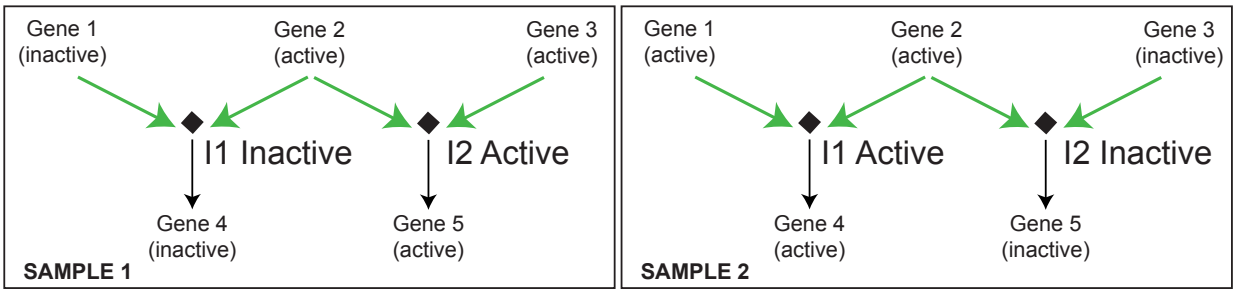
Supplementary Figures and Tables

InFlo: A Novel Systems Biology Framework Identifies cAMP-CREB1 Axis as a Key Modulator of Platinum Resistance in Ovarian Cancer

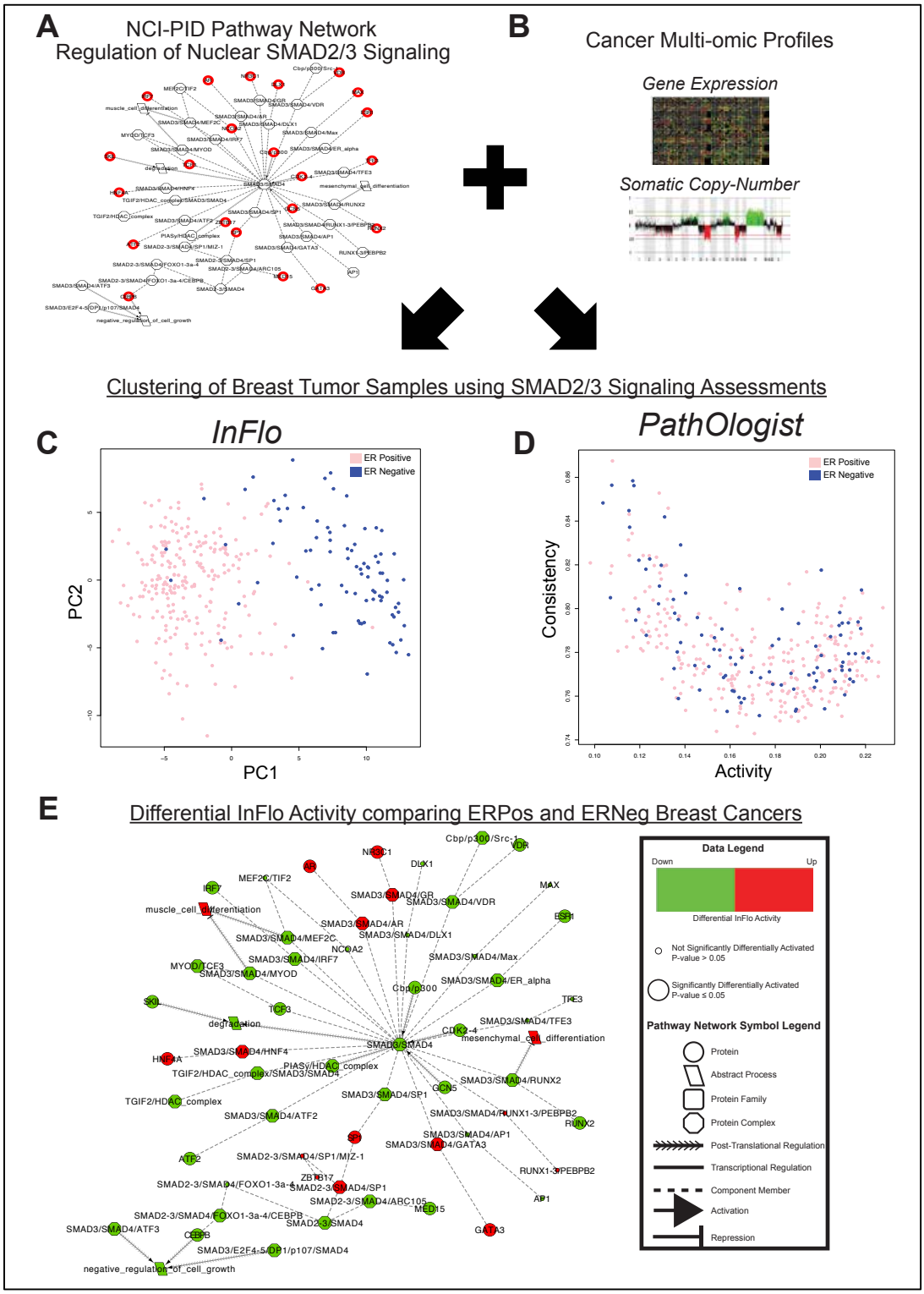
Nevenka Dimitrova, Anil Belur Nagaraj, Abolfazl Razi, Salendra Singh, Sitharthan Kamalakaran, Nilanjana Banerjee, Peronne Joseph, Alexander Mankovich, Prateek Mittal, Analisa DiFeo, Vinay Varadan

