Dissecting the pharmacological landscape of cancer cells to reveal novel target populations

Tzen S Toh [1,#], Nirmal Keshava [2,#], Jonathan Dry [3], Michael Menden [4,*], Dennis Wang [5,*]

1. The Medical School, University of Sheffield, Sheffield, UK
2. Constellation Analytics, LLC., Needham MA, USA
3. Oncology, IMED Biotech Unit, AstraZeneca R&D Boston, Waltham MA, USA
4. Oncology, IMED Biotech Unit, AstraZeneca, Cambridge, UK
5. NIHR Sheffield Biomedical Research Centre, University of Sheffield, Sheffield, UK

# equal contribution
*corresponding authors

Abstract

The effectiveness of a particular drug has predominantly been analysed in isolation and there lacks data-driven approaches to consider the full response pattern between multiple drugs to study biomarkers at the same time. To reveal subpopulations where the pharmacological response between compounds agree and diverge, we applied a novel population segmentation algorithm, POET, to compare 344 drug pairs targeting the MAPK and PI3K-AKT pathways across >800 genomically-diverse cancer cell lines. We show that POET was capable of integrating multiple measures of drug response to identify subpopulations that differentiate response to inhibitors of the same or different targets. MEK, BRAF and PI3K inhibitors with different sensitive subpopulations were shown to be effective as combined therapies, particularly when stratified for *BRAF* mutations. This data-driven approach paves a new way for patient stratification to identify novel cancer subpopulations, their genetic biomarkers, and effective drug combinations.

Introduction

Drug developers face a conundrum in predicting the efficacy of their investigational compound compared to existing drugs used as the standard of care treatment. Systematic screening of drug compounds across a variety of genomic backgrounds in cancer cell lines can improve clinical trial design and personalize treatments [1]. Following the pioneering NCI-60 screen comprised of 59 unique cell lines [2], modern high-throughput screens such as the Genomics of Drug Sensitivity in Cancer (GDSC) [3,4], the Cancer Cell Line Encyclopedia (CCLE) [5] and the Cancer Therapeutics Response Portal (CTRP) [6–8] have characterised >1,000 cancer cell lines with the goal of establishing the genetic landscape of cancer. The deep molecular characterisation of these large cell line panels is complemented with high-throughput drug screens, which enables the discovery of drug response biomarkers. For example, analysis of the drug PLX4720, SB590885 and CI-1040 reproduced drug sensitivity association with the *BRAF* mutation in melanoma, or afatinib with *ERBB2* amplifications in breast cancer [3,4,9]. These associations between genetic variants and treatment response have helped identify specific patient subpopulations who are most likely to benefit from treatment. In Phase III clinical trials, however, for new drugs to be successful, they must demonstrate a significant improvement over the existing standard of care. Accurately defining in which subpopulations a new drug demonstrates improved differential efficacy over other drugs targeting the same disease could lead to both better clinical outcomes as well as new targeted therapies.

While several methods have been proposed to identify drug response biomarkers in cell lines [4,510,11], there is a lack of systematic, data driven approaches to assess and compare monotherapy responses across multiple drugs and consecutively gain mechanistic insights from biomarkers. Prioritizing drugs that are more likely to succeed in clinical trials in the absence of well-established clinical biomarkers, therefore, becomes more challenging. Here, we comprehensively assess the differences in responses of a population of pan-cancer cell lines to multiple drugs, with the goal of discovering subpopulations in which cell lines possess a similar pattern of response across drugs, and from which, biomarkers distinguishing subpopulations can be derived. Accurately identifying distinct subpopulations of cell lines based on their joint pharmacological pattern of response and evaluating which subpopulations describe the independent action of single drugs could identify drug combinations [12].

We present results from experiments involving two or more drugs that successively highlight how an unsupervised machine learning technique identifies distinct pharmacological patterns of response and therapeutic biomarkers. Addressing the challenges in comparing the response of two drugs, the initial experiment assesses two gold standards with established clinical biomarkers, namely the differential response of a BRAF inhibitor and MEK inhibitor with anticipated *BRAF* and *KRAS* mutations, and an EGFR inhibitor and MEK inhibitor with expected biomarkers of *EGFR*, *ERBB2* and *KRAS* mutations. Next, we systematically investigated how multiple drugs targeting the MAPK and PI3K-AKT pathway yields different patterns of response within subpopulations and biomarker context. We test a previous study's hypothesis suggesting the benefit of drug combinations can be explained through independent action rather than probable synergy by examining subpopulations uniquely sensitive to a single drug [12]. Finally, we demonstrate how the discovery and analysis of subpopulations can guide the design of clinical trials by revealing which indications may respond best to a particular drug and what biomarkers identify target patient subpopulations.

Results

We used a novel unsupervised machine learning technique, POET (Population Outcome Enrichment Technique), to discover subpopulations of cell lines in which two or more compounds, possibly addressing the same disease state or even targeting the same genetic alteration, have a common pharmacological pattern of response. By further associating enriched genetic alterations in subpopulations with specific patterns of response, we also gain a deeper understanding of molecular mechanisms responsible for patient subpopulations that respond differently to two drugs.

Identifying subpopulations of differential drug response

To identify unique drug sensitive or resistant subpopulations for a portfolio of cancer therapies, we focused on agents targeting ERK or PI3K-AKT signaling, exploring up to 1,000 cancer cells derived from the GDSC database. Drug responses were quantified with the drug concentration required to reduce cell viability by half (IC_{50}) and the area under the dose-response curve (AUC; **Fig. 1a**). We applied POET, an unsupervised machine learning approach, to both drug response summary metrics. POET employs a multivariate similarity measures to compare the patterns of response for each distinct pair of cell lines, while requiring no *a priori* assumptions on the number or distribution of the subpopulations. The result is a diverse cell line population segmented into distinct subpopulations having homogeneous patterns of drug response (**Fig. 1b**). Here exemplified, we compare the drug response of 802 cell lines treated with either SB590885 (BRAF inhibitor) or CI-1040 (MEK inhibitor; **Fig. 1c**). Segmentation by POET resulted in 7 distinct subpopulations with a median size of 40 cell lines. The subpopulation sensitive to both inhibitors was significantly enriched for *BRAF* mutants (P=3.87e-14; **Fig. 1c**), while another subpopulation was exclusively sensitive to the MEK inhibitor and significantly enriched for *KRAS* mutations (P=0.00589).

