Briefings in Bioinformatics, 2026, 27, bbag242
https://doi.org/10.1093/bib/bbag242
Published: 24 May 2026
Problem Solving Protocol

OXFORD

A scalable, multi-resolution consensus clustering approach for prioritizing robust signals from high-throughput screens

Shaine Chenxin Bao[1,‡], Kathleen I. Pishas[2,‡], Karla J Cowley[3], Qiongyi Zhao[1], Emily C. A. Goodall[1], Ian R. Henderson[1], Xin Liu[2], Evanny Marinovic[2], Mark S. Carey[4,5], Ian G. Campbell[2,6], Kaylene J. Simpson[3,6], Dane Cheasley[2,6,§], Dalia Mizikovsky[1,§], Nathan J. Palpant[1,§,*]

[1]Institute for Molecular Bioscience, The University of Queensland, 306 Carmody Road, St Lucia, Brisbane, QLD, 4072, Australia
[2]Peter MacCallum Cancer Centre, 305 Grattan Street, Melbourne, VIC, 3000, Australia
[3]Victorian Centre for Functional Genomics, Peter MacCallum Cancer Centre, 305 Grattan Street, Melbourne, VIC, 3000, Australia
[4]Division of Gynecology Oncology, Department of Obstetrics and Gynecology, University of British Columbia, 2775 Laurel St, Vancouver, BC, V5Z 1M9, Canada
[5]Department of Clinical Research, BC Cancer, 600 West 10th Avenue, Vancouver, BC, V5Z 4E6, Canada
[6]The Sir Peter MacCallum Department of Oncology, The University of Melbourne, 305 Grattan Street, Parkville, VIC, 3000, Australia

*Corresponding author. Institute for Molecular Bioscience, The University of Queensland St. Lucia, QLD 4072, Australia. E-mail: n.palpant@imb.uq.edu.au
‡Shaine Chenxin Bao and Kathleen I. Pishas are co-first authors.
§Dane Cheasley, Dalia Mizikovsky and Nathan J. Palpant are co-senior authors.

Abstract

Modern biology increasingly relies on large-scale screening to generate high dimensional datasets with potential to accelerate discovery. However, analysing these complex datasets remains challenging, due to hierarchical biological structure, uncertainty in the true number of groups, and high dimensional noise that confounds genuine biological signal. Here we present an unsupervised consensus clustering tool, Untangled, that aggregates clustering solutions across granularities to construct a stability-based representation, followed by cluster number optimization and systematic evaluation of cluster robustness. Through extensive benchmarking on simulated datasets, ground truth datasets and diverse high-dimensional screening applications, we demonstrate that Untangled reliably recovers underlying relationships, resolves stable meaningful substructure, and more effectively prioritizes robust clusters with shared biological mechanisms and conserved phenotypic responses than alternative clustering approaches. These results establish Untangled as a scalable framework for cluster discovery to guide efficient follow-up investigation from high-dimensional biological datasets.

Keywords **high dimensional data analysis, consensus clustering, unsupervised learning, high-throughput screening, drug discovery, computational biology**

Introduction

Advancements in high-throughput technologies have generated unprecedented volumes of biological data [1, 2]. Traditional comparative analyses that require groups to be labelled *a priori* are inadequate for such complex datasets. Unsupervised clustering methods are therefore ideal to identify latent structure in high-dimensional feature spaces without prior assumptions. Clustering samples based on their molecular or phenotypic features enables data-driven approaches to discover meaningful relationships or shared mechanisms. Despite this, clustering high-dimensional data remains challenging, particularly in applications where the number of true biological groupings is unknown, and the signal is confounded by stochastic noise [3]. In addition, different algorithms, similarity definitions, and tuning parameters can yield markedly different partitions even on the same dataset, resulting in unstable cluster assignments.

Consensus clustering was developed to improve robustness by aggregating many clustering solutions and summarizing them as a co-clustering matrix, where each entry reflects how often two samples are assigned together across perturbations, before deriving a final partition from this consensus representation [4–6]. These approaches assume that robust biological relationships remain stable across varied hyperparameters or algorithms [7]. For example, SC3 performs consensus clustering by varying how cell–cell similarity is computed and represented in parameter space [5]; Resampling-based Sequential Ensemble Clustering (RSEC) explores combinations of different base algorithms and tuning parameters [8]; Ensemble methods such as Single-cell Aggregated (From Ensemble) (SAFE)

Received: July 17, 2025. **Revised:** April 1, 2026. **Accepted:** April 24, 2026

and Single-cell Aggregated Clustering via Mixture Model Ensemble (SAME) clustering aggregate outputs from several clustering methods [9, 10]; and ConsensusClusterPlus (ConClusPlus) estimates consensus structure through repeated subsampling of items and features [11].

