

ONCOGENOMICS

InFlo: a novel systems biology framework identifies cAMP-CREB1 axis as a key modulator of platinum resistance in ovarian cancer

N Dimitrova¹, AB Nagaraj², A Razi², S Singh², S Kamalakaran¹, N Banerjee¹, P Joseph², A Mankovich¹, P Mittal^{3,4}, A DiFeo^{2,4} and V Varadan^{2,4}

Characterizing the complex interplay of cellular processes in cancer would enable the discovery of key mechanisms underlying its development and progression. Published approaches to decipher driver mechanisms do not explicitly model tissue-specific changes in pathway networks and the regulatory disruptions related to genomic aberrations in cancers. We therefore developed InFlo, a novel systems biology approach for characterizing complex biological processes using a unique multidimensional framework integrating transcriptomic, genomic and/or epigenomic profiles for any given cancer sample. We show that InFlo robustly characterizes tissue-specific differences in activities of signalling networks on a genome scale using unique probabilistic models of molecular interactions on a per-sample basis. Using large-scale multi-omics cancer datasets, we show that InFlo exhibits higher sensitivity and specificity in detecting pathway networks associated with specific disease states when compared to published pathway network modelling approaches. Furthermore, InFlo's ability to infer the activity of unmeasured signalling network components was also validated using orthogonal gene expression signatures. We then evaluated multi-omics profiles of primary high-grade serous ovarian cancer tumours ($N=357$) to delineate mechanisms underlying resistance to frontline platinum-based chemotherapy. InFlo was the only algorithm to identify hyperactivation of the cAMP-CREB1 axis as a key mechanism associated with resistance to platinum-based therapy, a finding that we subsequently experimentally validated. We confirmed that inhibition of CREB1 phosphorylation potently sensitized resistant cells to platinum therapy and was effective in killing ovarian cancer stem cells that contribute to both platinum-resistance and tumour recurrence. Thus, we propose InFlo to be a scalable and widely applicable and robust integrative network modelling framework for the discovery of evidence-based biomarkers and therapeutic targets.

Oncogene (2017) 36, 2472–2482; doi:10.1038/onc.2016.398; published online 7 November 2016

INTRODUCTION

The etiology and progression of complex diseases typically involves coordinated dysregulation of multiple genes interacting in intricate ways that are not yet fully understood. Diseased cells exhibit significant perturbations at multiple genomic levels ranging from mutations and chromosomal aberrations in DNA to expression changes in mRNA and proteins. Genome-wide molecular profiling technologies have revealed genomic aberrations affecting hundreds of genes, each of which occurs infrequently in diseased tissues. Such heterogeneity in genomic aberrations at the level of individual genes necessitates the development of analytic approaches that can systematically integrate prior biological knowledge with molecular profiling data to delineate underlying biological mechanisms.¹ These approaches fall into three main categories: gene set enrichment, interaction subnetwork construction and signalling network modelling,¹ which utilize well-curated databases of gene sets,² protein–protein interactions and signalling pathway databases,^{3,4} respectively. Gene set enrichment or protein–protein interaction

subnetworks do not account for the complex regulatory interactions that underlie the functioning of signalling networks in contrast to pathway modelling frameworks that explicitly incorporate regulatory interaction information with multi-dimensional molecular profiles to identify functional activation of pathway subnetworks.^{1,5–7} However, published network modelling methods do not robustly infer the activity of unmeasured signalling components such as protein complexes and do not explicitly consider the possibility of systematic deviations in pathway network structures arising due to inadequate curation of tissue-specific regulatory relationships and/or changes in pathway regulatory interactions caused by genomic aberrations. Thus there is a clear need to develop robust approaches that can seamlessly integrate multi-dimensional molecular profiles of individual tumour samples while accounting for potential systematic structural deviations from the curated pathway networks in the specific tissue context.

Here we present a novel multi-omics systems biology framework, InFlo, for estimating signalling network activities through a

¹Department of Clinical Informatics Solutions and Services, Philips Research North America, Briarcliff Manor, NY, USA; ²Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA and ³Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. Correspondence: Dr A DiFeo or Dr V Varadan, Case Comprehensive Cancer Center Case Western Reserve University, 2103 Cornell Rd, WRB 3125, Cleveland, OH 44106, USA.

E-mail: analisa.difeo@case.edu or vinay.varadan@case.edu

⁴These authors contributed equally to this work.

Received 27 May 2016; revised 23 August 2016; accepted 18 September 2016; published online 7 November 2016

unique multidimensional model that explicitly accounts for potential tissue-specific deviations in pathway regulatory structures. We show using extensive cross-validation that InFlo robustly characterizes tissue-specific differences in activities of signalling networks on a genome scale, while also inferring the activities of unmeasured signalling pathway components. We then apply InFlo to a large ovarian cancer multi-omics dataset to identify signalling pathways associated with resistance to platinum-based chemotherapy, followed by experimental confirmation.

RESULTS

InFlo overview

The InFlo framework estimates activities of interactions within signalling pathways in individual patient samples through the integration of multiple molecular measurements while accounting for measurement noise as well as possible errors/disruptions in pathway model structures. The framework can be described in the following main conceptual steps as highlighted in Figure 1. The InFlo methodology contains the following four key computational steps: (1) integration of multi-omics profiles to infer differential gene activities per patient sample, (2) assessing activity of pathway interactions by integrating differential gene activities with pathway network structure, (3) capturing pathway network deregulations in individual tumour samples.

(1) *Integration of multi-omics profiles to infer differential gene activities:* The first step (Figure 1, Step A) in the InFlo methodology is designed specifically to integrate information from multi-omics profiles of individual samples to estimate the

differential activity levels of genes. Here we provide details on integrating somatic copy-number alterations and gene expression data but the model is naturally scalable to include other molecular measurements as well. The effects of differential dosage of gene copy-number and gene expression are combined into a two-element activity level vector denoted by $A(i)$. This activity vector is subsequently used to determine the gene state, which is a discrete value in the set $\Omega = \{-1, 0, 1\}$, respectively, corresponding to inactivated, neutral and activated states. These states are determined based on comparing the values of gene-level copy-number or expression in each tumour sample against a cohort of normal tissues. These are learned using class-conditional distributions such as Gaussian mixture models with three components, each for one of the three states for example $\sum_{j=1}^3 \pi_j N(\mu_j, \Sigma_j)$, where μ_j and Σ_j are the mean and the variance of the class-conditional normal distribution for gene expression and copy-number and π_j is the prior probability of class j , obtained by applying the Maximum Likelihood estimate. This is advantageous because, instead of having one discrete state for each gene in a given tumour sample, we have three probabilities for each state j of the gene i , denoted by $S_j(i)$ using Bayes rule, that is $S_j(i) = P(S(i) = j | A(i)) = \frac{\pi_j N(A(i); \mu_j, \Sigma_j)}{\sum_{l=1}^3 \pi_l N(A(i); \mu_l, \Sigma_l)}$. Therefore, the state of each gene with activity vector $A(i)$ is represented by a three-element probability vector of $S(i) = [S_1(i), S_2(i), S_3(i)]^T$. The maximum *a posteriori* probability estimate then provides the most likely state for the gene, that is $S_{MAP}(i) = \text{argmax}_j S_j(i) = \text{argmax}_j P(S(i) = j | A(i))$. Notably, other molecular factors with known impact on gene/protein expression can also be included in the activity vector $A(i)$, along with their expected