In another example, we compared selumetinib (MEK inhibitor) with afatinib (EGFR and ERBB2 dual inhibitor) across 812 overlapping cell lines (**Fig. 1d**). Strong sensitivity markers for selumetinib are subpopulations carrying *KRAS*, *NRAS* and *BRAF* mutations [4,5,13] (**Fig. 1e,d**). A less anticipated association is *APC* loss-of-function sensitivity to selumetinib, albeit concordant results are found with trametinib (another MEK inhibitor) in APC deficient mice [14]. We reproduced the well-established associations of afatinib with either *EGFR* and *ERBB2* amplifications [4,15], and surprisingly our unsupervised segmentation returned two subpopulations enriched for *EGFR* amplifications. The more sensitive subpopulation is solely enriched for *EGFR* amplifications, whilst the less sensitive subpopulations additional includes activating *PIK3CA* mutations. In concordance with recent literature, PI3K-AKT signaling drives acquired drug resistance to EGFR inhibitors in lung cancer [16].

Segmentation by POET resulted in 14 subpopulations with a median size of 38 (**Fig. 1d**). The subpopulation enriched for *EGFR*, *ERBB2* and *PI3KCA* variants, has an average $\log(IC_{50})$ of 0.9486μM for selumetinib and -0.596μM for afatinib. In contrast, the *BRAF* mutation was enriched in a subpopulation where the average $\log(IC_{50})$ for selumetinib was -1.061μM and 0.593μM for afatinib. The difference in response between afatinib and selumetinib was significantly greater (t-test P<0.01) between the subpopulations identified by POET and the total population of *PIK3CA* or *BRAF* mutant cell lines (**Fig. 1f, g**).

Distribution of subpopulations highlight distinct drug-drug relationships

We extracted 37 compounds from the GDSC database, which are either putatively targeting PI3K-AKT or ERK signalling (**Fig. 2a**), and systematically compared each drug pair targeting both key cancer pathways (**Fig 2b**). In total, we performed 342 pairwise comparisons of 18 PI3K-AKT and 19 MAPK pathway inhibitors. Each drug pair was classified into five categories based on the distribution of subpopulation drug responses: (i) no differential response, (ii) sensitive to both MAPK and PI3K-AKT pathway inhibitors (i.e. correlated response), (iii) preferential MAPK pathway sensitivity, (iv) preferential PI3K-AKT pathway sensitivity, (v) sensitive to either a MAPK pathway or a PI3K-AKT pathway inhibitor, i.e. anti-symmetric response (Supplementary Figure 2).

Subpopulations that have sensitivity to both, PI3K-AKT and MAPK pathway inhibitors, was observed in 29 instances. This phenotype was significantly enriched when comparing a CRAF inhibitor (TL-2-105) to PI3K-AKT signaling inhibitors (P=1.85e-6). The same trend was observed in inhibiting ERK (FR-180204) or RSK (FMK) in combination with any PI3K-AKT signaling agent (P=1.85e-6 and P=1.078e-7 respectively), but interestingly never encountered when combining with either BRAF or MEK inhibitors.

There were 71 drug pairs with a significantly high proportion of subpopulations ($P < 0.05$) exhibiting greater sensitivity to MAPK pathway inhibition. This phenotype is strongly pronounced in pairs with BRAF, ERK (FR-180204) and RSK (FMK) inhibitors (P=0.00344, P=0.00437 and P=0.000882, respectively; hypergeometric test). In contrast, 28 drug pairs were found with significantly high proportions of preferential PI3K-AKT pathway inhibition. In total, 28 drug pairs showed this phenotype, with an enrichment of 19 MEK inhibitors (hypergeometric test P=0.000114). MEK inhibitors were particularly enriched when paired with PI3K or PDK1 inhibitor (hypergeometric test P=0.000221).

In 55 cases, we observed drug pairs with sensitivity to either a MAPK pathway or a PI3K-AKT pathway inhibitor, i.e. anti-symmetric response. This response type was enriched for pairs of any PI3K-AKT pathway inhibitors and EGFR (erlotinib), BRAF (PLX4720-1 and PLX4720-2), or MEK inhibitors (P=9.57e-6, P=0.000384 and P=0.0297, respectively; hypergeometric test), while even more significant when focusing on AKT inhibitors in pair with either EGFR (erlotinib), BRAF (PLX4720-1 and PLX4720-2), or MEK inhibitors (P=0.0141, P=0.000292 and P=0.000379, respectively; hypergeometric test). Response patterns for all drug pairs are viewable from our portal (https://szen95.github.io/supp_poet/).

Anti-symmetric pattern of response predicts drug combination efficacy

Tumour heterogeneity among patients cloaks specific subpopulations that might show drug sensitivity, exhibiting the need to systematically stratify them and thereby enable the identification of suitable drug combinations. Previous studies have hypothesised that the efficacy of many approved drug combinations can be explained by the independent action of single agents on different patient subpopulations [12]. This suggests that revealing more specific patient subpopulations and designing drug combinations by optimizing their independent action can potentially improve drug combination efficacy.

