5*), consistent with experimental evidence that heterozygous expression of a strongly activating *PIK3CA* mutation alone is not sufficient to trigger cellular transformation in vitro or to accelerate tumorigenesis in vivo (reviewed in Ref. (*6*)). This is further supported by observations of people with disorders in the *PIK3CA*-related overgrowth spectrum (PROS). These are caused by postzygotic acquisition of an activating *PIK3CA* mutation, and are not known to be associated with excess adult malignancy (*6*).

We recently reported that human induced pluripotent stem cells (iPSCs) with homozygous knock-in of the cancer "hotspot" mutation $PIK3CA^{H1047R}$ exhibit striking phenotypic differences to isogenic cells that are heterozygous for the same variant (*5*). This suggested that major cellular consequences of genetic PI3K activation are exquisitely sensitive to the strength of the pathological PI3K signal in the context of endogenous expression. Human iPSCs, like embryonic stem cells from the inner cell mass of the pre-implantation blastocyst (collectively human pluripotent stem cells, hPSCs), can differentiate into derivatives of all three embryonic germ layers (*7*). They may also indefinitely self-renew (*7*), although this requires carefully controlled culture conditions (*8*). In the presence of homozygous $PIK3CA^{H1047R}$ expression, however, these cells exhibited robust, self-sustained stemness in vitro and in vivo (*5*).

The core hPSC pluripotency gene regulatory network features a feedforward, autoregulatory circuit comprising three transcription factors, namely SRY box 2 (SOX2), Octamer-binding transcription factor 3/4

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(OCT3/4; encoded by *POU5F1*), and the homeobox transcription factor NANOG (*9–11*). SOX2 helps sustain OCT3/4 expression, which is required for establishment and maintenance of the pluripotent state (*12*). However, even modest overexpression of OCT3/4 destabilizes the pluripotency network and triggers differentiation (*13*, *14*). In contrast, NANOG, while dispensable for maintenance of pluripotency (*15*), stabilizes the pluripotency gene regulatory network. Overexpression of NANOG by as little as 1.5-fold leads to sustained self-renewal (or "stemness") of murine and human PSCs (*16–19*). In hPSCs, *NANOG* expression is activated by the transcription factors SMAD2/3 (*20*), which in turn are activated by receptors binding TGFβ, Activin or NODAL (*21*). Overexpression of NODAL thus results in self-sustained stemness of hPSCs even in differentiation-promoting conditions (*22*, *23*).

NODAL expression is not only critical in embryogenesis, but is also more rarely seen in adult organs during extensive tissue remodeling, as in placenta, endometrium and the lactating mammary gland (*24*). There has also been growing awareness in recent years that aggressive cancers often re-express NODAL, "hijacking" its ability to promote cellular plasticity and epithelial-mesenchymal-transition (EMT), and recapitulating in a corrupted form some of its effects in embryonic stem cells (*24*). However, the mechanism(s) driving *NODAL* re-expression in neoplastic conditions and a possible relationship to known oncogenes remain unclear.

In this study, we used transcriptomics, proteomics and reverse phase protein arrays (RPPA) to investigate further the PI3K signaling threshold revealed by our prior studies, confirming near-binary changes in gene and protein expression around a sharp PI3K activity threshold defined by heterozygosity and homozygosity for the same strongly activating *PIK3CA* allele. We further demonstrate that the stemness phenotype of $PIK3CA^{H1047R/H1047R}$ iPSCs is maintained by self-sustained TGFβ signaling, in line with increased *PIK3CA*-mediated *NODAL* expression. Finally, we show that the allele dose-dependent link between $PIK3CA^{H1047R}$, *NODAL* expression and stemness applies to human breast cancers. By deploying pathway and network topology analysis and algorithms to identify the signaling rewiring underlying the 'locked' stemness induced by increased $PIK3CA^{H1047R}$ allele dosage, we raise the possibility that pharmacological targeting of TGFβ signaling in breast cancers stratified by $PIK3CA^{H1047R}$ allele dosage may limit cancer stem cell emergence and be a valuable therapeutic strategy.

Results

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A sharp PI3K activity threshold determines gene expression changes in *PIK3CA*^{H1047R} iPSCs

We previously generated isogenic human iPSCs with heterozygous or homozygous knock-in of the "hotspot" *PIK3CA*^{H1047R} mutation. Surprisingly, in heterozygous cells we found few discernible phenotypic changes and few differentially expressed protein-coding transcripts, despite the strong associations of *PIK3CA*^{H1047R} heterozygosity with cancer and PROS. In contrast, homozygous *PIK3CA*$^{H1047R/H1047R}$ cells exhibited marked morphological changes and altered gene expression, with strong enrichment for cancer-associated pathways (*5*).

To substantiate the apparent PI3K activity threshold manifest in *PIK3CA*^{H1047R}-driven gene expression changes, we undertook RNA sequencing at substantially greater depth, increasing the sample size to four previously unstudied, independently-derived iPSC cultures for each *PIK3CA* genotype. Homozygous mutant cells clearly separated from heterozygous and wild-type cells, which overlapped with each other on multidimensional scaling (**Fig. 1A**). Consistent with improved statistical power, we now detected reduced levels of 451 transcripts and increased levels of 710 transcripts in *PIK3CA*$^{WT/H1047R}$ hPSCs (**Fig. 1B**). This dropped to 149 and 343 transcripts, respectively, after applying an absolute fold-change cut-off of minimum 1.3 (**Fig. 1B and table S1**), consistent with the small magnitude of significant expression changes (**fig. S1A**). Use of the same cut-off, in sharp distinction, yielded 2873 and 2771 transcripts of decreased or increased abundance, respectively, in homozygous iPSC mutants (**Fig. 1B and table S2**). Not only was the number of gene expression changes higher by an order of magnitude in homozygous cells, but many expression changes were also large compared to wild-type controls (**fig. S1A**). The fold-changes in gene expression seen in *PIK3CA*$^{H1047R/H1047R}$ cells were highly correlated with our previously reported findings (Spearman's rho = 0.74) (**fig. S1B**), whereas the correlation was low (Spearman's rho = 0.1) for *PIK3CA*$^{WT/H1047R}$ iPSCs (**fig. S1C**).

Given previous reports that heterozygosity for *PIK3CA*^{H1047R} in breast epithelial cells extensively remodels gene expression (*25*, *26*), we undertook further transcriptional profiling in two unrelated cellular models of genetic *PIK3CA* activation. First, we examined iPSCs derived from a woman with clinically obvious but relatively mild PROS due to mosaicism for *PIK3CA*^{E418K} (**Fig. 1C**) (*27*). Heterozygous iPSCs were compared to wild-type lines established simultaneously from dermal fibroblasts derived from the same skin biopsy, which was possible due to genetic mosaicism of the sampled area of skin. Like *PIK3CA*$^{WT/H1047R}$ iPSCs, *PIK3CA*$^{WT/E418K}$ iPSCs closely clustered with the isogenic wild-type controls on MDS plotting (**Fig. 1D**), with only 30 differentially-expressed genes **(table S3**). We also studied previously reported *Pik3ca*$^{WT/H1047R}$ mouse embryonic fibroblasts (MEFs) 48 h after *Cre*-

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

mediated *Pik3ca*^{H1047R} induction (*28*). Wild-type and *Pik3ca*$^{WT/H1047R}$ MEFs were superimposable on an MDS plot (**Fig. 1E**), with only 192 downregulated and 77 upregulated genes (**table S4)**.

Collectively, these findings corroborate the existence of a threshold of PI3K pathway activity which determines the large majority of gene expression changes in *PIK3CA*$^{H1047R/H1047R}$ iPSCs in a near-binary manner. While higher depth of sequencing did reveal statistically significant gene expression changes in heterozygous iPSCs, effect sizes were modest and more variable. Similar findings in heterozygous MEFs suggest that this may be generalizable to differentiated cell types, irrespective of species.

To assess whether transcriptional changes observed in iPSCs were mirrored in the proteome, we applied label-free proteomics to the iPSC lines used in our first study (*5*). Around 4,600 protein ratios were obtained for both heterozygous *versus* wild-type and homozygous *versus* wild-type iPSC comparisons, as estimated using a novel Bayesian approach based on the Markov Chain Monte Carlo (MCMC) method (*29*). In contrast to other algorithms, the MCMC method generates an error estimate alongside each protein concentration which permits more confident determination of proteins with the most robust differential expression. The number of differentially expressed proteins correlated with *PIK3CA*^{H1047R} allele dosage, with 54 and 258 differentially-expressed proteins in *PIK3CA*$^{WT/H1047R}$ and *PIK3CA*$^{H1047R/H1047R}$ cells, respectively (**Fig. 1F, table S5 and S6)**. Of these, 27 proteins were differentially expressed in both heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs (**table S7**), with 16 changing in opposite directions (**Fig. 1F**). There was a good correlation between differentially- expressed proteins and corresponding transcripts in *PIK3CA*$^{H1047R/H1047R}$ iPSCs (**fig. S2A, S2B**), but not in heterozygous mutants (**fig. S2C, S2D**). As for the relatively poor correlation seen between transcriptomic experiments for heterozygous cells, this likely reflects the small magnitude of gene expression changes induced by heterozygous *PIK3CA*^{H1047R}.

Altogether, these data consolidate the view that only homozygosity for *PIK3CA*^{H1047R} results in robust and widespread transcriptional changes in otherwise normal, diploid cells, arguing against the "butterfly" effect of heterozygosity suggested in prior studies using a genetically abnormal breast epithelial cell line (*25*, *26*).

PIK3CA$^{H1047R/H1047R}$ hPSCs show evidence for signaling rewiring

We previously demonstrated a graded increase in AKT (S473) phosphorylation across heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs (*5*). To assess in more detail whether the near-binary gene expression difference between heterozygous and homozygous *PIK3CA*^{H1047R} cells is underpinned by corresponding differences in indices

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of PI3K pathway activation, we profiled phosphorylation of a wider repertoire of pathway components using reverse phase phosphoprotein arrays (RPPA). To mimic the physiological environment of the pluripotent epiblast, we studied cells in growth factor-replete conditions.

Observed changes in protein phosphorylation were surprisingly modest, with the maximal change a two- fold increase in AKT phosphorylation (on S473 and T308) in *PIK3CA*$^{H1047R/H1047R}$ cells. Contrasting with the near-binary response seen at the transcriptional level, heterozygous and homozygous *PIK3CA*^{H1047R} expression generally produced graded phosphorylation of PI3K pathway components, with slightly higher levels in homozygous iPSCs (**Fig. 2A**).

None of the mutant genotypes showed consistently increased mTORC1-dependent phosphorylation of P70S6K or its downstream substrate S6 (**fig. S3A**), perhaps reflecting saturation of this level of the pathway due to the presence of the additional stimuli in the medium (e.g. amino acids). When deprived of growth factors for 1 h prior to RPPA profiling, both heterozygous and homozygous mutant did exhibit increased P70S6K phosphorylation, whereas S6 phosphorylation remained similar to wild-type cells (**Fig. 2B**). Inhibition of PI3Kα activity with low-dose BYL719 for 24 hours fully reversed canonical PI3K signaling-related changes in phosphorylation of downstream proteins including AKT, GSK3, FOXO1, TSC2 and P70S6K (**Fig. 2B**). Other changes in protein phosphorylation remained unaffected in *PIK3CA*$^{H1047R/H1047R}$ hPSCs, including increased SMAD2 and ERK1/2 phosphorylation as well as increased expression of c-MYC and IGF1R (**Fig. 2B, fig. S3B**). This suggests signaling rewiring that is, unexpectedly, resistant to acute inhibition of the primary inducing event.

Pathway and network analyses implicate TGFβ signaling in *PIK3CA*^{H1047R} dose-dependent stemness

Pathway and network analyses were next applied to proteomic and transcriptomic data in an attempt to identify candidate mechanism(s) mediating *PIK3CA*^{H1047R} dose-dependent stemness. Consistent with our previous study, in which we showed strong induction of the TGFβ family member *NODAL* (*5*), TGFβ1 was again the most significant predicted upstream activator according to Ingenuity® Pathway Analysis (IPA) of the top 2000 upregulated and top 2000 downregulated transcripts in *PIK3CA*$^{H1047R/H1047R}$ iPSCs (**Fig. 3A**). TGFβ1 was also the most significant predicted upstream activator revealed by analysis of the *PIK3CA*$^{H1047R/H1047R}$ proteomic dataset (**Fig. 3B**). These data strongly suggest activation of the TGFβ pathway in homozygous *PIK3CA*^{H1047R} iPSCs.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Although *PIK3CA*$^{WT/H1047R}$ iPSCs showed around 10-fold fewer differentially-expressed genes than homozygous iPSC cells, IPA for upstream activators in the heterozygous iPSCs also revealed multiple TGFβ pathway-related stimuli among the most significant predicted activators (**Fig. 3C**). Moreover TGFβ1 was predicted as one of only two differential upstream activators when analysis was performed on genes that were concordantly differentially expressed (n = 180) in *PIK3CA*^{H1047R} mutant iPSCs *versus* wild-type controls (**Fig. 3C and table S8**). The other significant upstream regulator common to both heterozygous and homozygous *PIK3CA*^{H1047R} was MAPK1 (encoding ERK2), which is consistent with RPPA as well as immunoblot evidence of increased ERK kinase phosphorylation in *PIK3CA*^{H1047R} mutant iPSCs (Ref. (*5*) and schematic in **Fig. 2A**). Although TGFβ signaling-related predictions held in heterozygous *PIK3CA*^{H1047R} iPSCs, the significance of the effect (overlap p-value = 1.74e-05) was much lower than in homozygous (overlap p-value = 4.25e-21) mutants. This points towards a critical role for the TGFβ pathway in mediating the allele dose-dependent effect of *PIK3CA*^{H1047R} in human iPSCs.

To complement the IPA analysis, which is based on highly curated, proprietary datasets, we undertook non-hypothesis-based Weighted Gene Correlation Network Analysis (WGCNA) – a network-based data reduction method that seeks to determine gene correlation patterns across multiple samples, irrespective of the function of individual genes (*30*). Using all transcripts expressed across wild-type and both heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs (**Fig. 4A**), this analysis returned a total of 43 clusters (or modules) of highly interconnected genes (**Fig. 4B**). Of the two modules with the highest correlation to the homozygous trait, one showed enrichment for several KEGG pathway terms of relevance to the stemness phenotype of *PIK3CA*$^{H1047R/H1047R}$ iPSCs – most notably "Signaling pathways regulating pluripotency in stem cells" (**Fig. 4C**). Given prior predictions of strong activation of TGFβ signaling in homozygous mutants, we next constructed the minimal network of differentially expressed genes in *PIK3CA*$^{H1047R/H1047R}$ iPSCs that linked the pluripotency, PI3K and TGFβ signaling pathways within this network module (**Fig. 4D**). The resulting network exhibited high interconnectivity, with multiple shared nodes across all three pathways, suggesting close crosstalk between PI3K and TGFβ signaling in stemness regulation. Finally, the fact that most of the network nodes represented genes with increased expression in homozygous mutants strengthens the notion that strong oncogenic PI3Kα activation stabilizes the pluripotency network in human iPSCs.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Inhibition of TGFβ signaling destabilizes the pluripotency gene network in *PIK3CA*$^{H1047R/H1047R}$ iPSCs

TGFβ signaling plays a critical and well-established role in pluripotency regulation (*20*, *23*, *31*), and a differentiation-resistant phenotype has previously been reported in *NODAL*-overexpressing hPSCs (*22*). These observations, together with increased *NODAL* expression in homozygous *PIK3CA*^{H1047R} iPSCs and computational identification of enhanced TGFβ pathway activity in PI3K-driven "constitutive" stemness, led us to hypothesize that strong PI3Kα-dependent activation of *NODAL* expression underlies the establishment of the differentiation-resistant phenotype of homozygous mutants through stabilization of the pluripotency gene network. Specifically, we hypothesized that NODAL elicits autocrine enhancement of TGFβ signaling in *PIK3CA*$^{H1047R/H1047R}$ iPSCs, with downstream promotion of *NANOG* expression ultimately "locking" the cells in a state of perpetual stemness.

Testing this hypothesis in hPSCs is challenging for both biological and technical reasons. These include lack of specific pharmacological NODAL inhibitors and the difficulty in detecting the subtle early phenotypic consequences of partial destabilization of the hPSC pluripotency gene regulatory network. Moreover, the widely-adopted maintenance medium and coating substrate we used for cell culture both contain TGFβ ligands (*32*, *33*), which may mask effects of *NODAL* repression by PI3Kα-specific inhibition. To minimize these confounders, we prepared maintenance medium with and without recombinant NODAL supplementation, and assessed expression of *NODAL* itself and *NANOG* as surrogate markers of stemness over 72 h of culture.