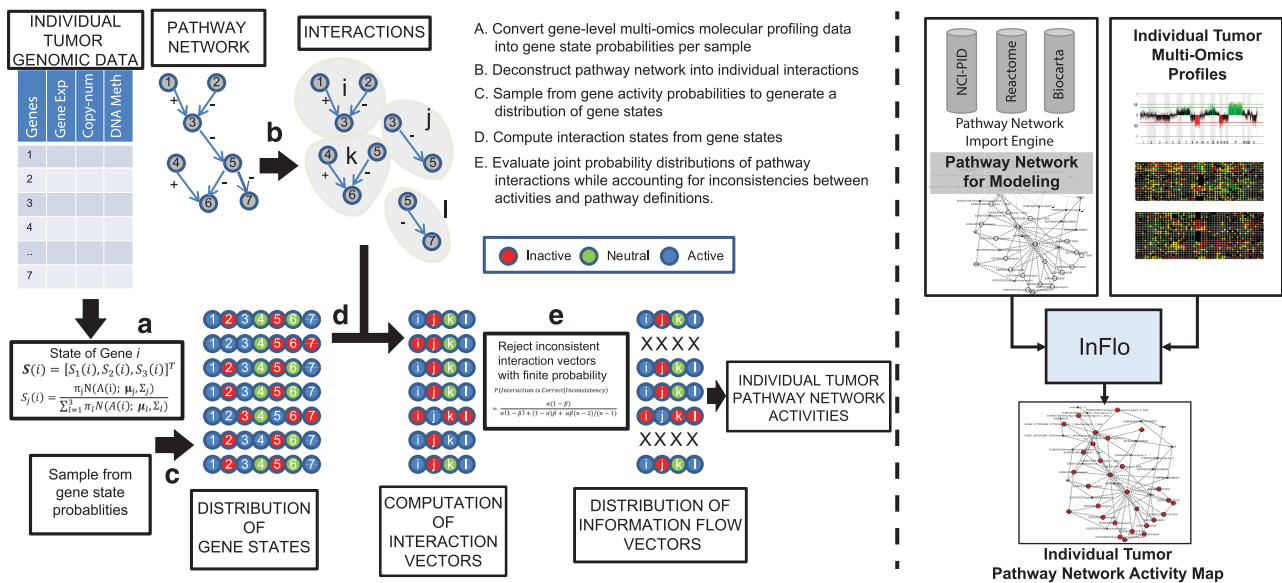


Figure 1. Overview of the InFlo computational framework. The left panel details the steps A–E in the InFlo computational pipeline. (a) InFlo integrates multi-omics data from an individual tumour and estimates the probability of each gene being in one of three states: active, inactive or neutral/unchanged as compared to tissue-matched normals. (b) InFlo parses pathway network databases and deconstructs each network into individual interactions, whose activities are to be inferred for the given tumour sample. (c) InFlo generates distribution of gene states by sampling from the gene activity probabilities. (d) InFlo computes the state of each interaction based on the states of the parent genes of the interaction. For example, interaction i is active when Gene1 (activator) is active and Gene2 (repressor) is inactive. However, if the state of Gene2 (repressor) is active, the interaction i is inactive. Similarly, interaction j is active when Gene3 is inactive, and vice versa. (e) InFlo then evaluates whether the activity status of a gene downstream of a given interaction is consistent with the interaction's activity status. Interaction vectors that are inconsistent with one or more downstream genes are rejected with finite probability. For example, in the second interaction vector, interaction l is active whereas the target of the interaction, Gene7, is inactive. Thus, this interaction vector is inconsistent with the gene state vector and is rejected. The resulting interaction vectors thus provide the joint-distribution of the pathway interaction activities in the individual tumour sample, while also accounting for inconsistencies between gene activities and pathway network definitions. The right panel details the data flow through the InFlo computational suite, integrating pathway network information from public databases and multi-omics tumour profiles to generate sample-specific pathway network activities.

effects on gene expression. For example, given that increased DNA methylation is expected to contribute negatively on gene expression, the measurements from the DNA methylation assay would be scaled by a negative factor before inclusion into A(i). This framework thus allows for scalable integration of multi-omics profiles on a per-sample basis. The resulting gene-level activity probabilities are then integrated with the pathway network model as detailed below.

(2) *Assessing activity of pathway interactions by integrating gene activities with pathway network structure:* Given that pathway network models are intended to capture mechanistic events that enable cells to integrate molecular information resulting in a functional cellular response, we developed InFlo to explicitly model the regulatory structure defined in the pathway network. InFlo defines the basic unit of information within a pathway as the activity of individual interactions among genes as captured in the pathway network annotation. This information is then captured as a vector of interaction activities for all the interactions defined in a particular pathway. Each interaction is defined by a set of parent genes that jointly regulate one or more children genes (Figure 1, Step B). An individual interaction activity is defined as an ensemble output of the activities of the parent genes of the interaction in the particular sample. In the simplest case, the predicted interaction activity is a linear combination of all the votes of the parents. The state vector of an interaction i denoted by li is given by $S(li) = (\sum_{k=1}^{p_i} r_{ik} * S(g_{ik}))$, where $S(g_{ik})$ is the state vector of the k th parent of the i th interaction and r_{ik} is the coefficient capturing the regulatory influence of the gene i_k towards this interaction. Likewise, p_i is the number of parents for interaction li . Considering equal influence from all parents, $|r_{ik}| = \frac{1}{p_i}$. Therefore if a parent gene is known to positively regulate the interaction, then the vote of the parent is $(r_{ik} = \frac{1}{p_i})$, whereas if a parent negatively regulates the interaction, $r_{ik} = -\frac{1}{p_i}$. Finally, the state of an interaction $S(li)$ is obtained using the similar maximum *a posteriori* probability estimation used for genes, that is $S_{MAP}(li) = \text{argmax}_j S_j(li)$.

Thus, InFlo explicitly models pathway deregulations as perturbations in the *information flow* within the signalling network. In other words, by focusing on interaction activity instead of gene activity levels, InFlo uniquely focuses on the information transmitted through the various arms of a signalling network's regulatory topology. The scalability of this modelling strategy is evident by the possibility to further extend this framework to non-equal voting strategies to account for differences in the influence of parent genes on a downstream interaction, when such prior biologic information is available. As an extreme example, this framework allows the incorporation of a snowballing strategy, where down-regulation of even one parent could result in complete disruption of complex-formation leading to abrogation of an interaction's activity.

(3) *Capturing pathway network deregulations in individual tumour samples:* In order to capture the pathway activity in a given patient sample, InFlo estimates the joint-probability distribution of activities of interactions through a generative process that incorporates a sampling framework⁸ that accounts for errors arising from measurements as well as pathway network disruptions arising from genomic aberrations.

For each patient and pathway, the sampling procedure generates a large number of instances of activity states of genes with associated measurements by sampling the background probabilities derived from gene activity model described earlier (Figure 1, Step C). Activity levels of pathway entities that do not have any measurement are derived by propagating the states of the measured entities through the pathway network. For each sampled instance of gene activity levels in a pathway, the

interaction states for the pathway are computed using the ensemble strategy for each interaction as defined earlier (Figure 1, Step D). Furthermore, at this stage a consistency check is performed between the states of an interaction and its children. Assuming reasonable probabilities of pathway model errors and measurement errors, we can estimate the likelihood of rejection of inconsistent pathway interaction vectors from the full set of generated interaction vectors (Figure 1, Step E). An interaction is deemed consistent if the activity states of the majority of children are equal to the predicted interaction activity state. Else, the interaction is deemed inconsistent. For instance, if the interaction k has three children, it is deemed consistent if the inferred state of interaction matches at least the state of two children. In general, if $C(k)$ is the set of the children of interaction k , then it is consistent if we have $\sum_{g \in C(k)} I(S_{MAP}(g) = S_{MAP}(k)) \geq \frac{1}{2}|C(k)|$ where $I(.)$ is the identity function and $|C(k)|$ is simply the cardinality of the set $C(k)$.

We thus compute the probability of correctness of an interaction configuration as a function of the probabilities that the underlying data and pathway models are correct. If we assume that the probability of measurement errors leading to a state change in the corresponding interaction is α and likewise the probability of incorrect prediction due to errors in the pathway model or underlying structural aberrations to be β , we obtain the probability of consistency of an interaction with n states as shown in Supplementary Table S1.