Supplementary Figures

Supplementary Tables

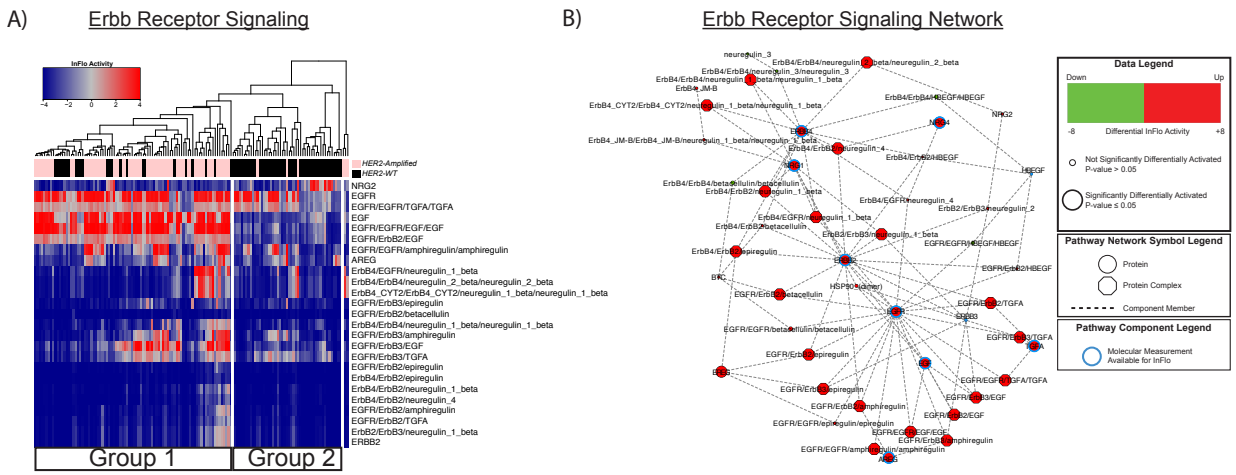


Supplementary Figure S1. **Theoretical example detailing the value of InFlo's approach to capturing pathway activity.** Shown is an example of a pathway with two interactions I1 and I2, involving a total of 5 genes. Interaction I1 captures the activity of Gene 1, Gene 2 and Gene 4. When Gene 1 and Gene 2 are both active, interaction I1 is active and Gene 4 is activated. Similarly, interaction I2 captures the activation of Gene 5 when both Gene 2 and Gene 3 are activated. Consider two separate tumor samples, Sample1 and Sample2. In Sample1, Gene1 is inactive whereas Gene2 and Gene3 are active. In contrast, in Sample2, Gene1 and Gene2 are active while Gene3 is inactive. Considering the wiring of the pathway network, in Sample1, Gene4 is inactive while Gene5 is active. Whereas, in Sample2, Gene4 is active and Gene5 is inactive. If the pathway activity were defined as the average of the gene activities or the average of the interaction activities, the pathway activity values for Sample1 and Sample2 would be identical. However, in reality, this pathway has been differently regulated, with different interactions in the pathway being activated in each of the two samples. Since InFlo uses all of the interaction activities in the pathway, it would correctly identify that this pathway is differentially regulated in the two samples. This is further clarified using a biologic example as detailed next in Supp. Fig. S2.

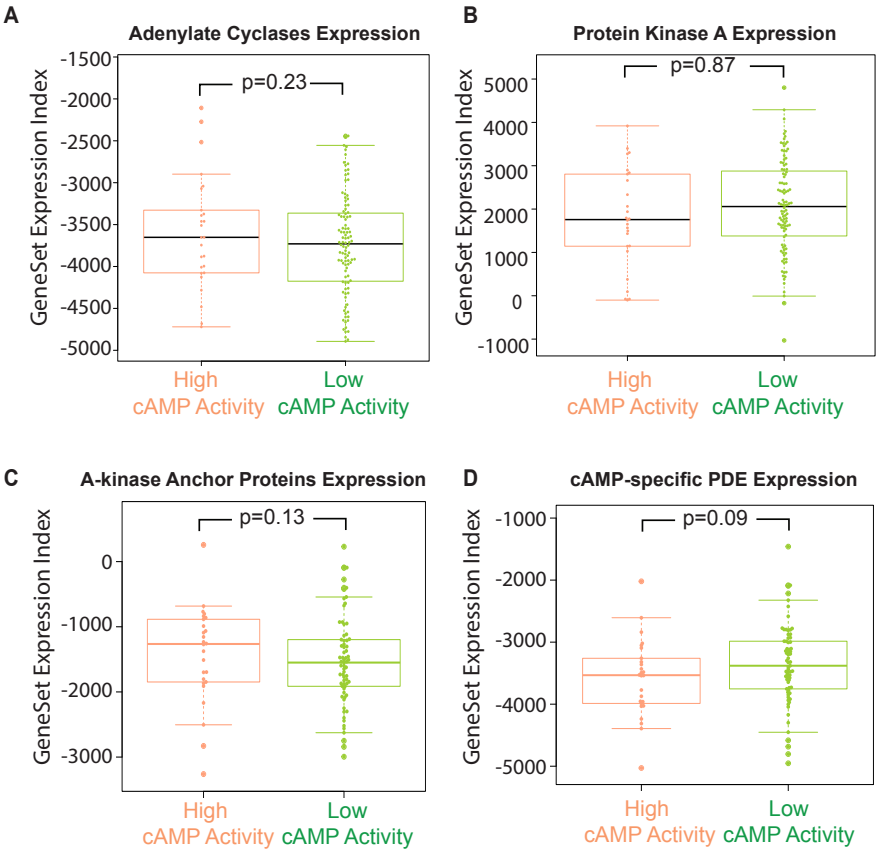


Supplementary Figure S2. Biologically relevant example contrasting InFlo's and PathOlogist's approaches to modeling signaling pathway activities. In order to contrast the differences in InFlo and PathOlogist in modeling pathway activities, we use as an example, the NCI-PID pathway network, *Regulation of Nuclear Smad2/3*