Within 48 h, exclusion of NODAL from the medium resulted in the expected downregulation of *NODAL* and *NANOG* expression in wild-type iPSCs, and this was greater still at 72 h (**Fig. 5 and fig. S5A**). In contrast, *NODAL* removal had no effect on the increased *NODAL* and *NANOG* expression in *PIK3CA*$^{H1047R/H1047R}$ iPSCs (**Fig. 5 and fig. S5A**), in line with a self-sustained stemness phenotype. Exposure of NODAL-free homozygous cultures to 250 nM BYL719 reduced *NODAL* expression within 24 h, and this continued to decrease subsequently (**Fig. 5**), consistent with NODAL's known ability to control its own expression through a feedforward loop (*34*). Despite a 55% reduction in *NODAL* mRNA after 72 h, however, little effect on *NANOG* expression was seen (**Fig. 5**). The lack of *NANOG* decrease may reflect the short time course studied, or the exquisite sensitivity of hPSCs to residual upregulation of *NODAL* in homozygous *PIK3CA*^{H1047R} iPSCs. This may further be compounded by residual low levels of TGFβ-like ligands in the coating substrate, or possibly by the increased expression of two other TGFβ

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

superfamily ligands, *GDF3* and *TFGB2*, which were detected in the transcriptome analysis of homozygous mutant cells.

Thus, to confirm that TGFβ signaling is required for maintenance of stemness in *PIK3CA*$^{H1047R/H1047R}$ iPSCs, we treated cells with SB431542 – a potent and specific inhibitor of TGFβ and NODAL type I receptors (*35*). This resulted in acute and complete repression of *NODAL* expression within 24 h, accompanied by downregulation of *NANOG* expression (**Fig. 5**). A similar effect was observed on *POU5F1* (*OCT3/4*) expression, consistent with destabilization of the pluripotency gene regulatory network in *PIK3CA*$^{H1047R/H1047R}$ iPSCs (**Fig. 5**). Confirming this, we used a lineage-specific gene expression array to demonstrate a similar reduction in the expression of several other well-established stemness markers (*MYC*, *FGF4*, *GDF3*) with increased expression in *PIK3CA*$^{H1047R/H1047R}$ iPSCs, performing the analysis after 48 h of TGFβ pathway inhibition (**fig. S4B**). Despite the short treatment duration, we also found evidence for the expected neuroectoderm induction upon inhibition of the TGFβ pathway (*22*, *36*), reflected by increased expression of *CDH9*, *MAP2*, *OLFM3* and *PAPLN* (**fig. S4B**).

Collectively, these data strongly suggest that the stemness phenotype of *PIK3CA*$^{H1047R/H1047R}$ iPSCs is mediated by self-sustained TGFβ signaling, most likely through PI3K dose-dependent increase in *NODAL* expression, and is amenable to reversal through pharmacological inhibition of the TGFβ pathway but not of PI3Kα.

Stage II human breast cancers with multiple *PIK3CA*^{H1047R} alleles exhibit increased *NODAL* expression

It has long been known that dedifferentiating tumor cells re-express embryonic markers (*37*), and that the TGFβ pathway promotes cancer stem cell maintenance, metastasis and drug resistance (*38–40*). The strong link between homozygosity for *PIK3CA*^{H1047R} and TGFβ signaling in human iPSCs prompted the question of whether this mechanism may be of relevance in human *PIK3CA*-associated cancers. Specifically, NODAL re-expression has previously been linked to aggressive features such as invasion and metastasis in a range of cancers (reviewed in Refs. (*24*, *41*)), and associated with poor prognosis in breast cancer (*42–45*). Given the high prevalence of *PIK3CA*^{H1047R} in breast cancers (35.4 %; cBioPortal non-overlapping breast cancer studies, accessed August 2019) (*3*), we next stratified the breast invasive carcinoma (BRCA) dataset in The Cancer Genome Atlas (TCGA) according to *PIK3CA*^{H1047R} allele dosage and conducted targeted analysis of *NODAL*, *NANOG*, *POU5F1* and *MYC* expression, comparing tumors with multiple *PIK3CA*^{H1047R} copies to those with a single copy.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

PIK3CA^{H1047R} was detected in 108 BRCA samples, 57 of which had more than one copy of the mutant allele. Of the remaining 51 samples with a single *PIK3CA*^{H1047R} allele, 10 harbored a second *PIK3CA* mutation and were thus grouped with cancers with multiple *PIK3CA*^{H1047R} copies (giving rise to a total of 67 samples classified as 'multiple'). MYC mRNA levels were robustly detected in both sample groups (**Fig. 6A**), consistent with ubiquitous expression. In contrast, mRNA expression of the embryonic markers *NODAL*, *NANOG* and *POU5F1* was low (< 1 count per million) (**Fig. 6A**), as expected if breast cancer stem cell-like cells only comprise a small proportion of the bulk tumor tissue (*46*). The low expression notwithstanding, there was evidence of increased expression of *MYC*, *POU5F1* and *NODAL* in several samples with multiple *PIK3CA*^{H1047R} copies (**Fig. 6A**).

Given that re-expression of *NODAL* has been linked with aggressive breast cancer (*42–45*), we next stratified the samples according to tumor stage and assessed the expression of the four stemness markers (**Fig. 6B**). Stage I denotes breast cancer that has not spread outside the breast and exhibits little or no involvement of proximal lymph nodes, whereas stage II and III tumors exhibit progressively higher spreading to lymph nodes (*47*). Stage IV denotes breast cancer that has spread to other organs (*47*). We excluded stage IV samples from the analysis due to low sample size (2 and 1 in the "single" and "multiple" *PIK3CA*^{H1047R} groups, respectively). Notably, in stage II breast tumors, for which the sample size was sufficient, *NODAL* expression increased as a function of *PIK3CA*^{H1047R} allele dosage (**Fig. 6B**). This trend was apparent despite the genetically heterogeneous nature of the samples and suggests that the allele dose-dependent signaling rewiring connecting *PIK3CA*^{H1047R}, *NODAL* expression and self-sustained stemness in human iPSCs may be relevant to breast cancer progression (**Fig. 6C**), with implications for therapeutic targeting of tumors with multiple *PIK3CA*^{H1047R} copies.

Discussion

The so-called "hotspot" variant *PIK3CA*^{H1047R} is the most common activating *PIK3CA* mutation in human cancers and in a group of largely benign overgrowth disorders known as PROS (*6*). We recently found that *PIK3CA*-associated cancers often harbor multiple mutant *PIK3CA* copies, and demonstrated that homozygosity but not heterozygosity for *PIK3CA*^{H1047R} leads to self-sustained stemness in human pluripotent stem cells (hPSCs) (*5*). Using a combination of computational analyses and targeted experiments, the current study adds further support for the existence of a PI3K signaling threshold in cells with oncogenic PI3Kα activation. We provide evidence for self-

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sustained TGFβ pathway activation as the main mechanism through which *PIK3CA*^{H1047R} homozygosity "locks" hPSCs in a differentiation-resistant state that becomes independent of the driver mutation and the associated PI3K pathway activation.

High-depth transcriptomics confirmed that heterozygosity for *PIK3CA*^{H1047R} fails to induce widespread substantial transcriptional remodeling, whether chronically modelled in CRISPR-edited iPSCs or acutely induced in mouse embryonic fibroblasts (MEFs) by *Cre* expression, despite induction of canonical PI3K pathway activation in both cases (current study and Ref. (*5*, *28*, *48*)). Similarly, iPSCs with heterozygous expression of *PIK3CA*^{E418K}, a "non-hotspot" mutation, were transcriptionally indistinguishable from their isogenic wild-type controls. In contrast to the mild transcriptional consequences of heterozygous *PIK3CA*^{H1047R} expression, homozygosity for *PIK3CA*^{H1047R} was associated with differential expression of nearly 1/3 of the hPSC transcriptome. This observed near-binary response is not a consequence of a similar quantitative difference in PI3K pathway activation assessed by phosphoprotein profiling, which instead showed a relatively modest and graded increase in homozygous *versus* heterozygous *PIK3CA*^{H1047R} hPSCs. This suggests that the apparent sharp PI3K signaling threshold that determines the cellular response in hPSCs hPSCs is "decoded" distal to the canonical pathway activation.

The increased expression of several proteins in homozygous *PIK3CA*^{H1047R} iPSCs, and their resistance to downregulation in response PI3Kα-specific inhibition with BYL719, further suggested that the phenotype of these cells may have become partially uncoupled from the oncogenic trigger event. In systems biology terms, the near-binary response observed with two *versus* one copy of *PIK3CA*^{H1047R}, and the inability to reverse the phenotype upon inhibition of the primary PI3K signaling defect, is consistent with a non-linear network topology characterized by bistability and an "all-or-nothing" response. This is usually mediated by indirect or direct positive feedback loops (*49*) such as those known to occur between the TGFβ pathway and the pluripotency network (*9*, *21*, *34*).

The well-known dose-dependent effects of TGFβ signaling in a developmental context (*34*) helps explain our previous observations of allele dose-dependent stemness effects of *PIK3CA*^{H1047R} in hPSCs (*5*). Specifically, our data suggest a model in which homozygosity but not heterozygosity for *PIK3CA*^{H1047R} promotes *NODAL* expression and thus increases TGFβ pathway activity to a level that is sufficient for increased *NANOG* expression and stabilization of the stem cell state (**Fig. 6C**), but insufficient to tip the balance towards mesendoderm differentiation (*5*). Such an indirect effect of PI3K activation on the pluripotency network in hPSCs is in line with the lack of

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

316 *NANOG* downregulation in response to PI3Kα-specific inhibition (this study and Ref. (*5*). Future studies are
317 warranted to determine whether PI3K-dependent regulation of *NODAL* expression contributes to its developmental
318 functions in vivo, which include roles in gastrulation and establishment of the body axes (*34*). It is notable that
319 patients with a germline activating *AKT2* mutation exhibit overgrowth that is predominantly left-sided (*50*), yet it
320 remains unclear whether this is linked to differential regulation of *NODAL* expression.

321 Since specific pharmacological inhibition of NODAL is not possible, our strongest evidence for its role
322 in the stemness phenotype of *PIK3CA*$^{H1047R/H1047R}$ hPSCs comes from the ability of these cells to sustain increased
323 expression of key stemness genes without exogenous NODAL supplementation. Moreover, the ability of SB431512
324 to collapse the stemness gene signature in *PIK3CA*$^{H1047R/H1047R}$ hPSCs unequivocally demonstrates that sustained
325 TGFβ pathway activation is required to maintain their stemness phenotype. Exactly how PI3K activation regulates
326 *NODAL* expression remains to be determined. A potential mechanism involves increased expression of the stem cell
327 reprogramming factor MYC, which was observed both at the mRNA and protein level in homozygous but not
328 heterozygous *PIK3CA*^{H1047R} iPSCs. Furthermore, MYC was the only node in the WGCNA-based network of
329 pluripotency, PI3K and TGFβ pathway components that was classified as a member of all three pathways. Finally,
330 increased MYC expression has previously been shown to exert oncogenic effects that depend on a sharp threshold
331 of MYC expression, reminiscent of the effects we observe for *PIK3CA* activation (*51*). As the first specific
332 pharmacological inhibitor of MYC was recently reported (*52*), its use in the *PIK3CA*^{H1047R} hPSC system may provide
333 additional insight into the potential role of MYC as a molecular link between oncogenic PI3Kα activation, *NODAL*
334 expression and stemness.

335 The close relationship between PI3K and TGFβ pathway-driven stemness has potentially important
336 implications for understanding PIK3CA-driven cancer. Breast cancers are enriched for *PIK3CA* mutations, and
337 studies in mice have revealed that *PIK3CA*^{H1047R} induces multipotency in differentiated mammary cells (*53*, *54*).
338 Conversely, advanced cancers typically exhibit increased phenotypic plasticity and stemness, which is closely linked
339 to drug resistance and the ability of cancer cells to undergo epithelial-mesenchymal-transition (EMT) and
340 metastasize (*40*, *55*). Prolonged exposure of human mammary epithelial cells to TGFβ was recently shown to trigger
341 irreversible EMT and PI3K/AKT-dependent stemness (*56*), illustrating the existence of a reciprocal relationship
342 between the two signaling networks in the context of pluripotent and cancer stem cell regulation. A previous study

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reported increased AKT-mediated phosphorylation of the pro-metastatic factor TWIST1 in invasive cancer cells and advanced human breast tumors, further demonstrating a positive feedback loop characterized by increased TGFβ2 expression, autocrine TGFβ signaling and potentiation of the PI3K signaling pathway (*57*). Overexpression of *PIK3CA*^{H1047R} in MCF10A breast epithelial cells has also been linked to induction of TGFβ1 expression, EMT and stem cell-like properties (*58*, *59*).

We now report that increased *NODAL* expression is discernible even in the context of cellularly heterogeneous human breast cancers, with the degree of upregulation related to the number of oncogenic *PIK3CA*^{H1047R} alleles present. This suggests that our findings in hPSCs have potential clinical relevance (**Figure 6c**). Monotherapy with pan- or isoform-specific PI3K inhibitors has had modest success in cancer (*60*), and it is thus noteworthy that PI3Kα-specific inhibition with BYL719 was unable to reverse the stemness phenotype of *PIK3CA*$^{H1047R/H1047R}$ iPSCs. Systematic studies in a larger number of breast cancer samples and breast cancer cell lines with defined *PIK3CA*^{H1047R} copy number are needed to confirm the phenomenon that we have observed in human iPSCs. Nevertheless, based on our current findings of increased *NODAL* expression in some human breast cancers with multiple *PIK3CA*^{H1047R} copies, we believe that TGFβ pathway inhibition is worthy of further exploration in this setting. Moreover, as induction of pro-tumorigenic TGFβ signaling has been reported in a range of human cancers (*24*, *61*), the allele dose-dependent effects of *PIK3CA*^{H1047R} may extend beyond breast cancer. Accordingly, it has previously been reported that *PIK3CA* amplification leads to TGFβ pathway-dependent EMT in a mouse model of oral carcinogenesis, with additional evidence that the degree of *PIK3CA* amplification correlates with tumor stage in human head and neck squamous cell carcinoma (*62*).

Finally, our study demonstrates the power of network-based approaches to study the highly context-dependent complexity of cell signaling and to identify potential therapeutic targets beyond those suggested by simply considering the presence/absence of a specific genetic defect (*63*–*65*). Given the dose-dependent effects of both TGFβ and PI3K pathway activation, a comprehensive understanding of their concerted actions will hinge upon adoption of network-based analyses and mathematical modeling capable of capturing context-dependent relationships in response to qualitative and quantitative signaling differences. This may result in tailored treatment strategies that take into account *PIK3CA* allele dose-dependent signaling reconfiguration in individual tumor cells.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Materials and Methods

Unless stated otherwise, standard chemicals were acquired from Sigma Aldrich, with details for the remaining reagents included in **Supplementary Key Resources Table**.

Induced pluripotent stem cell (iPSC) culture and treatments

Maintenance

The derivation of the iPSC lines, including associated ethics statements, has been described previously (*5*). All lines were grown at 37°C and 5% CO_2 in Essential 8 Flex (E8/F) medium on Geltrex-coated plates, in the absence of antibiotics. For maintenance, cells at 70-90% confluency were passaged as aggregates with ReLeSR, using E8 supplemented with RevitaCell (E8/F+R) during the first 24 h to promote survival. A detailed version of this protocol is available via protocols.io (doi: dx.doi.org/10.17504/protocols.io.4rtgv6n).

All cell lines were tested negative for mycoplasma and genotyped routinely to rule out cross-contamination during prolonged culture. Short tandem repeat profiling was not performed. All experiments were performed on cells within 10 passages since thawing.

Collection for RNA sequencing and total proteomics

For RNA sequencing and total proteomics, subconfluent cells were fed fresh E8/F 3 h prior to snap-freezing on dry ice and subsequent RNA or protein extraction. Relative to the results in Ref. (*5*), the current transcriptomic data of *PIK3CA*^{H1047R} were obtained more than 6 months following the first study, with cells at different passages, and were thus independent from one another. Moreover, sample collection for the second transcriptomics experiment was conducted over three days according to a block design, thus allowing us to determine transcriptional differences that are robust to biological variability.

Cell lysate collection for RPPA

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

For RPPA in growth factor-replete conditions, cells were fed fresh E8/F 3h before collection. To assess variability due to differences in collection timing, clones from each iPSC genotype were collected on each one of three days according to a block design, giving rise to a total of 22 cultures. To test the effect of the PI3Kα-specific inhibitor BYL719, cells were treated with 100 nM drug (or DMSO only as control treatment) for 24 h and exposed to growth factor removal within the last hour before collection. All cells were washed in DPBS prior to collection to rinse off residual proteins and cell debris.