We then obtain the probability of a predicted interaction state being correct in the case of the predicted interaction state being consistent with the state of its children as:

$$P(\text{Interaction is Correct} | \text{Consistency}) = \frac{P(\text{Consistency} | \text{Model is correct})P(\text{Model is correct})}{P(\text{Consistency} | \text{Model is correct})P(\text{Model is correct}) + P(\text{Consistency} | \text{Model is incorrect})P(\text{Model is incorrect})}$$

which, for $n=3$, becomes:

$$P(\text{Interaction is Correct} | \text{Consistency}) = \frac{(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta) + \alpha\beta/(n-1)} = \frac{(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta) + \alpha\beta/2} \quad (1)$$

Similarly, we can compute the probability of correctness of a predicted interaction state given an inconsistency between the states of the children and predicted interaction state as:

$$P(\text{Interaction is Correct} | \text{Inconsistency}) = \frac{P(\text{Inconsistency} | \text{Model is correct})P(\text{Model is correct})}{P(\text{Inconsistency} | \text{Model is correct})P(\text{Model is correct}) + P(\text{Inconsistency} | \text{Model is incorrect})P(\text{Model is incorrect})}$$

which, for $n=3$, becomes:

$$P(\text{Interaction is Correct} | \text{Inconsistency}) = \frac{\alpha(1-\beta)}{\alpha(1-\beta) + (1-\alpha)\beta + \alpha\beta(n-2)/(n-1)} = \frac{\alpha(1-\beta)}{\alpha(1-\beta) + (1-\alpha)\beta + \alpha\beta/2} \quad (2)$$

We observe that for small values of α and β , the probability of correctness of a consistent interaction tends to 1 whereas the probability of correctness of an inconsistent interaction remains low. Given that we expect the pathway annotations to be largely correct and somatic genomic aberrations impact only a relatively small subset of genes in the genome, we assume that β is typically small for most pathway interactions. Thus, we choose to reject inconsistent interactions with a low probability (0.05) while generating the joint-distribution of the information flow vectors, and accepting the consistent interactions with a high probability

(0.99). Using these probabilities, we use a generative model to capture the joint-distribution of states of interactions for a given pathway in a tissue sample.

Notably, published methods that attempt to model pathway activity take the average value of all gene activity levels to represent the pathway activity level in a given patient sample,⁷ thus completely missing the topological structure implicit in the pathway, as outlined in an example pathway with two interactions in Supplementary Figure S1. We therefore represent each patient's pathway activity as a joint-distribution of the activity states of all interactions in that pathway, which we term *information flow* vector since it captures each mechanistic information transfer event as defined in the pathway. The joint-distribution of interaction activities for each patient sample in a particular pathway can then be used to define *intersample distances*, using measures such as the Minkowski distance. These inter-patient distances can then be used in both supervised and unsupervised settings to identify clinically relevant patient subgroups as well as the most informative pathways for distinguishing them.

Robustness evaluation of InFlo

We first evaluated the ability of InFlo to robustly identify pathways discriminative of cancer phenotypes and compared InFlo's performance against two published pathway network modelling approaches highlighted in a recent review of pathway network approaches applied to cancer genomes,¹ PathOlogist⁷ and PARADIGM,⁶ although they significantly differ in their conceptual approaches. Specifically, we chose PathOlogist because it attempts to capture the inconsistency between the measurements and the curated signalling network, while PARADIGM attempts to smooth gene-level measurements to be consistent with the curated network. Given that both PathOlogist⁹ and PARADIGM⁶ were previously used to identify pathways associated with estrogen receptor (ER) status in breast cancer, we leveraged publicly available gene-expression and somatic copy-number profiles of breast cancers ($N=301$) from The Cancer Genome Atlas (TCGA) for comparative robustness analysis. We applied InFlo and the other two algorithms to assess the association of 183 pathways curated from the NCI-PID, Reactome and KEGG pathway databases with ER status of these breast cancers.

Accordingly, we employed a stringent double-loop cross-validation framework¹⁰ (Figure 2a) to assess whether the algorithms identify pathways discriminative of ER status that also validate on previously unseen data. InFlo identified a larger number of putative pathways discriminative of ER status, at higher frequencies of repeated identification in the discovery phase, as compared to either PathOlogist or PARADIGM (Figure 2b). InFlo also identified a higher number of pathways appearing at any given frequency in the discovery loop as compared to either PathOlogist or PARADIGM (Figure 2c), suggesting InFlo being more sensitive but also exhibiting higher reliability than the comparators in identifying pathways discriminative of ER status in breast cancer. The higher average sensitivity and specificity of InFlo as compared to PathOlogist or PARADIGM is evident from the receiver operating characteristics curves for each of the algorithms (Figure 2d), plotted by varying the silhouette threshold to select pathways discriminative of ER status. In addition to exhibiting higher performance metrics, InFlo revealed both previously known and novel pathways to be associated with ER status in breast cancer (Supplementary Tables S2-S4). InFlo's association of SMAD2/3 signalling in over 99% of the discovery and evaluation datasets is consistent with previous reports of cross-talk between TGF-beta signalling and the ER pathway,¹¹ an association that was not detected by PathOlogist (Supplementary Figure S2), thus highlighting the utility of InFlo's approach of

modelling all of the signalling pathway's interaction activities. Similarly, the FOXM1 transcription factor network was associated by InFlo with ER status in 92 and 85% of the discovery and evaluation cross-validation datasets (Supplementary Table S2), and was consistent with published reports of regulation of this pathway by ER-alpha in breast cancers.^{12,13} InFlo's inference of higher activity of the ErbB receptor signalling network in HER2-amplified breast cancers (Supplementary Figure S3) also highlights InFlo's ability to infer the effects of copy-number alterations on downstream signalling networks. Taken together, these findings strongly support InFlo as a robust methodology to discover signalling network deregulations associated with disease phenotypes.

Delineation of signalling pathways associated with progression-free survival of ovarian cancers using InFlo

Given the robustness of InFlo, we next applied InFlo to decipher potential mechanisms mediating platinum resistance in high-grade serous ovarian carcinomas (HGSOC). Ovarian cancer is an inherently difficult cancer to treat and the frontline treatment is the use of platinum-based chemotherapeutic agents. While some HGSOC patients respond briefly to platinum therapy, disease recurrence or progression is common, with 5-year overall-survival hovering around 30%.¹⁴ Therefore, there is a clear clinical need for biomarkers predictive of benefit from platinum therapy as well as new therapeutic targets that could enable the development of alternative interventions in this deadly disease.

We used InFlo to identify dysregulated pathways in individual HGSOC samples ($N=357$) when compared to a pool of normal fallopian tube samples within the TCGA's HGSOC dataset. Progression-free survival and additional clinical data (Supplementary Table S5) was available for a total of 267 samples in the ovarian dataset and was associated with the pathway-based clustering of patient samples. Using the average pathway interaction activity vectors for each given pathway, we estimated interpatient distances and performed hierarchical clustering to identify patient subgroups. Subsequently, we assessed differences in progression-free survival between subgroups controlling for potential confounding effects of clinico-pathological factors.

InFlo identified seven pathways as significantly associated with platinum resistance in the TCGA HGSOC dataset, with just two pathways remaining significant after adjusting for clinico-pathological factors (Table 1). InFlo was the only computational framework to identify regulation of p38-alpha to be associated with progression-free survival, consistent with published reports showing increased protein expression of phosphorylated-p38 MAPK in platinum-resistant ovarian carcinoma cell lines.^{15,16} While a majority of the pathways in Table 1 contained a large number of genes and involved multiple arms, we were particularly intrigued by the *Class IB PI3K non-lipid kinase events* pathway (P -value ≤ 0.001), given that it predicted higher cAMP activity to be associated with platinum resistance.

Figure 3a details the clustering of the patient samples using the InFlo-derived interaction activity levels (information flow vectors) in each of the patient samples, resulting in two major clusters exhibiting differential activity of the PDE3B-cAMP axis. Significant difference in progression-free survival was observed between High cAMP and Low cAMP clusters (P -value ≤ 0.001), with high inferred activity of cAMP being associated with the poorest progression-free survival (Figure 3b), and remained significant (P -value = 0.016) even in the multivariable setting, after adjusting for tumour stage, tumour grade, age at initial diagnosis and residual disease burden at surgery. Figure 3c details the InFlo-derived interaction activities in the pathway network, further highlighting that high cAMP activity is the primary contributor to poor disease-free survival in HGSOC treated with platinum-based chemotherapy.