Biological systems are frequently organized hierarchically, as a product of layered regulatory mechanisms in which broader processes subdivide into progressively more specific subprograms [12–14]. Therefore, meaningful biological variation can exist at multiple scales within the data, such that multiple clustering resolutions can simultaneously reflect the true underlying structure [15]. Despite this, existing consensus clustering approaches typically do not assess robustness across candidate cluster numbers to select a single best partition, implicitly assuming that the underlying biological structure is best captured at one optimal resolution. They typically perturb algorithms, distance metrics, or resampling approaches at fixed granularity [5, 8–11]. By contrast, quantifying the stability of groupings across increasing granularities can distinguish stable partitions representing biological substructure from transient partitions driven by noise and technical artefacts, enabling more reliable recovery of meaningful sub-structure.

Even when clustering is optimized, unsupervised analyses yield many candidate clusters that are not equally robust and may be differentially affected by sampling limitations or noise in the data. While the co-clustering matrix generated during consensus clustering serves as an empirical measure of how consistently samples cluster together across perturbations, most consensus approaches do not provide a framework for systematic evaluation, making it challenging to filter low-quality, stochastic clusters and prioritize robust clusters for downstream analyses [5, 8–11, 16, 17].

In this study, we present Untangled, an unsupervised consensus clustering pipeline that directly interrogates clustering resolution to quantify relationship strength. It iteratively clusters across increasing granularities to identify and prioritize stable data structures across biological scales. Through this process, Untangled explicitly assesses how finely the data can be meaningfully partitioned. When benchmarked across simulated and real datasets with ground-truth labels, Untangled consistently performed comparably to existing consensus and ensemble clustering approaches when clusters were independent, and outperformed existing methods when sub-structure was present. We demonstrated the practical utility of Untangled by applying it to high-throughput screening datasets characterizing drug effects using diverse modalities. In these contexts, Untangled effectively identified and prioritized drug clusters enriched for shared mechanisms and robust phenotypic responses. Overall, Untangled integrates multi-resolution consensus clustering with quantitative post-hoc evaluation to resolve hidden structure and prioritize robust, biologically coherent clusters, enabling efficient identification of actionable groupings in large-scale, complex datasets.

Results and discussion

Untangled workflow for multi-resolution consensus clustering

Using a pre-processed, dimensionality-reduced dataset as input, Untangled iteratively implements Seurat's graph algorithm *FindClusters* (default Louvain modularity optimization) across a finely spaced range of resolutions (default 0.2 to 20 in steps of 0.2), generating a clustering assignment for each object at each resolution. The resulting assignments are aggregated to derive a consensus matrix quantifying the pairwise co-clustering frequencies. This transforms raw distance into an interpretable stability measure quantifying relationship consistency across increasing granularities. To optimize the number of clusters, Ward's minimum-variance hierarchical clustering is applied to the consensus matrix to generate candidate partitions across a range of cluster numbers (default 2–300). The average silhouette score across clusters at each partition is calculated, where higher values indicate objects are consistently co-clustered within their assigned group and well-separated from other clusters. The optimal number of clusters is indicated by a plateau in the silhouette score, beyond which additional splits lead to no further stable partitions. The final clustering is obtained by cutting the hierarchical dendrogram at the optimal cluster number. Resulting clusters are stratified based on internal coherence (mean within-cluster correlation computed using original matrix), robustness (mean silhouette score computed using consensus matrix), and assessed for conservation across contexts where applicable (Fig. 1a).

Originally developed to identify gene programs from sparse, gene-trait association data [18], we implement Untangled as an R package that integrates multi-resolution clustering, cluster number optimization, and cluster evaluation and conservation metrics for effective analysis of diverse data modalities and experimental designs.

To illustrate the workflow, we first applied Untangled to Modified National Institute of Standards and Technology (MNIST) dataset, a standard dataset containing images of handwritten digits (0–9), numerically represented by pixel intensity (Fig. 1b) [19]. It is widely used as a machine learning benchmark and has been previously used to benchmark clustering methods in high-dimensional biological data [20, 21]. Applying Untangled to dimensionality-reduced MNIST data produced a sparse, block-structured consensus matrix that clearly separated digit classes (Fig. 1c). Untangled identified 106 clusters as the optimal cluster number using the silhouette plateau, markedly higher than the 10 'true' classes in the dataset (Fig. 1d-e). However, we showed that the average purity of clusters was very high (0.89 ± 0.12) and that sub-clusters of a single digit appeared to capture meaningful variation, such as distinct handwriting styles (Fig. 1f). To validate the cluster-evaluation framework in Untangled, clusters were stratified by the average silhouette score, indicative of cluster robustness, and the average pairwise correlation of pixel intensity features within each cluster (Fig. 1e). Clusters in the top quartile showed significantly higher mean purity (0.92 ± 0.08) than those in the bottom quartile (0.85 ± 0.15) (Wilcoxon test, $P = .001$), demonstrating effective prioritization (Fig. 1g).

