NODAL/TGF signalling mediates the self-sustained stemness induced by PIK3CAH1047R homozygosity in pluripotent stem cells

First posted online on 28 January 2021 as 10.1242/dmm.048298
Access the most recent version at http://dmm.biologists.org/lookup/doi/10.1242/dmm.048298

NODAL/TGFβ signalling mediates the self-sustained stemness induced by *PIK3CA*^{H1047R} homozygosity in pluripotent stem cells

Ralitsa R. Madsen[1,2,3,8,*], James Longden[4,5], Rachel G. Knox[2,3], Xavier Robin[4], Franziska Völlmy[4], Kenneth G. Macleod[6], Larissa S. Moniz[7], Neil O. Carragher[6], Rune Linding[4,5], Bart Vanhaesebroeck[7], Robert K. Semple[1*]

[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.
[8]Current Address: Cell Signalling, University College London Cancer Institute, Paul O'Gorman Building, University College London, London, UK.

*Corresponding authors: Ralitsa R. Madsen (R.R.M.), Robert K. Semple (R.K.S.)

Email: r.madsen@ucl.ac.uk (R.R.M.); rsemple@ed.ac.uk (R.K.S.)

0000-0001-8844-5167 (R.R.M.)

0000-0001-6539-3069 (R.K.S.)

Keywords

PI3K, *PIK3CA*, stemness, pluripotent stem cells

Summary statement

Transcriptomic and proteomic analyses of iPSCs with allele-dose dependent expression of the *PIK3CA*^{H1047R} oncogene confirm network rewiring in homozygous mutants, with self-sustained stemness likely driven by constitutive TGFβ/NODAL pathway activation.

Abstract

Activating *PIK3CA* mutations are known “drivers” of human cancer and developmental overgrowth syndromes. We recently demonstrated that the "hotspot" *PIK3CA*^{H1047R} variant exerts unexpected allele dose-dependent effects on stemness in human pluripotent stem cells (hPSCs). In the present study, we combine high-depth transcriptomics, total proteomics and reverse-phase protein arrays to reveal potentially disease-related alterations in heterozygous cells, and to assess the contribution of activated TGFβ signalling to the stemness phenotype of homozygous *PIK3CA*^{H1047R} cells. We demonstrate signalling rewiring as a function of oncogenic PI3K signalling strength, and provide experimental evidence that self-sustained stemness is causally related to enhanced autocrine NODAL/TGFβ signalling. A significant transcriptomic signature of TGFβ pathway activation in heterozygous *PIK3CA*^{H1047R} was observed but was modest and was not associated with the stemness phenotype seen in homozygous mutants. Notably, the stemness gene expression in homozygous *PIK3CA*^{H1047R} iPSCs was reversed by pharmacological inhibition of NODAL/TGFβ signalling, but not by pharmacological PI3Kα pathway inhibition. Altogether, this provides the first in-depth analysis of PI3K signalling in human pluripotent stem cells and directly links strong PI3K activation to developmental NODAL/TGFβ signalling. This work illustrates the importance of allele dosage and expression when artificial systems are used to model human genetic disease caused by activating *PIK3CA* mutations.

Introduction

Class IA phosphoinositide 3-kinases (PI3Ks) are evolutionarily conserved enzymes that catalyse formation of the membrane-bound second messenger phosphatidylinositol-3,4,5-trisphosphate (PIP_3). PI3Ks are activated downstream of receptor tyrosine kinases, with the ensuing increase in PIP_3 and its derivative $PI(3,4)P_2$ triggering a widespread signalling network, best known for the activation of the serine/threonine kinases AKT and mTORC1. PI3K activation promotes cell survival, glucose uptake, anabolic metabolism, cell proliferation and cell migration (Fruman *et al.*, 2017). 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 (Bilanges, Posor and Vanhaesebroeck, 2019).

Activating mutations in *PIK3CA* are among the most common somatic point mutations in cancer, together with inactivation or loss of the tumour suppressor *PTEN* (a negative regulator of PI3K) (Chang *et al.*, 2015; Sanchez-Vega *et al.*, 2018; Campbell *et al.*, 2020). The same mutations in *PIK3CA*, when acquired postzygotically during development, also cause a range of largely benign overgrowth disorders, for which the term *PIK3CA*-related overgrowth spectrum (PROS) has been coined (Madsen, Vanhaesebroeck and Semple, 2018). Motivated by the need to understand the role of PI3K signalling in a human developmental context, we previously generated an allelic series of human induced pluripotent stem cells (iPSCs) with heterozygous or homozygous expression of the activating mutation *PIK3CA*^{H1047R}, the most commonly observed *PIK3CA* mutation in both cancer and PROS (Madsen *et al.*, 2019). Despite the severe developmental disorders caused by heterozygosity for *PIK3CA*^{H1047R} in humans *in vivo*, we found little discernible effect on germ layer specification from heterozygous iPSCs. In sharp contrast, homozygosity for *PIK3CA*^{H1047R} led to self-sustained stemness and resistance to spontaneous differentiation *in vitro* and *in vivo* (Madsen *et al.*, 2019). This suggested a previously unappreciated quantitative relationship between the strength of PI3K signalling and the gene regulatory network (GRN) in pluripotent stem cells.

The core pluripotency GRN features a feedforward, autoregulatory circuit comprising three transcription factors, namely SRY box 2 (SOX2), Octamer-binding transcription factor 3/4 (OCT3/4; encoded by POU5F1), and the homeobox transcription factor NANOG (Boyer *et al.*, 2005; Loh *et al.*, 2006; Li and Belmonte, 2017). SOX2 helps sustain OCT3/4 expression, which is required for establishment and maintenance of the pluripotent state (Nichols *et al.*, 1998). However, even modest overexpression of OCT3/4 destabilises the pluripotency network and triggers differentiation (Niwa, Miyazaki and Smith, 2000; Radzisheuskaya *et al.*, 2013). In contrast, NANOG, while dispensable for maintenance of pluripotency (Chambers *et al.*, 2007), stabilises 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 (Chambers *et al.*, 2003; Mitsui *et al.*, 2003; Darr, 2006; Ivanova *et al.*, 2006). In hPSCs, NANOG expression is activated by the transcription factors SMAD2/3 (Xu *et al.*, 2008), which in turn are activated by receptors binding TGFβ, Activin or NODAL (Pauklin and Vallier, 2015). Overexpression of NODAL thus results in self-sustained stemness of hPSCs even in differentiation-promoting conditions (Vallier, Reynolds and Pedersen, 2004; Vallier, Alexander and Pedersen, 2005).

Given the unexpected and surprisingly mild phenotype caused by heterozygous *PIK3CA*^{H1047R} expression in iPSCs, we reasoned that more sensitive assays would allow us to discern small but disease-relevant alterations in these cells. Thus, in this study, we first applied high depth transcriptomics, and proteomics to seek evidence of disease-related phenotypes in heterozygous cells, and to investigate how high-dose PI3K signalling leads to self-sustained stemness in homozygous *PIK3CA*^{H1047R} iPSCs. We demonstrate that heterozygous cells do exhibit significant transcriptomic changes, although these are a weak echo of the widespread changes seen in homozygous cells. The mild transcriptional consequences of heterozygous expression of disease-relevant *PIK3CA* mutations were also validated in additional model systems and contrast with previous findings of major transcriptional rewiring in immortalised, non-transformed breast epithelial cells (Hart *et al.*, 2015; Kiselev *et al.*, 2015). We demonstrate that the stemness phenotype of *PIK3CA*$^{H1047R/H1047R}$ iPSCs is maintained by self-sustained NODAL/TGFβ signalling, in line with increased *PIK3CA*-mediated *NODAL* expression, and that it is not reversible by PI3Kα-specific inhibition. This work provides in-depth characterisation of the near-binary PI3K signalling effects seen in hPSCs, and provides evidence for PI3Kα-induced NODAL/TGFβ signalling as the mechanism for self-sustained stemness in homozygous *PIK3CA*^{H1047R} iPSCs. We discuss the implications of our findings for understanding and modelling developmental disorders and cancers driven by genetic PI3K activation.

Results

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, heterozygous cells showed few phenotypic changes and differentially expressed protein-coding transcripts. In contrast, homozygous *PIK3CA*$^{H1047R/H1047R}$ cells exhibited marked morphological changes and altered gene expression, with strong enrichment for cancer-associated pathways (Madsen *et al.*, 2019).

To substantiate the apparent PI3K activity threshold manifest in *PIK3CA*^{H1047R}-driven gene expression changes, and to look for further disease-related changes in heterozygous cells, we undertook RNA sequencing at substantially greater depth, also increasing the sample size to four and including previously unstudied iPSC clones for the wild-type and homozygous *PIK3CA* genotype. All clones were obtained from CRISPR/Cas9-editing of the same parental iPSC line (WTC11), which has been extensively validated and used for genetic engineering by the Allen Cell Collection (http://www.allencell.org/). As before, *PIK3CA*^{H1047R} homozygous mutant cells clearly separated from heterozygous and wild-type cells, which overlapped on multidimensional scaling (**Fig. 1A**), but we now detected a reduction in the levels of 451 transcripts and an increase in the levels of 710 transcripts in *PIK3CA*$^{WT/H1047R}$ iPSCs (**Fig. 1B**). This dropped to 149 and 343 transcripts, respectively, after applying a fold-change cut-off of 1.3 (**Fig. 1B and Dataset S1**), indicative of the small magnitude of many expression changes in heterozygous mutants (**Fig. S1A**). Use of the same cut-off of 1.3, in sharp distinction, yielded 2873 and 2771 transcripts of decreased or increased abundance, respectively,

in homozygous iPSC mutants (**Fig. 1B and Dataset S2**). Not only was the number of gene expression changes higher by an order of magnitude in homozygous cells, but many expression changes were large compared to wild-type controls (**Fig. S1A**). The magnitudes of gene expression changes in *PIK3CA*$^{H1047R/H1047R}$ cells correlated strongly with our previous findings (Spearman's rho = 0.74, p < 2e-16) (**Fig. S1B**), whereas correlation was low (Spearman's rho = 0.1, p < 2e-16) for *PIK3CA*$^{WT/H1047R}$ iPSCs (**Fig. S1C**), as expected given the smaller number and lower magnitude of observed gene expression changes in heterozygous cells, and the lower depth of previous transcriptomic studies.

Given prior reports that *PIK3CA*^{H1047R} heterozygosity in breast epithelial cells extensively remodels gene expression (Hart *et al.*, 2015; Kiselev *et al.*, 2015), 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 mild PROS due to mosaicism for *PIK3CA*^{E418K} (**Fig. 1C**) (Parker *et al.*, 2019). Heterozygous iPSCs were compared to wild-type lines established simultaneously from dermal fibroblasts from the same skin biopsy, which is possible due to genetic mosaicism of the sampled skin. Like *PIK3CA*$^{WT/H1047R}$ iPSCs, *PIK3CA*$^{WT/E418K}$ iPSCs closely clustered with isogenic wild-type controls on multidimensional scaling (MDS) plotting (**Fig. 1D**), with only 30 differentially expressed genes (**Dataset S3**). We also studied previously reported *Pik3ca*$^{WT/H1047R}$ mouse embryonic fibroblasts (MEFs) 48 h after *Cre*-mediated *Pik3ca*^{H1047R} induction (Moniz *et al.*, 2017). Wild-type and *Pik3ca*$^{WT/H1047R}$ MEFs were superimposable on an MDS plot (**Fig. 1E**), with only 192 downregulated and 77 upregulated genes (**Dataset S4**). Our findings suggest that there are *bona fide* transcriptional changes induced by heterozygosity for *PIK3CA*^{H1047R}, but these are dramatically smaller in number and magnitude than changes induced by homozygosity for *PIK3CA*^{H1047R}.

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 previous study (Madsen *et al.*, 2019). 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 (Robin *et al.*, 2019 preprint). 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, Datasets S5 and S6)**. Of these, 27 proteins were differentially expressed in both heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs (**Dataset S7**), with 16 changing in opposite directions (**Fig. 1F**). There was a strong 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 weak correlation seen between transcriptomic experiments for heterozygous cells, this likely reflects the small magnitude of gene expression changes induced by heterozygous *PIK3CA*^{H1047R} (**Fig. 1B, S1C**).

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 deeper sequencing did reveal statistically significant gene expression changes in heterozygous iPSCs, and while these changes may contribute to growth-related phenotypes in PROS when sustained across development, effect sizes were modest and more variable. Similar findings in heterozygous MEFs suggest that this may be generalisable to differentiated cell types, irrespective of species. This consolidates the view that only homozygosity for *PIK3CA*^{H1047R} results in robust and widespread transcriptional changes in otherwise normal, diploid cells, arguing against a universal "butterfly" effect of heterozygosity suggested based on studies of a genetically abnormal breast epithelial cell line (Hart *et al.*, 2015; Kiselev *et al.*, 2015).