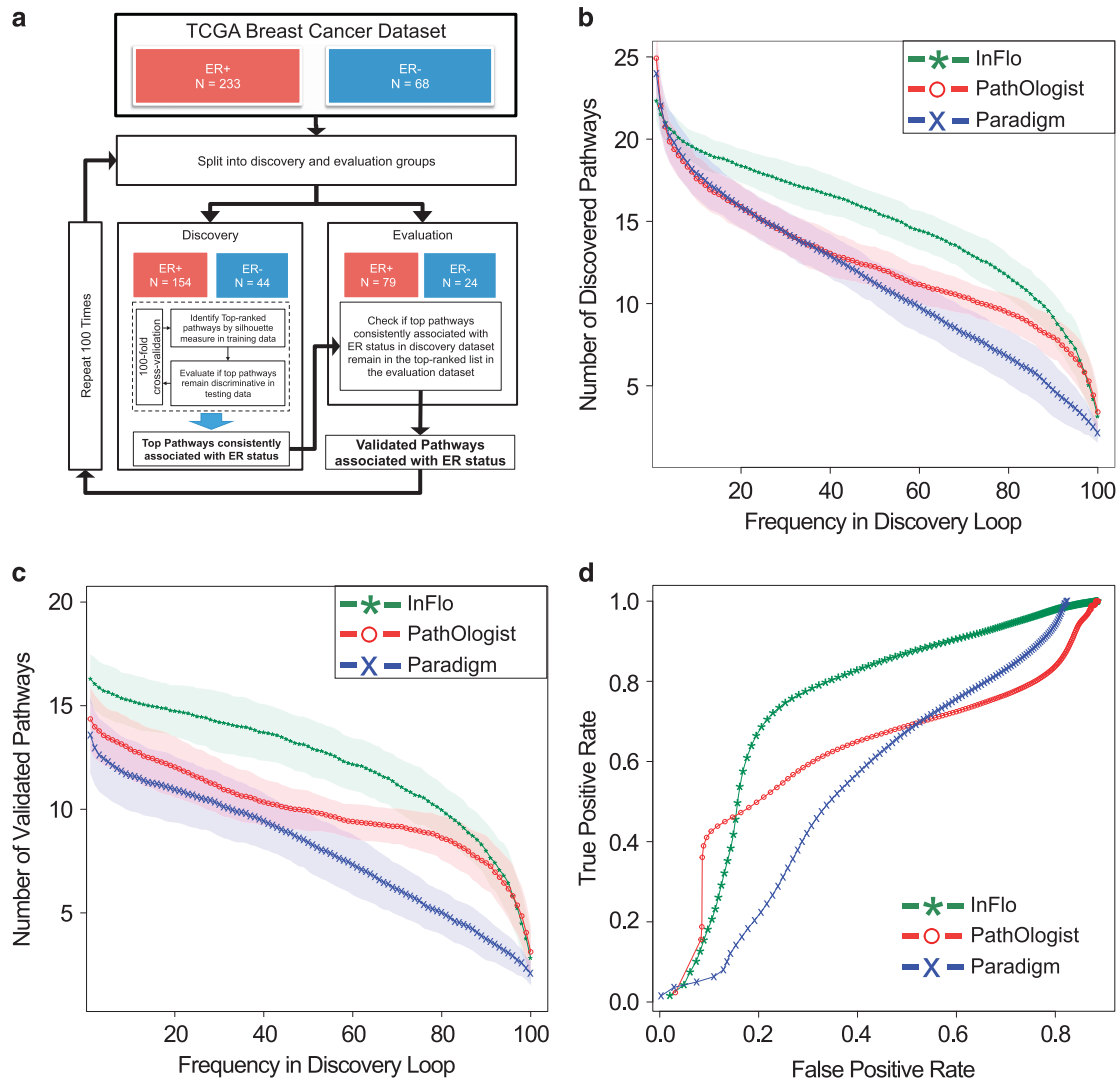


Figure 2. Robustness assessment of InFlo. **(a)** The double-loop cross-validation framework used to assess the robustness of algorithms in identifying pathways associated with ER status in breast cancer. For each iteration of the outer loop, the inner loop further subdivides the discovery group into training and testing datasets over 100 independent permutations, while maintaining the relative frequencies of ER+ versus ER- samples. For each algorithm, per iteration, pathways that are ranked in the top-20 by the Silhouette measure in both the training and testing datasets are denoted as putatively discriminative of ER status. The frequency with which a pathway is discriminative of ER status over the 100 iterations of the inner loop is a measure of its robustness. A putative discriminative pathway reported by the internal loop that is also ranked among the top-20 pathways in the outer loop evaluation dataset is considered to be validated. Otherwise, the putative discriminative pathway is deemed a false positive. An algorithm that identifies more validated pathways is considered superior in terms of sensitivity, with superior specificity associated with lower false positives. **(b)** The mean (solid lines) and standard-deviation (transparent bands) estimates of the number of putative ER-discriminative pathways as a function of how frequently they were observed within the discovery loop. The individual algorithms are denoted in green (InFlo), red (PathOlogist) and blue (PARADIGM). **(c)** The mean and standard deviation estimates of the number of ER-discriminative pathways that validated in the evaluation dataset as a function of how frequently they were observed within the discovery loop. **(d)** The receiver operating characteristics curves for each of the algorithms. The true positive rate is plotted against the false positive rate by varying the silhouette threshold to select pathways discriminative of ER status.

We then sought to evaluate whether the InFlo-inferred cAMP activity levels could be corroborated using downstream transcriptional effects of cAMP activity. The transcription factor CREB1 is the critical transcriptional regulator of cAMP responsive elements. The binding of CREB1 to promoter elements of cAMP-dependent target genes requires activation of CREB1 by cAMP-dependent protein kinase A (PKA) via the phosphorylation of CREB1 at Ser-133¹⁷ (Figure 3d). Accordingly, we ascertained the activity of CREB1 in each of the TCGA ovarian cancer samples by comparing the expression levels of 34 verified transcriptional targets of CREB1¹⁸ against the background expression of all other genes on the

microarray using the single sample gene set enrichment analysis methodology.¹⁹ We found that CREB1 activity was significantly (P -value ≤ 0.03) associated with the cAMP activity as inferred by InFlo (Figure 3e). Furthermore, the higher InFlo-inferred cAMP activity was not trivially explained by other factors that regulate cAMP production, hydrolysis and/or PKA expression (Supplementary Figure S4). Taken together, these findings provide orthogonal validation of InFlo's inference of cAMP activity modulation via the *Class IB PI3K non-lipid kinase events* pathway, thus suggesting high cAMP activity in HGSOc is a likely mechanism of platinum resistance (Figure 3f), which we next proceeded to validate experimentally.

Table 1. Pathways associated with progression-free survival after platinum-therapy in the TCGA high-grade serous ovarian carcinoma dataset

Pathway	Significance of survival difference (P-value)	Significance of pathway's Cox coefficient in multivariable setting (P-value)	Number of entities in pathway	Identified by Pathologist or PARADIGM
Class IB PI3K non-lipid kinase events	0.001	0.016	6	No
Regulation of p38-alpha and p38-beta	0.006	0.019	19	No
PDGFR-beta signalling pathway	0.02	0.084	95	No
Glypican 1 network	0.028	0.059	17	Yes
Insulin-mediated glucose transport	0.033	0.21	15	Yes
Plasma membrane estrogen receptor signalling	0.05	0.084	34	No
Arf6 downstream pathway	0.05	0.09	16	No

The significance of each pathway's ability to stratify patients based on progression-free survival was calculated using the Mantel–Haenszel test. Additionally, the association of each pathway with progression-free survival was also evaluated using a multivariable Cox proportional hazards model by controlling for tumour stage, tumour grade, age at diagnosis and extent of residual disease at surgery.