Untangled clustering robust to parameter selection

We further performed sensitivity analyses on two simulated and two real datasets (MNIST and high-content imaging) across varying graph-clustering backends, iterations, and resolution parameter choices (Supplementary Figs S1 and S2). We found the pipeline to be overall robust to parameter selection, with largely consistent agreement for all clusters (Adjusted Rand Index $0.23 \leq \text{ARI} \leq 0.96$), particularly in simulated data with clear, underlying structure (Fig. 1h-j, Supplementary Fig. S1). In real data, non-graph-based clustering methods (hierarchical clustering, GMM, K-means) showed reduced agreement with graph-based approaches (Louvain, Leiden, GraphSpectral, SLM) ($0.09 \leq \text{ARI} \leq 0.60$) and poorer clustering accuracy, suggesting that graph-based methods are optimal (Supplementary Figs S1 and S2).

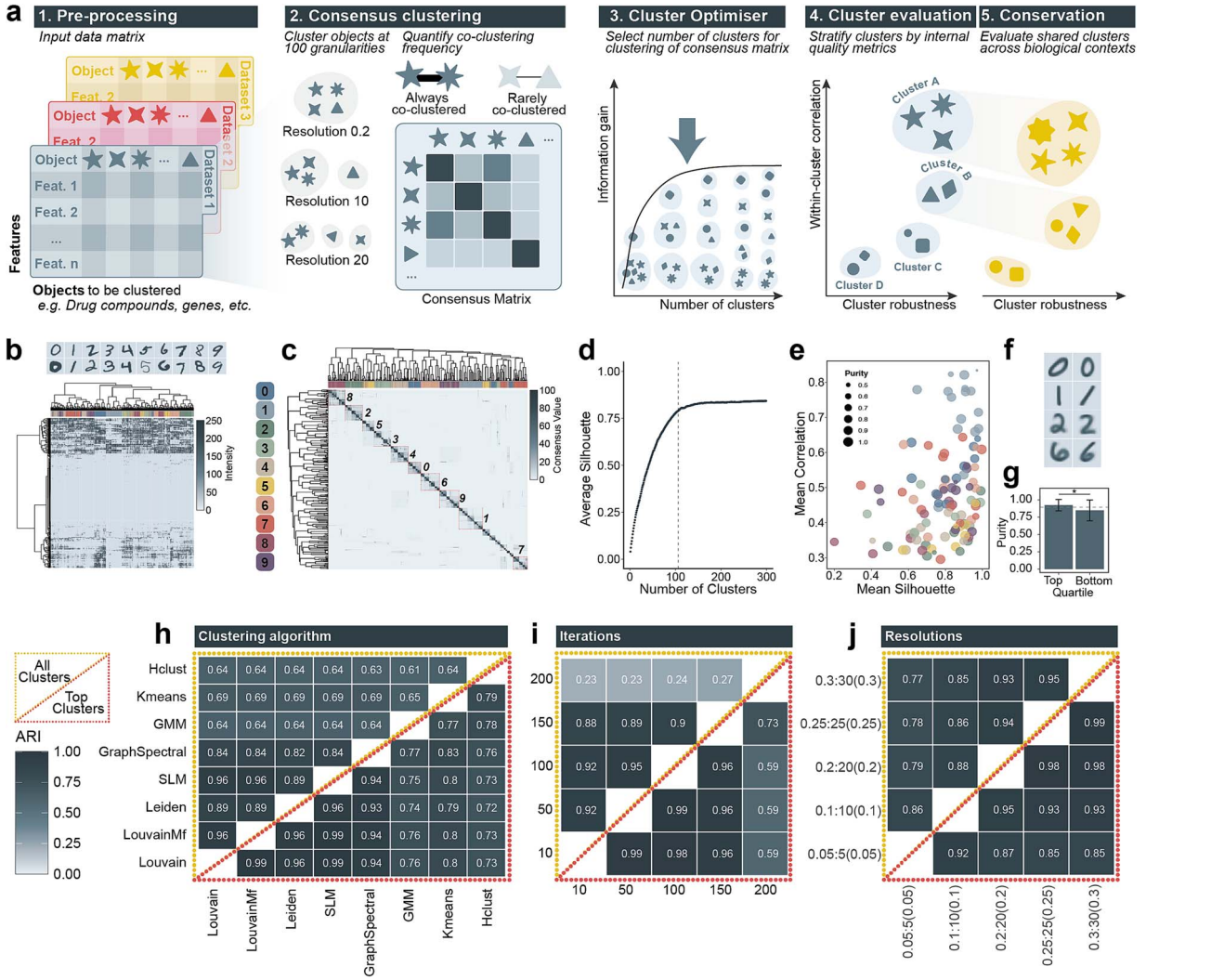

Figure 1 Untangled workflow and multi-resolution clustering of MNIST dataset. (a) Schematic overview of the Untangled workflow. (b) Representation of MNIST dataset, a collection of handwritten digits. (c) Consensus matrix generated by Untangled for 10 000 MNIST digits. The largest block structure for each digit is annotated. (d) Determination of optimal cluster number by plateau in average cluster robustness (silhouette) with increasing cluster number. (e) Untangled digit clusters stratified by their intra-cluster correlation and robustness. Colour indicates most prevalent digit. Size indicates cluster purity. (f) Averaged visualisations of digits in each Untangled cluster capture meaningful sub-digit variation. (g) Average cluster purity in the top and bottom quartile of clusters stratified by silhouette and correlation. $* p < 0.05$. Error bars represent standard deviation. Dashed line indicates average purity of all clusters. (h-j) Average ARI across benchmark (MNIST, morphology) and simulated datasets evaluating cluster robustness to (h) base clustering algorithms, (i) resolution density/iterations, (j) resolution ranges. Mean ARI summarized across all clusters (upper triangle) and top-ranked clusters (lower triangle).

Finally, we showed that top quartile clusters were highly preserved across parameter changes ($0.59 \leq \text{ARI} \leq 0.99$) (Fig. 1h–j), demonstrating that Untangled consistently prioritizes stable groupings irrespective of parameter selection.

Benchmarking untangled performance

We benchmarked Untangled against alternative consensus clustering methods (ConClusPlus, SAME, SAFE, SC3, RSEC) using datasets with ground truth labels (MNIST, Zhengmix8eq). With cluster labels available, we assessed the ability of each method to (i) accurately estimate the number of true groups, (ii) assign items to their labels accurately, measured through the ARI and the purity (proportion of each cluster consisting of a single ground-truth class). As previously discussed, Untangled overestimated the number of clusters in MNIST, leading to a poor ARI relative to all other consensus methods, except for RSEC (Fig. 2a, Supplementary Fig. S3a). However, this was expected as real benchmark datasets often contain additional sources of variation and noise that are not captured by the annotated classes. By