PIK3CA$^{H1047R/H1047R}$ iPSCs show evidence of signalling "rewiring"

We previously demonstrated a graded increase in AKT (S473) phosphorylation across heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs (Madsen *et al.*, 2019). 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 of PI3K pathway activation, we profiled phosphorylation of a wider repertoire of pathway components using reverse phase phosphoprotein array (RPPA) technology.

Changes in protein phosphorylation were surprisingly modest, with the largest 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 phosphorylation of the mTORC1 target P70S6K or its downstream substrate S6 (**Fig. S3A**), perhaps reflecting saturation at this level of the pathway due to other stimuli for mTORC1 in the complete culture medium (e.g. amino acids) (Valvezan and Manning, 2019). 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 the PI3Kα-selective inhibitor BYL719 for 24 h fully reversed canonical PI3K signalling-related changes in phosphorylation of downstream proteins including AKT, GSK3, FOXO1, TSC2 and P70S6K (**Fig. 2B**). Consistent with these signalling changes, we previously showed that the same dose of BYL719 (100 nM) abolishes the increased tolerance to growth factor deprivation-induced death conferred by heterozygous or homozygous *PIK3CA*^{H1047R} in iPSCs (Madsen *et al.*, 2019). Despite its effects on the primary PI3K signalling cascade, PI3Kα inhibition failed to reverse other changes observed in *PIK3CA*$^{H1047R/H1047R}$ iPSCs, including increased phosphorylation of SMAD2 and ERK1/2 and increased expression of c-MYC and IGF1R (**Fig. 2B, Fig. S3B**). This suggests signalling rewiring in *PIK3CA*$^{H1047R/H1047R}$ iPSCs that is partially resistant to relatively short-term inhibition of the inducing stimulus.

Pathway and network analyses implicate NODAL/TGFβ signalling in *PIK3CA^{H1047R}* allele dose-dependent stemness

Pathway and network analyses were next applied to proteomic and transcriptomic data to identify candidate mechanism(s) mediating *PIK3CA^{H1047R}* allele dose-dependent stemness. Consistent with our previous study (Madsen *et al.*, 2019), 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 upstream activator predicted by analysis of *PIK3CA$^{H1047R/H1047R}$* proteomic data (**Fig. 3B**). This is consistent with strong *NODAL* mRNA upregulation and increased pSMAD2 (S465/S467) in *PIK3CA$^{H1047R/H1047R}$* iPSCs in the current study (**Dataset S2** and RPPA data in **Fig. 2**, respectively), and with prior evidence of activation of the NODAL/TGFβ pathway in homozygous *PIK3CA^{H1047R}* iPSCs.

Although *PIK3CA$^{WT/H1047R}$* iPSCs showed around 10-fold fewer differentially expressed genes than homozygous iPSC cells, IPA in heterozygous iPSCs also revealed multiple TGFβ pathway-related stimuli among predicted upstream activators (**Fig. 3C**). Moreover, TGFβ1 was predicted as one of only two significant upstream activators when analysis was performed on genes concordantly differentially expressed (N = 180) in *PIK3CA^{H1047R}* mutant iPSCs *versus* wild-type controls (**Fig. 3C and Dataset S8**).

The other significant upstream regulator common to heterozygous and homozygous *PIK3CA^{H1047R}* was MAPK1 (also known as ERK2), consistent with RPPA findings and immunoblot evidence of increased ERK kinase phosphorylation in *PIK3CA^{H1047R}* mutant iPSCs (Ref. (Madsen *et al.*, 2019), **Fig. 2A and Fig. S3A**). The significance of predicted TGFβ activation in heterozygous *PIK3CA^{H1047R}* iPSCs (overlap p-value = 1.7e-05) was much lower than in homozygous (overlap p-value = 4.3e-21) mutants. This is in keeping with the much lower effect size in heterozygous cells, and consistent with a critical role for the NODAL/TGFβ pathway in mediating the allele dose-dependent effect of *PIK3CA^{H1047R}* in human iPSCs.

To complement 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 (Langfelder and Horvath, 2008). Using all transcripts expressed in wild-type, heterozygous and homozygous *PIK3CA^{H1047R}* iPSCs (**Fig. 4A**), this analysis returned 43 modules (or clusters) of highly interconnected genes (**Fig. 4B**). Of the two modules with the highest correlation with the homozygous trait, one showed enrichment for several KEGG pathway terms relevant to stemness of *PIK3CA$^{H1047R/H1047R}$* iPSCs, notably including “Signalling pathways regulating pluripotency in stem cells” (**Fig. 4C**).

Given prior evidence of strong activation of NODAL/TGFβ signalling in homozygous mutant cells, we next constructed the minimal network of differentially expressed genes in *PIK3CA$^{H1047R/H1047R}$* iPSCs that linked pluripotency, PI3K and TGFβ signalling pathways (**Fig. 4D**). This approach allowed us to navigate the signalling rewiring and to link strong PI3K pathway activation, stemness and NODAL/TGFβ signalling in an unbiased

manner. Indeed, the resulting network exhibited high interconnectivity, with multiple shared nodes across all three pathways, suggesting close crosstalk between PI3K and NODAL/TGFβ signalling in stemness regulation. That most nodes represented genes with increased expression in homozygous mutants strengthens the notion that strong oncogenic PI3Kα activation stabilises the pluripotency network in human iPSCs. The MYC oncogene stood out as the only network node intersecting with all three signalling pathways, suggesting it may comprise a key mechanistic link in the observed phenotype.

Inhibition of NODAL/TGFβ signalling destabilises the pluripotency gene network in *PIK3CA*$^{H1047R/H1047R}$ iPSCs

NODAL/TGFβ signalling plays a critical role in pluripotency regulation (Vallier, Alexander and Pedersen, 2005; Mesnard, Guzman-Ayala and Constam, 2006; Xu *et al.*, 2008), and a differentiation-resistant phenotype has been reported in *NODAL*-overexpressing iPSCs (Vallier, Reynolds and Pedersen, 2004). Together with increased *NODAL* expression in homozygous *PIK3CA*^{H1047R} iPSCs and computational identification of enhanced NODAL/TGFβ pathway activity in PI3K-driven "constitutive" stemness (Ref. (Madsen *et al.*, 2019) and current study), this led us to hypothesise that strong PI3Kα-dependent induction of *NODAL* underlies establishment of the differentiation-resistant phenotype of homozygous *PIK3CA*^{H1047R} iPSCs. Specifically, we hypothesised that autocrine NODAL enhances NODAL/TGFβ signalling in *PIK3CA*$^{H1047R/H1047R}$ iPSCs, with resulting increased *NANOG* expression "locking" the cells in perpetual stemness (Xu *et al.*, 2008).

Testing this hypothesis in iPSCs is challenging for biological and technical reasons, including lack of specific pharmacological inhibitors of NODAL, and difficulty in detecting subtle early phenotypic consequences of partial destabilisation of the iPSC pluripotency gene regulatory network. Moreover, the widely adopted maintenance medium and coating substrate we used for cell culture both contain TGFβ ligands (Vukicevic *et al.*, 1992; Chen *et al.*, 2011), which may mask effects of *NODAL* repression by PI3Kα-specific inhibition. We previously found that treatment of *PIK3CA*$^{H1047R/H1047R}$ iPSCs in this 'complete' maintenance medium with 500 nM BYL719 reduces *NODAL* mRNA expression within 24 h, but has no discernible effect on increased *NANOG* mRNA levels (Madsen *et al.*, 2019).

To minimise confounding effects of exogenous TGFβ ligands, we prepared medium with and without recombinant NODAL supplementation, and assessed expression of *NODAL* and *NANOG* as surrogate markers of stemness over 72 h of culture. We also reduced the BYL719 concentration to 250 nM given increased iPSC toxicity observed with 500 nM BYL719 (Madsen *et al.*, 2019); and pilot experiments (not shown) in which 24 h treatment with 250 nM but not 100 nM BYL719 in complete medium reduced *NODAL* mRNA expression in *PIK3CA*$^{H1047R/H1047R}$ iPSC clones. 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. S4A**). In *PIK3CA*$^{H1047R/H1047R}$ iPSCs, however, *NODAL* removal had no effect on the increased *NODAL* and *NANOG* expression (**Fig. 5 and Fig. S4A**), in line with a self-sustained stemness phenotype. Exposure of NODAL-free *PIK3CA*$^{H1047R/H1047R}$ cultures to 250 nM BYL719 had a visible colony

growth-inhibitory effect **(Fig. S5**) and decreased *NODAL* expression within 24 h, and this continued to decrease subsequently (**Fig. 5**). This is consistent with NODAL's known ability to control its own expression through a feed-forward loop (Hill, 2018). Despite a 55% reduction in *NODAL* mRNA after 72 h, however, little effect on *NANOG* expression was seen (**Fig. 5**). This may reflect the short time course studied (to avoid confounding effect of passaging), or the exquisite sensitivity of iPSCs to residual upregulation of *NODAL* in homozygous *PIK3CA*^{H1047R} iPSCs. This may be compounded by residual low levels of TGFβ-like ligands in the coating substrate, or possibly by increased expression of two other TGFβ superfamily ligands, *GDF3* and *TGFB2*, observed in homozygous mutant cells (**Dataset S2**).

To confirm that NODAL/TGFβ signalling is required for maintenance of stemness in *PIK3CA*$^{H1047R/H1047R}$ iPSCs, the cells were treated with SB431542 – a specific inhibitor of TGFβ and NODAL type I receptors (Inman *et al.*, 2002). This completely repressed *NODAL* expression within 24 h, accompanied by downregulation of *NANOG* and *POU5F1* expression (**Fig. 5**). Given that the pluripotent stem cell state is stabilised by an autoregulatory feedforward interaction between these transcription factors, their downregulation is expected to result in altered expression of multiple stemness and differentiation markers. Confirming this, we used a lineage-specific gene expression array to demonstrate similar reduction in mRNA expression of several other well-established stemness markers (*MYC*, *FGF4*, *GDF3*)**,** which were upregulated in *PIK3CA*$^{H1047R/H1047R}$ iPSCs at baseline (**Dataset S2**), performing the analysis after 48 h of NODAL/TGFβ pathway inhibition (**Fig. S4B**). Despite the short treatment, we also found evidence for the expected neuroectoderm induction upon inhibition of the NODAL/TGFβ pathway (Vallier, Reynolds and Pedersen, 2004; Vallier *et al.*, 2009), reflected by increased expression of *CDH9*, *MAP2*, *OLFM3* and *PAPLN* (**Fig. S4B**).

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

Discussion

PIK3CA^{H1047R} is the most common activating *PIK3CA* mutation in human cancers and in PROS (Madsen, Vanhaesebroeck and Semple, 2018). We recently found that *PIK3CA*-associated cancers often harbour multiple mutated *PIK3CA* copies, and demonstrated that homozygosity but not heterozygosity for *PIK3CA*^{H1047R} leads to self-sustained stemness in human pluripotent stem cells (hPSCs) (Madsen *et al.*, 2019). High-depth transcriptomics in this study confirmed that heterozygosity for *PIK3CA*^{H1047R} induces significant but very modest transcriptional changes; observed both in CRISPR-edited hPSCs with long-term *PIK3CA*^{H1047R} expression and in mouse embryonic fibroblasts (MEFs) upon acute *PIK3CA*^{H1047R} induction by *Cre*, with canonical PI3K pathway activation seen in both cases (current study and Ref. (Berenjeno *et al.*, 2017; Moniz *et al.*, 2017; Madsen *et al.*, 2019)). Similarly, hPSCs 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 these heterozygous variants, however, homozygosity for *PIK3CA*^{H1047R} was associated with differential expression of nearly a third of the hPSC transcriptome, suggesting widespread epigenetic reprogramming. This near-binary response is not a consequence of a similar quantitative difference in PI3K pathway activation, as assessed by phosphoprotein profiling, which instead showed a relatively modest and graded increase in homozygous *versus* heterozygous *PIK3CA*^{H1047R} hPSCs. This implies that the apparent PI3K signalling threshold that determines the cellular response in hPSCs is "decoded" distal to the canonical pathway activation.

Using a combination of computational analyses and targeted experiments, the current study further provides evidence for self-sustained NODAL/TGFβ pathway activation as the main mechanism through which *PIK3CA*^{H1047R} homozygosity "locks" hPSCs in a differentiation-resistant state that has become independent of the driver mutation and the associated PI3K pathway activation. We suggest that homozygosity but not heterozygosity for *PIK3CA*^{H1047R} promotes sufficient NODAL/TGFβ pathway activity to induce increased *NODAL* and downstream *NANOG* expression to levels that stabilise the stem cell state, yet are not high enough to tip the balance towards mesendoderm differentiation (Madsen *et al.*, 2019). Exactly how PI3K activation regulates *NODAL* expression remains unknown. A potential mechanism involves increased expression of the stem cell reprogramming factor MYC, which was observed at both mRNA and protein level in homozygous but not heterozygous *PIK3CA*^{H1047R} iPSCs. Furthermore, MYC was the only node in the WGCNA-based network of pluripotency, PI3K and TGFβ pathway components that was shared by all three pathways (**Fig. 4D**). MYC has previously been shown to exert oncogenic effects that depend on a sharp threshold of MYC expression, reminiscent of the effects we observe for allele dose-dependent *PIK3CA* activation (Murphy *et al.*, 2008). Elevated MYC has also been shown to allow *PIK3CA*^{H1047R}-induced murine breast cancers to become independent of continuous *PIK3CA*^{H1047R} expression (Liu *et al.*, 2011).