Inhibition of cAMP activity reverses platinum resistance in ovarian cancer cells

Based on InFlo's analysis, we hypothesized that CREB1 activity is required to maintain drug resistance and survival in platinum-resistant HGSO cells, and thus inhibiting CREB1 activity should potentially result in decreased survival of platinum-resistant cells. Thus, we set out to validate InFlo's novel finding using primary HGSO patient-derived platinum-resistant cells (OV81.2-CP10) as well as platinum-resistant ovarian tumour-initiating cells (TICs) (ALDH^{POS} CP70) isolated from the platinum-resistant cell line, A2780-CP70.²⁰ Given the growing evidence suggesting that platinum-based therapies are very efficient at eradicating differentiated cancer cells but are unable to effectively eliminate TICs, this allowed us to assess whether PKA inhibition would also be able to eradicate this subpopulation of cells.^{21–23}

We began by directly measuring cAMP levels in non-transformed fallopian tube epithelial cells (FTSE), ovarian surface epithelial cells (IOSE), OV81.2-CP10 and ALDH^{POS} CP70 cells. cAMP concentrations were significantly upregulated in both platinum-resistant cells compared to the non-transformed cell lines (Figure 4a). Next, in order to determine whether inhibition of CREB1 would induce cell death in the platinum-resistant cells we utilized the H-89 inhibitor, which has been shown to inhibit the phosphorylation of CREB1 at Ser-133 by PKA.²⁴ We first confirmed that H-89 inhibits activity of CREB1 in both platinum-resistant OV81.2-CP10 and ALDH^{POS} CP70 cells. H-89 treatment antagonized the increase in phospho-Ser 133-CREB1 induced upon cAMP-PKA axis activation by the cAMP agonist forskolin, thus confirming that H-89 inhibits CREB1 activity (Figures 4b and c). In turn, H-89 robustly decreased the survival of the platinum-resistant OV81.2-CP10 and ALDH^{POS} CP70 cells (Figures 4d and e). In addition, cell cycle analysis revealed that H-89 induced G2-M cell cycle arrest (Figure 4f). Given that ovarian tumour cells are reported to grow as spheroids under non-adherent culture conditions and these tumour spheres constitute various aspects of ovarian cancer pathology including stem-like properties, metastasis, drug resistance and tumour recurrence, we also assessed the ability of H-89 to eradicate these ovarian TICs under stem-like culture conditions. H-89 significantly reduced tumour sphere formation in both OV81.2-CP10 and ALDH^{POS} CP70 cells (Figure 4g).

Next, we found that the combination of H-89 and cisplatin resulted in significantly greater cell death as compared to treatment with single agents alone (Figure 5). 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) analysis of H-89 and cisplatin combination treatment in OV81.2-CP10 and ALDH^{POS} CP70 cells showed a significant decrease in the IC50 of cisplatin upon combining with H-89 (Figure 5a). Furthermore, Annexin-V analysis revealed that the increase in apoptosis induced by the

combination was significantly higher compared to H-89 or cisplatin alone in both OV81.2-CP10 and ALDH^{POS}CP70 TICs, which correlated with decreased p-CREB1 protein levels (in both OV81.2-CP10 and ALDH^{POS}CP70 TICs) and increased cleaved caspase-3 protein level in ALDH^{POS}CP70 TICs (Figure 5b), suggesting that inhibiting CREB1 activity could potentially synergize with cisplatin therapy by eradicating ovarian adherent and TICs. Overall, our results show that H-89 decreases survival of platinum-resistant ovarian tumour cells in both adherent and non-adherent stem-like conditions, thus experimentally validating InFlo's assessment that high activity of the cAMP-CREB1 axis is associated with low progression-free survival in HGSO.

DISCUSSION

We have developed a novel systems biology framework, InFlo, which infers deregulation of pathway subnetworks in individual biological samples by integrating genomic profiling data with detailed regulatory information derived from pathway network annotations. InFlo generates probabilistic models of activities of signalling network interactions on a per-sample basis. We showed that InFlo exhibited higher sensitivity and specificity in detecting pathways associated with disease phenotypes as compared to published pathway network modelling approaches. We then applied InFlo to identify pathways associated with progression-free survival in HGSO, and showed that tumours with high cAMP activity have low progression-free survival. We confirmed InFlo's inference of cAMP activity by assessing the expression levels of downstream transcriptional targets of CREB1, a well-known key mediator of cAMP activity. Finally, we experimentally validated this novel finding using platinum-resistant cell line models of HGSO, although further studies confirming our findings using *in vivo* platinum-resistant HGSO PDX models are necessary to validate the efficacy of this combination therapy.

A key insight in the development of the InFlo framework is the evaluation of the consistency between InFlo's predicted interaction activity and the activity of the gene regulated by the interaction (Figure 1, Step E). InFlo utilizes this consistency check to exclude interaction states that are inconsistent with the pathway definition with a small probability, thus essentially denoising the information flow vectors that define the pathway activity in a given tumour sample. We evaluated the contribution of this key step of InFlo by excluding this key step while deriving the information flow interaction vectors for the pathways and then evaluating their association with progression-free survival in the TCGA HGSO dataset. We find that the pathways that were previously found to be significantly associated with progression-