Using POET, we show a category of joint pharmacological response that occurs when the sensitivity of cell lines to two drugs (where one may be the existing standard-of-care) is strongly divergent, or anti-symmetric. Two drugs are deemed to have anti-symmetric

response when there is a high proportion of subpopulations sensitive for either drug, but not both. We observed this phenomenon in 55 drug pair comparisons, including a MEK inhibitor (RDEA119-2) showing anti-symmetric responses to four PI3K inhibitors (PI-103, GSK2126458, ZSTK474, and PIK-93; Supplementary Figure 3). Drug pairs with anti-symmetric response are also observed in cell lines treated with PLX4720-1 (BRAF inhibitor) and PI-103 (MEK inhibitor) after undergoing segmentation by POET (**Fig. 3a**). Two subpopulations identified with greater sensitivity to PI-103 contained a high proportion of cells with the *BRAF* mutation (**Fig. 3b**).

Collectively, our results suggest that drug pairs exhibiting anti-symmetric response in a specific subpopulation may be efficacious in treating tumors when given in combination. We tested our hypothesis in cell lines and patient-derived tumor xenograft models (PDXs) from two independent studies (**Methods**) to investigate whether the drug pairs with anti-symmetric response exhibited any efficacy as a combination treatment. Based on our observations of anti-symmetric response in subpopulations, we tested whether pairs of drugs with anti-symmetric response, given in combination, are efficacious in cancer cell lines [17] as well as patient-derived tumor xenograft (PDX) models [18] (**Fig. 3c**).

Our results in both cell lines and PDXs confirm subpopulations with anti-symmetric behavior for a specific drug combination lead to greater efficacy. Using the synergy score as a measurement of efficacy in cell lines, the MEK/PI3K inhibitor combination yielded a greater synergistic effect compared to all other drug combinations (t-test P=2.8e-9, **Fig. 3d**, Supplementary Figure 4). The synergistic effect is more pronounced in cell lines with the *BRAF* mutation compared to all cell lines (t-test P=0.0102). Similarly in PDXs, we observed a smaller increase in tumour volume of the BRAF/PI3K inhibitor combination compared to all other combinations (t-test P=1.93e-8; **Fig. 3e**, Supplementary Figure 4). *BRAF* mutant tumours experience an even smaller increase in tumour volume when treated with the BRAF/PI3K inhibitor combination compared to all tumours (t-test P=0.00201).

Discussion

The ability to identify enriched mutations in distinct subpopulations is the basis for therapeutic biomarkers, which ultimately may increase the likelihood of successful clinical trials [19,20]. First, we investigated gold standards with well-established clinical biomarkers by comparing the response patterns for BRAF (SB590885) and MEK (CI-1040) inhibition, which expectedly reproduced subpopulations sensitive to both enriched for *BRAF* mutants (**Fig. 1c**) [21–23]. In another example, when comparing EGFR/ERBB2 dual (afatinib) and MEK (selumetinib) inhibition (**Fig. 1d,e**), we found the expected biomarkers such as *BRAF*, *KRAS* and *NRAS* mutations for selumetinib [24–27], and afatinib associated with *EGFR* and *ERBB2* amplifications [28,29]; however, interestingly, we observe two subpopulations enriched for *EGFR* and *ERBB2* amplifications, whilst the more afatinib resistant population had enriched an additional *PI3KCA* activating mutation, which may cause acquired resistance (**Fig. 1d,e**) [16].

After analysing gold standards, consecutively we systematically compared 19 and 18 drugs targeting the MAPK and PI3K pathway respectively (**Fig. 2a**). We identified five different response patterns; these are (i) no differential response, (ii) subpopulations that are concordantly sensitive, (iii, iv) subpopulations exclusively sensitive to one drug, and (v) an anti-symmetric response type (**Fig. 2b**). Surprisingly, we observed concordant sensitivity when inhibiting CRAF, ERK or RSK across all levels of the PI3K-AKT pathway, but not

enriched for any other kinase in the MAPK signalling which are either up- or downstream of *CRAF*, *ERK* or *RSK* [30]. Intuitively BRAF inhibitors are enriched for MAPK pathway sensitivity which is strongly driven by *BRAF* mutations [21], while unexpected MEK inhibition has a preference for PI3K-AKT signalling sensitivity [25]. EGFR, BRAF and MEK inhibitors show an enrichment for anti-symmetric response, particularly when targeting downstream of the PI3K-AKT pathway. Investigating the MAPK and PI3K-AKT pathway based on drug response profiles highlights how intertwined those two pathways are in pharmacology space [30].

Arguably, the anti-symmetric response type is the most exciting, since it may explain independent drug action in effective drug combinations [12]. Here exemplified, we showed that PI3K inhibitors combined with either BRAF or MEK inhibitors increase *in vitro* synergy and reduce tumour volume of PDX models [31,32]. Furthermore, we were able to show that this synergy and tumour shrinkage can be further enhanced by the correct biomarker indication, in this instance, *BRAF* mutant subpopulations [33,34], which were systematically identified with our segmentation and biomarker pipeline [12].

A “bottom up” approach of hierarchical clustering has been utilized routinely to attribute molecular markers to differences in subpopulation drug response and outcomes [35,36]. Yet, continuing concern about the sensitivity and inherent suboptimality of clustering techniques [37,38], have motivated alternative techniques to identify intrinsic subpopulations. Because of their success in other industries [39,40] and their natural amenability to matrix decomposition techniques, network-based approaches have emerged as viable alternatives for discovering distinct subpopulations [41–43]. Deeper interpretations of matrix subspaces may provide further insight into the linkage between subpopulations of cancer cell lines and drugs.

Our work demonstrates several important insights about the pharmacological pattern of response of different cancer drugs by applying a new machine learning platform, POET, to segment populations from a large pan-cancer *in vitro* pharmacology data set. By organizing cell lines by similar pharmacological patterns of response, we identified distinct, intrinsic subpopulations sensitive to one drug but resistant to others, and in some cases identified genetic alterations that are biomarkers for those subpopulations. In the context of analytical frameworks for increasing drug R&D productivity by sharpening the focus of drugs [44], our work demonstrates the value of advanced analytical methods in combination with pre-clinical data to enable decision making that is more data-driven and less ambiguous. Moreover, by analyzing pharmacological responses with POET and interpreting its outputs in the context of the underlying genetics and molecular pathways, we have created a multi-faceted landscape for developing and assessing new cancer therapies.