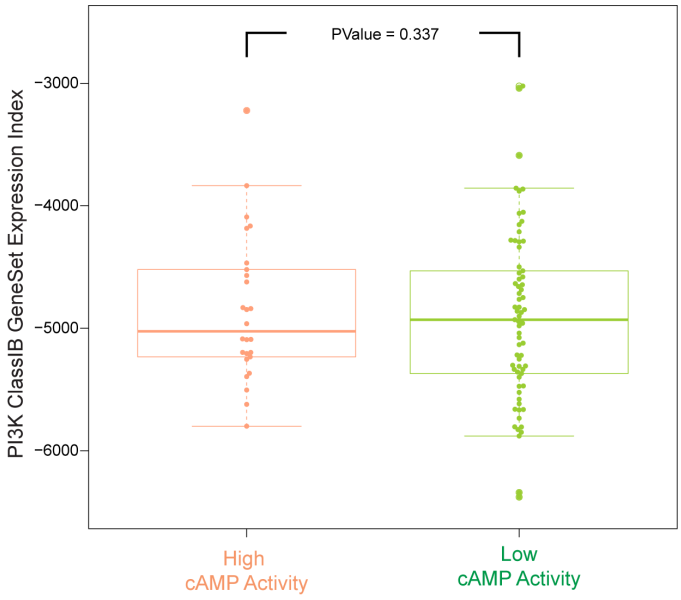
Signaling. A) The network structure of the pathway titled "*Regulation of Nuclear Smad2/3 Signaling*" as parsed by InFlo from the NCI-PID database. The nodes circled in red are genes for whom multi-omic measurements (B) are available to the InFlo engine, corresponding to about 35% of all nodes in the pathway network. The InFlo framework propagates the activities of the measured nodes to infer the activity of all nodes in the pathway network across each of the breast tumor samples. C) We applied InFlo and PathOlogist to model the activity of the Smad2/3 signaling pathway in a total of 301 breast tumor samples (233 ER-positive and 68 ER-negative). For each tumor sample, InFlo derived the activities of all the interactions in the pathway. A scatter plot of the samples using the top-two principal components (PC1 and PC2) of the InFlo interaction activities reveals a clear separation of the ER-positive and ER-negative tumors. D) Scatter plot of the two values (Activity and Consistency) that PathOlogist generates to chracterize the state of the Smad2/3 signaling pathway in each tumor sample. Note that PathOlogist is unable to distinguish the differences in Smad2/3 signaling activities between ER-positive and ER-negative breast cancers. E) Plot detailing the differential activity of the Smad2/3 signaling network components between the ER-positive and ER-negative breast cancers as inferred by InFlo. Of note, each node is colored according to the differential activity of the interaction immediately upstream, with higher activities in ER-positive breast cancers denoted as red, while lower activities denoted as green. Importantly, InFlo's assessment of down-regulation of Smad2/3 signaling in ER-positive breast cancers is consistent with the known cross-talk between these two pathways.