TGFβ/NODAL signaling studies

Wild-type or homozygous *PIK3CA*^{H1047R} iPSCs were seeded in 12-well plates all coated with Geltrex from the same lot (#2052962; diluted in DMEM/F12 lot #RNBH0692). Cells were processed for seeding at a ratio of 1:15 according to the standard maintenance protocol. One day after seeding, individual treatments were applied to triplicate wells. Briefly, cells were first washed twice with 2 and 1 ml of Dulbecco's PBS (DPBS) to remove residual growth factors. The base medium for individual treatments was Essential 6 supplemented with 10 ng/ml heat-stable FGF2. This was combined with one of the following reagents or their diluent equivalents: 100 ng/ml NODAL (diluent: 4 mM HCl), 250 nM BYL719 (diluent: DMSO), 5 μM SB431542 (diluent: DMSO). Cells were snap-frozen on dry ice after 24, 48 and 72 h following a single DPBS wash. Individual treatments were replenished daily at the same time of day to limit temporal confounders.

Mouse embryonic fibroblast (MEF) culture

The derivation and culture of the wild-type and *PIK3CA*$^{WT/H1047R}$ MEFs used in this study have been reported previously (*28*). Cell pellets were collected on dry ice 48 h after induction of heterozygous *PIK3CA*^{H1047R} expression, without prior starvation.

RNA sequencing

Induced pluripotent stem cell lysates were collected in QIAzol and processed for RNA extraction with the DirectZol Kit as per the manufacturer's instructions. The final RNA was subjected

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to quantification and quality assessment on an Agilent Bioanalyzer using the RNA 6000 Nano Kit, confirming that all samples had a RIN score of 10. For *PIK3CA*^{H1047R} iPSCs and corresponding wild-types, an Illumina TruSeq Stranded mRNA Library Prep Kit was used to synthesize 150-bp-long paired-end mRNA libraries, followed by sequencing on an Illumina HiSeq 4000, with average depth of 70 million reads per sample. *PIK3CA*$^{WT/E418K}$ and isogenic control iPSCs were subjected to 50-bp-long single-end RNA sequencing (RNAseq) at an average depth of 20 million reads per sample.

MEF RNA was extracted using Qiagen's RNAeasy miniprep (with QIAshredder). All samples had a confirmed Agilent Bioanalyzer RIN score of 10. An Illumina TruSeq Unstranded mRNA kit was used to prepare 100-bp-long paired-end libraries, followed by Illumina HiSeq 2000 sequencing.

Data analysis

The raw reads were mapped to the human genome build GRCh38 (for iPSC RNAseq) or the mouse genome build GRCm38 (for MEF RNAseq), and gene level counts were performed using Spliced Transcripts Alignment to a Reference (STAR) v2.5 (*66*). Subsequent data processing was performed using the open-source R software according to the *limma-voom* method (*67*). Briefly, raw counts were converted to counts per million (cpm) using the *cpm()* function in *edgeR* (*68*), followed by normalization according the trimmed mean of M (TMM) method (*69*). The mean-variance relationship was modelled with *voom()*, followed by linear modelling and computation of moderated t-statistics using the *lmFit()* and *eBayes()* functions in the *limma* package (*67*). The associated p-value for assessment of differential gene expression was adjusted for multiple comparisons with the Benjamini-Hochberg method at false-discovery rate (FDR) $\leq$ 5% (*70*). The function *duplicateCorrelation()* was applied to correct for the use of replicate iPSC clones.

Correlations between corresponding transcriptomics and/or proteomics data were calculated using Spearman's rank-order correlation test for non-normally distributed data.

Label-free total proteomics

Sample preparation

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cells were cultured to subconfluence in Geltrex-coated T175 flasks, and protein was harvested by lysis in 3 ml modified RIPA buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA) supplemented with phosphatase inhibitors (5 mM ß-glycerophosphate, 5 mM NaF, 1 mM Na_3VO_4) and protease inhibitors (Roche cOmplete ULTRA Tablets, EDTA-free). The lysates were sonicated on ice (4x 10s bursts, amplitude = 60%; Bandelin Sonopuls HD2070 sonicator) and spun down for 20 min at 4300g. Ice-cold acetone was added to the supernatant to achieve a final concentration of 80% acetone, and protein was left to precipitate overnight at -20°C. Precipitated protein was pelleted by centrifugation at 2000g for 5 min and solubilized in 6 M urea, 2 M thiourea, 10 mM HEPES pH 8.0. Protein was quantified using the Bradford assay and 8 mg of each sample were reduced with 1 mM dithiothritol, alkylated with 5 mM chloroacetamide and digested with endopeptidase Lys-C (1:200 v/v) for 3 h. Samples were diluted to 1 mg/ml protein using 50 mM ammonium bicarbonate and incubated overnight with trypsin (1:200 v/v). Digested samples were acidified and urea removed using SepPak C18 cartridges. Peptides were eluted, and an aliquot of 100 µg set aside for total proteome analysis. The peptides were quantified using the Pierce quantitative colorimetric peptide assay. The equalized peptide amounts were lyophilized and resolubilized in 2% acetonitrile and 1% trifluoroacetic acid in order to achieve a final 2 µg on-column peptide load.

Mass spectrometry (MS) data acquisition

All spectra were acquired on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) operated in data-dependent mode coupled to an EASY-nLC 1200 liquid chromatography pump (Thermo Fisher Scientific) and separated on a 50 cm reversed phase column (Thermo Fisher Scientific, PepMap RSLC C18, 2 µM, 100A, 75 µm x 50 cm). Proteome samples (non-enriched) were eluted over a linear gradient ranging from 0-11% acetonitrile over 70 min, 11-20% acetonitrile for 80 min, 21-30% acetonitrile for 50 min, 31-48% acetonitrile for 30 min, followed by 76% acetonitrile for the final 10 min with a flow rate of 250 nl/min.

Survey-full scan MS spectra were acquired in the Orbitrap at a resolution of 120,000 from m/z 350-2000, automated gain control (AGC) target of $4x10^5$ ions, and maximum injection time of 20 ms.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Precursors were filtered based on charge state (≥2) and monoisotopic peak assignment, and dynamic exclusion was applied for 45s. A decision tree method allowed fragmentation for ITMS2 via electron transfer dissociation (ETD) or higher-energy collision dissociation (HCD), depending on charge state and m/z. Precursor ions were isolated with the quadrupole set to an isolation width of 1.6 m/z. MS2 spectra fragmented by ETD and HCD (35% collision energy) were acquired in the ion trap with an AGC target of 1e4. Maximum injection time for HCD and ETD was 80 ms for proteome samples.

Whole-exome sequencing (WES) and FASTA file generation

WES was performed on a single clone per genotype to generate cell-specific databases for downstream mass spectrometry searchers. Genomic DNA was extracted with Qiagen's QIAamp DNA Micro Kit according to the manufacturer's instructions, followed by quantification using the Qubit dsDNA High Sensitivity Assay Kit and by dilution to 5 ng/µl in the supplied TE buffer. The samples were submitted for library preparation and sequencing by the SMCL Next Generation Sequencing Hub (Academic Laboratory of Medical Genetics, Cambridge). Sequencing was performed on an Illumina HiSeq 4000 with 50X coverage across more than 60% of the exome in each sample. Raw reads were filtered with Trimmomatic (*71*) using the following parameters: headcrop = 3, minlen = 30, trailing = 3. The trimmed reads were aligned to the human reference genome (hg19 build) with BWA (*72*), followed by application of GATK base quality score recalibration, indel realignment, duplicate removal and SNP/indel discovery with genotyping (*73*). GATK Best Practices standard hard filtering parameters were used throughout (*74*).

In order to find non-reference, mutated peptides in the MS data, we increased the search FASTA file with mutations affecting the protein sequence, as detected by WES with a high sensitivity filter: QD ≤ 1.5, FS ≥ 60, MQ ≥ 40, MQRankSum ≤ -12.5, ReadPosRankSum ≤ -8.0, and average DP ≥ 5 per sample. The Ensembl Variant Effect Predictor (VEP) with Ensembl v88 was used to predict the effect of the mutations on the protein sequence (*75*). For every variant with an effect on the protein sequence we added the predicted mutated tryptic peptide at the end of the protein sequence.

Mass spectrometry searches

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Raw files were processed using MaxQuant 1.5.0.2 (*76*) with all searches conducted using cell-specific databases (see *Whole-exome sequencing and FASTA file generation*), where all protein sequence variants were included in addition to the reference (Ensemble v68 human FASTA). Methionine oxidation, protein N-terminal acetylation and serine/threonine/tyrosine phosphorylation were set as variable modifications and cysteine carbamidomethylation was set as a fixed modification. False discovery rates were set to 1% and the "match between runs" functionality was activated. We filtered out peptides that were associated with multiple identifications in the MaxQuant *msms.txt* file, had a score < 40, were identified in the reverse database or came from known contaminants. Analysis of the observed peptides passing these filters was performed using a Monte Carlo Markov Chain model as described previously (*29*). Briefly, the model predicted the average ratio (sample versus control) of a peptide as a function of the observed protein concentration (obtained from the MaxQuant *evidence.txt* file). Combined with a noise model, a distribution of likely values for the parameters was obtained. The mean and standard deviation of this resulting distribution was used to calculate a z-score which was used together with the fold-change (FC) for subsequent filtering for differentially expressed proteins ($|z| \geq 1.2; |\ln(FC)| \geq \ln(1.2)$

Reverse phase protein array (RPPA)

For RPPA, snap-frozen cells were lysed in ice-cold protein lysis buffer containing: 50 mM HEPES, 150 mM NaCl, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 1% (v/v) TritonX-100, 1 mM EGTA, 100 mM NaF, 10 mM $Na_4P_2O_7$, 2 mM Na_3VO_4 (added fresh), 1X EDTA-free protease inhibitor tablet, 1X PhosStop tablet. Protein concentrations were measured using BioRad's DC protein assay, and all concentrations were adjusted to 1 mg/ml with lysis buffer and 1X SDS sample buffer (10% glycerol, 2% SDS, 62.5 mM Tris-HCl pH 6.8) supplemented with 2.5% β-mercaptoethanol.

The protein lysates were processed for slide spotting and antibody incubations as described previously (*77*). Briefly, a four-point dilution series was prepared for each sample and printed in triplicate on single pad Avid Nitrocellulose slides (Grace Biolabs) consisting of 8 arrays with 36x12 spots each. Next, slides were blocked and incubated in primary and secondary antibodies. The processed arrays were imaged using an Innopsys 710 slide scanner. Non-specific signals were determined for each

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

slide by omitting primary antibody incubation step. For normalization, sample loading on each array was determined by staining with Fast Green dye and recording the corresponding signal at 800 nm. Details for all primary and secondary RPPA antibodies are included in **Supplementary Key Resources Table**.

Data analysis

Slide images were analyzed using Mapix software (Innopsys), with the spot diameter of the grid set to 270 µm. Background signal intensity was determined for each spot individually and subtracted from the sample spot signal. A test for linearity was performed from the four-point dilution series, according to a flag system where $R^2 > 0.9$ was deemed good, $R^2 > 0.8$ was deemed acceptable and $R^2 < 0.8$ was poor (excluded from subsequent analyses). Median values from the four-point dilution series were calculated for each technical replicate and normalized to the corresponding Fast Green value to account for differences in protein loading. For each sample and protein target, a mean expression value was calculated from the remaining technical replicates and normalized to the corresponding mean of the wild-type group. All phosphoprotein signals were also normalized to the corresponding total protein values.

A statistical test for differential expression was performed on datasets with more than three samples per group, using the *limma* package to apply the *limma-trend* method with *lmFit()* and *eBayes()*, specifying collection time as blocking factor (*67*). Phosphoprotein and total protein lists were processed separately. The associated p-value for assessment of differential gene expression was adjusted for multiple comparisons with the Benjamini-Hochberg method at FDR $\leq$ 5% (*70*). The function *duplicateCorrelation()* was applied to correct for the use of replicate iPSC clones on the same day. Heatmaps were generated using the *heatmap.2()* function within the *gplots* package in R, using target-wise correlation for dendrogram construction.

Reverse transcription-quantitative PCR (RT-qPCR)

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cellular RNA was extracted as described above for RNA Sequencing, and 200 ng used for complementary DNA (cDNA) synthesis with Thermo Fisher's High-Capacity cDNA Reverse Transcription Kit. Subsequent SYBR Green-based qPCRs were performed on 2.5 ng total cDNA. A 5-fold cDNA dilution series was prepared and used as standard curve for relative quantitation of gene expression. *TBP* was used as normalizer following confirmation that its gene expression remained unaffected by the tested conditions. Melt curve analyses were used to confirm amplification of a single product by each primer. All primers had amplification efficiencies 95%-105%. Individual samples were loaded in duplicate in 384-well plates.

TaqMan hPSC Scorecards (384-well) were used according to the manufacturer's instructions with the following modifications. From each sample diluted to 20 ng/µl, two 50 µl RT reactions were set up, with 500 ng RNA sample in each. Next, the two RT replicates were combined to obtain 1 µg cDNA in a total volume of 100 µl (final concentration: 10 ng/µl). This was subsequently diluted to 0.715 ng/µl and 10 µl loaded into each Scorecard well. All Ct values were mapped to their corresponding genes using the TaqMan hPSC Scorecard analysis software provided by the manufacturer. Genes with Ct values < 15 were excluded from further analyses. To be considered for downstream processing, genes were also required to have Ct values < 30 in at least two out of the eight samples. Next, Ct values were linearized (antilog) under the assumption of 100 % primer amplification efficiency. The geometric expression mean of the control gene assays was used for subsequent normalization of individual gene expression values.

All qPCR data were acquired on a Quant Studio™ 5 Real-Time PCR System (Thermo Fisher Scientific). The thermocycling conditions (SYBR Green reactions) were as follows (ramp rate 1.6°C/s for all): 50°C for 2 min, 95°C for 10 min, 40 cycles at 95°C for 15 sec and 60°C for 1 min, followed by melt curve analysis (95°C for 15 sec, 60°C for 1 min, and 95°C for 15 min with ramp rate 0.075°C/sec). The TaqMan hPSC Scorecard thermocycling conditions were as specified by the manufacturer in the accompanying template.

All relevant primer sequences are included in **Supplementary Key Resource Table**.

Bioinformatic analyses

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ingenuity® Pathway Analysis (IPA)

The list of differentially expressed total proteins (*PIK3CA*$^{H1047R/H1047R}$ *vs* wild-type) was subjected to IPA (build version: 448560M; content version: 36601845) against the Ingenuity Knowledge Base, considering only relationships where confidence was classified as "Experimentally Observed". Following exclusion of chemicals and drugs, the Upstream Regulators list was used for generation of Volcano plots of the respective activation z-scores and overlap p-values.

IPA was also used to analyze the lists of differentially expressed genes in both heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs (IPA build version: 484108M; content version: 45868156) and MEFs (IPA build version: 486617M; content version: 46901286), using the Ingenuity Knowledge Base and considering only relationships where confidence was classified as "Experimentally Observed". Chemicals and diseases were excluded from Node Types. For the iPSC datasets, differentially expressed genes were only considered for IPA analysis if having an absolute log2(fold-change) $\geq$ log2(1.3). The choice of this relatively permissive log fold-change choice was guided by the *limma()* tutorial for RNAseq (*67*), and the assumption that small fold-changes in the expression of genes that act within the same pathway may be sufficient to elicit a functionally important response and thus should not be omitted. This consideration is of particular relevance for transcriptomic data from heterozygous *PIK3CA*^{H1047R} iPSCs where fold-changes were relatively small. Due to the high number of differentially expressed transcripts in *PIK3CA*$^{H1047R/H1047R}$ iPSCs, the analysis was conducted using the top 2000 up- and top 2000 downregulated transcripts.

The IPA Upstream Regulator Analysis is based on the proprietary Ingenuity Knowledge Base which is used to compute two scores based on user-specified data: an enrichment score (Fisher's exact test p-value) that measures overlap between observed and predicted regulated gene sets; a z-score that assesses the match between observed and predicted up/down regulation patterns (*78*). The results of the Upstream Regulators Analysis were extracted for downstream Volcano plotting of overlap p-values and associated activation z-scores. Note that for heterozygous *PIK3CA*^{H1047R} iPSCs, a bias-corrected activation z-score was used for plotting to take into account any bias arising from a larger number of upregulated *vs* downregulated genes in these cells.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Weighted Gene Correlation Network Analysis (WGCNA)

RNA sequencing counts from all 12 samples were converted to reads per kilobase million (RPKM). A threshold of 10 RPKM was used to filter out low-expression genes, followed by removing any genes with missing values caused by this filtering. The RPKM values for the remaining 16,823 genes were log2 transformed and taken forward for network analysis using the *WGCNA* R package (*30*, *79*), with a soft power threshold of 28 (chosen to maximize scale independence and minimize mean connectivity) and a minimum module size of 30 genes.