Methods

Pharmacology data

The discovery pharmacology dataset was extracted from the The Genomics of Drug Sensitivity in Cancer (GDSC) database [3,4], while leads from the analysis were validated with the Cancer Cell Line Encyclopedia (CCLE) [5] and the Cancer Therapeutics Response Portal (CTRP) [6–8]. Furthermore, suggested drug combinations were validated with cell line responses from the AstraZeneca-DREAM challenge dataset [17] and patient derived xenograft (PDX) models from Gao *et al.* [18].

For a given cell line in GDSC, the drug response was fitted with a sigmoid curve [45] and consecutively quantified as area under the curve (AUC) or the concentration required to reduce cell viability by half (IC_{50}). GDSC contains 265 compounds tested in 990 cell lines, whilst we focus on a subset of 38 drugs targeting either the PI3K-AKT or ERK signalling, which leads to 344 experiments considered for evaluation.

Deep molecular characterisation of the cancer cell lines

The GDSC resource provides the characterisation of >1,000 cell lines including whole exome sequencing and SNP6.0 arrays, which enabled to quantify gene-level mutational and copy number variation status. Additional, 10 key fusion genes were included in this analysis, which is summarized in the binary event matrix (BEM) from Iorio *et al.* [4].

POET – workflow from AUC/IC_{50} values to cell subpopulations

Cell lines pharmacology measurements were analyzed using a new population segmentation framework, POET [46], based upon network models that recursively segment a population whose members are described by multiple, possibly heterogeneous variables, into distinct subpopulations by identifying optimal cuts for graph bisection. Traditional approaches, such as agglomerative, "bottoms-up" hierarchical clustering and iterative *K*-Means clustering are greedy algorithms that are inherently sub-optimal in constructing clusters and consequently may not identify the most distinct clusters. Moreover, these approaches frequently require *a priori* estimates of the number of sub-populations for which many heuristics exist but in practice is commonly estimated using trial and error.

POET builds network models by calculating a similarity value between two compounds using two important measurements extrapolated from the cell line pharmacology screens: the IC_{50} value and the AUC of the dose-response curve (Supplementary Table 1) observed when one compound is applied *in vitro* to a single cell line sample at successively greater concentrations. When two compounds are compared, the inputs to POET comprise a length-4 dose-response profile (DRP) for that cell line. The similarity between any two DRPs is calculated on a range between 0 and 1 by a multivariate quasi-Gaussian comparison that differences the elements of the DRPs but also weighs the differences by a combination of local and global network statistics. The framework is generalizable to include more variables and different measures of similarity.

Recursive segmentation of a local population is terminated when the similarity between the resulting sub-populations does not meet the criteria of a significance test, but other practical constraints such as sub-population size and tree depth can modulate segmentation. When employed recursively, POET segments a population, *P*, consisting of *N* members, each described by *M* variables, into *K* sub-populations, $P_k,\ k = 1, ..., K$, where $P = \bigcup_{k=1} P_k$.

Successive segmentation results in increasingly homogeneous sub-populations in which the input variables span narrower ranges of values than the overall population, and the cell lines consequently demonstrate a common pattern of response across compounds. Successive segmentation of subpopulations can be halted by one or more criteria. In our experiments we required subpopulations to have at least 40 members in order to be segmented. Further, we required that both resulting subpopulations have 20 or more members in order to be retained.

Because genetic alterations in each cell lines are known, each subpopulation can be evaluated by significance tests to identify enriched alterations that may be attributed to patterns of sensitivity or resistance in that subpopulation to an individual compound.

To visualize output from POET, we first calculated the average log(IC_{50}) values for subpopulations generated based on their response to the tested drug pairs (Supplementary Table 2). We then plotted the values on a scatter plot using an open-source Python library called Matplotlib. Dashed lines indicative of 20th percentile of log(IC_{50}) values for each drug were also plotted on the scatter plot.

Tree visualization of segmentations

We utilized tree diagrams to visualize the data generated by our algorithm. The tree diagrams illustrate how the algorithm segments the cancer cell lines into different subpopulations, based on whether they are sensitive or resistant to the drugs that are being tested. The tree diagrams were generated through an open-source Python library called Graphviz. The style of each component of the tree diagram was first initialized through a class. This included the colours, shapes, and fonts of the edges and nodes of the tree diagram. A method was created to handle the creation of the tree diagrams. It took parameters that included the number of vertices and leaves, the labels for the leaves, and the tree diagram filename. The tree diagram is finally generated and saved by calling the method.

Enrichment of features to nominate biomarkers

For each subpopulation, we measured the number of cell lines in the subpopulation with a particular gene mutation, and the number of cell lines outside of the subpopulation with the mutation. A 2X2 contingency table was generated from the cell line counts of with/without mutation and inside/outside of subpopulation. Significance of observed enrichment of mutations within subpopulations were calculated using the Fisher's exact test. The resulting p-values were corrected for multiple testing using the Benjamini and Hochberg (BH) procedure (Supplementary Table 3).

Classification of pair-wise drug responses

POET made 342 pairwise comparisons of drugs targeting the MAPK and PI3K-AKT pathways. Based on the distribution of log(IC_{50}) values across all cell lines tested with both drugs, we determined the 20th-percentile of log(IC_{50}) values for each drug. The 20th percentile cutoffs P_{20} for drugs A and B was used to categorise the average log(IC_{50}) y of each subpopulation i into four categories:

$y_i < P_{20,A}$ and $y_i < P_{20,B}$ = *sensitive to drugs A and B*

$y_i < P_{20,A}$ and $y_i \geq P_{20,B}$ = *more sensitive to drug A*

$y_i \geq P_{20,A}$ and $y_i < P_{20,B}$ = *more sensitive to drug B*

$y_i \geq P_{20,A}$ and $y_i \geq P_{20,B}$ = *resistant to drugs A and B*

The number of subpopulations in each category were recorded in an 2X2 contingency matrix and normalised by the number of cell lines in each subpopulation. For each drug pair, a binomial test was performed to test whether the number of subpopulations in each category is greater than what would be expected.