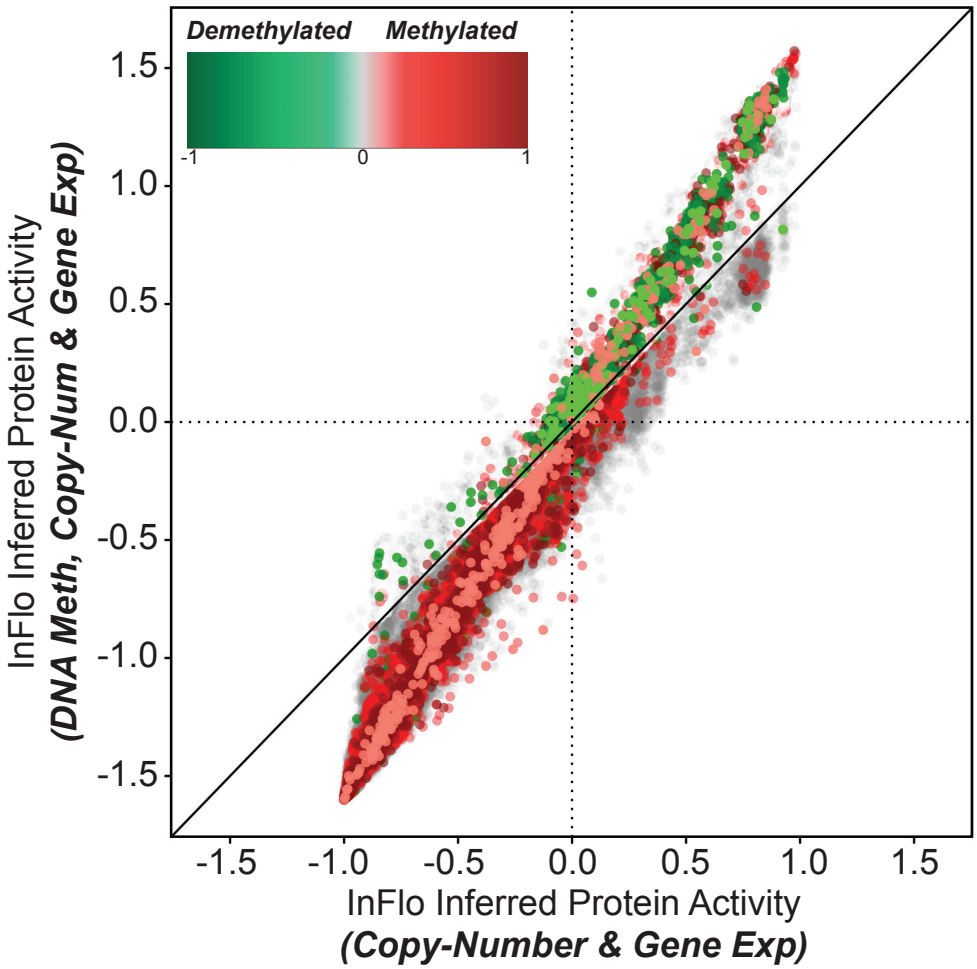
Supplementary Figure S3. **Example detailing InFlo's ability to infer the effects of copy-number alterations on downstream signaling networks.** Here we highlight InFlo's ability to capture the impact of copy-number alterations on downstream signaling networks by evaluating whether InFlo captures the effect of amplifications in the *Erbb2* locus on the activity of components in the *Erbb Receptor Signaling* pathway (NCI-PID). Using InFlo, we integrated gene expression and copy-number data on a total of 140 breast cancers from the TCGA dataset, of which 70 harbored amplifications in the *Erbb2* locus (SNP-array segmental logratio >= 0.5), the remainder of the samples being wild-type for amplifications in this locus. **A)** Unsupervised clustering of the 140 tumor samples using InFlo-derived activities of the *Erbb Receptor Signaling* pathway revealed two major clusters, Group 1 and Group 2 exhibiting significant segregation of HER2-Amplified tumors (Fisher's Exact Test Pvalue << 0.0001). **B)** Plot detailing significantly higher InFlo-derived activities of the *Erbb Receptor Signaling* pathway components in *HER2-Amplified* breast cancers as compared to *HER2-WT* breast cancers. Of note, InFlo had access to molecular measurements (gene expression and copy-number) for only 7 genes/nodes in this network, using which InFlo inferred the activities of all interactions/components.