Stabilisation of the stemness phenotype in hPSCs by strong genetic PI3K pathway activation may be generalisable beyond the iPSC model system. BYL719 (alpelisib; Novartis), the PI3Kα-selective inhibitor used in our cellular studies, was recently approved for use in combination with anti-estrogen therapy in ER-positive breast cancers (André *et al.*, 2019). In a separate study focusing on human breast cancer, we have described use of computational analyses to demonstrate a strong, positive relationship between a transcriptomically-derived PI3K activity score, stemness gene expression and tumour grade in breast cancer (Madsen *et al.*, 2020 preprint). Prior reports have suggested a role for NODAL in driving breast cancer stemness and aggressive disease (Bar-Eli, 2012; Margaryan *et al.*, 2019), with potential links to mTORC1 activation (Katsuno *et al.*, 2019; Jewer *et al.*, 2020). Our findings that BYL719 fails to fully reverse the increased *NODAL* and stemness gene expression in homozygous *PIK3CA*^{H1047R} iPSCs suggests that inhibition of TGFβ signalling as a pro-differentiation therapy warrants investigation as co-therapy with PI3K inhibitors in breast tumours with strong PI3K pathway activation. The lack of widespread transcriptional changes upon heterozygous expression of mutant *PIK3CA* in otherwise genetically normal cell models may explain the low oncogenicity of this genotype in isolation *in vivo*.

Finally, our observations are important for future studies seeking to model human PIK3CA-related diseases. The modest changes observed in heterozygous *PIK3CA*^{H1047R} cells, in sharp contrast to the radical transcriptional alterations in homozygous cells, emphasise the importance of careful allele dose titration when artificial overexpression systems are used to model disorders caused by genetic *PIK3CA* activation. Our findings in heterozygous cells are also a reminder that very small effect sizes in cellular systems may summate and result in major human phenotypes over a life course. That such minor changes are found in a cellular study of a rare and severe disorder emphasises the challenges of modelling much more subtle disease susceptibility conferred by GWAS-detected genetic associations, where cellular effect sizes are likely to be smaller still.

Materials and Methods

All cell lines used in this study are listed in **Table S1**. Unless stated otherwise, standard chemicals were acquired from Sigma Aldrich, with details for the remaining reagents included in **Table S2**.

iPSC culture and treatments

Maintenance

The derivation of the iPSC lines, including associated ethics statements, has been described previously (Madsen *et al.*, 2019). All lines were grown at 37°C and 5% CO2 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. (Madsen *et al.*, 2019), the current transcriptomic data of PIK3CAH1047R 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

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.

NODAL/TGFβ signalling studies

Wild-type or homozygous PIK3CAH1047R 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^{-1} heat-stable FGF2. This was combined with one of the following reagents or their diluent equivalents: 100 ng ml^{-1} 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 PIK3CAWT/H1047R MEFs used in this study have been reported previously (Moniz *et al.*, 2017). Cell pellets were collected on dry ice 48 h after induction of heterozygous PIK3CAH1047R 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 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 PIK3CAH1047R iPSCs and corresponding wild-types, an Illumina TruSeq Stranded mRNA Library Prep Kit was used to synthesise 150-bp-long paired-end mRNA libraries, followed by sequencing on an Illumina HiSeq 4000, with average depth of 70 million reads per sample. PIK3CAWT/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.

Details of the subsequent data analyses (raw read mapping, counting, statistical testing, pathway and network analyses) are provided in Supplementary Information.

Label-free total proteomics

Sample preparation

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 Na3VO4) 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 solubilised 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^{-1} 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 equalised peptide amounts were lyophilised and resolubilised 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 4x105 ions, and maximum injection time of 20 ms. 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 ion trap MS2 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.

Details of the subsequent data analyses (FASTA file generation, mass spectrometry searches) are provided in Supplementary Information.

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 MgCl2, 10% (v/v) glycerol, 1% (v/v) TritonX-100, 1 mM EGTA, 100 mM NaF, 10 mM Na4P2O7, 2 mM Na3VO4 (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^{-1} 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 (Macleod, Serrels and Carragher, 2017). 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 slide by omitting primary antibody incubation step. For normalisation, 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 Table S3.

Details of all subsequent data analyses, including statistical testing and evaluation of antibody/assay performance, are provided in Supplementary Information.

Reverse transcription-quantitative PCR (RT-qPCR)

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. TaqMan hPSC Scorecards (384-well) were used according to the manufacturer's instructions with minor modifications. Further details on protocol modifications and all data analysis steps are provided in Supplementary Information.

Statistical analyses

Bespoke statistical analyses are specified in the relevant sections above and in Supplementary Information. No statistical methods were used to predetermine sample size.

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 iPSC culturing facilities and to Pamela Brown (SURF Biomolecular Core, University of Edinburgh) for access to qPCR facilities.

Competing interests

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

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. Work in the laboratory of B.V. is supported by Cancer Research UK (C23338/A25722) and the UK NIHR University College London Hospitals Biomedical Research Centre. 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). 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.

Data availability

Raw data and bespoke RNotebooks containing guided scripts and plots are available via the Open Science Framework (doi: 10.17605/OSF.IO/MUERY). The original RNAseq data have been deposited to the Gene Expression Omnibus (GEO), under accession numbers: GSE134076 (H1047R iPSC data), GSE138161 (E418K iPSC data), GSE135046 (MEF data). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al.*, 2019) 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 (r.madsen@ucl.ac.uk) or Robert K. Semple (rsemple@ed.ac.uk).

Author Contributions

Overall conceptualisation and study design by R.R.M. and R.K.S., with important contributions from B.V. (PI3K signalling biology, 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 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., X.R., J.L., F.V. and R.G.K. were responsible for data curation. R.R.M., B.V. and R.K.S. wrote the manuscript. R.L., J.L., N.C., X.R., L.M., and F.V. reviewed and edited the final version.

References

André, F. *et al.* (2019) 'Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer', *New England Journal of Medicine*, 380(20), pp. 1929–1940. doi: 10.1056/NEJMoa1813904.

Bar-Eli, M. (2012) 'Back to the embryonic stage: Nodal as a biomarker for breast cancer progression', *Breast Cancer Research*, 14(3), p. 105. doi: 10.1186/bcr3177.

Berenjeno, I. M. *et al.* (2017) 'Oncogenic PIK3CA induces centrosome amplification and tolerance to genome doubling.', *Nature Communications*, 8(1), p. 1773. doi: 10.1038/s41467-017-02002-4.

Bilanges, B., Posor, Y. and Vanhaesebroeck, B. (2019) 'PI3K isoforms in cell signalling and vesicle trafficking', *Nature Reviews Molecular Cell Biology*. Springer US, 20(9), pp. 515–534. doi: 10.1038/s41580-019-0129-z.

Boyer, L. *et al.* (2005) 'Core transcriptional regulatory circuitry in human embryonic stem cells', *Cell*, 122(6), pp. 947–956. doi: 10.1016/j.cell.2005.08.020.

Campbell, P. J. *et al.* (2020) 'Pan-cancer analysis of whole genomes', *Nature*, 578(7793), pp. 82–93. doi: 10.1038/s41586-020-1969-6.

Chambers, I. *et al.* (2003) 'Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells', *Cell*, 113(5), pp. 643–655. doi: 10.1016/S0092-8674(03)00392-1.

Chambers, I. *et al.* (2007) 'Nanog safeguards pluripotency and mediates germline development', *Nature*, 450(7173), pp. 1230–1234. doi: 10.1038/nature06403.

Chang, M. T. *et al.* (2015) 'Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity', *Nature Biotechnology*. Nature Publishing Group, 34(2), pp. 155–163. doi: 10.1038/nbt.3391.

Chen, G. *et al.* (2011) 'Chemically defined conditions for human iPSC derivation and culture.', *Nature methods*, 8(5), pp. 424–429. doi: 10.1038/nmeth.1593.

Darr, H. (2006) 'Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features', *Development*, 133(6), pp. 1193–1201. doi: 10.1242/dev.02286.

Fruman, D. A. *et al.* (2017) 'The PI3K Pathway in Human Disease', *Cell*, 170(4), pp. 605–635. doi: 10.1016/j.cell.2017.07.029.

Hart, J. R. *et al.* (2015) 'The butterfly effect in cancer: A single base mutation can remodel the cell', *Proceedings of the National Academy of Sciences*, 112(4), pp. 1131–1136. doi: 10.1073/pnas.1424012112.

Hill, C. S. (2018) 'Spatial and temporal control of NODAL signaling', *Current Opinion in Cell Biology*, 51, pp. 50–57. doi: 10.1016/j.ceb.2017.10.005.

Inman, G. J. *et al.* (2002) '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.', *Molecular pharmacology*, 62(1), pp. 65–74. doi: 10.1124/mol.62.1.65.

Ivanova, N. *et al.* (2006) 'Dissecting self-renewal in stem cells with RNA interference', *Nature*, 442(7102), pp. 533–538. doi: 10.1038/nature04915.

Jewer, M. *et al.* (2020) 'Translational control of breast cancer plasticity', *Nature Communications*. Springer US, 11(1), p. 2498. doi: 10.1038/s41467-020-16352-z.

Katsuno, Y. *et al.* (2019) 'Chronic TGF-b exposure drives stabilized EMT, tumor stemness, and cancer drug resistance with vulnerability to bitopic mTOR inhibition', *Science Signaling*, 12(570), p. eaau8544. doi: 10.1126/scisignal.aau8544.

Kiselev, V. Y. *et al.* (2015) 'Perturbations of PIP3 signalling trigger a global remodelling of mRNA landscape and reveal a transcriptional feedback loop.', *Nucleic acids research*, 43(20), pp. 9663–79. doi: 10.1093/nar/gkv1015.

Langfelder, P. and Horvath, S. (2008) 'WGCNA: an R package for weighted correlation network analysis.', *BMC bioinformatics*, 9, p. 559. doi: 10.1186/1471-2105-9-559.

Li, M. and Belmonte, J. C. I. (2017) 'Ground rules of the pluripotency gene regulatory network.', *Nature reviews. Genetics*, 18(3), pp. 180–191. doi: 10.1038/nrg.2016.156.

Liu, P. *et al.* (2011) 'Oncogenic PIK3CA-driven mammary tumors frequently recur via PI3K pathway-dependent and PI3K pathway-independent mechanisms.', *Nature Medicine*, 17(9), pp. 1116–1120. doi: 10.1038/nm.2402.

Loh, Y. H. *et al.* (2006) 'The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells', *Nature Genetics*, 38(4), pp. 431–440. doi: 10.1038/ng1760.

Macleod, K. G., Serrels, B. and Carragher, N. O. (2017) *Proteomics for Drug Discovery*. Edited by I. M. Lazar, M. Kontoyianni, and A. C. Lazar. New York, NY: Springer New York (Methods in Molecular Biology). doi: 10.1007/978-1-4939-7201-2.

Madsen, R. R. *et al.* (2019) 'Oncogenic PIK3CA promotes cellular stemness in an allele dose-dependent manner', *Proceedings of the National Academy of Sciences*, 116(17), pp. 8380–8389. doi: 10.1073/pnas.1821093116.

Madsen, R. R. *et al.* (2020) 'Relationship between stemness and transcriptionally-inferred PI3K activity in human breast cancer', *bioRxiv*, p. 2020.07.09.195974. doi: 10.1101/2020.07.09.195974.

Madsen, R. R., Vanhaesebroeck, B. and Semple, R. K. (2018) 'Cancer-Associated PIK3CA Mutations in Overgrowth Disorders', *Trends in Molecular Medicine*, 24(10), pp. 856–870. doi: 10.1016/j.molmed.2018.08.003.

Margaryan, N. V. *et al.* (2019) 'The stem cell phenotype of aggressive breast cancer cells', *Cancers*, 11(3), pp. 1–11. doi: 10.3390/cancers11030340.

Mesnard, D., Guzman-Ayala, M. and Constam, D. B. (2006) 'Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning', *Development*, 133(13), pp. 2497–505. doi: 10.1242/dev.02413.

Mitsui, K. *et al.* (2003) 'The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells', *Cell*, 113(5), pp. 631–642. doi: 10.1016/S0092-8674(03)00393-3.

Moniz, L. S. *et al.* (2017) 'Phosphoproteomic comparison of Pik3ca and Pten signalling identifies the nucleotidase NT5C as a novel AKT substrate', *Scientific Reports*, 7, p. 39985. doi: 10.1038/srep39985.

Murphy, D. J. *et al.* (2008) 'Distinct Thresholds Govern Myc's Biological Output In Vivo', *Cancer Cell*, 14(6), pp. 447–457. doi: 10.1016/j.ccr.2008.10.018.