To identify modules associated with homozygosity for *PIK3CA*^{H1047R}, we used the correlation between a gene's module membership (eigengene) and significance for differential expression in homozygous *PIK3CA*$^{H1047R/H1047R}$ iPSCs. The top two most significant modules for the homozygocity trait were selected for functional enrichment analysis. CytoScape plugin ClueGO (version 2.5.4) (*80*) was used to perform pathway analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) ontology (build 27.02.19) (*81*). All settings were kept at default values. Only pathways with p-value $\leq$ 0.05 were selected, and a custom reference gene set was used as background (the 16,823 genes analysed using WGCNA). Network visualization was performed using Cytoscape (*82*).

TCGA Breast Cancer Data Analysis

The number of *PIK3CA*^{H1047R} copies in The Cancer Genome Atlas (TCGA) breast invasive carcinoma (BRCA) samples was retrieved by filtering previously published copy number information from this dataset (*5*). Samples were classified as having multiple *PIK3CA*^{H1047R} copies ("multiple") when the mutant allele dosage (*mut.multi*) $\geq$ 1.5. Samples with a single *PIK3CA*^{H1047R} copy and an additional *PIK3CA* variant were also classified as "multiple". The associated RNA sequencing and clinical data were retrieved using the R package TCGAbiolinks (version: 2.12.3) according to the accompanying vignette (*83*). The raw sequencing counts were converted to cpm and normalized as described for iPSC/MEF RNAseq analysis above. A filtering step was applied, requiring $\geq$ 1 cpm in more than one tumor sample for a gene to be considered expressed. Copy number data, RNAseq and clinical information were merged based on the TCGA sample barcode and filtered according to genes

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of interest. Candidate gene expression data were plotted either as a function of the number of *PIK3CA*^{H1047R} allele or upon additional stratification according to tumor stage.

Statistical analyses

Bespoke statistical analyses are specified in the relevant sections above. Statistical analyses were restricted to -omics datasets to correct appropriately for multiple comparisons and to take advantage of target-wise variances as outlined in Ref. (*67*). In line with recent (ATOMIC) recommendations by the American Statistical Association (*84*), we have avoided arbitrary use of "statistical significance" applied to data from small-scale cell culture experiments which violate assumptions of the most widely used statistical tests. Instead, we present all data from multiple orthogonal experiments, alongside complete information on experimental replicates, independent clones and replicate cultures.

The same reasoning applies to our analysis of TCGA breast cancer data, where frequentist tests do not take into account the prior hypothesis that *NODAL* may have higher expression in breast cancer samples with multiple *PIK3CA*^{H1047R} copies. A Bayesian approach would be more appropriate, but it remains unclear what the prior distribution should be in the absence of data that complement our human iPSC studies. We acknowledge the relatively small sample size used in Discussion and note that systematic testing in future studies with large sample size and bespoke breast cancer cell models is needed. Our communication of this analysis in its current format is guided by the possibility of a potentially clinically relevant finding worthy of further studies.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Materials:

Fig. S1, related to Fig. 1. Fold-change distribution and transcriptome correlations.

Fig. S2, related to Fig. 1. Transcriptome-proteome correlations.

Fig. S3, related to Fig. 2. Additional RPPA data.

Fig. S4, related to Fig. 5. Alternative representation of the experimental data in Fig. 5A and additional RT-qPCR-based profiling of lineage-specific markers.

Table S1. List of differentially expressed genes in *PIK3CA*$^{WT/H1047R}$ vs wild-type hPSCs after applying an absolute fold-change cut-off of minimum 1.3.

Table S2. List of differentially expressed genes in *PIK3CA*$^{H1047R/H1047R}$ vs wild-type hPSCs after applying an absolute fold-change cut-off of minimum 1.3.

Table S3. List of differentially expressed genes in *PIK3CA*$^{WT/E418K}$ vs wild-type hPSCs.

Table S4. List of differentially expressed genes in *PIK3CA*$^{WT/H1047R}$ vs wild-type MEFs.

Table S5. List of differentially expressed proteins in *PIK3CA*$^{WT/H1047R}$ vs wild-type hPSCs.

Table S6. List of differentially expressed proteins in *PIK3CA*$^{H1047R/H1047R}$ vs wild-type hPSCs.

Table S7. List of differentially expressed proteins in both heterozygous and homozygous *PIK3CA*^{H1047R} hPSCs vs wild-type controls.

Key Resources Table.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. D. A. Fruman, H. Chiu, B. D. Hopkins, S. Bagrodia, L. C. Cantley, R. T. Abraham, The PI3K Pathway in Human Disease. *Cell*. **170**, 605–635 (2017).
2. B. Bilanges, Y. Posor, B. Vanhaesebroeck, PI3K isoforms in cell signalling and vesicle trafficking. *Nat. Rev. Mol. Cell Biol.* **In press** (2019), doi:10.1038/s41580-019-0129-z.
3. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
4. T. L. Yuan, L. C. Cantley, PI3K pathway alterations in cancer: variations on a theme. *Oncogene*. **27**, 5497–5510 (2008).
5. R. R. Madsen, R. G. Knox, W. Pearce, S. Lopez, B. Mahler-Araujo, N. McGranahan, B. Vanhaesebroeck, R. K. Semple, Oncogenic PIK3CA promotes cellular stemness in an allele dose-dependent manner. *Proc. Natl. Acad. Sci.* **116**, 8380–8389 (2019).
6. R. R. Madsen, B. Vanhaesebroeck, R. K. Semple, Cancer-Associated PIK3CA Mutations in Overgrowth Disorders. *Trends Mol. Med.* **24**, 856–870 (2018).
7. P. Cahan, G. Q. Daley, Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* **14**, 357–68 (2013).
8. K. M. Loh, B. Lim, A precarious balance: Pluripotency factors as lineage specifiers. *Cell Stem Cell*. **8**, 363–369 (2011).
9. M. Li, J. C. I. Belmonte, Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* **18**, 180–191 (2017).
10. L. Boyer, T. I. Lee, M. Cole, S. Johnstone, S. Levine, J. Zucker, M. Guenther, R. Kumar, H. Murray, R. Jenner, D. Gifford, D. Melton, R. Jaenisch, R. Young, Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. **122**, 947–956 (2005).
11. Y. H. Loh, Q. Wu, J. L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K. Y. Wong, K. W. Sung, C. W. H. Lee, X. D. Zhao, K. P. Chiu, L. Lipovich,

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

V. A. Kuznetsov, P. Robson, L. W. Stanton, C. L. Wei, Y. Ruan, B. Lim, H. H. Ng, The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440 (2006).

12. J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Scholer, A. Smith, Formation of pluripotent stem cells in the mammalian embryo dependes on the POU transcription factor Oct4. *Cell*. **95**, 379–391 (1998).

13. H. Niwa, J. Miyazaki, A. G. Smith, Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376 (2000).

14. A. Radzisheuskaya, G. Le Bin Chia, R. L. Dos Santos, T. W. Theunissen, L. F. C. Castro, J. Nichols, J. C. R. Silva, A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat. Cell Biol.* **15**, 579–590 (2013).

15. I. Chambers, J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, A. Smith, Nanog safeguards pluripotency and mediates germline development. *Nature*. **450**, 1230–1234 (2007).

16. N. Ivanova, R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. Decoste, X. Schafer, Y. Lun, I. R. Lemischka, Dissecting self-renewal in stem cells with RNA interference. *Nature*. **442**, 533–538 (2006).

17. K. Mitsui, Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, S. Yamanaka, The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell*. **113**, 631–642 (2003).

18. I. Chambers, D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, A. Smith, Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*. **113**, 643–655 (2003).

19. H. Darr, Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development*. **133**, 1193–1201 (2006).

20. R. H. Xu, T. L. Sampsell-Barron, F. Gu, S. Root, R. M. Peck, G. Pan, J. Yu, J. Antosiewicz-Bourget, S. Tian, R. Stewart, J. A. Thomson, NANOG Is a Direct Target of TGFβ/Activin-Mediated SMAD Signaling in Human ESCs. *Cell Stem Cell*. **3**, 196–206 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

21. S. Pauklin, L. Vallier, Activin/Nodal signalling in stem cells. *Development*. **142**, 607–619 (2015).

22. L. Vallier, D. Reynolds, R. A. Pedersen, Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev. Biol.* **275**, 403–421 (2004).

23. L. Vallier, M. Alexander, R. A. Pedersen, Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* **118**, 4495–4509 (2005).

24. D. F. Quail, G. M. Siegers, M. Jewer, L. M. Postovit, Nodal signalling in embryogenesis and tumourigenesis. *Int. J. Biochem. Cell Biol.* **45**, 885–898 (2013).

25. J. R. Hart, Y. Zhang, L. Liao, L. Ueno, L. Du, M. Jonkers, J. R. Yates, P. K. Vogt, The butterfly effect in cancer: A single base mutation can remodel the cell. *Proc. Natl. Acad. Sci.* **112**, 1131–1136 (2015).

26. V. Y. Kiselev, V. Juvin, M. Malek, N. Luscombe, P. Hawkins, N. Le Novère, L. Stephens, Perturbations of PIP3 signalling trigger a global remodelling of mRNA landscape and reveal a transcriptional feedback loop. *Nucleic Acids Res.* **43**, 9663–79 (2015).

27. V. E. R. Parker, K. M. Keppler-Noreuil, L. Faivre, M. Luu, N. L. Oden, L. De Silva, J. C. Sapp, K. Andrews, M. Bardou, K. Y. Chen, T. N. Darling, E. Gautier, B. R. Goldspiel, S. Hadj-Rabia, J. Harris, G. Kounidas, P. Kumar, M. J. Lindhurst, R. Loffroy, L. Martin, A. Phan, K. I. Rother, B. C. Widemann, P. L. Wolters, C. Coubes, L. Pinson, M. Willems, C. Vincent-Delorme, P. Vabres, R. K. Semple, L. G. Biesecker, Safety and efficacy of low-dose sirolimus in the PIK3CA-related overgrowth spectrum. *Genet. Med.* **In Press**, 1–10 (2018).

28. L. S. Moniz, S. Surinova, E. Ghazaly, L. G. Velasco, S. Haider, J. C. Rodríguez-Prados, I. M. Berenjeno, C. Chelala, B. Vanhaesebroeck, Phosphoproteomic comparison of Pik3ca and Pten signalling identifies the nucleotidase NT5C as a novel AKT substrate. *Sci. Rep.* **7**, 39985 (2017).

29. X. Robin, F. Voellmy, J. Ferkinghoff-Borg, C. Howard, T. Altenburg, M. Engel, C. D. Simpson, G. Saginc, S. Koplev, E. Klipp, J. Longden, R. Linding, Probability-based detection of phosphoproteomic uncertainty reveals rare signaling events driven by oncogenic kinase gene fusion. *bioRxiv*, 621961 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

30. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**, 559 (2008).

31. D. Mesnard, M. Guzman-Ayala, D. B. Constam, Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development*. **133**, 2497–505 (2006).

32. G. Chen, D. R. Gulbranson, Z. Hou, J. M. Bolin, V. Ruotti, M. D. Probasco, K. Smuga-Otto, S. E. Howden, N. R. Diol, N. E. Propson, R. Wagner, G. O. Lee, J. Antosiewicz-Bourget, J. M. C. Teng, J. A. Thomson, Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods*. **8**, 424–429 (2011).

33. S. Vukicevic, H. K. Kleinman, F. P. Luyten, A. B. Roberts, N. S. Roche, A. H. Reddi, Identification of multiple active growth factors in basement membrane Matrigel suggests caution in interpretation of cellular activity related to extracellular matrix components. *Exp. Cell Res.* **202**, 1–8 (1992).

34. C. S. Hill, Spatial and temporal control of NODAL signaling. *Curr. Opin. Cell Biol.* **51**, 50–57 (2018).

35. G. J. Inman, F. J. Nicolás, J. F. Callahan, J. D. Harling, L. M. Gaster, A. D. Reith, N. J. Laping, C. S. Hill, SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* **62**, 65–74 (2002).

36. L. Vallier, S. Mendjan, S. Brown, Z. Chng, A. Teo, L. E. Smithers, M. W. B. Trotter, C. H.-H. Cho, A. Martinez, P. Rugg-Gunn, G. Brons, R. a Pedersen, Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development*. **136**, 1339–49 (2009).

37. M. Monk, C. Holding, Human embryonic genes re-expressed in cancer cells. *Oncogene*. **20**, 8085–8091 (2001).

38. C. J. David, J. Massagué, Contextual determinants of TGFβ action in development, immunity and cancer. *Nat. Rev. Mol. Cell Biol.* **19**, 1–17 (2018).

39. C. Hadjimichael, K. Chanoumidou, N. Papadopoulou, P. Arampatzi, J. Papamatheakis, A. Kretsovali, Common stemness regulators of embryonic and cancer stem cells. *World J Stem*

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cells. **7**, 1150–1184 (2015).

40. N. Oshimori, D. Oristian, E. Fuchs, TGF-β Promotes Heterogeneity and Drug Resistance in Squamous Cell Carcinoma. *Cell*. **160**, 963–976 (2015).

41. T. M. Bodenstine, G. S. Chandler, R. E. B. Seftor, E. A. Seftor, M. J. C. Hendrix, Plasticity underlies tumor progression: role of Nodal signaling. *Cancer Metastasis Rev.* **35**, 21–39 (2016).

42. L. Strizzi, K. M. Hardy, N. V. Margaryan, D. W. Hillman, E. A. Seftor, B. Chen, X. J. Geiger, E. A. Thompson, W. L. Lingle, C. A. Andorfer, E. A. Perez, M. J. C. Hendrix, Potential for the embryonic morphogen Nodal as a prognostic and predictive biomarker in breast cancer. *Breast Cancer Res.* **14** (2012), doi:10.1186/bcr3185.

43. M. Bar-Eli, Back to the embryonic stage: Nodal as a biomarker for breast cancer progression. *Breast Cancer Res.* **14**, 105 (2012).

44. N. V. Margaryan, H. H. Jenkins, M. A. Salkeni, M. B. Smolkin, J. A. Coad, S. Wen, E. A. Seftor, R. E. B. Seftor, M. J. C. Hendrix, The stem cell phenotype of aggressive breast cancer cells. *Cancers (Basel)*. **11**, 1–11 (2019).

45. W. Gong, B. Sun, H. Sun, X. Zhao, D. Zhang, T. Liu, N. Zhao, Q. Gu, X. Dong, F. Liu, Nodal signaling activates the Smad2/3 pathway to regulate stem cell-like properties in breast cancer cells. *Am. J. Cancer Res.* **7**, 503–517 (2017).

46. M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, M. F. Clarke, Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci.* **100**, 3983–3988 (2003).

47. A. E. Giuliano, J. L. Connolly, S. B. Edge, E. A. Mittendorf, H. S. Rugo, L. J. Solin, D. L. Weaver, D. J. Winchester, G. N. Hortobagyi, Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA. Cancer J. Clin.* **67**, 290–303 (2017).

48. I. M. Berenjeno, R. Piñeiro, S. D. Castillo, W. Pearce, N. McGranahan, S. M. Dewhurst, V. Meniel, N. J. Birkbak, E. Lau, L. Sansregret, D. Morelli, N. Kanu, S. Srinivas, M. Graupera, V. E. R. Parker, K. G. Montgomery, L. S. Moniz, C. L. Scudamore, W. A. Phillips, R. K.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Semple, A. Clarke, C. Swanton, B. Vanhaesebroeck, Oncogenic PIK3CA induces centrosome amplification and tolerance to genome doubling. *Nat. Commun.* **8**, 1773 (2017).

49. V. Doldán-Martelli, D. G. Míguez, Drug treatment efficiency depends on the initial state of activation in nonlinear pathways. *Sci. Rep.* **8**, 1–10 (2018).

50. K. Hussain, B. Challis, N. Rocha, F. Payne, M. Minic, A. Thompson, A. Daly, C. Scott, J. Harris, B. J. L. Smillie, D. B. Savage, U. Ramaswami, P. De Lonlay, S. O'Rahilly, I. Barroso, R. K. Semple, An Activating Mutation of AKT2 and Human Hypoglycemia. *Science (80-.).* **334**, 474–474 (2011).

51. D. J. Murphy, M. R. Junttila, L. Pouyet, A. Karnezis, K. Shchors, D. A. Bui, L. Brown-Swigart, L. Johnson, G. I. Evan, Distinct Thresholds Govern Myc's Biological Output In Vivo. *Cancer Cell*. **14**, 447–457 (2008).