preserving co-clustering relationships across resolutions, the algorithm is sensitive to within-digit substructures. Indeed, clusters identified by Untangled had the highest cluster purity compared to alternative methods, indicating strong between-digit distinction (Fig. 2b). Moreover, the sub-clusters of each digit identified by Untangled captured more within-digit variance than any other consensus method in digit shape features, such as slant (Untangled $pR^2 = 0.74$, benchmark methods $pR^2 = 0.23$–0.58), vertical symmetry (Untangled $pR^2 = 0.39$, benchmark methods $pR^2 = 0.11$–0.28), and aspect-ratio (Untangled $pR^2 = 0.60$, benchmark methods $pR^2 = 0.14$–0.37) (Fig. 2c).

Figure 2 Benchmarking of untangled against other consensus clustering approaches. (a-c) Application of consensus clustering methods to MNIST. (a) Optimal number of clusters identified by each method across increasing numbers of images used. (b) Average cluster purity of each method relative to a random control at 10,000 images. (c) Proportion of within-digit variance captured by each method at 10 000 images. (d) Schematic of simulated data with key parameters and outcomes indicated. (e) Deviation between optimal number of clusters identified by each method and the true number of clusters. (f-g) Performance of each method for (f) ARI and (g) average cluster purity across parameter variations in simulated datasets. (h) Schematic of simulated data with hierarchically structured clusters and multi-level evaluation. (i) Deviation between optimal number of clusters identified by each method and the true number of clusters. (j-k) Performance of each method for (j) ARI and (k) cluster purity at each hierarchy level summarized across multiple data parameters. Error bars represent standard deviation.

Given that many consensus methods were developed for Ribonucleic acid-sequencing (RNA-seq) data, we additionally benchmarked methods on a gold-standard single-cell RNA-seq peripheral blood mononuclear cell (PBMC) dataset [22], where Untangled similarly achieved higher cluster purity relative to other methods despite a lower ARI (Supplementary Fig. S3b-d). We further examined Untangled-defined subclusters within canonical PBMC cell types (Supplementary Fig. S4a). Within each parent lineage, sibling subclusters showed distinct transcriptomic profiles, with modest pairwise correlation in average gene expression of variable genes and limited overlap in enriched STRING and GO terms (Supplementary Fig. S4b-e). More importantly, many enriched terms identified in child subclusters were absent from the corresponding parent cell type, demonstrating its ability to resolve biological heterogeneity embedded in broader cell identity classes (Supplementary Fig. S4f). Additionally, Untangled showed stronger performance as dataset size increased, showing higher cluster purity across both MNIST and PBMC datasets and explaining greater variance in MNIST digit-shape features (Supplementary Fig. S5). Runtime and memory usage across increasing sample sizes further indicated that Untangled scales in a predictable manner, supporting its use on larger datasets, despite a current consensus matrix construction bottleneck at very large sample sizes (>100 k) (Supplementary Fig. S6).