Nichols, J. *et al.* (1998) 'Formation of pluripotent stem cells in the mammalian embryo dependes on the POU transcription factor Oct4.', *Cell*, 95(3), pp. 379–391. doi: 10.1016/S0092-8674(00)81769-9.

Niwa, H., Miyazaki, J. and Smith, A. G. (2000) 'Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.', *Nature genetics*, 24(4), pp. 372–376. doi: 10.1038/74199.

Parker, V. E. R. *et al.* (2019) 'Safety and efficacy of low-dose sirolimus in the PIK3CA-related overgrowth spectrum', *Genetics in Medicine*. Springer US, 21(5), pp. 1189–1198. doi: 10.1038/s41436-018-0297-9.

Pauklin, S. and Vallier, L. (2015) 'Activin/Nodal signalling in stem cells', *Development*, 142(4), pp. 607–619. doi: 10.1242/dev.091769.

Perez-Riverol, Y. *et al.* (2019) 'The PRIDE database and related tools and resources in 2019: Improving support for quantification data', *Nucleic Acids Research*. doi: 10.1093/nar/gky1106.

Radzisheuskaya, A. *et al.* (2013) 'A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages', *Nature Cell Biology*. Nature Publishing Group, 15(6), pp. 579–590. doi: 10.1038/ncb2742.

Robin, X. *et al.* (2019) 'Probability-based detection of phosphoproteomic uncertainty reveals rare signaling events driven by oncogenic kinase gene fusion', *bioRxiv*, p. 621961. doi: 10.1101/621961.

Sanchez-Vega, F. *et al.* (2018) 'Oncogenic Signaling Pathways in The Cancer Genome Atlas', *Cell*, 173(2), pp. 321-337.e10. doi: 10.1016/j.cell.2018.03.035.

Vallier, L. *et al.* (2009) 'Activin/Nodal signalling maintains pluripotency by controlling Nanog expression.', *Development*, 136(8), pp. 1339–49. doi: 10.1242/dev.033951.

Vallier, L., Alexander, M. and Pedersen, R. A. (2005) 'Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells', *Journal of Cell Science*, 118(19), pp. 4495–4509. doi: 10.1242/jcs.02553.

Vallier, L., Reynolds, D. and Pedersen, R. A. (2004) 'Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway', *Developmental Biology*, 275(2), pp. 403–421. doi: 10.1016/j.ydbio.2004.08.031.

Valvezan, A. J. and Manning, B. D. (2019) 'Molecular logic of mTORC1 signalling as a metabolic rheostat', *Nature Metabolism*. Springer US. doi: 10.1038/s42255-019-0038-7.

Vukicevic, S. *et al.* (1992) 'Identification of multiple active growth factors in basement membrane Matrigel suggests caution in interpretation of cellular activity related to extracellular matrix components.', *Experimental cell research*, 202(1), pp. 1–8. doi: 10.1016/0014-4827(92)90397-q.

Xu, R. H. *et al.* (2008) 'NANOG is a Direct Target of TGFβ/Activin-Mediated SMAD Signaling in Human ESCs', *Cell Stem Cell*, 3(2), pp. 196–206. doi: 10.1016/j.stem.2008.07.001.

Figures

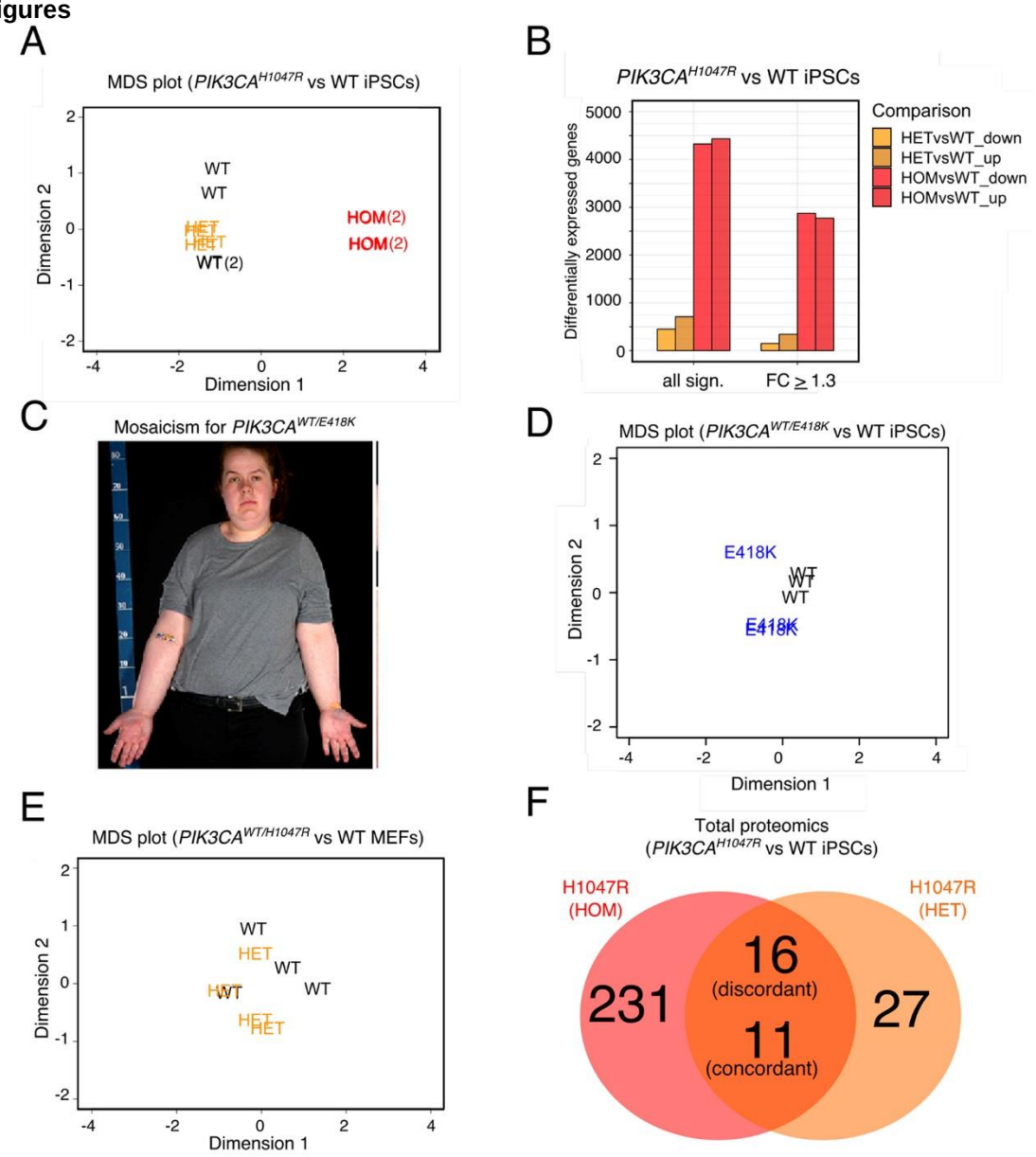

Figure 1. Transcriptomic and proteomic analyses of human and mouse cell lines with endogenous expression of oncogenic *PIK3CA*. (A) Multidimensional scaling (MDS) plot of the transcriptomes of wild-type (WT), *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) human iPSCs. The numbers in 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 ≥ 1.3 (FDR ≤ 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. (Parker *et al.*, 2019). **(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.

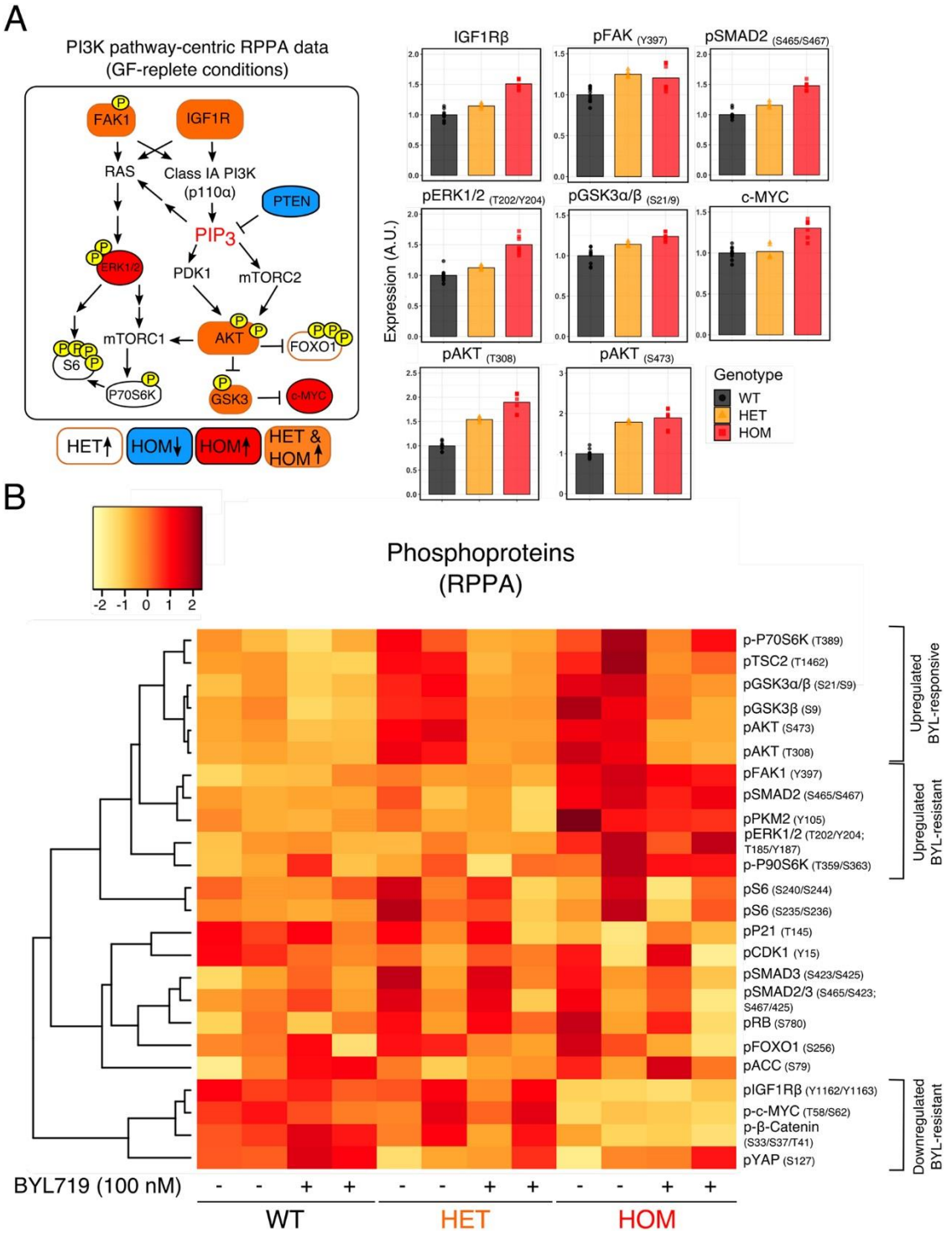

Figure 2. Reverse Phase Protein Array (RPPA) of *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) human iPSCs. (A) Left: Diagram of PI3K pathway-related phosphorylated proteins, with colour code used to signify differentially expressed targets in *PIK3CA*^{H1047R} mutant iPSCs *versus* isogenic wild-type controls. Color-coded targets were significant at FDR $\leq$ 0.05 (Benjamini-Hochberg). Right: Barplots show representative examples of differentially expressed protein targets, revealing relatively modest quantitative changes. Phosphorylated proteins were normalised 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 colours correspond to Fast Green-normalised expression values in units of standard deviation (z-score) from the mean (centred 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.