Supplementary Figure S4. **Association of cAMP-related geneset expression indices with InFlo-inferred cAMP activity.** Shown are the GeneSet Expression indices of **A)** adenylate cyclases; **B)** protein kinase A; **C)** A-kinase Anchor Proteins; and **D)** cAMP-specific phosphodiesterases, in primary HGSOC samples, categorized by InFlo-derived cAMP activity levels. The expression index of adenylate cyclases that catalyze the conversion of ATP to cAMP was calculated using the expression levels of the gene set defined as *ADCY1*, *ADCY2*, *ADCY3*, *ADCY6*, *ADCY7*, *ADCY8*, *ADCY9* and *ADCY10*. The expression index of protein kinase A was defined using the gene set *PRKACA* and *PRKACB*, which are the two cAMP-dependent protein kinase catalytic subunits critical in mediating the phosphorylation of CREB1 in the presence of cAMP. The expression index of A-kinase anchor proteins that bind PKA was calculated using the expression levels of the 13-gene set including *AKAP1* through *AKAP13*. Finally, the expression index of cAMP-specific phosphodiesterases included *PDE4A*, *PDE4B*, *PDE4D*, *PDE7A*, *PDE7B*, *PDE8A* and *PDE8B*, which are known to be cAMP-selective hydrolases. Low and high cAMP activity groups correspond to samples with InFlo activity scores ≤ -0.25 and ≥ 0.25, respectively and P-values were based on two-sided Student's t test. Notably, the InFlo-inferred cAMP-activity and associated CREB1 target activity (Figure 3) were not trivially explained by either Adenylase Cyclase, or protein kinase A, or A-kinase anchor protein expression. However, the observed lower expression index of the cAMP-specific PDEs in the InFlo-derived High cAMP Activity group was in accordance with expectation but did not achieve statistical significance.