52. M. E. Beaulieu, T. Jauset, D. Massó-Vallés, S. Martínez-Martín, P. Rahl, L. Maltais, M. F. Zacarias-Fluck, S. Casacuberta-Serra, E. S. Del Pozo, C. Fiore, L. Foradada, V. C. Cano, M. Sánchez-Hervás, M. Guenther, E. R. Sanz, M. Oteo, C. Tremblay, G. Martín, D. Letourneau, M. Montagne, M. Á. M. Alonso, J. R. Whitfield, P. Lavigne, L. Soucek, Intrinsic cell-penetrating activity propels omomyc from proof of concept to viable anti-myc therapy. *Sci. Transl. Med.* **11**, 1–14 (2019).

53. S. Koren, L. Reavie, J. P. Couto, D. De Silva, M. B. Stadler, T. Roloff, A. Britschgi, T. Eichlisberger, H. Kohler, O. Aina, R. D. Cardiff, M. Bentires-Alj, PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. *Nature*. **525**, 114–8 (2015).

54. A. Van Keymeulen, M. Y. Lee, M. Ousset, S. Brohée, S. Rorive, R. R. Giraddi, A. Wuidart, G. Bouvencourt, C. Dubois, I. Salmon, C. Sotiriou, W. a. Phillips, C. Blanpain, Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature*. **525**, 119–23 (2015).

55. T. Brabletz, R. Kalluri, M. A. Nieto, R. A. Weinberg, EMT in cancer. *Nat. Rev. Cancer*. **18**, 128–134 (2018).

56. Y. Katsuno, D. S. Meyer, Z. Zhang, K. M. Shokat, R. J. Akhurst, K. Miyazono, R. Derynck, Chronic TGF-b exposure drives stabilized EMT, tumor stemness, and cancer drug resistance

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with vulnerability to bitopic mTOR inhibition. *Sci. Signal*. **12**, eaau8544 (2019).

57. G. Xue, D. F. Restuccia, Q. Lan, D. Hynx, S. Dirnhofer, D. Hess, C. Rüegg, B. A. Hemmings, Akt/PKB-Mediated Phosphorylation of Twist1 Promotes Tumor Metastasis via Mediating Cross-Talk between PI3K/Akt and TGF-β Signaling Axes. *Cancer Discov.* **2**, 248–259 (2012).

58. A. Chakrabarty, S. Surendran, N. E. Bhola, V. S. Mishra, T. H. Wani, K. S. Baghel, C. L. Arteaga, R. Garg, G. Chowdhury, The H1047R PIK3CA oncogene induces a senescence-like state, pleiotropy and acute HSP90 dependency in HER2+ mammary epithelial cells. *Carcinogenesis* (2019), doi:10.1093/carcin/bgz118.

59. C. Thomas, W. Henry, B. G. Cuiffo, A. Y. Collmann, E. Marangoni, V. Benhamo, M. K. Bhasin, C. Fan, L. Fuhrmann, A. S. Baldwin, C. Perou, A. Vincent-Salomon, A. Toker, A. E. Karnoub, Pentraxin-3 is a PI3K signaling target that promotes stem cell–like traits in basal-like breast cancers. *Sci. Signal.* **10**, eaah4674 (2017).

60. F. Janku, T. A. Yap, F. Meric-Bernstam, Targeting the PI3K pathway in cancer: are we making headway? *Nat. Rev. Clin. Oncol.* **15**, 273–291 (2018).

61. L. Zhang, F. Zhou, P. ten Dijke, Signaling interplay between transforming growth factor-β receptor and PI3K/AKT pathways in cancer. *Trends Biochem. Sci.* **38**, 612–620 (2013).

62. L. Du, X. Chen, Y. Cao, L. Lu, F. Zhang, S. Bornstein, Y. Li, P. Owens, S. Malkoski, S. Said, F. Jin, M. Kulesz-Martin, N. Gross, X.-J. Wang, S.-L. Lu, Overexpression of PIK3CA in murine head and neck epithelium drives tumor invasion and metastasis through PDK1 and enhanced TGFβ signaling. *Oncogene*. **35**, 4641–4652 (2016).

63. E. H. Wilkes, C. Terfve, J. G. Gribben, J. Saez-Rodriguez, P. R. Cutillas, Empirical inference of circuitry and plasticity in a kinase signaling network. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7719–24 (2015).

64. B. Samyn, K. Sergeant, P. Castanheira, C. Faro, J. Van Beeumen, Pathway and network analysis of cancer genomes. *Nat. Methods*. **12**, 615–621 (2015).

65. M. Uhlen, C. Zhang, S. Lee, E. Sjöstedt, L. Fagerberg, G. Bidkhori, R. Benfeitas, M. Arif, Z. Liu, F. Edfors, K. Sanli, K. von Feilitzen, P. Oksvold, E. Lundberg, S. Hober, P. Nilsson, J. Mattsson, J. M. Schwenk, H. Brunnström, B. Glimelius, T. Sjöblom, P.-H. Edqvist, D.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Djureinovic, P. Micke, C. Lindskog, A. Mardinoglu, F. Ponten, A pathology atlas of the human cancer transcriptome. *Science (80-.).* **357**, eaan2507 (2017).

66. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

67. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

68. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2009), doi:10.1093/bioinformatics/btp616.

69. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

70. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B-Methodological*. **57**, 289–300 (1995).

71. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (2014), doi:10.1093/bioinformatics/btu170.

72. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. **26**, 589–595 (2010).

73. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

74. M. A. Depristo, E. Banks, R. Poplin, K. V Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–501 (2011).

75. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cunningham, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

76. S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).

77. K. G. Macleod, B. Serrels, N. O. Carragher, *Proteomics for Drug Discovery* (Springer New York, New York, NY, 2017; http://link.springer.com/10.1007/978-1-4939-7201-2), vol. 1647 of *Methods in Molecular Biology*.

78. A. Krämer, J. Green, J. Pollard, S. Tugendreich, Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*. **30**, 523–530 (2014).

79. B. Zhang, S. Horvath, A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* **4** (2005), doi:10.2202/1544-6115.1128.

80. G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. **25**, 1091–1093 (2009).

81. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).

82. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–504 (2003).

83. A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).

84. R. L. Wasserstein, A. L. Schirm, N. A. Lazar, Moving to a World Beyond " $p < 0.05$." *Am. Stat.* **73**, 1–19 (2019).

85. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ternent, A. Brazma, J. A. Vizcaíno, The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* (2019), doi:10.1093/nar/gky1106.

Acknowledgements: We thank Dominique McCormick and Ineke Luijten for help with RNA extraction and cDNA synthesis, Cornelia Gewert for technical support, and Marcella Ma, Brian Lam and Michelle Dietzen for technical support with RNA sequencing and genomics analyses, respectively. We thank Evelyn K. Lau for help with GEO upload of the MEF RNAseq data. We are grateful to Prof Siddhartan Chandran and his group for providing space for iPSC culturing and to Pamela Brown (SURF Biomolecular Core, University of Edinburgh) for access to qPCR facilities. **Funding:** R.R.M. and R.K.S. are supported by the Wellcome Trust (105371/Z/14/Z, 210752/Z/18/Z) and United Kingdom (UK) NIHR Cambridge Biomedical Research Centre, and R.R.M. by a Boak Student Award from Clare Hall. B.V. is supported by Cancer Research UK (C23338/ A25722), PTEN Research, and the UK NIHR University College London Hospitals Biomedical Research Centre. Metabolic Research Laboratories Core facilities are supported by the Medical Research Council Metabolic Diseases Unit (MC_UU_12012/5) and a Wellcome Major Award (208363/Z/17/Z). N.M. is Sir Henry Dale Fellow funded by the Wellcome Trust and the Royal Society (211179/Z/18/Z). N.M. receives funding from Cancer Research UK, and Rosetrees Trust, the NIHR BRC at University College London Hospitals, and the CRUK University College London Experimental Cancer Medicine Centre. R.L. is funded by a Lundbeck Foundation Fellowship. K.G.M., N.O.C. and the University of Edinburgh RPPA facility is supported by a Cancer Research UK Centre award. **Author contributions:** Overall conceptualization and study design was by R.R.M. and R.K.S., with important contributions from B.V. (*PIK3CA* cancer biology and MEF data), R.L. and J.L. (total proteomics and WGCNA). R.R.M. and R.G.K. performed all hPSC experiments. F.V. performed the mass spectrometry experiments., and X.R. performed MCMC computational analysis. L.M. performed the MEF experiment. K.M. and N.C. were responsible for RPPA design and sample processing, and R.R.M carried out statistical analysis. R.R.M. performed all RNA sequencing quantitation and IPA analyses on transcriptomic and proteomic datasets. R.R.M. performed the TCGA breast cancer data analysis with supervision from N.M. R.R.M., X.R., J.L., F.V. and R.G.K. were responsible for data curation. R.R.M. and R.K.S. wrote the manuscript. B.V., R.L., J.L., N.C., N.M., X.R., L.M., X.R. and F.V. reviewed and edited the final version. **Competing interests:** B.V. is a consultant for Karus Therapeutics (Oxford, UK), iOnctura

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Geneva, Switzerland) and Venthera (Palo Alto, US) and has received speaker fees from Gilead Sciences (Foster City, US). N.O.C. is a director of Ther-IP Ltd (Edinburgh, UK) and founder shareholder and advisor for PhenoTherapeutics Ltd (Edinburgh, UK) and a member of the advisory board and shareholder of Amplia Therapeutics Ltd (Melbourne, Australia). **Data and materials availability:** Raw data and bespoke RNotebooks containing guided scripts and plots are available *via* the Open Science Framework (public link to be added upon acceptance; view-only link to be provided to reviewers). Individual scripts include information on the name and version of applied R packages. The original RNAseq data have been deposited to the Gene Expression Omnibus (GEO), under accession numbers: GSE134076 (H1047R iPSC data), GSEXXXXXX (under submission), GSE135046 (MEF data). View-only link to be provided to reviewers before public release. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (*85*) partner repository with the dataset identifier PXD014719 (password to be provided to reviewers before public release). Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding authors, Ralitsa R. Madsen (rmadsen@ed.ac.uk) or Robert K. Semple (rsemple@ed.ac.uk).

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Legends

Fig. 1. Transcriptomic analyses of human and mouse cell lines with endogenous expression of oncogenic *PIK3CA*. (A) Multidimensional scaling (MDS) plot of the transcriptomes of wild-type (WT), *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) human iPSCs. The numbers in brackets indicate the presence of two closely overlapping samples. **(B)** The number of differentially expressed genes in iPSCs heterozygous or homozygous for *PIK3CA*^{H1047R} before and after application of an absolute fold-change cut-off $\geq$ 1.3 (FDR $\leq$ 0.05, Benjamini-Hochberg). The data are based on four iPSCs cultures from minimum two clones per genotype. See also fig. S1. **(C)** Woman with asymmetric overgrowth caused by mosaicism for cells with heterozygous expression of *PIK3CA*^{E418K}. Skin biopsies obtained from unaffected and affected tissues were used to obtain otherwise isogenic dermal fibroblasts for subsequent reprogramming into iPSCs. This image was reproduced from Ref. (*27*) **(D)** MDS plot of the transcriptomes of wild-type (WT) and *PIK3CA*$^{WT/E418K}$ iPSCs (based on 3 independent mutant clones and 3 wild-type cultures from 2 independent clones). **(E)** MDS plot of the transcriptomes of wild-type (WT) and *PIK3CA*$^{WT/H1047R}$ (HET) mouse embryonic fibroblasts (MEFs) following 48 h of mutant induction (n = 4 independent clones per genotype). **(F)** Venn diagram showing the number of differentially expressed proteins in *PIK3CA*$^{H1047R/H1047R}$ (HOM) and *PIK3CA*$^{WT/H1047R}$ (HET) iPSCs relative to wild-type controls, profiled by label-free total proteomics on three clones per genotype. An absolute fold-change and z-score $\geq$ 1.2 were used to classify proteins as differentially expressed. The number of discordant and concordant changes in the expression of total proteins detected in both comparisons are indicated. See also fig. S2.

Fig. 2. Reverse Phase Protein Array (RPPA) of *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) human iPSCs. (A) Left: Diagram of PI3K pathway-related phosphorylated proteins, with color code used to signify differentially expressed targets in *PIK3CA*^{H1047R} mutant iPSCs *versus* isogenic wild-type controls. Color-coded targets were significant at FDR $\leq$ 0.05 (Benjamini-Hochberg). Right: Barplots show representative examples of differentially expressed phosphorylated protein targets, revealing relatively modest quantitative changes. Phosphorylated proteins were normalized to the corresponding total protein when available. The data are based on 10 wild-type cultures (3 clones), 5 *PIK3CA*$^{WT/H1047R}$ cultures (3 clones) and 7 *PIK3CA*$^{H1047R/H1047R}$ cultures (2 clones) as indicated. See also fig. S3A. **(B)** Unsupervised hierarchical clustering based on target-wise correlations of RPPA

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

data from wild-type (WT), *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) iPSCs following short-term growth factor removal (1 h), +/- 100 nM BYL719 (PI3Kα inhibitor) for 24 h. The data are from two independent experiments, each performed using independent clones. For each row, the colors correspond to Fast Green-normalized expression values in units of standard deviation (z-score) from the mean (centered at 0) across all samples (columns). Groups of phosphorylated proteins exhibiting a consistent expression pattern in BYL719-treated *PIK3CA*$^{H1047R/H1047R}$ iPSCs are specified. See also fig. S3B.

Fig. 3. Ingenuity® pathway analyses (IPA) predict activation of TGFβ signaling in heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs. (A) IPA of upstream regulators using the list of top 2000 upregulated and top 2000 downregulated mRNA transcripts in *PIK3CA*$^{H1047R/H1047R}$ iPSCs (for RNAseq details, see Fig. 1B). Red points signify transcripts with absolute predicted activation z-score > 2 and overlap p-value < 0.001 (Fisher's Exact Test). The red rectangle highlights the most significant upstream regulator, TGFβ1. **(B)** As in (A), but using the list of differentially-expressed proteins identified by total proteomics and red-coloring targets with predicted activation z-score > 2 and overlap p-value < 0.05 (Fisher's Exact Test). **(C)** As in (A), but using the list of differentially expressed total proteins in *PIK3CA*$^{WT/H1047R}$ iPSCs and red-coloring upstream regulators with absolute predicted bias-corrected z score > 2 and overlap p-value < 0.05 (Fisher's Exact Test). Red rectangles highlight the two upstream regulators (TGFβ1 and MAPK1) with absolute predicted bias-corrected z score > 2 that remained significant (overlap p-value < 0.05) when the analysis was repeated using the list of shared and concordant differentially expressed genes (n = 180) in heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs *vs* wild-type controls.

Fig. 4. Weighted gene correlation network analysis (WGCNA) identifies links among pluripotency components, TGFβ and PI3K signaling. (A) Schematic of the WGCNA workflow and subsequent data selection for visualization. (**B)** Unsigned WGCNA modules identified using the list of transcripts expressed in wild-type, *PIK3CA*$^{WT/H1047R}$ and *PIK3CA*$^{H1047R/H1047R}$ iPSCs (for RNAseq details, see Fig. 1B). **(C)** The two gene network modules with genes whose module membership correlated strongest with differential expression in homozygous *PIK3CA*^{H1047R} iPSCs. The color of each module corresponds to its color in the module dendrogram in (B). Representative KEGG pathways with significant enrichment in each gene network module are listed

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(hypergeometric test with two-sided uncorrected $p < 0.05$). **(D)** The minimal network connecting KEGG pluripotency, PI3K/AKT and TGFβ pathway components within the turquoise gene network module. Fill color and shape are used to specify direction of differential mRNA expression in *PIK3CA*$^{H1047R/H1047R}$ iPSCs and pathway membership, respectively. Fill color saturation represents gene expression fold-change (FC; log2) in *PIK3CA*$^{H1047R/H1047R}$ (HOM) *vs* wild-type (WT) iPSCs.

Fig. 5. TGFβ signaling-dependent regulation of stemness in *PIK3CA*$^{H1047R/H1047R}$ iPSCs. Gene expression time course of *NODAL*, *NANOG* and *POU5F1* in wild-type (WT) or *PIK3CA*$^{H1047R/H1047R}$ iPSCs following the indicated treatments for 24 h, 48 h or 72 h. B250: 250 nM BYL719 (PI3Kα-selective inhibitor); E6/FGF2: Essential 6 medium supplemented with 10 ng/ml basic fibroblast growth factor 2 (FGF2). SB431542 is a specific inhibitor of the NODAL type I receptors ALK4/7 and the TGFβ type I receptor ALK5; used at 5 μM. When indicated, cultures were supplemented with 100 ng/ml NODAL. The data are from two independent experiments, with each treatment applied to triplicate cultures of three wild-type and two homozygous iPSC clones. To aid interpretation, gene expression values are normalized to the E6/FGF2 condition within each genotype and time-point. An alternative visualization that illustrates the differential expression of *NODAL* and *NANOG* between mutant and wild-type cells is shown in fig. S4A. For analysis of additional lineage markers, see fig. S4B. A.U., arbitrary units.