After classification of pair-wise drug responses, we assessed whether a drug was significantly enriched for one category in comparisons with all other drugs. Testing was carried out using the hypergeometric test (*phyper* R package).

References

1. Weinstein, J. N. Cell lines battle cancer. *Nature* **483,** 544–545 (2012).
2. Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6,** 813–823 (2006).
3. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483,** 570–575 (2012).
4. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166,** 740–754 (2016).
5. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
6. Rees, M. G. *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12,** 109–116 (2016).
7. Seashore-Ludlow, B. *et al.* Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov.* **5,** 1210–1223 (2015).
8. Basu, A. *et al.* An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154,** 1151–1161 (2013).
9. Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41,** D955–61 (2013).
10. Costello, J. C. *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32,** 1202–1212 (2014).
11. Menden, M. P. *et al.* Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One* **8,** e61318 (2013).
12. Palmer, A. C. & Sorger, P. K. Combination Cancer Therapy Can Confer Benefit via

Patient-to-Patient Variability without Drug Additivity or Synergy. *Cell* **171,** 1678–1691.e13 (2017).

13. Jain, N. *et al.* Phase II study of the oral MEK inhibitor selumetinib in advanced acute myelogenous leukemia: a University of Chicago phase II consortium trial. *Clin. Cancer Res.* **20,** 490–498 (2014).
14. Fujishita, T., Kajino-Sakamoto, R., Kojima, Y., Taketo, M. M. & Aoki, M. Antitumor activity of the MEK inhibitor trametinib on intestinal polyp formation in Apc(Δ716) mice involves stromal COX-2. *Cancer Sci.* **106,** 692–699 (2015).
15. De Grève, J. *et al.* Clinical activity of afatinib (BIBW 2992) in patients with lung adenocarcinoma with mutations in the kinase domain of HER2/neu. *Lung Cancer* **76,** 123–127 (2012).
16. Jacobsen, K. *et al.* Convergent Akt activation drives acquired EGFR inhibitor resistance in lung cancer. *Nat. Commun.* **8,** 410 (2017).
17. Menden, M. P. *et al.* A cancer pharmacogenomic screen powering crowd-sourced advancement of drug combination prediction. (2017). doi:10.1101/200451
18. Gao, H. *et al.* High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21,** 1318–1325 (2015).
19. Long, G. V. *et al.* Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N. Engl. J. Med.* **371,** 1877–1888 (2014).
20. Van Allen, E. M. *et al.* The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov.* **4,** 94–109 (2014).
21. Holderfield, M., Deuker, M. M., McCormick, F. & McMahon, M. Targeting RAF kinases for cancer therapy: BRAF-mutated melanoma and beyond. *Nat. Rev. Cancer* **14,** 455–467 (2014).
22. Hall, R. D. & Kudchadkar, R. R. BRAF mutations: signaling, epidemiology, and clinical experience in multiple malignancies. *Cancer Control* **21,** 221–230 (2014).

23. Abdel-Wahab, O. *et al.* Efficacy of intermittent combined RAF and MEK inhibition in a patient with concurrent BRAF- and NRAS-mutant malignancies. *Cancer Discov.* **4,** 538–545 (2014).
24. Patel, S. P. *et al.* Clinical responses to selumetinib (AZD6244; ARRY-142886)-based combination therapy stratified by gene mutations in patients with metastatic melanoma. *Cancer* **119,** 799–805 (2013).
25. Catalanotti, F. *et al.* Phase II trial of MEK inhibitor selumetinib (AZD6244, ARRY-142886) in patients with BRAFV600E/K-mutated melanoma. *Clin. Cancer Res.* **19,** 2257–2264 (2013).
26. Metro, G. *et al.* Selumetinib: a promising pharmacologic approach for KRAS-mutant advanced non-small-cell lung cancer. *Future Oncol.* **9,** 167–177 (2013).
27. Kim, D. W. & Patel, S. P. Profile of selumetinib and its potential in the treatment of melanoma. *Onco. Targets. Ther.* **7,** 1631–1639 (2014).
28. Paz-Ares, L. *et al.* Afatinib versus gefitinib in patients with EGFR mutation-positive advanced non-small-cell lung cancer: overall survival data from the phase IIb LUX-Lung 7 trial. *Ann. Oncol.* **28,** 270–277 (2017).
29. Takezawa, K. *et al.* HER2 amplification: a potential mechanism of acquired resistance to EGFR inhibition in EGFR-mutant lung cancers that lack the second-site EGFRT790M mutation. *Cancer Discov.* **2,** 922–933 (2012).
30. Mendoza, M. C., Er, E. E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* **36,** 320–328 (2011).
31. Haagensen, E. J., Kyle, S., Beale, G. S., Maxwell, R. J. & Newell, D. R. The synergistic interaction of MEK and PI3K inhibitors is modulated by mTOR inhibition. *Br. J. Cancer* **106,** 1386–1394 (2012).
32. Sweetlove, M. *et al.* Inhibitors of pan-PI3K Signaling Synergize with BRAF or MEK Inhibitors to Prevent BRAF-Mutant Melanoma Cell Growth. *Front. Oncol.* **5,** 135 (2015).

33. Temraz, S., Mukherji, D. & Shamseddine, A. Dual Inhibition of MEK and PI3K Pathway in KRAS and BRAF Mutated Colorectal Cancers. *Int. J. Mol. Sci.* **16,** 22976–22988 (2015).

34. Villanueva, J., Vultur, A. & Herlyn, M. Resistance to BRAF inhibitors: unraveling mechanisms and future treatment options. *Cancer Res.* **71,** 7137–7140 (2011).