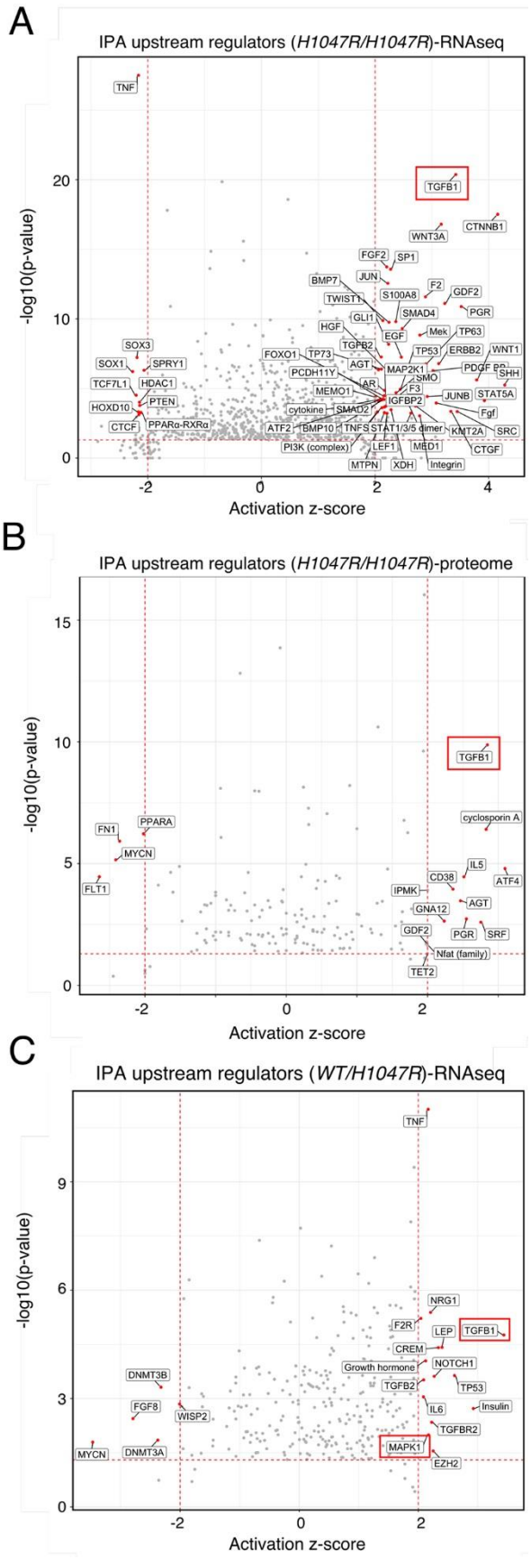

Figure 3. Ingenuity® pathway analyses (IPA) predict activation of TGFβ signalling in heterozygous and homozygous *PIK3CA*^{H1047R} iPSCs. (A) IPA of upstream regulators using the list of top 2000 upregulated and top 2000 downregulated mRNA transcripts in *PIK3CA*$^{H1047R/H1047R}$ iPSCs (for RNAseq details, 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. Note that this analysis does not differentiate between NODAL, TGFβ1 and Activin, which all act through the same pathway in human pluripotent stem cells. **(B)** As in (A) but using the list of differentially-expressed proteins identified by total proteomics and red-colouring 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-colouring 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.

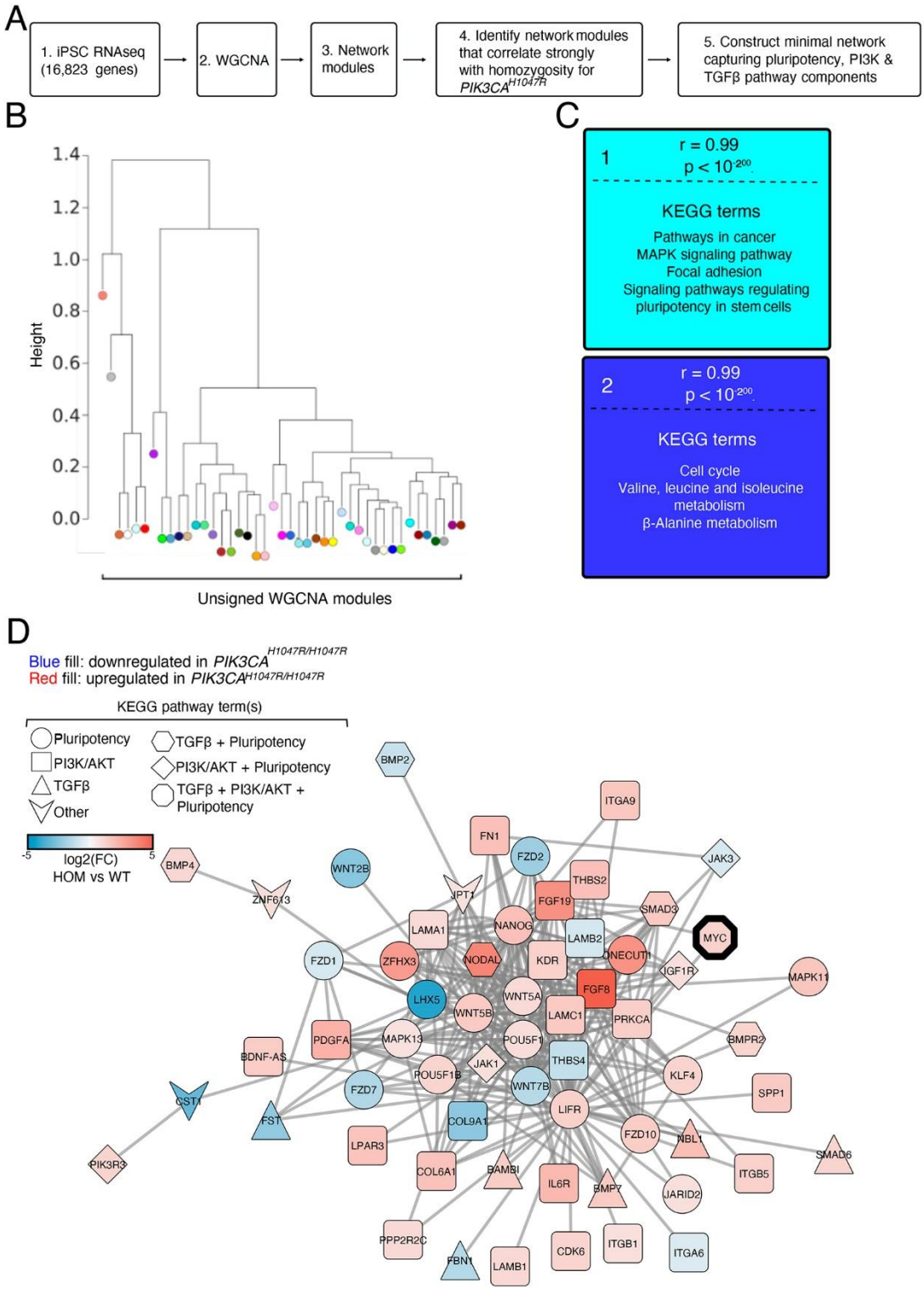

Figure 4. Weighted gene correlation network analysis (WGCNA) identifies links among pluripotency components, TGFβ and PI3K signalling. (A) Schematic of the WGCNA workflow and subsequent data selection for visualisation. (**B)** Unsigned WGCNA modules identified using the list of transcripts expressed in wild-type, *PIK3CA*$^{WT/H1047R}$ and *PIK3CA*$^{H1047R/H1047R}$ iPSCs (for RNAseq details, see Fig. 1B). **(C)** The two network modules with genes whose module membership correlated strongest with differential expression in

homozygous *PIK3CA*^{H1047R} iPSCs. The colour of each module corresponds to its colour 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 colour and shape are used to specify direction of differential mRNA expression in *PIK3CA*$^{H1047R/H1047R}$ iPSCs and pathway membership, respectively. Fill colour saturation represents gene expression fold-change (FC; log2) in *PIK3CA*$^{H1047R/H1047R}$ (HOM) *vs* wild-type (WT) iPSCs. MYC is highlighted as the only network component intersecting all three KEGG pathways, suggesting it may comprise a key mechanistic link in the observed phenotype.

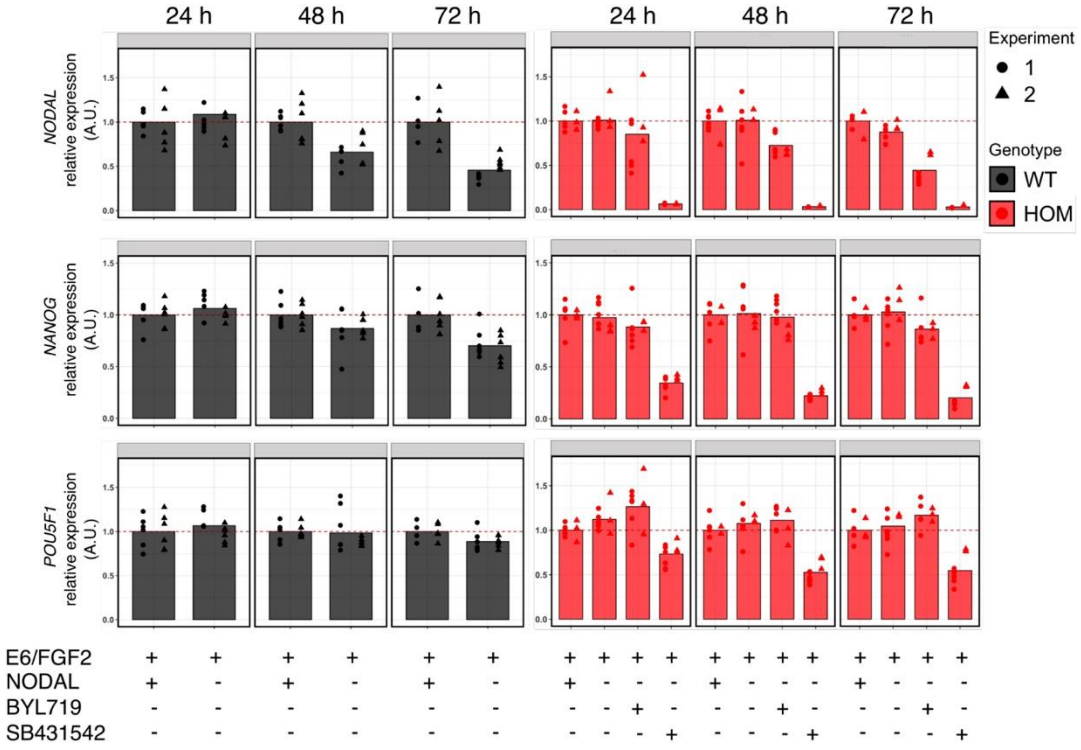

Figure 5. NODAL/TGFβ signalling-dependent regulation of stemness in *PIK3CA*$^{H1047R/H1047R}$ iPSCs. Gene expression time course of *NODAL*, *NANOG* and *POU5F1* in wild-type (WT) or *PIK3CA*$^{H1047R/H1047R}$ iPSCs following the indicated treatments for 24 h, 48 h or 72 h. B250: 250 nM BYL719 (PI3Kα-selective inhibitor); E6/FGF2: Essential 6 medium supplemented with 10 ng ml^{-1} 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^{-1} 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 normalised to the E6/FGF2 condition within each genotype and time-point. An alternative visualisation 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. For representative micrographs of *PIK3CA*$^{H1047R/H1047R}$ iPSCs exposed to the different treatments, see Fig. S5. A.U., arbitrary units.

Supplementary Material Text

iPSC processing for individual experiments

Cell lysate 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. (Madsen *et al.*, 2019), 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

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 signalling 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^{-1} heat-stable FGF2. This was combined with one of the following reagents or their diluent equivalents: 100 ng ml^{-1} 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.

RNA sequencing data analyses

Raw read mapping, counting and differential expression

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 (Dobin *et al.*, 2013). Subsequent data processing was performed using open-source R software according to the *limma-voom* method (Ritchie *et al.*, 2015). Briefly, raw counts were converted to counts per million (cpm) using the *cpm()* function in the *edgeR* package (Robinson, McCarthy and Smyth, 2009), followed by normalisation according the trimmed mean of M (TMM) method (Robinson and Oshlack, 2010). 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 (Ritchie *et al.*, 2015). 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% (Benjamini and Hochberg, 1995). 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.

Pathway and network analyses

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 analyse 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 (Ritchie *et al.*, 2015), 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 (Krämer *et al.*, 2014). 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.

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 (Zhang and Horvath, 2005; Langfelder and Horvath, 2008), with a soft power threshold of 28 (chosen to maximise scale independence and minimise 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 homozygosity trait were selected for functional enrichment analysis. CytoScape plugin ClueGO (version 2.5.4) (Bindea *et al.*, 2009) was used to perform pathway analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) ontology (build 27.02.19) (Kanehisa *et al.*, 2017). 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 visualisation was performed using Cytoscape (Shannon *et al.*, 2003).

Mass spectrometry data acquisition and analyses

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 4x105 ions, and maximum injection time of 20 ms. 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 ion trap MS2 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 (Bolger, Lohse and Usadel, 2014) using the following parameters: headcrop = 3, minlen = 30, trailing = 3. The trimmed reads were aligned to the human reference genome (hg19 build) with BWA (Li and Durbin, 2010), followed by application of GATK base quality score recalibration, indel realignment, duplicate removal and SNP/indel discovery with genotyping (McKenna *et al.*, 2010). GATK Best Practices standard hard filtering parameters were used throughout (Depristo *et al.*, 2011).

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 $\leq$ 1.5, FS $\geq$ 60, MQ $\geq$ 40, MQRankSum $\leq$ -12.5, ReadPosRankSum $\leq$ -8.0, and average DP $\geq$ 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 (McLaren *et al.*, 2016). 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

Raw files were processed using MaxQuant 1.5.0.2 (Tyanova, Temu and Cox, 2016) 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 (Robin *et al.*, 2019 preprint). 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 ($|z| \geq 1.2$; $|\ln(FC)| \geq \ln(1.2)$).

RPPA data analyses

Slide images were analysed 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 antibody 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 normalised 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 normalised to the corresponding mean of the wild-type group. All phosphoprotein signals were also normalised to the corresponding total protein values.