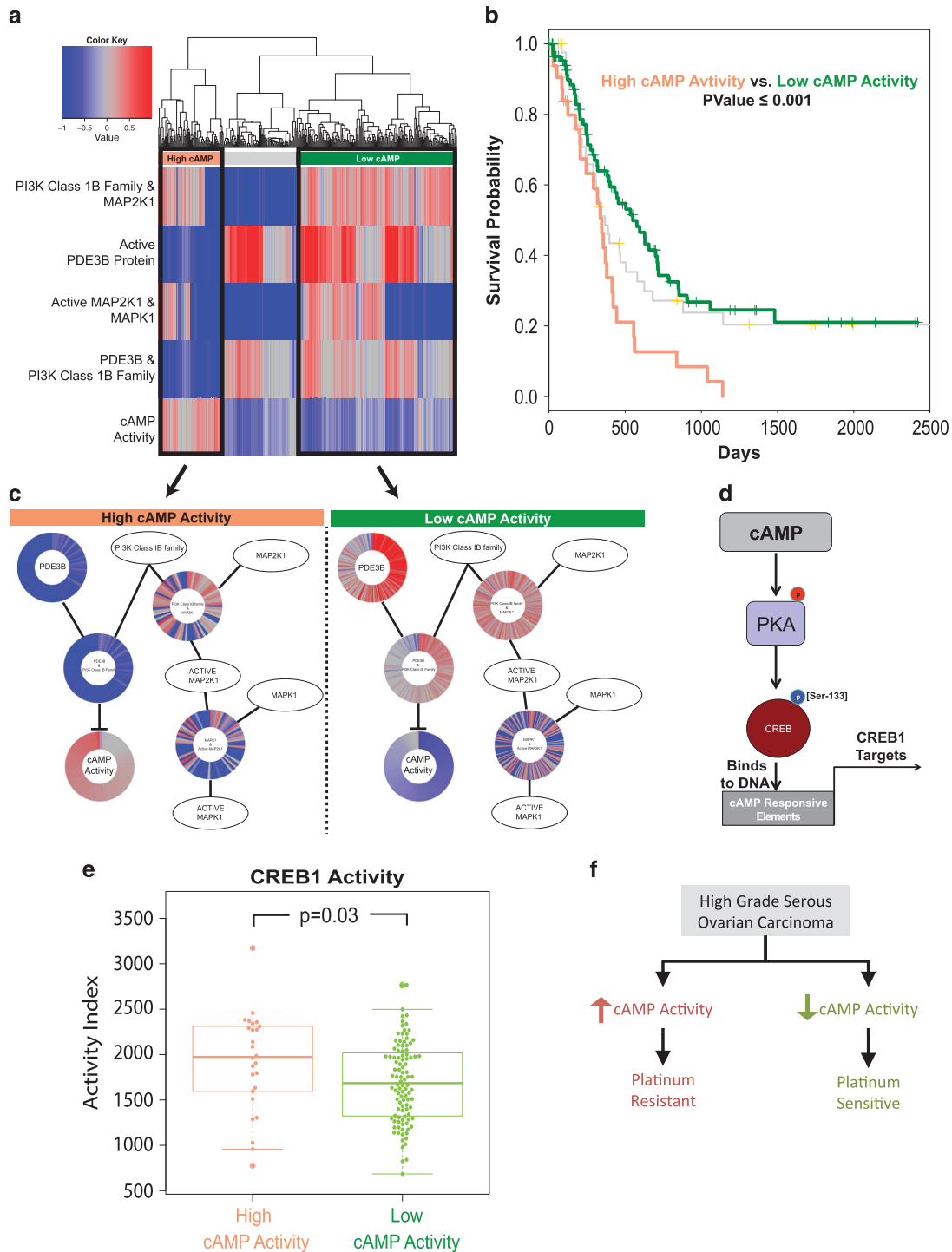


Figure 3. InFlo-based association of the cAMP-PKA-CREB1 axis with progression-free survival in HGSOc. **(a)** Hierarchical clustering of TCGA HGSOc samples ($N=357$) using the InFlo-estimated activities of pathway interactions by integrating gene expression and copy-number profiles of genes in the pathway. Rows correspond to InFlo-estimated activity of interactions and columns denote individual tumour samples. The major clusters identified are indicated by pink (High cAMP), grey and green (Low cAMP). **(b)** Platinum-based progression-free survival of HGSOc patients in each of the clusters. Difference in progression-free survival between High cAMP and Low cAMP clusters (Mantel-Haenszel test P -value ≤ 0.001), remained statistically significant (significance of multivariable Cox regression coefficient, P -value = 0.016) even after controlling for tumour stage, tumour grade, age at diagnosis and extent of residual disease at surgery. **(c)** Pathway interaction activities across ovarian tumours, organized according to the pathway network structure, are plotted per sample in each of the two major clusters. Each ray in the individual ray graphs denotes the InFlo-estimated activity of the pathway interaction in an individual ovarian tumour sample. **(d)** The cAMP-PKA-CREB1 signalling axis. **(e)** CREB1 activity distribution estimated using 34 known transcriptional targets of CREB1 plotted across InFlo-inferred cAMP activity groups. No discernable difference in PI3K Class 1B kinase expression was observed between the Low and high cAMP activity groups (Supplementary Figure S5). Low and high cAMP activity groups correspond to samples with InFlo activity scores ≤ -0.25 and ≥ 0.25 , respectively. **(f)** InFlo-derived model of high cAMP activity as a likely mechanism of platinum resistance in HGSOc.

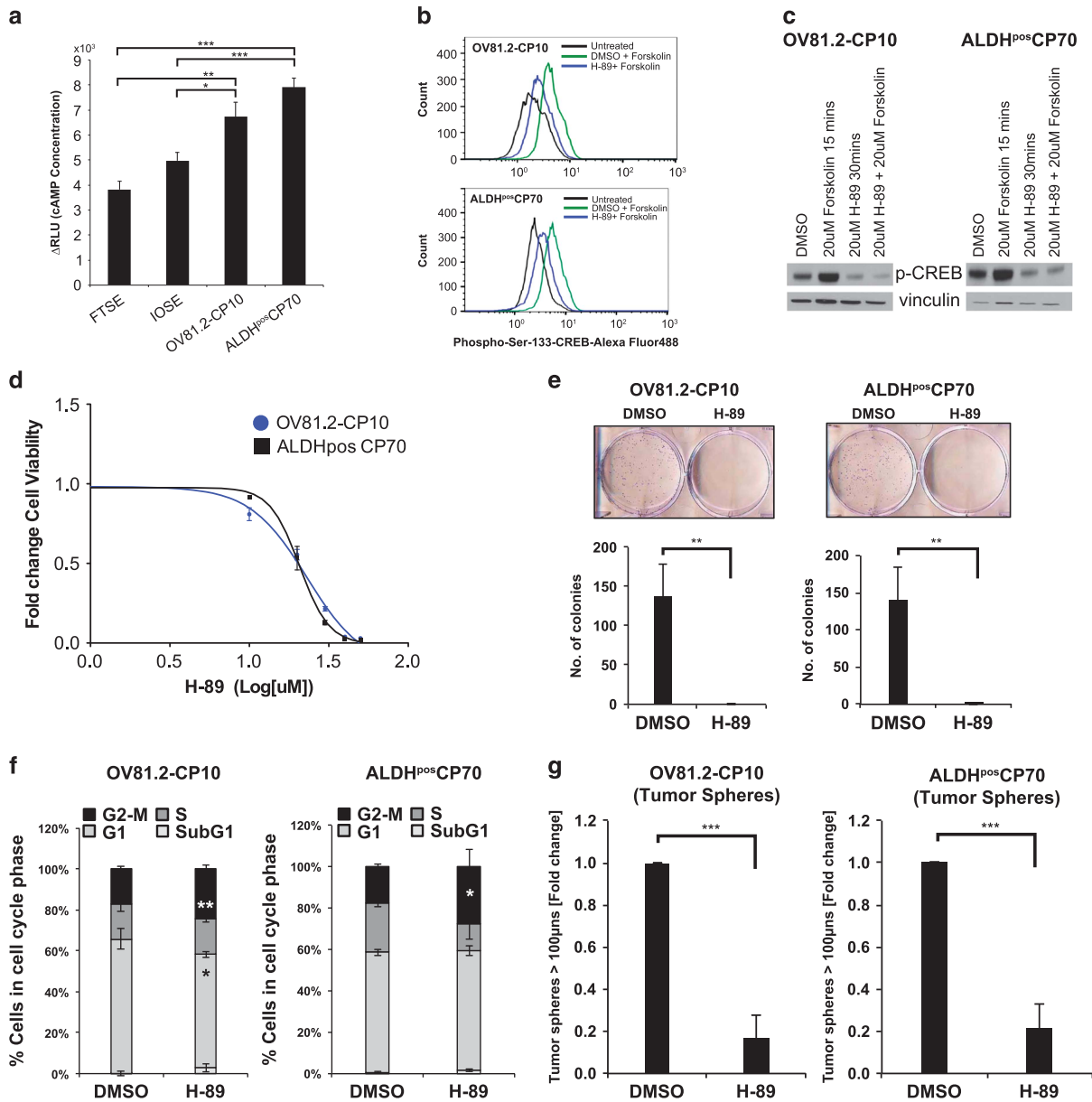


Figure 4. Functional assessment of the role of cAMP axis in HGSOc platinum-resistant models. **(a)** cAMP assay showing increased cAMP concentrations in OV81.2-CP10 and ALDH^{pos} CP70 cells as compared to non-transformed FTSE and IOSE cells in response to 30 min stimulation by 20 μM forskolin. Values are plotted as differences in relative luciferase units (ΔRLU), which is indicative of cAMP concentration **(b)** Flow cytometry analysis and **(c)** western blotting showing H-89 (20 μM-30 min) decreases phosphorylation of CREB1 at ser-133 residue either uninduced or induced in response to forskolin (20 μM-15 min) in both OV81.2-CP10 and ALDH^{pos} CP70 cells. **(d)** 48-h MTT assay showing H-89 decreases viability in both OV81.2-CP10 and ALDH^{pos} CP70 cells. **(e)** Clonogenic assay on day 7 showing decreased survival in response to H-89 treatment in these cells. The individual wells treated with H-89 and DMSO control are shown for both OV81.2-CP10 and ALDH^{pos}CP70 cells after 7 days of treatment (top row). The bar graphs represent the mean number of colonies in the H-89 treatment versus DMSO control for both cell types (bottom row). **(f)** 48-h cell cycle analysis by propidium iodide staining shows that H-89 induces G2-M transition arrest in OV81.2-CP10 and ALDH^{pos} CP70 cells. Bar graphs represent the mean percentage of cells in cell cycle phases. **(g)** 10× light microscopy and 10 × 10 integrated metamorph software analysis showing decreased tumour sphere formation by day 6 upon H-89 treatment in OV81.2-CP10 and ALDH^{pos} CP70 cells. Bar graphs represent the mean fold-change in number of tumour spheres upon treatment with H-89 as compared to control. (All mean values were estimated over three independent replicates, with the error bars representing standard errors of the mean, with the Student's t-test significance denoted as **P* < 0.05, ***P* < 0.005, ****P* < 0.0005).

free survival are no longer significant, while biologically unlikely pathways are now found to be significant (Supplementary Table S6). We also evaluated the stability of estimation of the joint probability distribution of interaction activities for each patient sample in a particular pathway (Figure 1, Step E) by varying the number of information flow vectors generated per pathway and

patient sample in the TCGA HGSOc dataset from 100 to 1000 and found no significant changes in pathways associated with progression-free survival (Supplementary Table S7).