35. Gholami, A. M. *et al.* Global proteome analysis of the NCI-60 cell line panel. *Cell Rep.* **4,** 609–620 (2013).

36. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14863–14868 (1998).

37. Stefano Monti, Pablo Tamayo, Jill Mesirov, Todd Golub. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach. Learn.* **52,** 91–118 (2003).

38. Ramoni, M. F., Sebastiani, P. & Kohane, I. S. Cluster analysis of gene expression dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 9121–9126 (2002).

39. Sarwar, B., Karypis, G., Konstan, J. & Reidl, J. Item-based collaborative filtering recommendation algorithms. in *Proceedings of the tenth international conference on World Wide Web - WWW '01* (2001). doi:10.1145/371920.372071

40. Scott, J. *Social Network Analysis*. (SAGE, 2017).

41. Wang, B. *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* **11,** 333–337 (2014).

42. Li, L. *et al.* Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.* **7,** 311ra174 (2015).

43. Nicolau, M., Levine, A. J. & Carlsson, G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7265–7270 (2011).

44. Cook, D. *et al.* Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **13,** 419–431 (2014).

45. Vis, D. J. *et al.* Multilevel models improve precision and speed of IC50 estimates. *Pharmacogenomics* **17,** 691–700 (2016).

46. Keshava, N. POET: A Precision Medicine Platform for Data Integration, Population Segmentation, Patient Stratification, and Personalized Prediction (under review).

Supplementary Information

1. Supplementary Information
2. Supplementary Table 1
3. Supplementary Table 2
4. Supplementary Table 3
5. Supplementary Website

Acknowledgements

We would like to thank Ben Sidders (AstraZeneca plc.) and Winston Hide (Harvard University) for useful discussions. This work is supported by funding from the National Institute for Health Research Sheffield Biomedical Research Centre.

Author contributions

NK, TST, JD, MM, and DW contributed to the conceptualization of the project. NK, TST, and DW were responsible for data curation. Formal analysis was performed by NK, TST, MM, and DW. DW was solely responsible for funding acquisition. The methodology was developed by NK, TST, MM, and DW. TST and DW were responsible for administration of the project. Resources for the experiments were prepared by NK. Various software for the project was developed and implemented by NK, TST, and DW. JD, MM, and DW oversaw supervision of the project. NK, TST, JD, MM, and DW wrote the manuscript.

Competing interests

M.M and J.D. are employees of AstraZeneca plc., and N.K. is with Constellation Analytics, LLC.