Fig. 6. Stage II human primary breast tumors with multiple *PIK3CA*^{H1047R} copies have increased *NODAL* gene expression. (A) Violin plots of *NODAL*, *POU5F1*, *NANOG* and *MYC* mRNA expression in counts per million (cpm) in *PIK3CA*-associated human primary breast tumors (n = 108) stratified according to the presence of a single (41 tumors) or multiple copies (67 tumors) of *PIK3CA*^{H1047R} and/or a second *PIK3CA* mutation. **(B)** Boxplots of the data in (a) stratified according to tumor stage information if available. The number of samples containing a single *PIK3CA*^{H1047R} allele and tumor stage information were: 9 (stage i), 21 (stage ii) and 5 (stage iii); the corresponding numbers in samples with multiple copies were: 9 (stage i), 42 (stage ii) and 14 (stage iii).

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figures

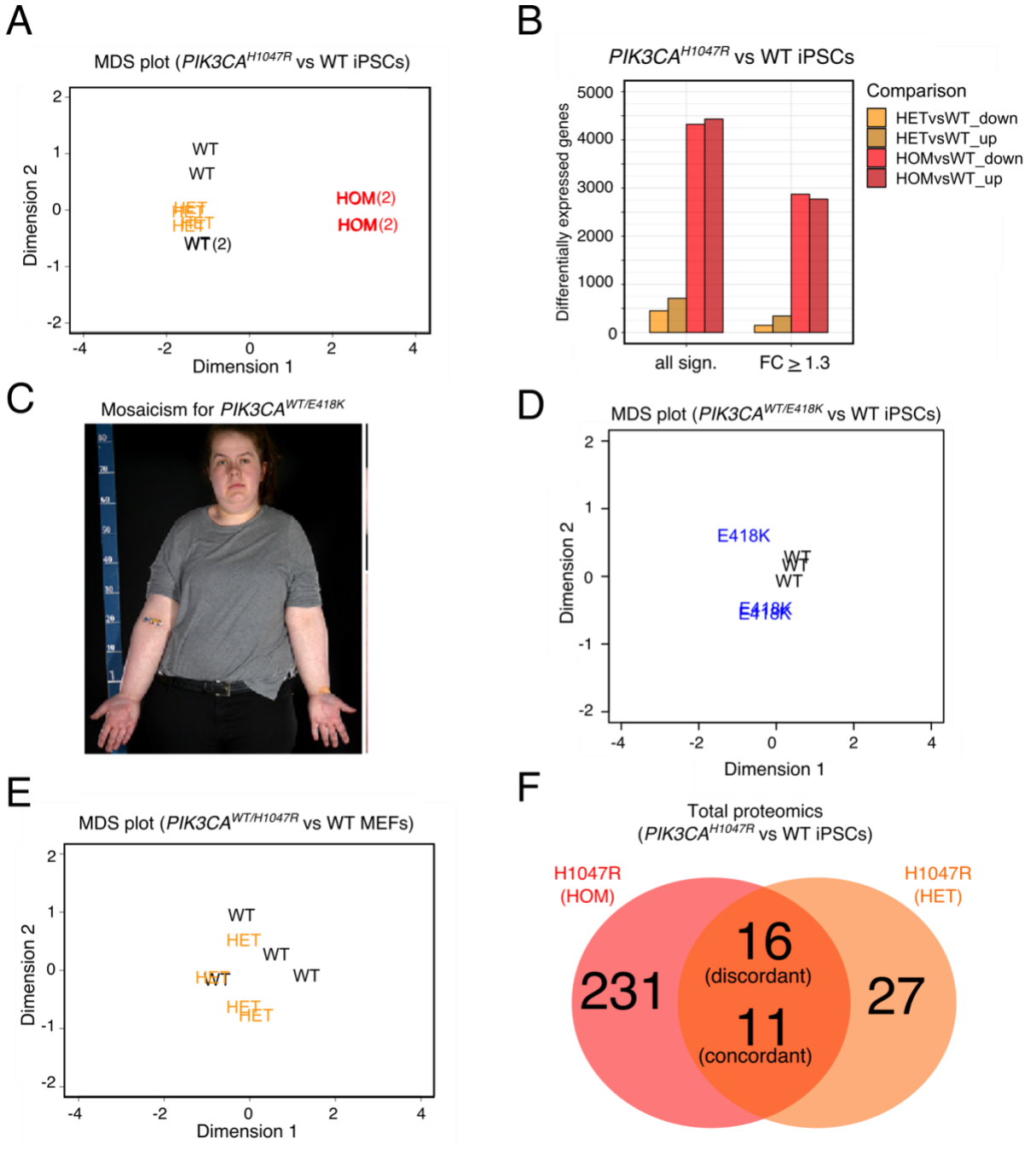

Fig. 1. Transcriptomic analyses of human and mouse cell lines with endogenous expression of oncogenic *PIK3CA*. (A) Multidimensional scaling (MDS) plot of the transcriptomes of wild-type (WT), *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) human iPSCs. The numbers in brackets indicate the presence of two closely overlapping samples. **(B)** The number of differentially expressed genes in iPSCs heterozygous or homozygous for *PIK3CA*^{H1047R} before and after application of an absolute fold-change cut-off $\geq$ 1.3 (FDR $\leq$ 0.05, Benjamini-

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hochberg). The data are based on four iPSCs cultures from minimum two clones per genotype. See also fig. S1. **(C)** Woman with asymmetric overgrowth caused by mosaicism for cells with heterozygous expression of *PIK3CA*^{E418K}. Skin biopsies obtained from unaffected and affected tissues were used to obtain otherwise isogenic dermal fibroblasts for subsequent reprogramming into iPSCs. This image was reproduced from Ref. (*27*) **(D)** MDS plot of the transcriptomes of wild-type (WT) and *PIK3CA*$^{WT/E418K}$ iPSCs (based on 3 independent mutant clones and 3 wild-type cultures from 2 independent clones). **(E)** MDS plot of the transcriptomes of wild-type (WT) and *PIK3CA*$^{WT/H1047R}$ (HET) mouse embryonic fibroblasts (MEFs) following 48 h of mutant induction (n = 4 independent clones per genotype). **(F)** Venn diagram showing the number of differentially expressed proteins in *PIK3CA*$^{H1047R/H1047R}$ (HOM) and *PIK3CA*$^{WT/H1047R}$ (HET) iPSCs relative to wild-type controls, profiled by label-free total proteomics on three clones per genotype. An absolute fold-change and z-score $\geq$ 1.2 were used to classify proteins as differentially expressed. The number of discordant and concordant changes in the expression of total proteins detected in both comparisons are indicated. See also fig. S2.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

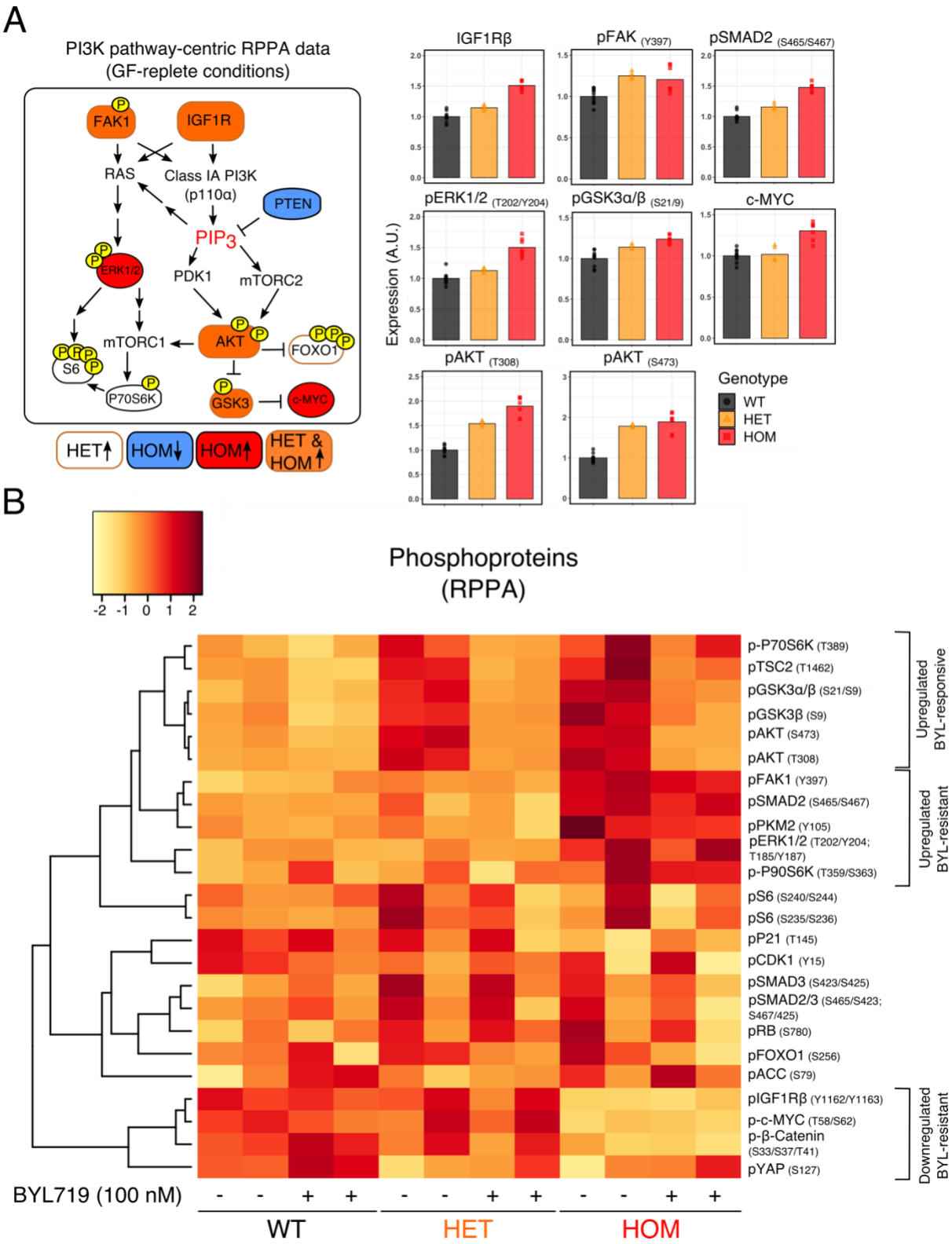

Fig. 2. Reverse Phase Protein Array (RPPA) of *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) human iPSCs. (A) Left: Diagram of PI3K pathway-related phosphorylated proteins, with color code used to signify differentially expressed targets in *PIK3CA*^{H1047R} mutant iPSCs *versus* isogenic wild-type controls. Color-coded

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

targets were significant at FDR $\leq$ 0.05 (Benjamini-Hochberg). Right: Barplots show representative examples of differentially expressed phosphorylated protein targets, revealing relatively modest quantitative changes. Phosphorylated proteins were normalized to the corresponding total protein when available. The data are based on 10 wild-type cultures (3 clones), 5 *PIK3CA*$^{WT/H1047R}$ cultures (3 clones) and 7 *PIK3CA*$^{H1047R/H1047R}$ cultures (2 clones) as indicated. See also fig. S3A. **(B)** Unsupervised hierarchical clustering based on target-wise correlations of RPPA data from wild-type (WT), *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) iPSCs following short-term growth factor removal (1 h), +/- 100 nM BYL719 (PI3Kα inhibitor) for 24 h. The data are from two independent experiments, each performed using independent clones. For each row, the colors correspond to Fast Green-normalized expression values in units of standard deviation (z-score) from the mean (centered at 0) across all samples (columns). Groups of phosphorylated proteins exhibiting a consistent expression pattern in BYL719-treated *PIK3CA*$^{H1047R/H1047R}$ iPSCs are specified. See also fig. S3B.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

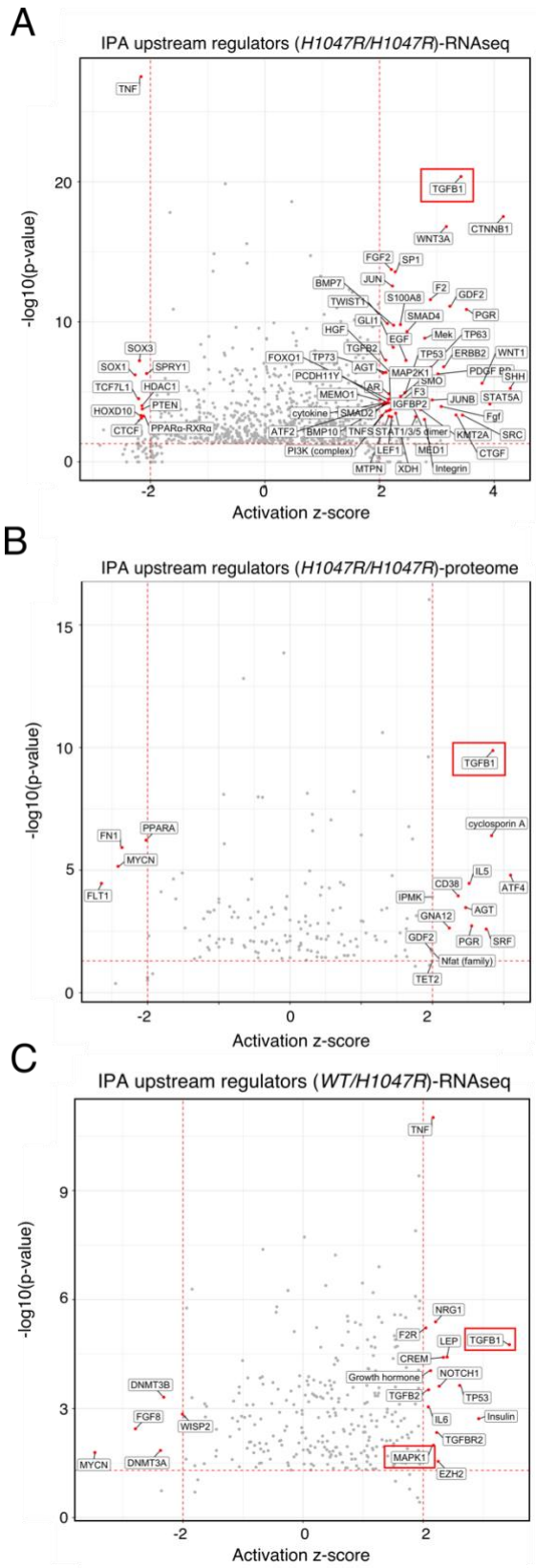

Fig. 3. Ingenuity® pathway analyses (IPA) predict activation of TGFβ signaling in heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs. (A) IPA of upstream regulators using the list of top 2000 upregulated and top

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2000 downregulated mRNA transcripts in *PIK3CA*$^{H1047R/H1047R}$ iPSCs (for RNAseq details, see Fig. 1B). Red points signify transcripts with absolute predicted activation z-score > 2 and overlap p-value < 0.001 (Fisher's Exact Test). The red rectangle highlights the most significant upstream regulator, TGFβ1. **(B)** As in (A), but using the list of differentially-expressed proteins identified by total proteomics and red-coloring targets with predicted activation z-score > 2 and overlap p-value < 0.05 (Fisher's Exact Test). **(C)** As in (A), but using the list of differentially expressed total proteins in *PIK3CA*$^{WT/H1047R}$ iPSCs and red-coloring upstream regulators with absolute predicted bias-corrected z score > 2 and overlap p-value < 0.05 (Fisher's Exact Test). Red rectangles highlight the two upstream regulators (TGFβ1 and MAPK1) with absolute predicted bias-corrected z score > 2 that remained significant (overlap p-value < 0.05) when the analysis was repeated using the list of shared and concordant differentially expressed genes (n = 180) in heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs *vs* wild-type controls.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