To test the effect of common data variables on Untangled performance, we simulated a latent-pathway model with independent clusters, where the number of samples (n), features (p), clusters (k), latent pathways (q), cluster variability (σ_q), and data noise (σ_e) could be explicitly defined (Fig. 2d). Under this simulation, where clusters had no sub-structure, Untangled minimally overestimated the true cluster number (deviation from $k = 3.1 \pm 2.6$), performing comparably to other methods (Fig. 2e). We then systematically evaluated the effect of permutations in all key simulation variables (n, p, k, q, σ_q, σ_e), on all consensus methods summarizing performance using the area under the curve (AUC) for both ARI and cluster purity

(Supplementary Table S1). Across perturbations, Untangled achieved AUCs comparable to the best-performing consensus methods, with consistently high performance for both ARI (AUC = 0.91–0.98) and purity (AUC = 0.96–0.99), while maintaining robustness under increasing noise (ARI AUC = 0.79; purity AUC = 0.88) (Fig. S2f-g, Supplementary Table S2). To test the ability of Untangled to detect underrepresented groups, we simulated rare clusters that varied in size relative to a common cluster and in their degree of separation relative to other clusters. Across both $n = 1000$and $n = 15\,000$, precision and recall remained high when the rare cluster was more than approximately one-fifth the size of the common cluster. Performance declined as the rare cluster became increasingly small, particularly at lower separation, whereas more distinct rare clusters were recovered reliably even at low abundance (Supplementary Fig. S7).

Many biological systems are hierarchically organized, in which broad upstream programs branch into more specific downstream pathways and signalling processes, as exemplified by the Gene Ontology Resource [23]. To model this phenomenon, we simulated data with a hierarchical cluster structure, so that true groups could be present at multiple levels (Fig. 2h, Supplementary Table S3). Under this simulation, Untangled was the only method to accurately estimate the true number of clusters in the data at the finest level (deviation from $k = 0.67 \pm 3.4$) whereas all other methods, except for RSEC, markedly under-estimated cluster number (deviation from $k < -8.89$) (Fig. 2i). Compared to other methods, Untangled also achieved higher ARI at deeper levels of the hierarchical tree (levels 4–5) (Untangled ARI = 0.63, benchmark methods ARI = 0.44–0.58), and maintained the highest cluster purity across all hierarchy levels (Untangled purity = 0.92, benchmark methods purity = 0.75–0.91) (Fig. 2j-k, Supplementary Table S4). These results demonstrate that Untangled is well-suited to recovering the true structure in hierarchical data where meaningful groupings exist across multiple scales.

Applying untangled to high-content morphology drug screens

Large-scale screening datasets often contain complex response patterns shaped by biological heterogeneity, context-dependent effects, and technical variation, which further complicate cluster interpretation and target prioritization. By aggregating clustering solutions across granularities, Untangled reduces noise-induced associations while prioritizing robust drug sets with shared mechanisms. We demonstrate these properties across various large-scale screens, including a high-content imaging drug screen of ovarian cancer cell lines [24], a gene deletion library in *Escherichia coli* [25], and profiling of transcriptional responses to drug perturbations [1] (Fig. 3a).

First, we analyzed high-content imaging data measuring the morphological response of 3 low-grade serous ovarian carcinoma cell lines with distinct genetic profiles (AOCS-2, VOA-6406, SLC58) [24, 26, 27], and one normal ovarian surface epithelial cell line (IOSE-523) following exposure to a diverse library of 5596 drugs at 3 different concentrations ($10, 1, 0.1\ \mu$M) (Supplementary Table S5) [24]. We tested whether Untangled could identify and prioritize drug clusters that induced shared morphological changes across 2175 imaging features and reflected common underlying biological mechanisms. In the original high-dimensional drug-by-imaging matrix, drugs appeared highly similar, obscuring the true number of groupings (Fig. 3b). Reconfiguring the dense drug-by-imaging matrix into a sparse drug–drug consensus matrix revealed clear co-response structure and enabled meaningful assessment of cluster separation and stability (Fig. 3b). Through this transformation, Untangled identified the optimal number of clusters at the silhouette curve plateau, yielding ∼200 clusters in each of the four cell lines (Fig. 3c).

Morphology-based drug clusters conserved across cell lines and enriched for shared mechanisms

Clusters were next stratified by silhouette and correlation scores to prioritize drug groups with robust, shared morphological responses (Supplementary Fig. S8a-d). We were particularly interested in identifying drug mechanisms that induced consistent phenotypic responses irrespective of genetic profile, given the genetic heterogeneity of ovarian cancer. To assess this, we evaluated the consistency of drug clusters across cell lines using a conservation metric implemented in Untangled (Methods). Many of the drug clusters prioritized by Untangled's internal metrics were significantly conserved across all ovarian cancer cell lines, reinforcing their validity and highlighting the ability of Untangled to effectively prioritize reproducible clusters (Fig. 3d). As large-scale screens are inherently noisy, we simulated noise in the morphology data to test whether the conservation metric can distinguish biologically meaningful clusters from noise-induced artefacts. As expected, increasing noise reduced within-cluster correlation, silhouette score, and similarity to matched unperturbed clusters, while conservation across cell lines remained relatively robust (Supplementary Fig. S9a). Noisy clusters were then classified as biologically retained if they matched a significantly enriched baseline cluster from the unperturbed data at a specific recall threshold. Across noise levels, biologically retained clusters consistently showed higher conservation scores than artefactual clusters (Supplementary Fig. S9b-c).