Figures

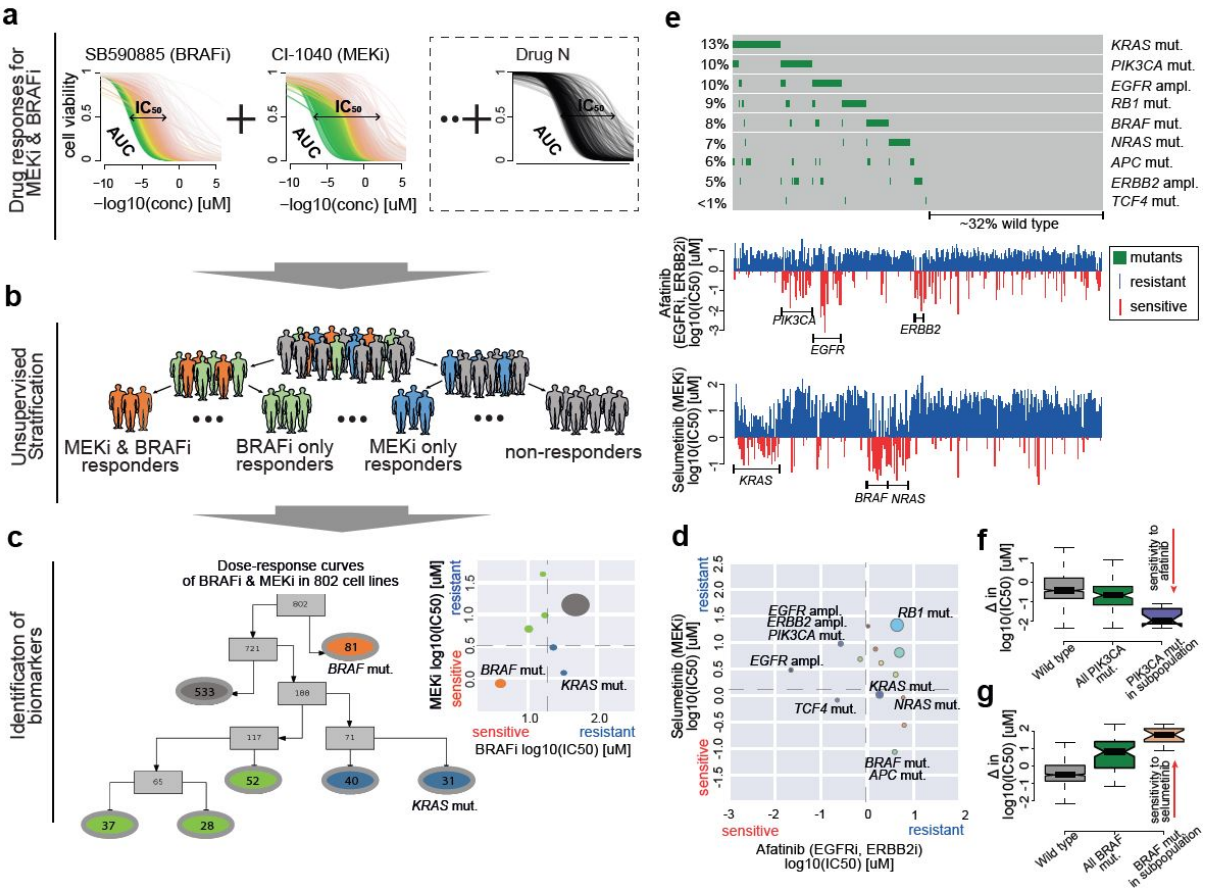

Figure 1: Segmentation of a population based on pharmacological patterns of response discovers subpopulations with differential sensitivity. a Dose response curves of two or more drugs are measured across a population of up to 1,000 cancer cell lines. **b** The population is segmented into distinct and homogeneous subpopulations based on their response to multiple drugs. When comparing two drugs, subpopulations can be categorised based on their mean log(IC_{50}s): sensitive to both drugs (orange), sensitive to drug A but not drug B (green), sensitive to drug B but not drug A (blue), resistant to both drugs (grey). **c** Segmentation results for a BRAF inhibitor (SB590885) and a MEK inhibitor (CI-1040) using POET. Tree nodes contain the number of cell lines and are colored based upon their category of response. Significance testing reveals subpopulations enriched for *BRAF* and *KRAS* mutations. **d** Scatter plot showing POET-derived subpopulations based on their pharmacological responses for afatinib and selumetinib. Dashed lines indicate 20th percentile of log(IC_{50}) values for each drug. *PIK3CA, EGFR, ERBB2, KRAS, NRAS, BRAF, APC, TCF4* and *RB1* mutations were found enriched in the associated subpopulations. **e** OncoPrint visualizing the percentage of mutations of selected genes in cell line panel treated with either afatinib (EGFR inhibitor) or selumetinib (MEK inhibitor). The waterfall plots compare response of the cell lines to afatinib and to selumetinib. **f** Boxplot of difference in log(IC_{50}) values between afatinib and selumetinib response for wild-type cell lines, all cell lines with *PI3KCA* mutation and cell lines in POET-derived subpopulations with enriched *PI3KCA* mutation. **g** Same as panel **f**, but for *BRAF* mutations.

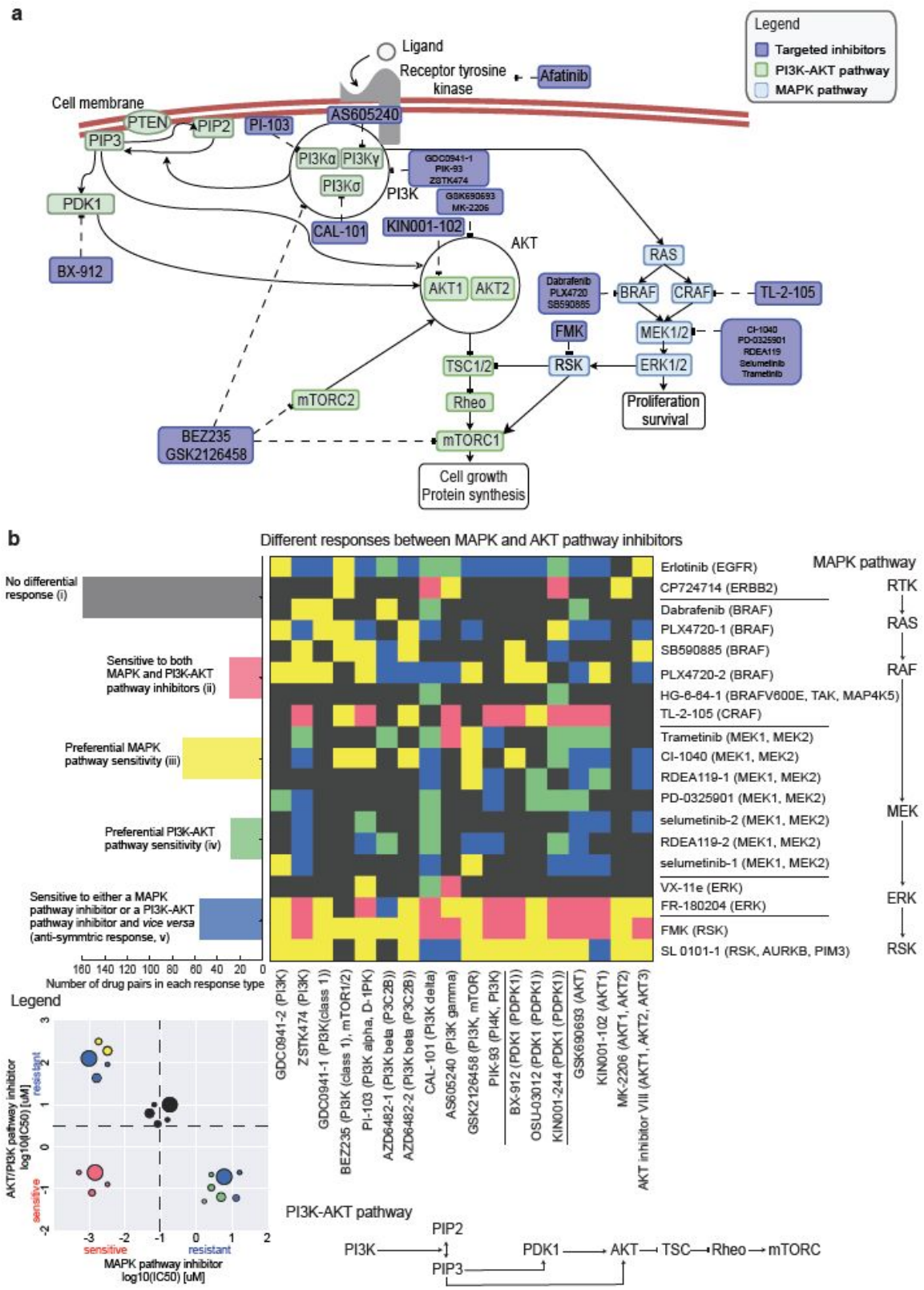

Figure 2: Distinct drug response types after unsupervised segmentation of pharmacological response pattern for targeting ERK or PI3K-AKT signaling. a MAPK and PI3K-AKT pathways illustrating drugs in purple boxes which were assessed by POET, and their different gene targets in the pathway. Genes in the green boxes are involved in the PI3K-AKT pathway while genes in the blue boxes are

involved in the MAPK pathway. **b** Heatmap illustrating pairwise comparison of responses for 19 different inhibitors targeting the MAPK pathway and 18 different inhibitors targeting the PI3K-AKT pathway. It illustrates five possible classes of differential response when comparing two drugs: (i) no differential response (black), (ii) subpopulations sensitive to both MAPK and PI3K-AKT pathway inhibitors (pink), (iii) preferential MAPK pathway sensitivity (yellow), (iv) preferential PI3K-AKT pathway sensitivity (green) or (v) sensitive to either a MAPK pathway inhibitor or a PI3K-AKT pathway inhibitor and *vice versa* (anti-symmetric response, blue).