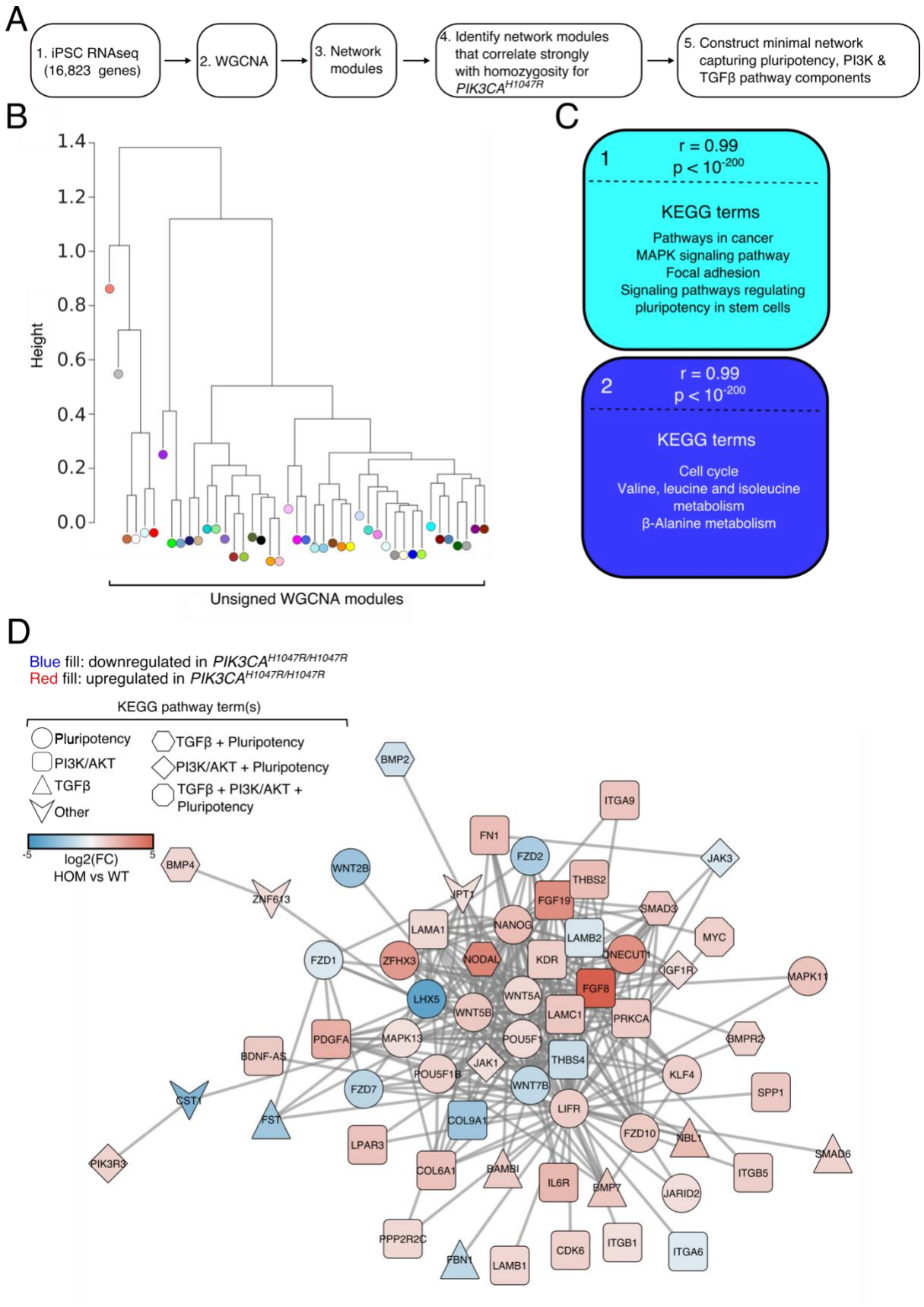

Fig. 4. Weighted gene correlation network analysis (WGCNA) identifies links among pluripotency components, TGFβ and PI3K signaling. (A) Schematic of the WGCNA workflow and subsequent data selection

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

for visualization. (**B**) Unsigned WGCNA modules identified using the list of transcripts expressed in wild-type, *PIK3CA*$^{WT/H1047R}$ and *PIK3CA*$^{H1047R/H1047R}$ iPSCs (for RNAseq details, see Fig. 1B). (**C**) The two gene network modules with genes whose module membership correlated strongest with differential expression in homozygous *PIK3CA*^{H1047R} iPSCs. The color of each module corresponds to its color in the module dendrogram in (B). Representative KEGG pathways with significant enrichment in each gene network module are listed (hypergeometric test with two-sided uncorrected $p < 0.05$). (**D**) The minimal network connecting KEGG pluripotency, PI3K/AKT and TGFβ pathway components within the turquoise gene network module. Fill color and shape are used to specify direction of differential mRNA expression in *PIK3CA*$^{H1047R/H1047R}$ iPSCs and pathway membership, respectively. Fill color saturation represents gene expression fold-change (FC; log2) in *PIK3CA*$^{H1047R/H1047R}$ (HOM) *vs* wild-type (WT) iPSCs.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

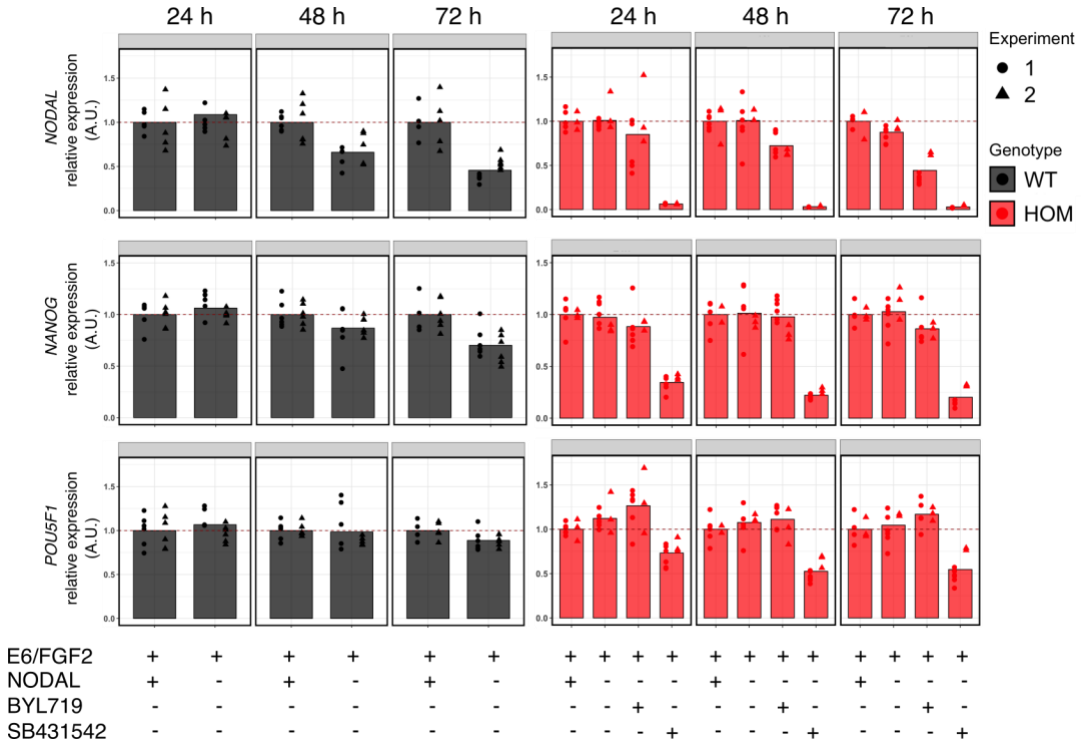

Fig. 5. TGFβ signaling-dependent regulation of stemness in *PIK3CA*$^{H1047R/H1047R}$ iPSCs. Gene expression time course of *NODAL*, *NANOG* and *POU5F1* in wild-type (WT) or *PIK3CA*$^{H1047R/H1047R}$ iPSCs following the indicated treatments for 24 h, 48 h or 72 h. B250: 250 nM BYL719 (PI3Kα-selective inhibitor); E6/FGF2: Essential 6 medium supplemented with 10 ng/ml basic fibroblast growth factor 2 (FGF2). SB431542 is a specific inhibitor of the NODAL type I receptors ALK4/7 and the TGFβ type I receptor ALK5; used at 5 μM. When indicated, cultures were supplemented with 100 ng/ml NODAL. The data are from two independent experiments, with each treatment applied to triplicate cultures of three wild-type and two homozygous iPSC clones. To aid interpretation, gene expression values are normalized to the E6/FGF2 condition within each genotype and time-point. An alternative visualization that illustrates the differential expression of *NODAL* and *NANOG* between mutant and wild-type cells is shown in fig. S4A. For analysis of additional lineage markers, see fig. S4B. A.U., arbitrary units.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

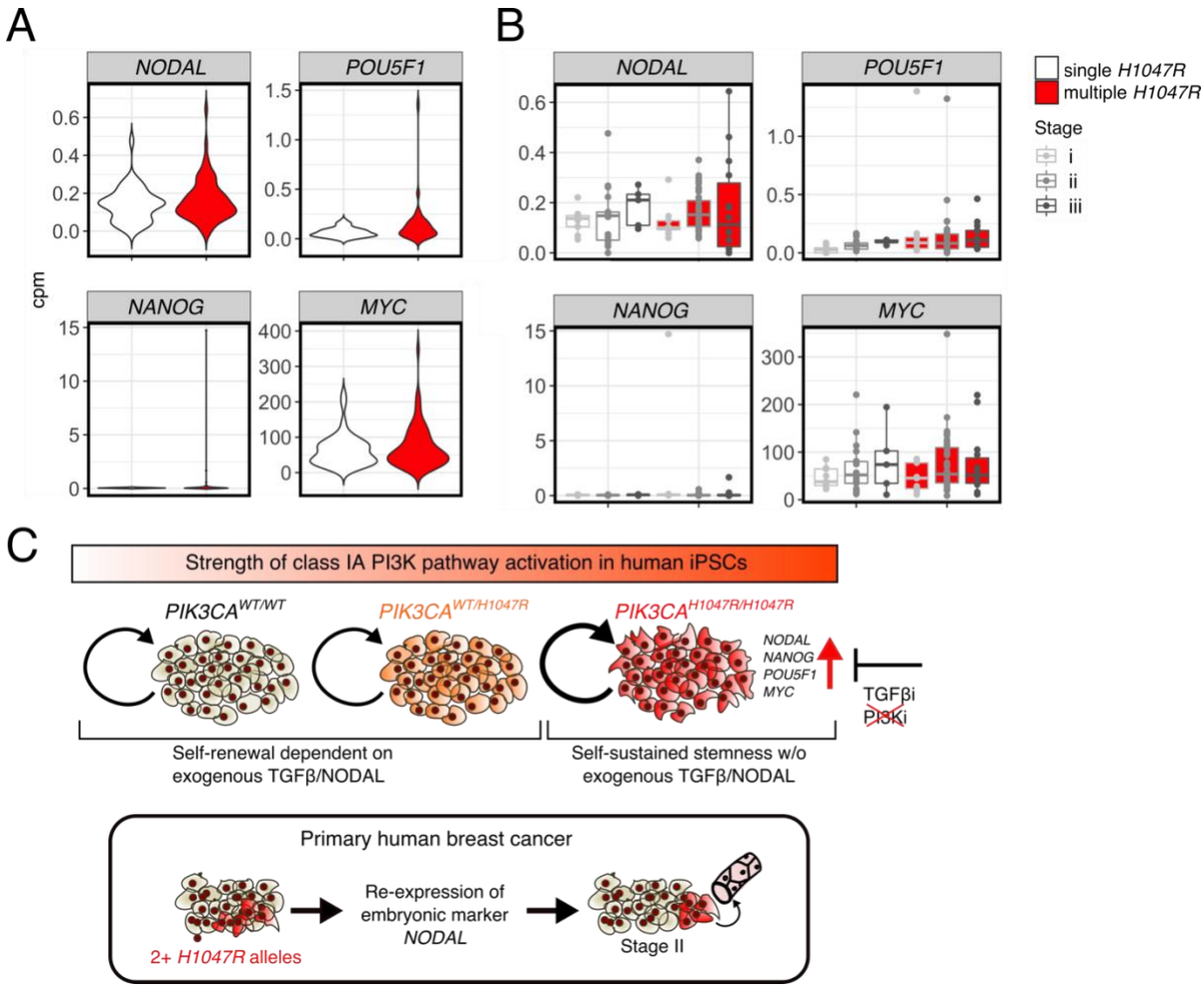

Fig. 6. Stage II human primary breast tumors with multiple *PIK3CA*^{H1047R} copies have increased *NODAL* gene expression. (A) Violin plots of *NODAL*, *POU5F1*, *NANOG* and *MYC* mRNA expression in counts per million (cpm) in *PIK3CA*-associated human primary breast tumors (n = 108) stratified according to the presence of a single (41 tumors) or multiple copies (67 tumors) of *PIK3CA*^{H1047R} and/or a second *PIK3CA* mutation. **(B)** Boxplots of the data in (a) stratified according to tumor stage information if available. The number of samples containing a single *PIK3CA*^{H1047R} allele and tumor stage information were: 9 (stage i), 21 (stage ii) and 5 (stage iii); the corresponding numbers in samples with multiple copies were: 9 (stage i), 42 (stage ii) and 14 (stage iii).

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Material for

High-dose oncogenic *PIK3CA* drives constitutive cellular stemness through self-sustained TGFβ pathway activation

Ralitsa R. Madsen*, James Longden, Rachel G. Knox, Xavier Robin, Franziska Völlmy, Kenneth G. Macleod, Larissa Moniz, Neil O. Carragher, Nicholas McGranahan, Rune Linding, Bart Vanhaesebroeck, Robert K. Semple*

* Corresponding author. Email: rmadsen@ed.ac.uk (R.R.M.); rsemple@ed.ac.uk (R.K.S.)

This document includes:

Fig. S1, related to Fig. 1. Fold-change distribution and transcriptome correlations.

Fig. S2, related to Fig. 1. Transcriptome-proteome correlations.

Fig. S3, related to Fig. 2. Additional RPPA data.

Fig. S4, related to Fig. 5. Alternative representation of the experimental data in Fig. 5A and additional RT-qPCR-based profiling of lineage-specific markers.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

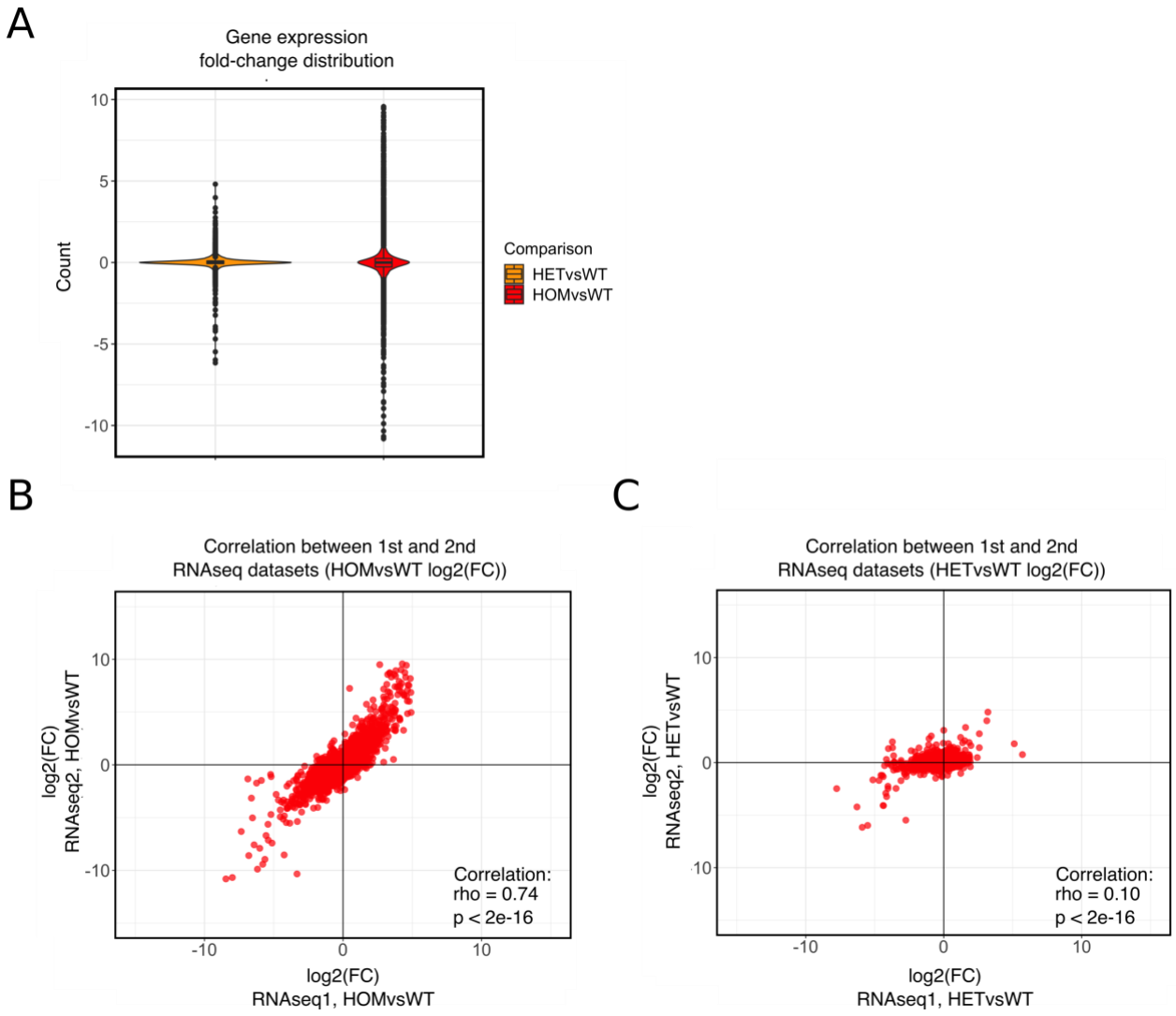

Fig. S1, related to Fig. 1. Fold-change distribution and transcriptome correlations. (A) Combined violin-boxplot representation of the fold-change (log2) distribution of gene expression changes between heterozygous and homozygous *PIK3CA*^{H1047R} and wild-type iPSCs as indicated. **(B) and (C)** Correlation plot of the log2 expression fold-changes (FC) of significant total proteins *vs* the corresponding mRNA transcripts in *PIK3CA*$^{H1047R/H1047R}$ (HOM) (**B**) and *PIK3CA*$^{WT/H1047R}$ (HET) **(C)** iPSCs. Spearman's rho and the corresponding p-values are indicated on all plots. The first transcriptomic study (RNAseq1) was based on three independent cultures from three different clones per genotype; the second transcriptomic study (RNAseq2) used four independent cultures per genotype, from minimum two independent clones each.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