Taken together, these findings underscore InFlo's ability to extract biologically meaningful information by integrating and denoising multi-omics genomic profiling data using regulatory

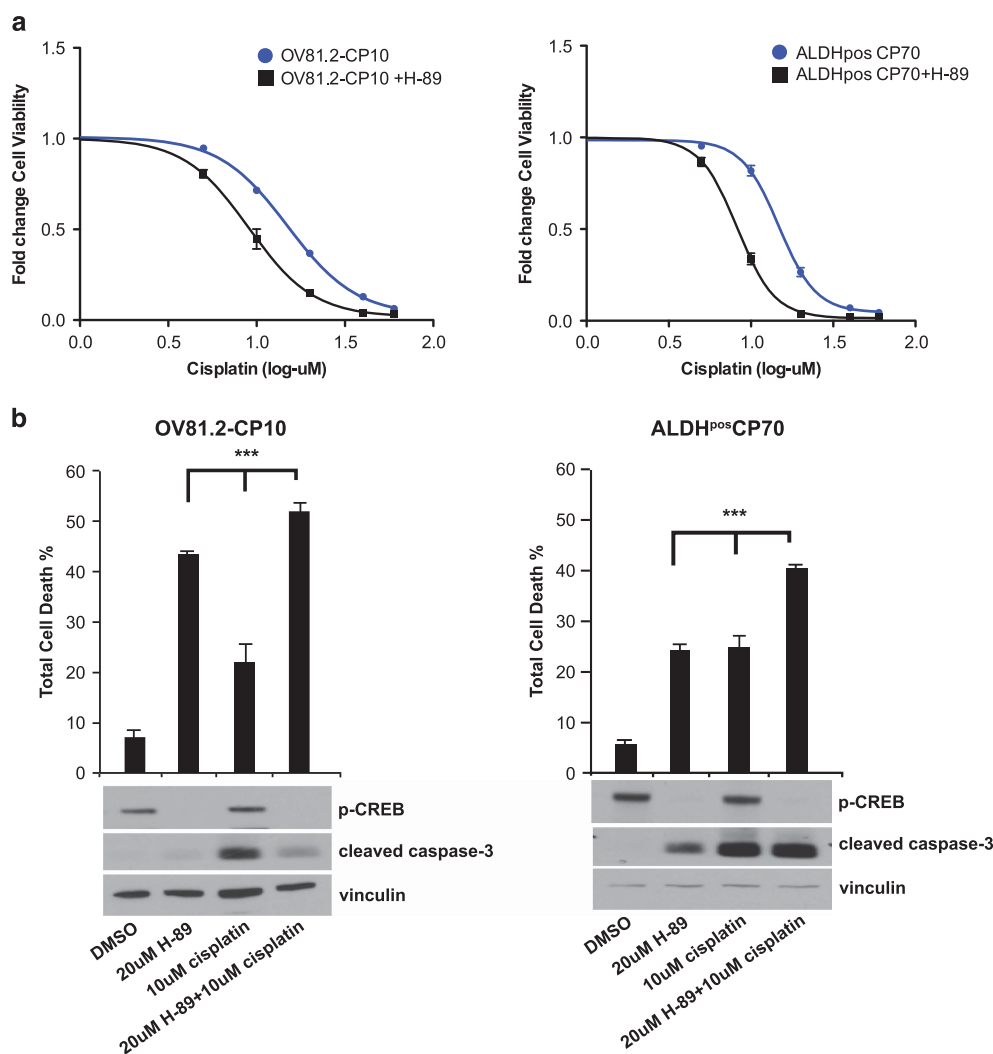


Figure 5. Effects of H-89 and cisplatin combination treatment in HGSOc platinum-resistant cells. **(a)** 48-h MTT analysis of H-89 and cisplatin combination treatment in OV81.2-CP10 (left) and ALDH^{pos} CP70 (right) cells showing decrease in cisplatin IC₅₀ in these cells upon combining with H-89. Plotted are the mean fold-changes in cell viability across varying concentrations of cisplatin alone and in combination with H-89. **(b)** Annexin-PI staining (48 h) and western blotting (24 h) showing that H-89 and cisplatin combo induces significantly higher cell death in both OV81.2-CP10 and ALDH^{pos} CP70 (Annexin-PI) and decreases p-CREB1 protein levels in both OV81.2-CP10 and ALDH^{pos} CP70 (western blotting) and increases cleaved caspase-3 protein levels in ALDH^{pos} CP70 (western blotting) as compared to individual drugs alone. (All mean values were estimated over three independent replicates, with the error bars representing standard errors of the mean, with the Student's *t*-test significance denoted as ****P* < 0.0005).

information obtained from pathway network annotations. We propose InFlo as a robust systems biology approach for integrative analysis of multi-omics data to characterize complex biological signalling network activities in any given biological sample. InFlo has been implemented as a modular and scalable computational platform to integrate multi-omics profiles (Supplementary Figure S6) in a computationally efficient manner, thus delineating genome scale pathway network activities. Additionally, the inconsistency check incorporated within InFlo can also be used to potentially estimate the functional impact of somatic mutations on downstream targets of pathways. We expect InFlo to be widely applicable to reliably delineate key molecular determinants of disease progression, thus enabling the discovery of evidence-based biomarkers and therapeutic targets, as well as for facilitating selection of tailored therapies in individual patients.

MATERIALS AND METHODS

The Cancer Genome Atlas datasets

Level 3 RMA-normalized gene expression data and somatic copy-number profiles^{25,26} were obtained for breast (*N* = 301) and ovarian (*N* = 357) cancer samples from TCGA portal (<https://tcga-data.nci.nih.gov/tcga/>). All TCGA data were used in accordance with TCGA policies.

Identification of pathways discriminative of ER status in the TCGA breast cancer dataset

The breast cancer dataset was processed using InFlo, Pathologist and PARADIGM according to their individual requirements as follows. The Pathologist⁷ and PARADIGM⁶ frameworks were run using default parameters according to the instructions provided by the developers. Of note, Pathologist only accepts gene

expression data as input and generated two scores (activation and consistency) per pathway, resulting in a two-element vector for each pathway, which was used to compute intersample distances, using the Euclidean distance measure. PARADIGM accepts both gene expression and copy-number data on a per-gene level and generates integrated pathway activity levels for each component of a given biological pathway network in a given breast cancer sample. Accordingly, intersample distances was computed as the Euclidean distance between the vectors representing the integrated pathway activity levels of all of the components of the given pathway. InFlo was run according to the details provided in the description of the algorithm. All three computational frameworks were run on the same set of 183 curated pathways downloaded from the NCI-PID, Reactome and KEGG pathway databases. For each algorithm, a pathway's ability to discriminate ER status in breast cancer was estimated by using the Silhouette measure²⁷ based on the algorithm-derived pairwise distances within ER+ or ER- samples (intra-cluster) and between ER+ and ER- samples (inter-cluster).

Cell culture and reagents

Ovarian cancer cell lines were maintained in culture as previously described²⁰ and tested for mycoplasma contamination. Platinum-resistant HGSOc PDX derived OV81.2-CP10 cells were generated and maintained as previously described.²⁰ Cisplatin was purchased from Mount Sinai Hospital Pharmacy (New York, NY, USA). 10 mM stock solutions of H-89 (Tocris Biosciences, Minneapolis, MN, USA) were prepared in DMSO (Fisher Scientific, Pittsburgh, PA, USA) and stored at -20 °C.