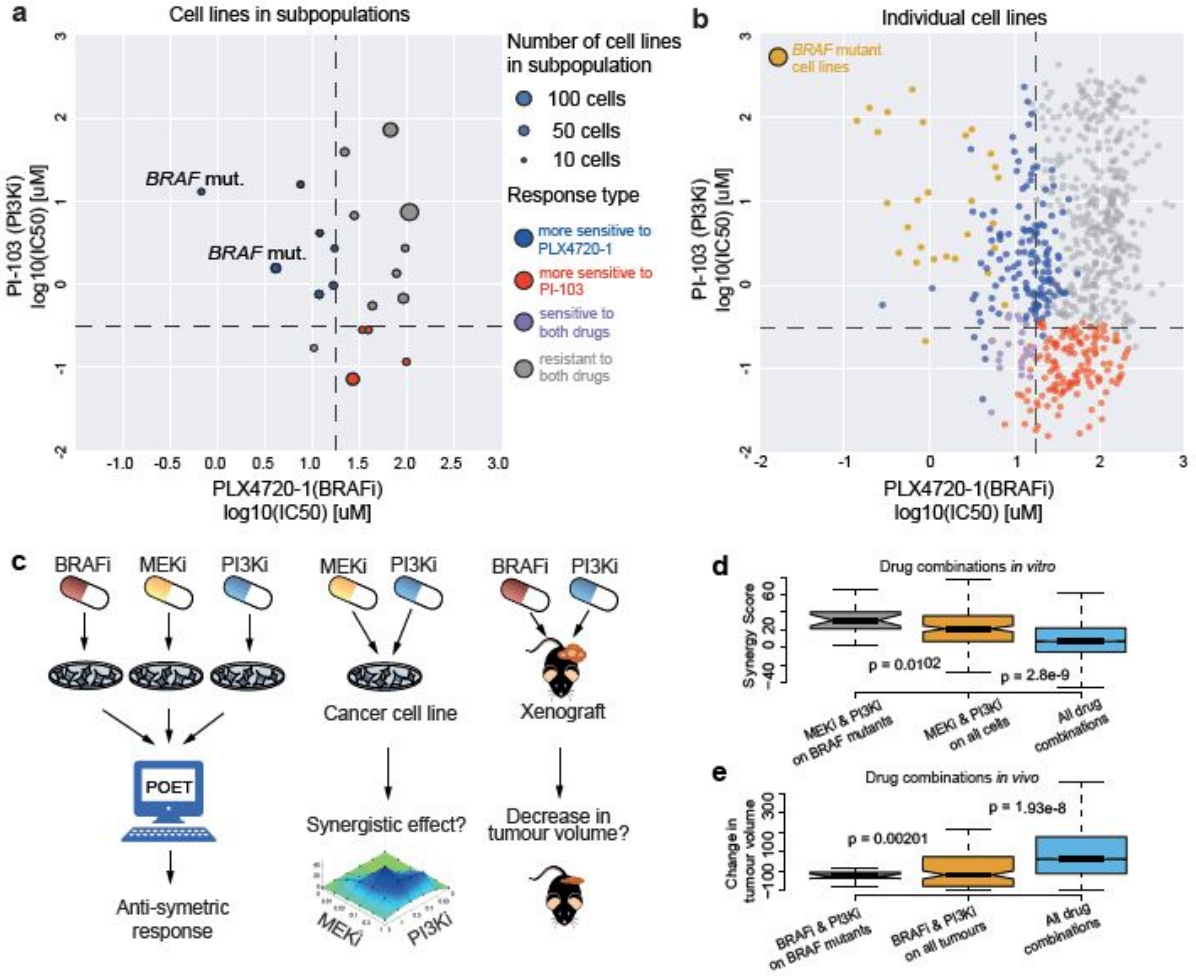

Figure 3: Anti-symmetric response exemplified with PI3K inhibition in comparison to MEK and BRAF inhibitors. a Scatter plot showing POET-derived subpopulations that exhibit anti-symmetric pharmacological response to PLX4720-1 (BRAF inhibitor) and PI-103 (PI3K inhibitor). Dashed lines indicate 20th percentile of log(IC_{50}) values for each drug. *BRAF* mutations were found enriched in the associated subpopulations. **b** Same as panel **a**, but showing individual cell lines and those in gold are cell lines with the *BRAF* mutation. **c** Workflow illustrating cell lines being tested with individual inhibitors and their joint pharmacological patterns of response analysed by POET. Drug pairs with anti-symmetric response suggest suitable drug combination therapies, which are validated in cell lines [17] and patient-derived tumor xenograft (PDX) models [18]. **d** *In vitro* synergistic effect of combining the MEK and PI3K inhibitors over all drug combinations (t-test P=2.8e-9), and combining the MEK and PI3K inhibitors in *BRAF* mutant cell lines over all cell

lines (t-test P=0.0102). **e** *In vivo* effect of combining BRAF and PI3K inhibitors over all drug combinations in preventing tumour growth (t-test P=1.93e-8), and combining BRAF and PI3K inhibitors in *BRAF* mutant tumours over all tumours (P=0.00201, t-test).