Supplementary Figure S5. **Association of PI3K Class IB kinase expression with InFlo-inferred cAMP activity.** Shown are the expression indices of the PI3K Class IB kinases gene set, including *PIK3R5* and *PIK3CG*, in primary HGSOC samples, categorized by InFlo-derived cAMP activity levels. Notably, no significant difference is observable between the two groups, further supporting *PDE3B* as the likely reason of cAMP activity modulation in HGSOC.



Supplementary Figure S6. **Example detailing scalability of the InFlo framework to integrate multi-omics profiles.** Here we highlight InFlo's scalability by detailing how InFlo captures the impact of promoter DNA Methylation on protein activity in addition to incorporating copy-number and gene expression profiles. We obtained normalized DNA Methylation profiles from the TCGA portal derived using the Illumina Infinium HumanMethylation27 BeadChip assay. We then employed InFlo in two separate instances, wherein we either integrated a) gene expression and copy-number profiles or b) DNA Methylation, copy-number and gene expression profiles. In each instance, InFlo inferred the activity of proteins in the pathway networks across all ovarian tumors (*N* = 291). Shown is the scatter plot comparing the InFlo-inferred protein activities across all samples in each of the two InFlo instances. Each point in the scatter plot denotes a particular protein's activity in a given tumor sample, with the color of each point denoting the extent of promoter methylation (red) or demethylation (green) as compared to the normal fallopian tube samples. Notably, InFlo's estimates of protein activity upon integration of DNA Methylation data were in accordance with expectation, where promoter demethylation contributed to increased protein activity while promoter methylation resulted in lower protein activity as compared to InFlo's estimates based on copy-number and gene expression alone.

Supplementary Table S1. **Calculating the probability of interaction consistency**. The term 1/(n-1) accounts for the probability that the state change from the correct state to any of n-1 wrong states which incurred by model error is compensated by data error; while 1-1/(n-1) is the probability of the ultimate state being incorrect.

	Data is ok: 1-α	Data Error: α
Model is ok: 1-β	P(consistency)=(1- α)(1- β) P(Inconsistency) =0	P(consistency)=0 P(Inconsistency) =(α)(1- β)
Model Error: β	P(consistency)=0 P(Inconsistency) =(1- α)β	P(consistency)= αβ.1/(n-1) P(Inconsistency) = αβ.(n-2)/(n-1)

Supplementary Table S2. Pathways identified by InFlo as being discriminative of ER status in breast cancer using the double-loop cross-validation framework. Listed are the pathways that were identified by InFlo to be discriminative of estrogen receptor status in the TCGA breast cancer dataset. Highlighted in grey are pathways that were associated with ER status in over 80% of the discovery and evaluation loops. Pathways in red were identified by InFlo, PARADIGM and PathOlogist.

Pathway	Fraction of Iterations in which Pathway Ranked in Top-20	
	Discovery Loop	Evaluation Loop
Regulation Of Nuclear SMAD2/3 Signaling	0.9994	0.9994
IL2 Signaling Events Mediated By STAT5	0.9861	0.9861
Syndecan-1-Mediated Signaling Events	0.9636	0.9436
Coregulation Of Androgen Receptor Activity	0.9506	0.9207
PLK2 And PLK4 Events	0.9469	0.9269
FOXM1 Transcription Factor Network	0.9225	0.8526
PLK3 Signaling Events	0.916	0.886
Atypical NF-Kappab Pathway	0.9108	0.8908
Ras Signaling In The CD4+ TCR Pathway	0.8927	0.8527
Direct P53 Effectors	0.8316	0.7533
FOXA1 Transcription Factor Network	0.7959	0.6697
IL1-Mediated Signaling Events	0.7062	0.5089
Regulation Of RAC1 Activity	0.7052	0.6
Reelin Signaling Pathway	0.5845	0.3011
Glypican 3 Network	0.5651	0.3058
Syndecan-3-Mediated Signaling Events	0.5603	0.3418
IL6-Mediated Signaling Events	0.4728	0.2876
IL12 Signaling Mediated By STAT4	0.3572	0.1176
ATF-2 Transcription Factor Network	0.2607	0.0368
Erbb Receptor Signaling Network	0.1584	0.0095