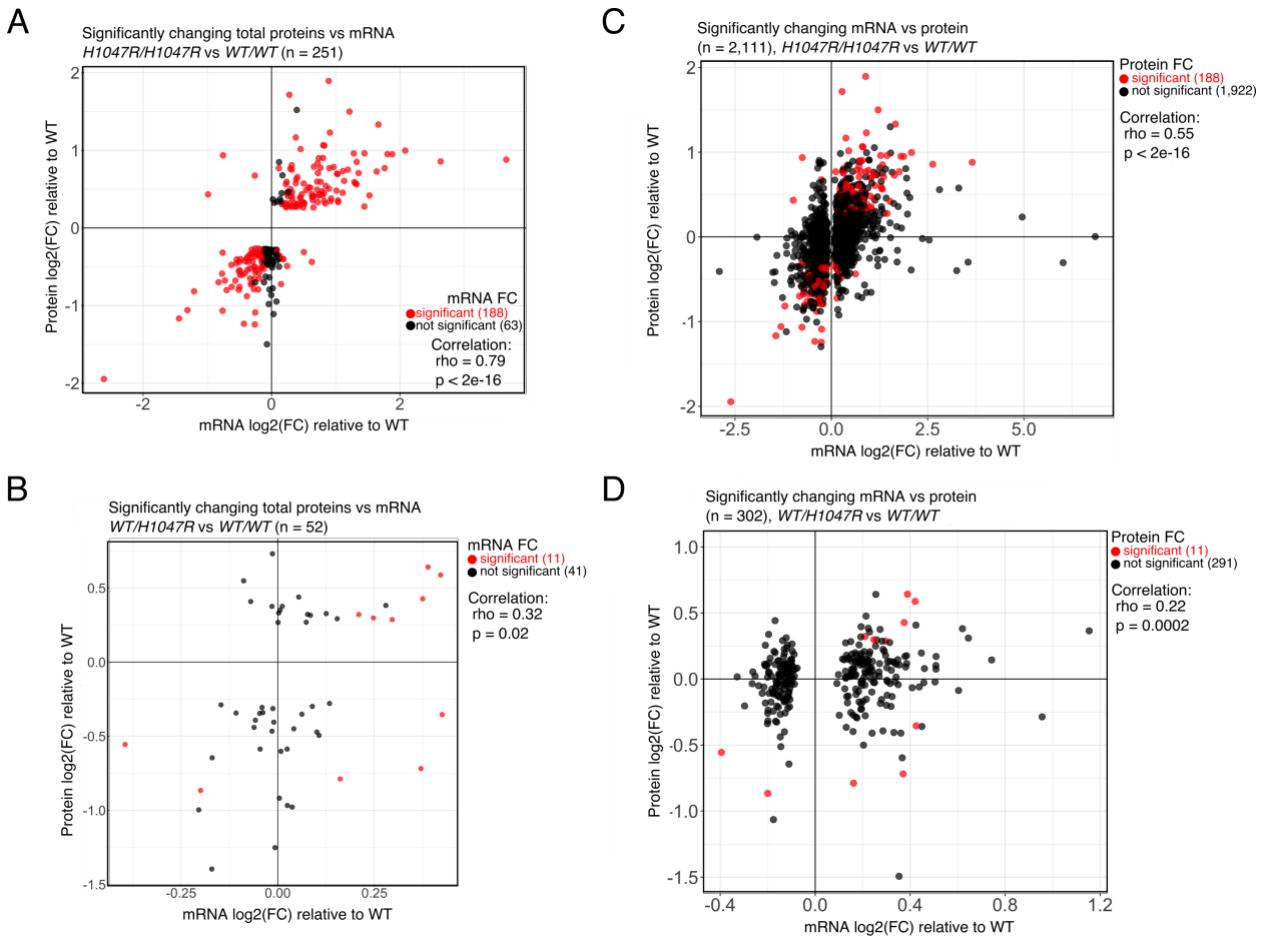

Fig. S2, related to Fig. 1. Transcriptome-proteome correlations. (A) Correlation plot of the log2 expression fold-changes (FC) of differentially expressed total proteins ($|z| \geq 1.2$ and $|\ln(FC)| \geq 1.2$) in *PIK3CA*$^{H1047R/H1047R}$ iPSCs and the corresponding mRNA transcripts in an independent set of cultures. If identified as differentially expressed following statistical analysis (FDR $\leq$ 0.05), mRNA transcripts are highlighted in red irrespective of absolute fold-change. **(B)** As in (A), but starting with all differentially expressed mRNA transcripts (FDR $\leq$ 0.05 irrespective of fold-change magnitude) and plotting them to the corresponding protein identified by total proteomics. If differentially expressed (see (a)), the matched proteins are highlighted in red. **(C) and (D)** As in (A) and (B), respectively, but using the data for *PIK3CA*$^{WT/H1047R}$ iPSCs. All proteomic data were obtained from 3 independent clones per genotype using cultures at passages P47-P52, corresponding to the cultures used in our previous study (Madsen et al., 2019). The high-depth transcriptomic data were obtained from 4 independent cultures (minimum 2 independent clones) per genotype using cultures at passages P55-P59. Spearman's rho and the corresponding p-values are indicated on all plots.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

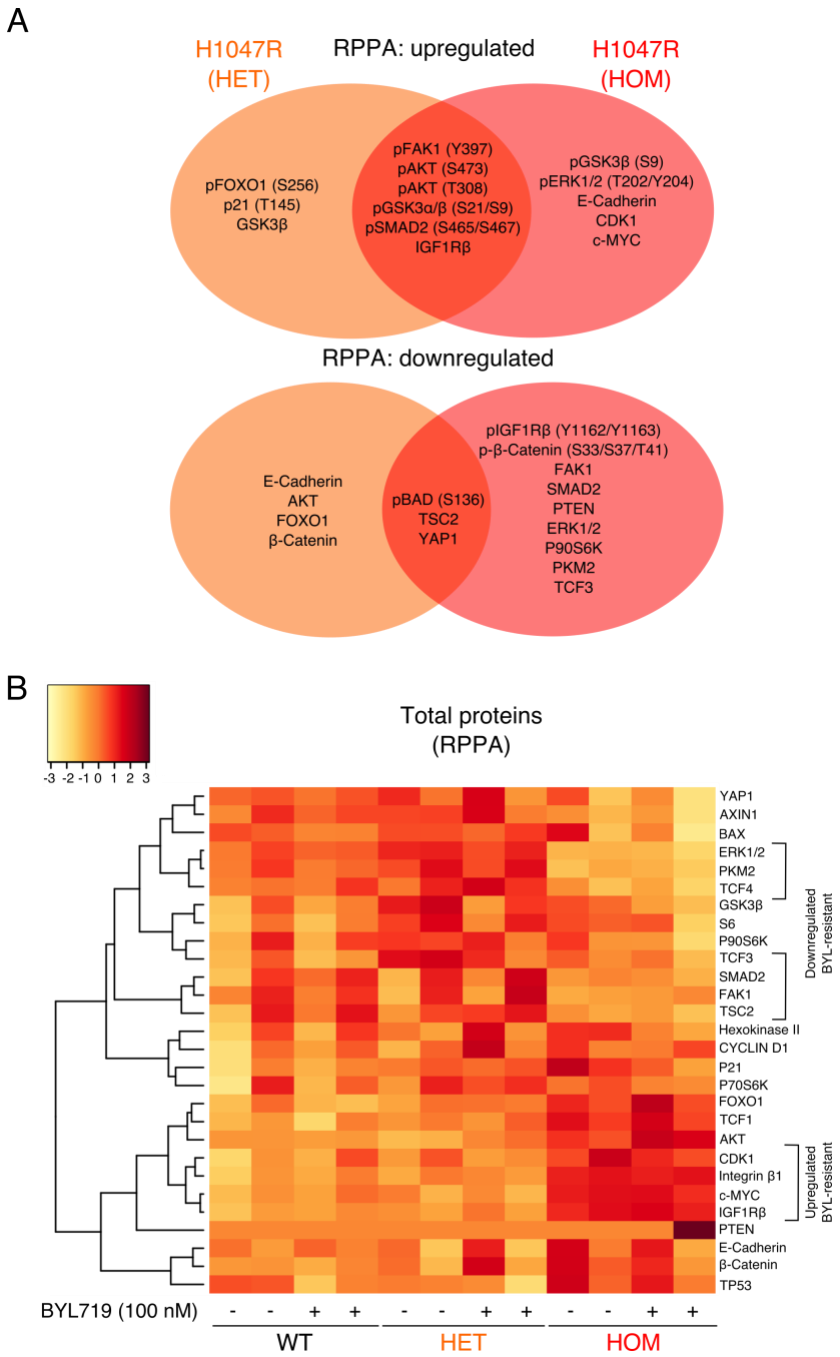

Fig. S3, related to Fig. 2. Additional RPPA data. (A) Venn-diagrams specifying differentially expressed phosphorylated and total proteins in *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) iPSCs relative to wild-type controls, based on RPPA profiling of cells cultured in growth factor-replete conditions. The data are based on a total of 10 wild-type cultures, 5 *PIK3CA*$^{WT/H1047R}$ cultures and 7 *PIK3CA*$^{H1047R/H1047R}$ cultures, and all shown targets were differentially expressed at a false-discovery rate (FDR) $\leq$ 0.05. Following quality checks (see Materials and Methods), the RPPA data included 21 phosphorylated and 21 total proteins. **(B)** Heatmap of total proteins from

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

RPPA profiling of wild-type (WT), *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) iPSCs following short-term growth factor removal (1 h), +/- 100 nM BYL719 (PI3Kα inhibitor) for 24 h. Groups of total proteins exhibiting a consistent expression pattern in BYL719-treated *PIK3CA*$^{H1047R/H1047R}$ iPSCs are specified.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

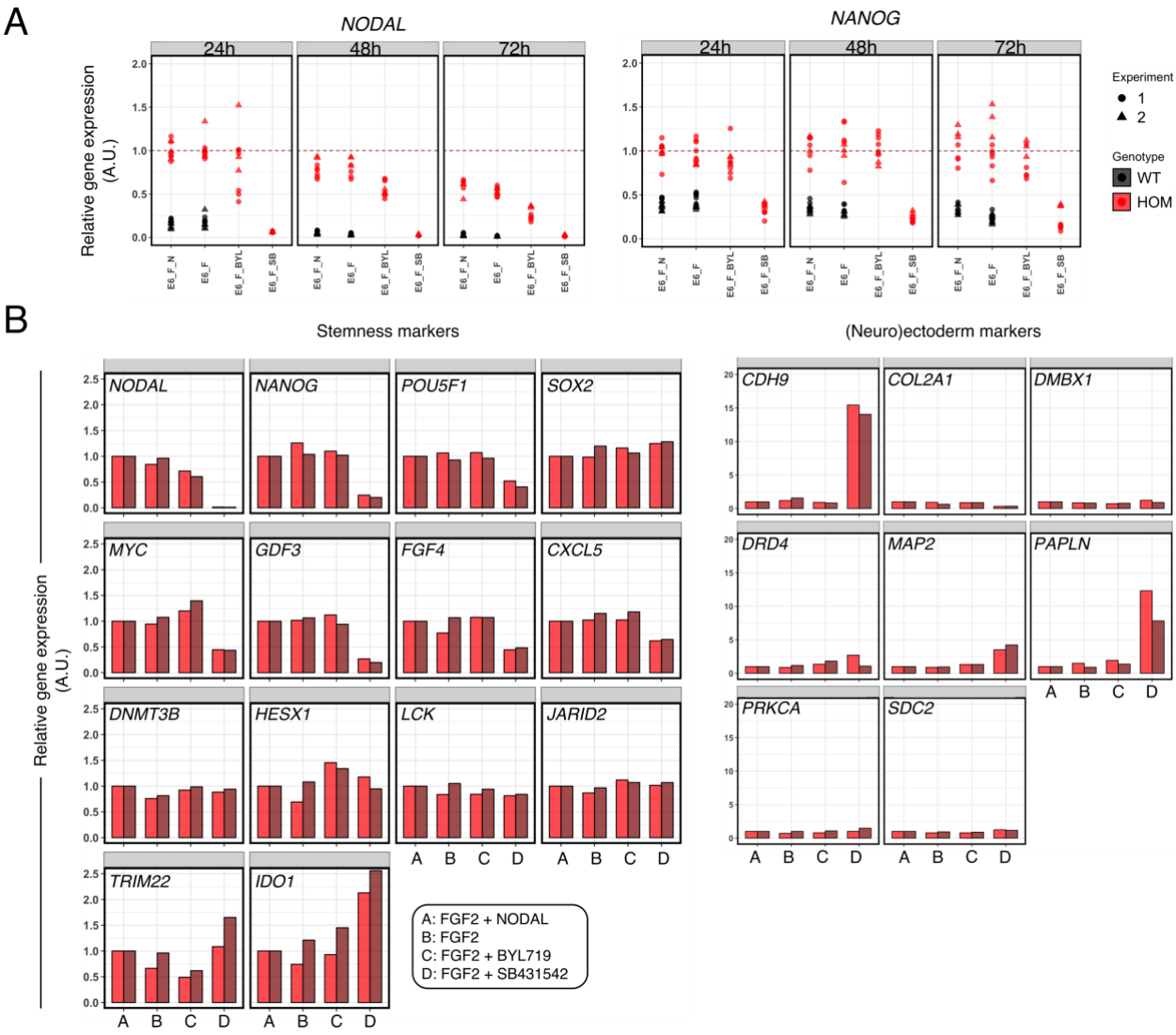

Fig. S4, related to Fig. 5. Alternative representation of the experimental data in Fig. 5A and additional RT-qPCR-based profiling of lineage-specific markers. (A) As in Fig. 5A, but representing individual gene expression values for *NODAL* and *NANOG* scaled to the mean expression in *PIK3CA*$^{H1047R/H1047R}$ after 24 h in E6 medium supplemented with FGF2 and NODAL. A.U., arbitrary units. **(B)** TaqMan hPSC Scorecards were used to profile a set of stemness and (neuro)ectoderm markers in *PIK3CA*$^{H1047R/H1047R}$ iPSCs following the indicated treatments for 48 h. Each bar corresponds to a single sample, with colors specifying the use of two independent clones per treatment. For each gene and clone, values are scaled to the expression value in cells cultured in E6 medium with FGF2 and NODAL.

bioRxiv preprint doi: https://doi.org/10.1101/753830; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Additional files

Table S1. List of differentially expressed genes in *PIK3CA*$^{WT/H1047R}$ vs wild-type hPSCs after applying an absolute fold-change cut-off of minimum 1.3.

Table S2. List of differentially expressed genes in *PIK3CA*$^{H1047R/H1047R}$ vs wild-type hPSCs after applying an absolute fold-change cut-off of minimum 1.3.

Table S3. List of differentially expressed genes in *PIK3CA*$^{WT/E418K}$ vs wild-type hPSCs.

Table S4. List of differentially expressed genes in *PIK3CA*$^{WT/H1047R}$ vs wild-type MEFs.

Table S5. List of differentially expressed proteins in *PIK3CA*$^{WT/H1047R}$ vs wild-type hPSCs.

Table S6. List of differentially expressed proteins in *PIK3CA*$^{H1047R/H1047R}$ vs wild-type hPSCs.

Table S7. List of differentially expressed proteins in both heterozygous and homozygous *PIK3CA*^{H1047R} hPSCs vs wild-type controls.

Key Resources Table.