To assess the biological significance of identified clusters, we evaluated the enrichment of annotated pathways in each drug cluster. Across all four cell lines, clusters were significantly enriched for cancer-related pathways (Supplementary Fig. S8). For instance, the top-ranked drug clusters from cell line SLC58 were strongly enriched for key oncogenic pathways, including *MAPK/ERK* [26], *PI3K/AKT/mTOR* [28], and Cell cycle/DNA damage [29] (Fig. 3d). To reinforce these findings, we examined representative microscopy images of cells treated with drugs from selected clusters. Prioritized clusters exhibited more consistent morphological changes in treated cells versus DMSO control, contrasting with inconsistent and weak morphologies in poorly ranked clusters (Supplementary Fig. S10). Notably, the observed morphological changes closely aligned with the enriched pathways for these clusters. For example, drugs in top-ranked Cluster 200 from cell line SLC58 were enriched for the MAPK/ERK pathway and induced morphological changes indicative of apoptosis, including cell shrinkage, membrane blebbing, and fragmented nuclei (Fig. 3e). These observations align with the known role of the MAPK/ERK pathway in regulating cell proliferation and survival, particularly in low-grade serous ovarian cancer pathogenesis [26, 30], where its disruption can lead to apoptosis [31].

To test the ability of Untangled to recover hierarchical structure in real screening data, we used drug annotations organized as pathway and target hierarchies and evaluated enrichment across a grid of clustering granularities (Supplementary Fig. S11a). As the granularity increased, Untangled clusters became enriched for deeper, more specific annotation levels (Fig. S3f, Supplementary Fig. S11b-c). Together,