RT-qPCR set-up and data analyses

For SYBR Green-based qPCRs, A 5-fold cDNA dilution series was prepared and used as standard curve for relative quantitation of gene expression. *TBP* was used as normaliser following confirmation that its gene expression remained unaffected by the tested conditions. Melt curve analyses confirmed amplification of a single product by each primer pair. All primers had amplification efficiencies 95%-105%. Samples were loaded in duplicate in 384-well plates.

The TaqMan hPSC Scorecard was set up according to the manufacturer's instructions with the following modifications. From each cDNA 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 linearised (antilog) under the assumption of 100 % primer amplification efficiency. The geometric expression mean of the control gene assays was used for subsequent normalisation 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 **Table S4**.

Colony size quantitation from light micrographs

Routine cell culture light micrographs were acquired on an EVOS FL digital inverted microscope (AMF4300, Thermo Fisher Scientific) using the 4X or 10X objective (final magnification 40X and 100X, respectively). For quantitation, 4X images were used for colony segmentation with Definiens Developer XD software. Background was detected using a contrast threshold; for this each pixel was compared to those in the surrounding 24 pixels (i.e. a 5x5 pixel box), and pixels with low contrast (between -50 and +50) were classified as background. Remaining pixels were classified as colonies, and any holes (pixels that were not initially classified as being part of the colony due to low contrast) were filled. Edges of the resulting colonies were smoothened by shrinking and then growing the colonies by 2 pixels. Finally, colonies less than 2000 pixels in size were reclassified as background. The area of the resulting colonies could then be measured and averaged over each field of view.

Statistical analyses

In line with recent (ATOMIC) recommendations by the American Statistical Association (Wasserstein, Schirm and Lazar, 2019), 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.

For RPPA data in Fig. 2A, 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 (Ritchie *et al.*, 2015). 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% (Benjamini and Hochberg, 1995). 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.

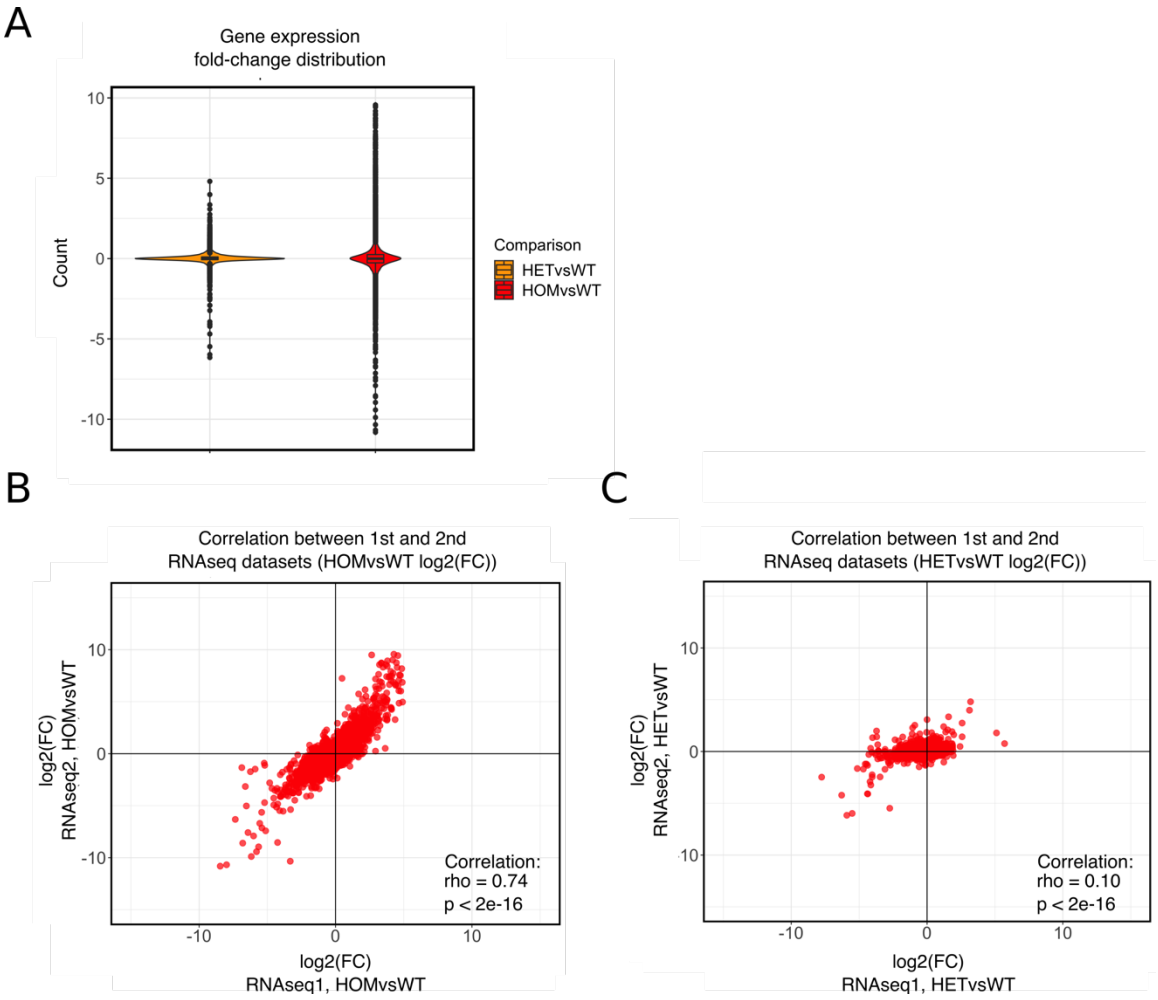

Fig. S1, related to Fig. 1. Fold-change distribution and transcriptome correlations. (A) Combined violin-boxplot representation of the fold-change (log2) distribution of gene expression changes between heterozygous and homozygous *PIK3CA*^{H1047R} and wild-type iPSCs as indicated. (B) and (C) Correlation 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 at least two independent clones each.

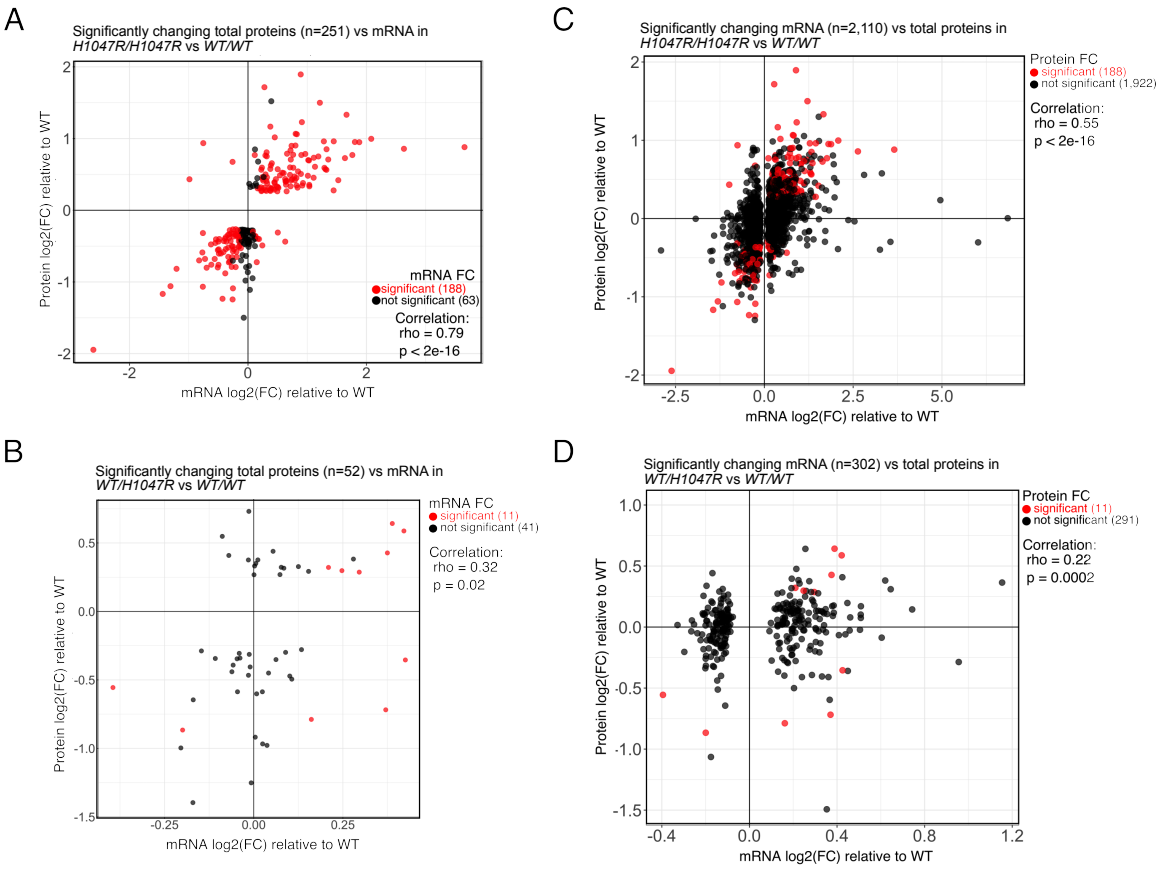

Fig. S2, **related to Fig. 1. Transcriptome-proteome correlations.** (A) Correlation plot of the log2 expression fold-changes (FC) of differentially expressed total proteins ($|z| \geq 1.2$ and $|\ln(FC)| \geq 1.2$) in *PIK3CA*$^{H1047R/H1047R}$ iPSCs and the corresponding mRNA transcripts in an independent set of cultures. If identified as differentially expressed following statistical analysis (FDR ≤ 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 ≤ 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.

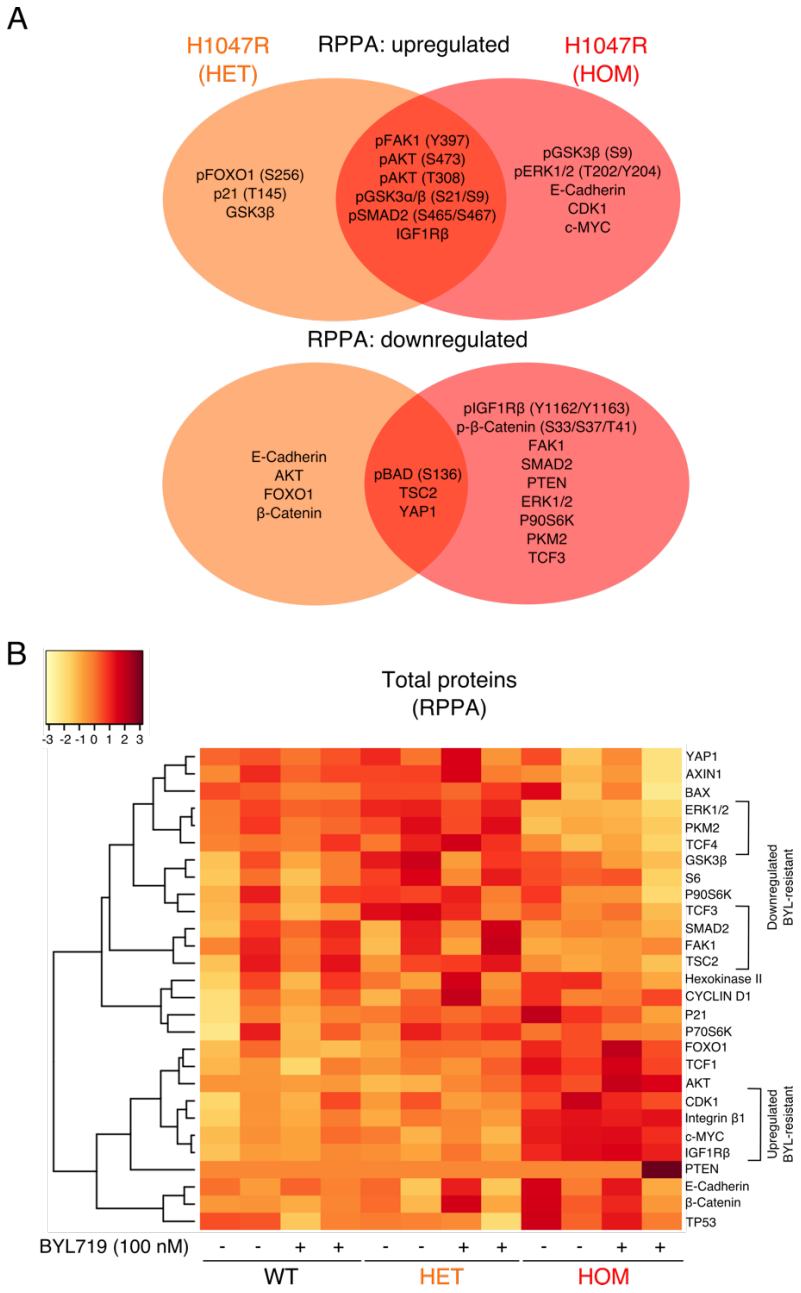

Fig. S3, **related to Fig. 2. Additional RPPA data.** (A) Venn-diagrams specifying differentially expressed phosphorylated and total proteins in *PIK3CA*$^{WT/H1047R}$ (HET) and *PIK3CA*$^{H1047R/H1047R}$ (HOM) iPSCs relative to wild-type controls, based on RPPA profiling of cells cultured in growth 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 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.