Supplementary Table S3. Pathways identified by PARADIGM as being discriminative of ER status in breast cancer using the double-loop cross-validation framework. Listed are the pathways that were identified by PARADIGM to be discriminative of estrogen receptor status in the TCGA breast cancer dataset. Highlighted in grey are pathways that were associated with ER status in over 80% of the discovery and evaluation loops. Pathways in red were identified by InFlo, PARADIGM and PathOlogist.

Pathway	Fraction of Iterations in which Pathway Ranked in Top-20	
	Discovery Loop	Evaluation Loop
P75(Ntr)-Mediated Signaling	0.9988	0.9988
Lpa Receptor Mediated Events	0.9085	0.8985
Alternative Nf-KappaB Pathway	0.8878	0.8679
FOXA1 Transcription Factor Network	0.8816	0.7397
Bmp Receptor Signaling	0.7753	0.6339
Nongenotropic Androgen Signaling	0.7197	0.6245
Downstream Signaling In Naiive Cd8+ T Cells	0.708	0.5205
Cd40/Cd40l Signaling	0.6345	0.4966
Aurora A Signaling	0.5792	0.4184
Epo Signaling Pathway	0.5463	0.362
Coregulation Of Androgen Receptor Activity	0.5076	0.3208
Direct P53 Effectors	0.5041	0.2607
Arf1 Pathway	0.4891	0.2667
Canonical Nf-Kappab Pathway	0.3999	0.1878
Glucocorticoid Receptor Regulatory Network	0.2759	0.0936
Amb2 Integrin Signaling	0.2071	0.0065
Foxa Transcription Factor Networks	0.1853	0.0089
Androgen-Mediated Signaling	0.1561	0.0229
Fas Signaling Pathway (Cd95)	0.1305	0.0113
Cdc42 Signaling Events	0.1189	0.0006

Supplementary Table S4. Pathways identified by PathOlogist as being discriminative of ER status in breast cancer using the double-loop cross-validation framework. Listed are the pathways that were identified by PathOlogist to be discriminative of estrogen receptor status in the TCGA breast cancer dataset. Highlighted in grey are pathways that were associated with ER status in over 80% of the discovery and evaluation loops. Pathways in red were identified by InFlo, PARADIGM and PathOlogist.

Pathway	Fraction of Iterations in which Pathway Ranked in Top-20	
	Discovery Loop	Evaluation Loop
Foxa2 And Foxa3 Transcription Factor Networks	0.9994	0.9994
Coregulation Of Androgen Receptor Activity	0.9861	0.9861
Regulation Of Androgen Receptor Activity	0.9506	0.9207
Sphingosine 1-Phosphate (S1p) Pathway	0.9469	0.9269
Hedgehog Signaling Events Mediated By Gli Proteins	0.9225	0.8526
Ephrin A Reverse Signaling	0.916	0.886
Rapid Glucocorticoid Signaling	0.9108	0.8908
FOXA1 Transcription Factor Network	0.9016	0.8738
Glypican 1 Network	0.8927	0.8527
Regulation Of Rhoa Activity	0.8316	0.7533
P38 Mapk Signaling Pathway	0.7959	0.6697
Aurora A Signaling	0.7062	0.5089
Signaling Events Mediated By Ptp1b	0.7052	0.6
Retinoic Acid Receptors-Mediated Signaling	0.5845	0.3011
Signaling Mediated By P38-Gamma And P38-Delta	0.5651	0.3058
Foxo Family Signaling	0.5603	0.3418
Lkb1 Signaling Events	0.4728	0.2876
Ephrin B Reverse Signaling	0.3572	0.1176
Paxillin-Dependent Events Mediated By A4b1	0.2607	0.0368
Signaling Events Mediated By Hdac Class Iii	0.1584	0.0095