Flow cytometry analysis

For phospho-Ser-133 flow cytometry analysis, phospho-Ser-133-Alexa flour 488 conjugate (Cell Signaling Technology Inc., Danvers, MA, USA) and forskolin (Tocris Biosciences) was used and the data were acquired by Coulter Epics XL machine (Beckman Coulter Inc., Brea, CA, USA). ALDH^{POS} CP70 cells were sorted from CP70 cell line as previously described.²⁰ Cell viability (MTT) assays, clonogenic assays, cell cycle assessment, Annexin V assays and tumour sphere formation assays were performed as previously described.²⁰

Western blotting

p-CREB1(1:250), cleaved caspase-3(1:250), cleaved PARP(1:250) and vinculin (1:500) were purchased from (Cell Signaling Technology Inc.) and immunoblotting was done as previously described.²⁰

cAMP assay

cAMP concentrations in response to forskolin stimulation were measured by cAMP-Glo assay kit (Promega Corporation, Madison, WI, USA) and the values were plotted as difference in relative luciferase units (RLU) between untreated and treated samples, which is indicative of cAMP concentration in the cells.

Availability of data and materials

InFlo is implemented in C++ along with a collection of shell scripts to enable easy application of the algorithm on new datasets, and is available for academic use. InFlo is available for download from GitHub at <http://varadanlab.github.io/InFlo/>. Additionally, we have used InFlo to integrate pan-cancer multi-omics datasets from the TCGA and have provided the results as a resource for the community at <http://varadanlab.org/InFlo>.

CONFLICT OF INTEREST

Vinay Varadan reports receiving commercial research funding from Philips Healthcare, and is a consultant for Curis, Inc. The other authors have no potential conflicts of interest to declare.

ACKNOWLEDGEMENTS

The results published here are in part based on data generated by TCGA project established by the NCI and NHGRI. The data were used in accordance to the TCGA's informed consent and data access policies (<http://cancergenome.nih.gov/abouttcga/policies>). Grant Support: This research was supported, in part, by PHS awards: Career Development Program of Case GI SPORE (P50 CA150964) award to Vinay Varadan; Career Development Program in Computational Genomic Epidemiology of Cancer (R25T CA094186) award to Vinay Varadan In addition, this research was supported, in part, by research funding from Philips Healthcare to Vinay Varadan: the Rosalie and Morton Cohen Family Memorial Genomics Fund of University Hospitals to Vinay Varadan: Ohio Cancer Research award to Vinay Varadan: VelaSano Bike for Cure Funds award to Analisa DiFeo and Vinay Varadan: and the Norma I and Al G Geller Endowment in Ovarian Cancer Research (Analisa DiFeo).

AUTHOR CONTRIBUTIONS

VV, PM, ND conceived and developed the InFlo methodology. PM implemented and optimized the InFlo workflow: ABN, PJ performed all of the HGSOc cell line functional validation experiments under the supervision of AD who also provided the previously characterized HGSOc cell lines: AR performed the InFlo robustness analysis: SK, NB, AM performed assessment of TCGA ovarian cancer platinum resistance: SS developed the InFlo software package: VV wrote the manuscript with contributions from all other authors. VV, AD, PM designed the study and supervised all analysis.

REFERENCES

- 1 Mutation C, Pathway Analysis working group of the International Cancer Genome C, Pathway and network analysis of cancer genomes. *Nat Methods* 2015; **12**: 615–621.
- 2 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011; **27**: 1739–1740.
- 3 Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 2004; **32**: D277–280.
- 4 Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T *et al*. PID: the Pathway Interaction Database. *Nucleic Acids Res* 2009; **37**: D674–679.
- 5 Varadan V, Mittal P, Vaske CJ, Benz SC. The integration of biological pathway knowledge in cancer genomics: a review of existing computational approaches. *IEEE Signal Process Mag* 2012; **29**: 35–50.
- 6 Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J *et al*. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* 2010; **26**: i237–245.
- 7 Greenblum SI, Efroni S, Schaefer CF, Buetow KH. The Pathologist: an automated tool for pathway-centric analysis. *BMC Bioinformatics* 2011; **12**: 133.
- 8 Liu JS. *Monte Carlo Strategies in Scientific Computing*. Springer: New York, 2001.
- 9 Efroni S, Schaefer CF, Buetow KH. Identification of key processes underlying cancer phenotypes using biologic pathway analysis. *PLoS One* 2007; **2**: e425.
- 10 Wessels LF, Reinders MJ, Hart AA, Veenman CJ, Dai H, He YD *et al*. A protocol for building and evaluating predictors of disease state based on microarray data. *Bioinformatics* 2005; **21**: 3755–3762.
- 11 Figueroa JD, Flanders KC, Garcia-Closas M, Anderson WF, Yang XR, Matsuno RK *et al*. Expression of TGF-beta signaling factors in invasive breast cancers: relationships with age at diagnosis and tumor characteristics. *Breast Cancer Res Treat* 2010; **121**: 727–735.
- 12 Millour J, Constantinidou D, Stavropoulou AV, Wilson MS, Myatt SS, Kwok JM *et al*. FOXM1 is a transcriptional target of ERalpha and has a critical role in breast cancer endocrine sensitivity and resistance. *Oncogene* 2010; **29**: 2983–2995.
- 13 Sanders DA, Ross-Innes CS, Beraldi D, Carroll JS, Balasubramanian S. Genome-wide mapping of FOXM1 binding reveals co-binding with estrogen receptor alpha in breast cancer cells. *Genome Biol* 2013; **14**: R6.
- 14 Atlas TCG. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011; **474**: 609–615.
- 15 Benedetti V, Perego P, Luca Beretta G, Corna E, Tinelli S, Righetti SC *et al*. Modulation of survival pathways in ovarian carcinoma cell lines resistant to platinum compounds. *Mol Cancer Ther* 2008; **7**: 679–687.

- 16 Xie Y, Peng Z, Shi M, Ji M, Guo H, Shi H. Metformin combined with p38 MAPK inhibitor improves cisplatin sensitivity in cisplatinresistant ovarian cancer. *Mol Med Rep* 2014; **10**: 2346–2350.
- 17 Michael LF, Asahara H, Shulman AI, Kraus WL, Montminy M. The phosphorylation status of a cyclic AMP-responsive activator is modulated via a chromatin-dependent mechanism. *Mol Cell Biol* 2000; **20**: 1596–1603.
- 18 Conkright MD, Guzman E, Flechner L, Su AI, Hogenesch JB, Montminy M. Genome-wide analysis of CREB target genes reveals a core promoter requirement for cAMP responsiveness. *Mol Cell* 2003; **11**: 1101–1108.
- 19 Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF *et al*. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 2009; **462**: 108–112.
- 20 Nagaraj AB, Joseph P, Kovalenko O, Singh S, Armstrong A, Redline R *et al*. Critical role of Wnt/beta-catenin signaling in driving epithelial ovarian cancer platinum resistance. *Oncotarget* 2015; **6**: 23720–34.
- 21 Steg AD, Bevis KS, Katre AA, Ziebarth A, Dobbin ZC, Alvarez RD *et al*. Stem cell pathways contribute to clinical chemoresistance in ovarian cancer. *Clin Cancer Res: An Official Journal of the American Association for Cancer Research* 2012; **18**: 869–881.
- 22 Latifi A, Abubaker K, Castrechini N, Ward AC, Liongue C, Dobill F *et al*. Cisplatin treatment of primary and metastatic epithelial ovarian carcinomas generates residual cells with mesenchymal stem cell-like profile. *J Cell Biochem* 2011; **112**: 2850–2864.
- 23 Kuroda T, Hirohashi Y, Torigoe T, Yasuda K, Takahashi A, Asanuma H *et al*. ALDH1-high ovarian cancer stem-like cells can be isolated from serous and clear cell adenocarcinoma cells, and ALDH1 high expression is associated with poor prognosis. *PLoS One* 2013; **8**: e65158.
- 24 Yang S, Alkayed NJ, Hurn PD, Kirsch JR. Cyclic adenosine monophosphate response element-binding protein phosphorylation and neuroprotection by 4-phenyl-1-(4-phenylbutyl) piperidine (PPBP). *Anesth Analg* 2009; **108**: 964–970.
- 25 Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011; **12**: R41.
- 26 Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B *et al*. Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013; **45**: 1134–1140.
- 27 Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1986; **20**: 53–65.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2017

Supplementary Information accompanies this paper on the Oncogene website (<http://www.nature.com/onc>)