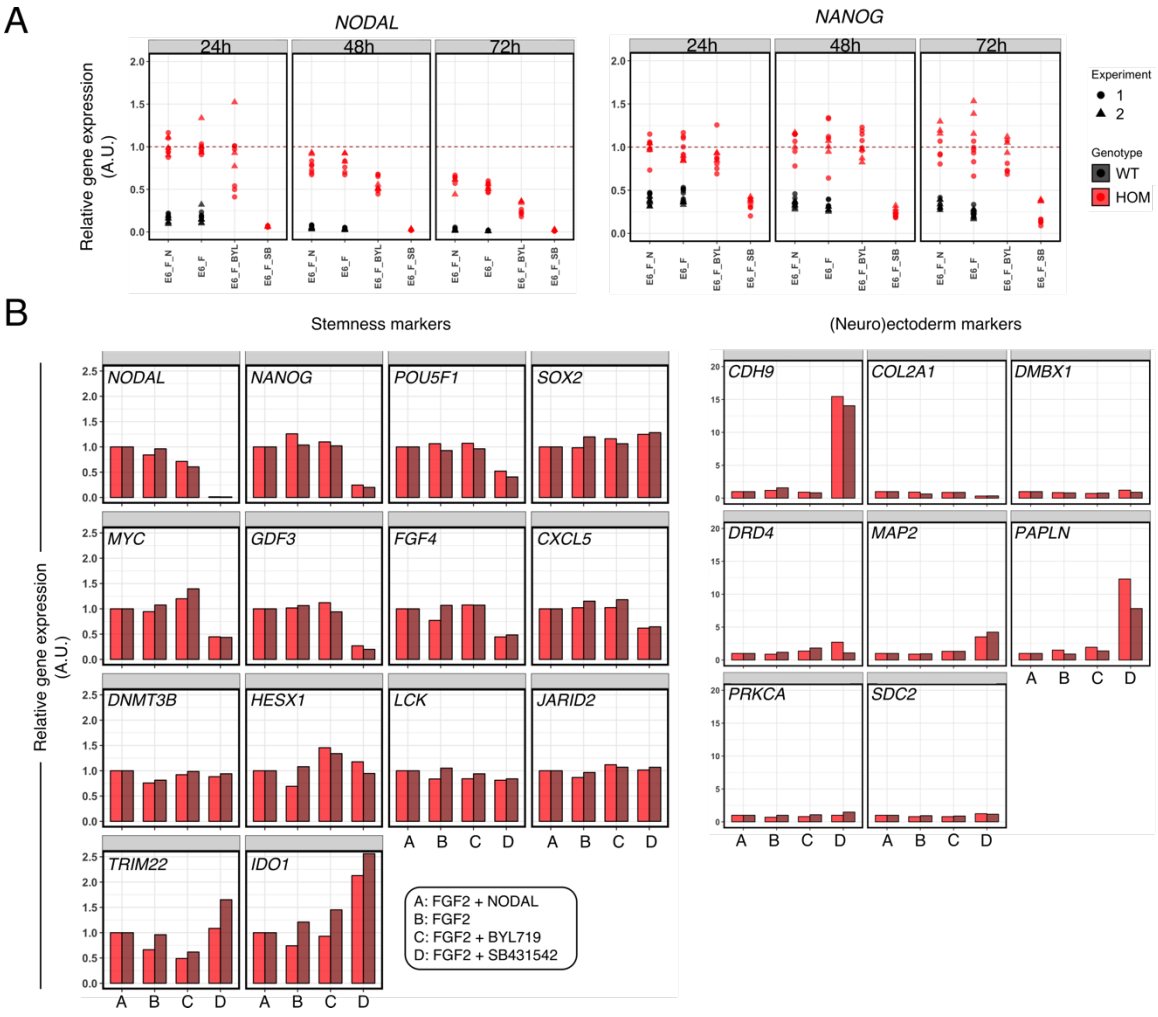

Fig. S4, **related to Fig. 5. Alternative representation of the experimental data in Fig. 5A and additional RT-qPCR-based profiling of lineage-specific markers.** (A) As in Fig. 5A but representing individual gene expression values for *NODAL* and *NANOG* scaled to the mean expression in *PIK3CA*$^{H1047R/H1047R}$ 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. Note that *FGF4, GDF3, MYC* and *JARID2* were manually classified as stemness markers based on previous large scale characterisations of human pluripotent stem cells (Adewumi *et al.*, 2007). Each bar corresponds to a single sample, with colours 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.

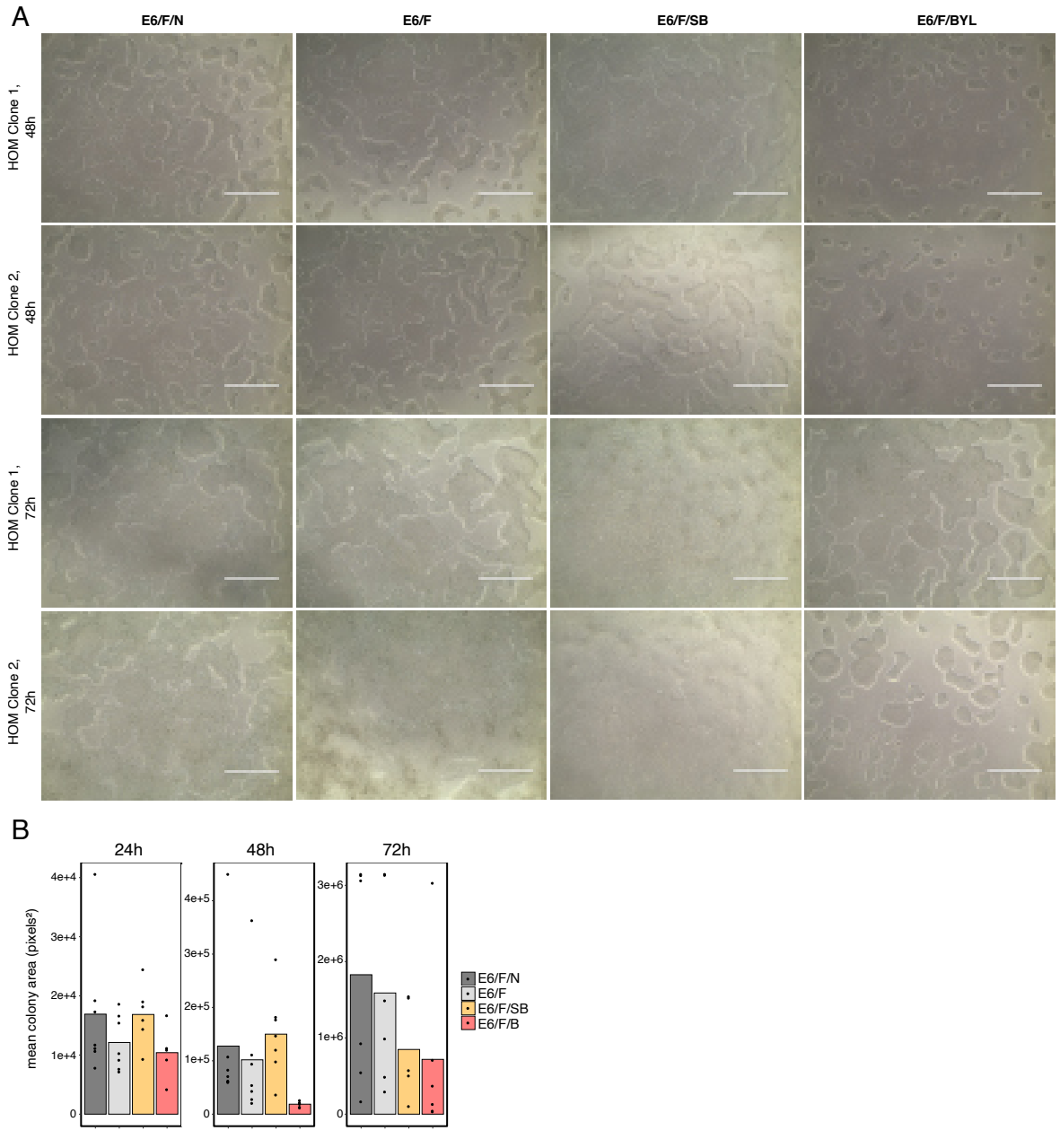

Fig. S5, related to Fig. 5. Light micrographs of *PIK3CA*$^{H1047R/H1047R}$ iPSC clones exposed to different media formulations +/- specific inhibitors. A. Each row corresponds to images of parallel cultures of one of two independent iPSC clones exposed to different treatments for 48 or 72 h as indicated. E6/F/N: Essential 6 medium supplemented with 10 ng ml^{-1} FGF2 and 100 ng ml^{-1} NODAL; E6/F: Essential 6 medium supplemented with 10 ng ml^{-1} FGF2; E6/F/SB: Essential 6 medium supplemented with 10 ng ml^{-1} FGF2 and 5 µM SB431542; E6/F/BYL: Essential 6 medium supplemented with 10 ng ml^{-1} FGF2 and 250 nM BYL719. The images are representative of 3 independent experiments. Scale bar: 1000 µm. The original images are available online on https://osf.io/hbf7x/. **B.** Quantification of colony size based on the acquired light micrographs; individual points corresponding to separate images and bars represent the overall mean across the three experimental replicates.

Cell line	Source	Identifier
WTC11 male parental iPSC line	Coriell Institute	GM25256
WTC11-derived male wild-type post-CRISPR editing	Madsen *et al.* PNAS 2019 (Madsen *et al.*, 2019)	N/A
WTC11-derived male *PIK3CA*$^{WT/H1047R}$	Madsen *et al.* PNAS 2019 (Madsen *et al.*, 2019)	N/A
WTC11-derived male *PIK3CA*$^{H1047R/H1047R}$	Madsen *et al.* PNAS 2019 (Madsen *et al.*, 2019)	N/A
M98-WT female iPSC	Madsen *et al.* PNAS 2019 (Madsen *et al.*, 2019)	N/A
M98-E418K (*PIK3CA*$^{WT/E418K}$) female iPSC	Madsen *et al.* PNAS 2019 (Madsen *et al.*, 2019)	N/A
Primary MEFs *PIK3CA*$^{WT/WT}$ (H1047R uninduced)	Moniz *et al.* Sci Reports 2017 (Moniz *et al.*, 2017)	N/A
Primary MEFs *PIK3CA*$^{WT/H1047R}$ (induced for 48h)	Moniz *et al.* Sci Reports 2017 (Moniz *et al.*, 2017)	N/A

Table S1. Cell lines.

Item	Catalogue no.	Vendor	Application
100 ml DMSO, sterile-filtered	D2650	Sigma Aldrich	Cells
2X PowerUp SYBR Green Master Mix	A25742	Thermo Fisher Scientific	Molbio
Acetonitrile	1.00029	Sigma Aldrich	Protein work
Amonium Bicarbonate	09830	Sigma Aldrich	Protein work
Bioanalyzer RNA 6000 Nano Kit	5067-1511	Agilent	Molbio
BYL719 (lot #2598155)	B9700-1mg	Cambridge Bioscience	Cells
Chloroacetamide	C0267	Sigma Aldrich	Protein work
cOmplete ULTRA Tablets, EDTA-free	589279100 1	Roche	Protein work
cOmplete ULTRA Tablets, Mini, EasyPack, PhosStop	490684500 1	Roche	Protein work
DC Protein Assay Reagent A	5000113	Biorad	Protein
DC Protein Assay Reagent B	5000114	Biorad	Protein
DC Protein Assay Reagent S	5000115	Biorad	Protein
DirectZol RNA Miniprep Kit	R2051	ZymoResearch	Molbio
Dithiothritol	101977770 01	Sigma Aldrich	Protein work
DMEM/F12 without L-Glutamine, HEPES 500 ml	D6421	Sigma Aldrich	Cells
DPBS without Ca^{2+} and Mg^{2+}	D8537	Sigma Aldich	Cells
EndoFree Plasmid Maxi Kit (10)	12362	Qiagen	Molbio
Endopeptidase Lys-C	129-02541	Wako	Protein work
Essential 6 Medium	A1516401	Thermo Fisher Scientific	Cells
Essential 8 Flex Medium Kit	A2858501	Thermo Fisher Scientific	Cells
Geltrex LDEV Free hESC Quality 5 ml	A1413302	Thermo Fisher Scientific	Cells
Gibco Heat Stable Recombinant Human bFGF (for prolonged culture), 100 ug (lot #2046190)	PHG0369	Thermo Fisher Scientific	Cells
HEPES	H4034	Sigma Aldrich	Protein work
High-Capacity cDNA Reverse Transcription Kit	4368814	Thermo Fisher Scientific	Molbio
InSolution™ TGF-b RI Kinase Inhibitor VI, SB431542 (lot #3075261)	616464-5MG	Merck Millipore	Cells
Nuclease-free H_2O	AM9932	Thermo Fisher Scientific	Molbio/Portein work
Nunc™ Delta Cell-Culture Treated Multidishes; 6-well	140685	Thermo Fisher Scientific	Cells
Nunc™ Delta Cell-Culture Treated Multidishes; 12-well	150628	Thermo Fisher Scientific	Cells
Pierce Quantitative Colorimetric Peptide Assay	23275	Thermo Fisher Scientific	Protein work