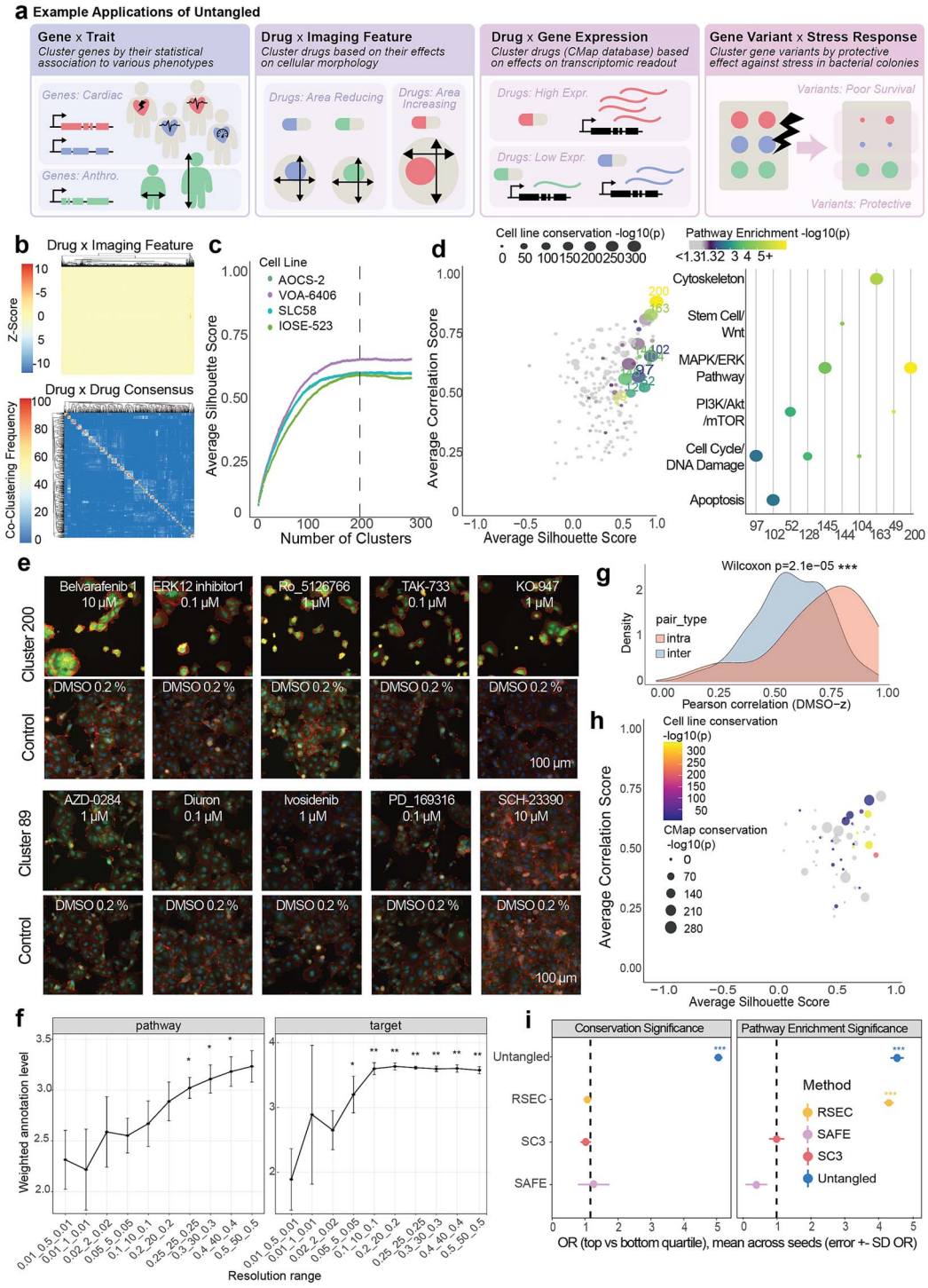

Figure 3 Untangled reveals biologically meaningful drug clusters based on drug-induced cellular morphology changes in ovarian cancer. (a) Overview of four applications of Untangled across diverse biological data modalities. (b) Transformation of raw drug-by-imaging matrix (top) into a drug–drug consensus similarity matrix (bottom). (c) Determination of optimal cluster number by average silhouette score plateau for all four cell lines. The dashed line indicates the selected cluster number, K = 200. (d) Scatter plot of drug clusters from cell line SLC58 stratified by average cluster silhouette score and correlation score. The most significant pathway enrichment for the 10 top-ranked clusters is shown. (e) Representative microscopy images of cells (1 field per well) from cell line SLC58 treated with example drug conditions from top-ranked cluster 200 and bottom-ranked cluster 89. Images shown with DMSO control and 100 μm scale bar. Cells were stained with DAPI (nucleus), CellMask (plasma membrane), and phalloidin/rhodamine (F-actin). (f) Shift in annotation specificity across clustering resolution ranges, shown as the mean weighted annotation level for enriched pathway (left) and target (right) terms across cell lines.Points show the mean across cell lines, and error bars indicate $\pm$ SD. Asterisks denote significance versus the coarsest reference range from paired t-test ($*$ p adj < 0.05, $**$ p adj < 0.01, $***$ p adj < 0.001. (g) Matched multiplexed RNA-seq validation of morphology-defined clusters. Density distribution of DMSO Z-score-normalised expression correlation between intra-cluster drug pairs and inter-cluster drug pairs for a subset of compounds profiled at matched concentration. (h) Orthogonal validation using independent transcriptomic dataset, CMap. Morphology-based clusters stratified with point size indicating conservation significance with CMap expression-derived drug clusters and colour indicating cross-cell line conservation significance. (i) Logistic regression quantifying enrichment of significant clusters in the top quartile for cell-line conservation (left) and pathway enrichment (right) across untangled and alternative clustering methods. Points indicate the mean odds ratio pooled across cell lines and random seeds and error bars indicate $\pm$ SD.

the results reinforce that Untangled can leverage resolution as a proxy for biological scales to identify meaningful hierarchical structure, allowing broad phenotypic programs to be decomposed into more specific mechanistic subgroups.

Morphology-based drug clusters exhibit shared transcriptional signatures

We further demonstrated that morphology-derived drug clusters also elicit concordant transcriptomic changes through three independent analyses that reinforce shared underlying biological mechanisms. First, we analysed high-content imaging datasets for AOCS-2 and four additional cell lines (VOA-4698, VOA-10841, VOA-3448, iOvCa241) along with matched multiplexed RNA-seq (MAC-seq) profiles for a subset of compounds present in these datasets [32]. We applied Untangled to the full imaging dataset for each cell line to derive morphology-based drug clusters. Then, for compounds available in MAC-seq, we computed Pearson correlations between DMSO-normalized RNA-seq signatures for compound pairs assigned to the same morphology-based cluster ('intra') versus pairs assigned to different clusters ('inter'). Intra-cluster pairs showed significantly higher transcriptomic correlation than inter-cluster pairs (Wilcoxon $P = 2.1 \times 10^{-5}$; Fig. 3g), supporting that compounds clustered by Untangled also share concordant transcriptional programs.