Supplementary Table S5. **Clinical characteristics of the TCGA serous ovarian cancer cohort.** Listed are the summaries of clinical characteristics of patients whose tumor profiles were included in the InFlo analysis to identify pathways associated with platinum resistance in ovarian cancer. Briefly, we only included patients who received at least three cycles of adjuvant platinum-based therapy and who were coded as complete responders in the primary therapy outcome success category. Progression-free interval was calculated as the number of days from the time adjuvant treatment ended to the time of disease progression or recurrence. Patients who did not suffer recurrence or progression until the time to last follow-up were subsequently censored for progression-free survival analysis, resulting in progression-free survival information for a total of 267 patients in the TCGA HGSOC dataset. Clinical data including tumor stage, tumor grade, age at initial diagnosis and residual disease burden at the time of surgery was also obtained for use as potential confounders in the survival analyses.

Clinical Characteristics (*N* = 267)	Distribution (Percentage of Cohort)
Tumor Stage	Stage I (2.6%) Stage II (5.9%) Stage III (76.7%) Stage IV (14.6%)
Tumor Grade	Grade 1 (1.12%) Grade 2 (10.9%) Grade 3 (85%) Grade 4 (<1%)
Median age at initial diagnosis	57 (Range 26-87)
Residual Disease Burden at surgery	Largest nodule >10 mm (17.9%) Largest nodule <10 mm (73.8%)
Median number of cycles of platinum-based adjuvant chemotherapy	6 (Range 3-26)

Supplementary Table S6. **Evaluation of the impact of InFlo's consistency check on identification of pathways associated with progression-free survival in ovarian cancer.** Listed are the pathways that were identified by InFlo to be significantly associated with progression-free survival in the TCGA HGSOC dataset. Hierarchical clustering (Pearson correlation distance and average linkage) was performed using the per-sample average information flow vectors derived by InFlo. The samples were then divided into three clusters, followed by evaluation of differences in progression-free survival evaluated between the clusters using the LogRank test. Reported are the LogRank Pvalues for the most significant difference between the clusters per pathway. InFlo was run in two different modes: with and without the rejection of information flow vectors inconsistent with the pathway definition.

Pathway	Difference in Progression-Free Survival (LogRank Pvalue)	
	InFlo with Consistency Evaluation	InFlo without Consistency Evaluation
Class IB PI3K non-lipid kinase events	0.001	0.074
Regulation of p38-alpha and p38-beta	0.006	0.151
PDGFR-beta signaling pathway	0.02	0.076
Glypican 1 network	0.028	0.140
Insulin-mediated glucose transport	0.033	0.243
Plasma membrane estrogen receptor signaling	0.05	0.074
Arf6 downstream pathway	0.05	0.071
Visual signal transduction: Cones	0.131	0.002
LKB1 signaling events	0.271	0.03

Supplementary Table S7. **Evaluation of the impact of the number of information flow vectors on InFlo's ability to robustly infer pathway activities.** Listed are the pathways that were identified by InFlo to be significantly associated with progression-free survival in the TCGA HGSOC dataset. Hierarchical clustering (Pearson correlation distance and average linkage) was performed using the per-sample average information flow vectors derived by InFlo. The samples were then divided into three clusters, followed by evaluation of differences in progression-free survival evaluated between the clusters using the LogRank test. Reported are the LogRank Pvalues for the most significant difference between the clusters per pathway. Separate InFlo runs involving a) generating 100 information flow vectors per sample and b) generating 1000 information flow vectors per sample, resulted in almost no differences in significant pathways.

Pathway	Difference in Progression-Free Survival (LogRank Pvalue)	
	N = 100	N = 1000
Class IB PI3K non-lipid kinase events	0.001	0.004
Regulation of p38-alpha and p38-beta	0.006	0.012
PDGFR-beta signaling pathway	0.02	0.018
Glypican 1 network	0.028	0.011
Insulin-mediated glucose transport	0.033	0.007
Plasma membrane estrogen receptor signaling	0.05	0.039
Arf6 downstream pathway	0.05	0.058