QIAamp DNA Micro Kit (50)	56304	Qiagen	Molbio
QIAshredder	79654	Qiagen	Molbio
QIAzol Lysis Reagent	79306	Qiagen	Molbio
Quibit dsDNA HS Assay Kit	Q32854	Thermo Fisher Scientific	Molbio
Recombinant human NODAL protein, 25 ug (lot #OLF1115091)	3218-ND/CF	R&D	Cells
ReLeSR 100 ml	5872	Stem Cell Technologies	Cells
RevitaCell Suppplement 5 ml	A2644501	Thermo Fisher Scientific	Cells
RNeasy Mini Kit	74104	Qiagen	Molbio
SepPak C18 cartridges	WAT020515	Waters	Protein work
SuperScript IV Fist Strand Synthesis System	18091050	Thermo Fisher Scientific	Molbio
Urea	15604	Sigma Aldrich	Protein work
TaqMan® hPSC Scorecard™ Panel, 384-well	A15870	Thermo Fisher Scientific	Molbio
Thiourea	T7875	Sigma Aldrich	Protein work
Trifluoroacetic acid	T6508	Sigma Aldrich	Protein work
TrueSeq Stranded mRNA Library Prep Kit	20020594	Illumina	Molbio
TruSeq RNA Library Prep Kit v2	RS-930-2002	Illumina	Molbio
TruSeq RNA Single Indexes Set A	20020492	Illumina	Molbio
Trypsin	T6567	Sigma	Protein work
VenorGeM Classic Mycoplasma Kit	11-1005	Minerva Biolabs	Molbio

Table S2, Materials list. Molbio, molecular biology.

Primary antibody (clone if mAb)	Vendor	Catalogue #	Lot #	Species
4E-BP1	CST	9452	10	rabbit
4E-BP1 P Ser65	CST	9451	14	rabbit
4E-BP1 P Ser65 (174A9)	CST	9456	5	rabbit
4E-BP1 P Thr37,Thr46	CST	9459	8	rabbit
Acetyl-CoA carboxylase P Ser79	CST	3661	10	rabbit
AKT	CST	9272	24	rabbit
AKT P Ser473	CST	9271	13	rabbit
AKT P Thr308	CST	9275	19	rabbit
AXIN1 (C76H11)	CST	2087	4	rabbit
β -Catenin	CST	9562	11	rabbit
β -Catenin P Ser33,Ser37,Thr41	CST	9561	12	rabbit
β -Catenin P Thr41,Ser45	CST	9565	2	rabbit
BAD	CST	9292		rabbit
BAD P Ser112	CST	9291	8	rabbit
BAD P Ser136	CST	9295	11	rabbit
BAX	CST	2772	8	rabbit
c-MYC	CST	5605	8	rabbit
c-MYC P Thr58,Ser62	Epitomics	1203-1		rabbit
CDK1	CST	9112	5	rabbit
CDK1 P Tyr15	CST	9111	12	rabbit
Cyclin D1	CST	2926	10	mouse
Cyclin D1 P Thr286	CST	3300	3	rabbit
E-Cadherin	CST	3195	13	rabbit
ERK1/2	CST	9102	26	rabbit
ERK1/2 P CST Thr202/Thr185,Tyr204/Tyr187	CST	4370	12	rabbit
FAK1	CST	3285	9	rabbit
FAK1 P Y397	CST	3283	2	rabbit
FOXO1 P S256	CST	9461	7	rabbit
FOXO1 (C29H4)	CST	2880	10	rabbit
GSK3α/β P Ser21/Ser9	CST	9331	15	rabbit
GSK3β	CST	9315	12	rabbit
GSK3β P Ser9	CST	9336	12	rabbit
Hexokinase II	CST	2867	3	rabbit
IGF1Rβ	CST	3027	13	rabbit
IGF1Rβ P Tyr1162,Tyr1163	Invitrogen (Biosource)	44-804G	633298A	rabbit
Integrin β 1 (EP1041Y)	abcam	ab52971		rabbit
Lamin A/C	CST	2032	5	rabbit
LEF1 (C12A5)	CST	2230	5	rabbit
P21	CST	2946	14	mouse
P21 p Thr145	Santa Cruz	sc-20220-R	A2811	rabbit
p110α	CST	4249	7	rabbit
PKM2 XP(R)	CST	4053	3	rabbit
PKM2 P Tyr105	CST	3827	3	rabbit
PTEN	CST	9552	3	rabbit
RB P Ser780	CST	9307	13	rabbit
Rictor P T1135	CST	3806		rabbit
RPS6KA (Rsk1-3)	Santa Cruz	sc-231	I0110	rabbit
RPS6KA (Rsk1-3) P Thr359,Ser363	CST	9344	8	rabbit
S6	CST	2217	5	rabbit

S6 P Ser235,Ser236	CST	2211	20	rabbit
S6 P Ser240,Ser244	CST	2215	14	rabbit
S6K1	CST	9202	15	rabbit
S6K1 P Thr389	CST	9205S	16	rabbit
SLUG (C19G7)	CST	9585	5	rabbit
SMAD2 (C86F7)	CST	3122	8	rabbit
SMAD2 P Ser465,Ser467	CST	3108	7	rabbit
SMAD2/3 P Ser465/Ser423, Ser467/Ser425	CST	8828	6	rabbit
SMAD3 P Ser423,Ser425	CST	9520	13	rabbit
TCF1 (C63O9)	CST	2203	6	rabbit
TCF3 (D15G11)	CST	2883	4	rabbit
TCF4 (C9B9)	CST	2565	3	rabbit
TGFβ (56E4)	CST	3709	4	rabbit
TP53	CST	9282	4	rabbit
TSC2	CST	3612	3	rabbit
TSC2 P Thr1462	CST	3617	4	rabbit
YAP P Ser127	CST	4911	2	rabbit
YAP1 (EP1674Y)	Abcam	ab52771		rabbit
Secondary antibody				
Anti Rabbit Dylight 800 conjugated	CST	5151		goat
Anti Mouse Dylight 800 conjugated	CST	5257		goat

Table S3. Primary and secondary antibodies used for RPPA. All primary antibodies were diluted 1:250 in Superblock T20 (Thermo Fisher Scientific #37536). Secondary antibodies were also diluted in Superblock, at 1:2500. CST, Cell Signaling Technology. mAb, monoclonal antibody.

Target	Accession ID (RefSeq)	Fwd primer	Rev primer	Amplicon size (bp)
NANOG	NM_024865.3	CAGTCTGGACACTGGCTGAA	CTCGCTGATTAGGCTCCAAC	149
NODAL	NM_018055.4	CAGTACAACGCCTATCGCTGT	TGCATGGTTGGTCGGATGAAA	75
POU5F1 (OCT3/4)	NM_002701.5	TGTACTCCTCGGTCCCTTTC	TCCAGGTTTTCTTTCCCTAGC	150
TBP	NM_003194.4	TAATCCCAAGCGGTTTGC	TAGCTGGAAAACCCAACTTCT	170

Table S4. SYBR Green qPCR primers. Bp, base pairs; Fwd, forward; Rev; reverse. Only a single accession ID is given due to space constraints, but all primers detect additional splice isoforms of their specific target.

Dataset 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.

Click here to Download Data S1

Dataset 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.

Click here to Download Data S2

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

Click here to Download Data S3

Dataset S4. List of differentially expressed genes in PIK3CAWT/H1047R vs wild-type MEFs

Click here to Download Data S4

Dataset S5. List of differentially expressed proteins in PIK3CAWT/H1047R vs wild-type hPSCs.

Click here to Download Data S5

Dataset S6. List of differentially expressed proteins in PIK3CAH1047R/H1047R vs wild-type hPSCs.

Click here to Download Data S6

Dataset S7
iPSCs_H1047Rhom_H1047Rhet_total_proteome_shared_differential_expression

Click here to Download Data S7

Dataset S8. List of differentially expressed transcripts changing in the same direction in both heterozygous and homozygous PIK3CAH1047R hPSCs vs wild-type controls.

Click here to Download Data S8

Source data deposited on the Open Science Framework

- High_depth_iPSC_H1047R_RNAseq (related to Fig. 1, 3; Fig. S1, S2)
- M98_iPSCs_E418K_RNaseq (related to Fig. 1)
- MEF_H1047R_RNAseq (related to Fig. 1)
- Total_proteomics (related to Fig. 1, 3; Fig. S2)
- GF_replete_iPSCs_H1047R_RPPA_study (related to Fig. 2; Fig. S3)
- BYL719_iPSCs_H1047R_RPPA_study (related to Fig. 2; Fig. S3)
- WGCNA_iPSC_RNAseq_analysis (related to Fig. 4)
- NODAL_experiment (related to Fig. 5; Fig. S4)

References

Adewumi, O. *et al.* (2007) 'Characterization of human embryonic stem cell lines by the International Stem Cell Initiative', *Nature Biotechnology*, 25(7), pp. 803–816. doi: 10.1038/nbt1318.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society Series B-Methodological*, 57, pp. 289–300.

Bindea, G. *et al.* (2009) 'ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks', *Bioinformatics*, 25(8), pp. 1091–1093. doi: 10.1093/bioinformatics/btp101.

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

Depristo, M. A. *et al.* (2011) 'A framework for variation discovery and genotyping using next-generation DNA sequencing data', *Nature Genetics*, 43(5), pp. 491–501. doi: 10.1038/ng.806.

Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics (Oxford, England)*, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.

Kanehisa, M. *et al.* (2017) 'KEGG: New perspectives on genomes, pathways, diseases and drugs', *Nucleic Acids Research*, 45(D1), pp. D353–D361. doi: 10.1093/nar/gkw1092.

Krämer, A. *et al.* (2014) 'Causal analysis approaches in ingenuity pathway analysis', *Bioinformatics*, 30(4), pp. 523–530. doi: 10.1093/bioinformatics/btt703.

Langfelder, P. and Horvath, S. (2008) 'WGCNA: an R package for weighted correlation network analysis.', *BMC bioinformatics*, 9, p. 559. doi: 10.1186/1471-2105-9-559.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows–Wheeler transform', *Bioinformatics*, 26(5), pp. 589–595. doi: 10.1093/bioinformatics/btp698.

Madsen, R. R. *et al.* (2019) 'Oncogenic PIK3CA promotes cellular stemness in an allele dose-dependent manner', *Proceedings of the National Academy of Sciences*, 116(17), pp. 8380–8389. doi: 10.1073/pnas.1821093116.

McKenna, A. *et al.* (2010) 'The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research*, 20(9), pp. 1297–1303. doi: 10.1101/gr.107524.110.

McLaren, W. *et al.* (2016) 'The Ensembl Variant Effect Predictor', *Genome Biology*, 17(1), p. 122. doi: 10.1186/s13059-016-0974-4.

Moniz, L. S. *et al.* (2017) 'Phosphoproteomic comparison of Pik3ca and Pten signalling identifies the nucleotidase NT5C as a novel AKT substrate', *Scientific Reports*, 7, p. 39985. doi: 10.1038/srep39985.

Ritchie, M. E. *et al.* (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research*, 43(7), pp. e47–e47. doi: 10.1093/nar/gkv007.

Robin, X. *et al.* (2019) 'Probability-based detection of phosphoproteomic uncertainty reveals rare signaling events driven by oncogenic kinase gene fusion', *bioRxiv*, p. 621961. doi:

10.1101/621961.

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

Robinson, M. D. and Oshlack, A. (2010) 'A scaling normalization method for differential expression analysis of RNA-seq data', *Genome Biology*, 11(3), p. R25. doi: 10.1186/gb-2010-11-3-r25.

Shannon, P. *et al.* (2003) 'Cytoscape: a software environment for integrated models of biomolecular interaction networks.', *Genome research*, 13(11), pp. 2498–504. doi: 10.1101/gr.1239303.

Tyanova, S., Temu, T. and Cox, J. (2016) 'The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.', *Nature protocols*. Nature Publishing Group, 11(12), pp. 2301–2319. doi: 10.1038/nprot.2016.136.

Wasserstein, R. L., Schirm, A. L. and Lazar, N. A. (2019) 'Moving to a World Beyond " p < 0.05"', *The American Statistician*. Taylor & Francis, 73(sup1), pp. 1–19. doi: 10.1080/00031305.2019.1583913.

Zhang, B. and Horvath, S. (2005) 'A General Framework for Weighted Gene Co-Expression Network Analysis', *Statistical Applications in Genetics and Molecular Biology*, 4(1). doi: 10.2202/1544-6115.1128.