Next, we assessed whether this concordance generalizes outside of ovarian cancer using the Connectivity Map (CMap) [1], an independent dataset characterizing gene expression response to drug perturbations, focusing on breast, prostate cancer, and leukaemia cell lines. Applying Untangled to CMap yielded 175 drug clusters. Conservation analysis between CMap transcriptomic-based clusters and morphology-based clusters showed significant cluster conservation across all four ovarian cell lines, indicating consistent transcriptomic changes in an orthogonal dataset, with stronger conservation observed in clusters prioritized by Untangled (Fig. 3h, Supplementary Fig. S12). Notably, stronger cross-cell-line conservation remained significantly associated with greater CMap concordance after adjustment for cluster size and compound overlap ($\beta = 0.063$; Linear $P = 3.6 \times 10^{-2}$).

Finally, using independent pharmacogenomic dataset, Cancer Therapeutics Response Portal (CTRP), we investigated if morphology-defined clusters share concordant drug sensitivity profiles across independent cancer cell-line panels. Using overlapping drug compounds (Supplementary Fig. S13a), we show that drug compound pairs assigned to the same morphology-defined cluster have significantly higher correlation than those assigned to different clusters by bootstrapping (Bootstrap $P = 3.3 \times 10^{-2}$; Supplementary Fig. S13b). This indicates that compounds grouped by morphology exhibit more similar sensitivity profiles across diverse cancer cell lines than expected by chance.

Benchmarking in drug screening dataset

To evaluate Untangled's performance relative to other consensus clustering methods in this complex, screening dataset, we evaluated whether clusters prioritized by each method were more likely to be enriched for biological signal. In the absence of ground-truth labels, conservation across cell lines and pathway enrichment were used as proxies. Untangled consistently showed the strongest enrichment of significant clusters in the top quartile relative to the bottom quartile across cell-line conservation (Untangled Odds Ratio (OR) = 5.9 ± 0.1, benchmark methods OR = 0.91–1.10) and pathway enrichment (OR = 4.9 ± 0.2, benchmark methods OR = 0.40–4.31), outperforming other consensus methods (Fig. 3i). The proportion of conserved or enriched clusters between the top and bottom quartiles showed the same pattern (Supplementary Fig. S14a), demonstrating that Untangled's internal ranking more reliably prioritized clusters that reflect true underlying biological programs. Cluster prioritization enrichment was also robust to parameter selection (Supplementary Fig. S14b). Lastly, Untangled's ability to prioritize biologically significant clusters also improved with dataset scale (Supplementary Fig. S14c).

Application to additional high-dimensional screens

We further demonstrated the ability of Untangled to identify and prioritize robust clusters on two additional high-dimensional screening data. First, when applied to a genome-wide *E. coli* single gene-deletion library screened against 324 stress conditions [25], Untangled effectively grouped gene-deletion mutants based on shared phenotypic responses to chemical and environmental stressors, identifying highly robust gene clusters enriched for shared biological processes and condition clusters enriched for shared drug targets (Supplementary Fig. S15). Likewise, in analysing CMap's transcriptional responses to over 1000 drug perturbations across three other tumour cell lines (Luminal A breast cancer, prostate cancer, and leukaemia) [1], Untangled identified drug clusters enriched for shared mechanisms of action and disease indications that were conserved across cancer types (Supplementary Fig. S16).

Conclusion

Unsupervised clustering is widely used to discover latent biological structure, yet most screening datasets contain nested or hierarchical structures, where broad cellular programs subdivide into specific pathway states, through which they shape drug response, cellular morphology, or transcriptional landscape. While typically treated as a tunable hyperparameter, resolution can reflect biological scales where data group meaningfully at each level. Coarse partitions can reflect shared broader upstream processes, while finer partitions capture downstream subprograms. Current consensus and ensemble clustering frameworks largely cluster at a single optimal resolution, improving cluster robustness mainly by aggregating perturbations in algorithm, hyperparameter, sampling, or label [5, 8–11, 16, 17]. As a result, they do not capture strength of relationships across granularity, limiting their ability to separate stable biological substructure from noise-induced partitions.

Untangled addresses this limitation by modelling clustering behaviour across resolutions as a natural proxy for biological scales, rather than optimizing a single resolution. It builds consensus from pairwise co-occurrence across increasing resolution and then ranks consensus-derived clusters using measures of internal coherence and robustness. Using benchmark and simulated datasets, Untangled resolved meaningful clusters with higher purity and more accurately recovered nested structure than benchmark methods, highlighting its utility for partitioning hierarchically structured biological data. Other methods that do not exploit relationship stability were less likely to recover distinct sub-clusters confined in hierarchical nodes. While RSEC performed comparably in nested simulation, its performance declined as real datasets scaled up, likely due to greater susceptibility to noise.

Importantly, Untangled's strength in screening lies not only in cluster recovery but in cluster prioritization. When applied in real-life drug screens, Untangled's internal ranking more reliably identified biologically significant clusters than other consensus methods, prioritizing latent drug groupings with shared pathway perturbations that converge on similar cellular morphologies. This prioritization leverages multi-resolution stability to distinguish meaningful substructure from noise-driven associations. In contrast, other frameworks were less effective at prioritizing biologically significant clusters. In a screening context, failing to triage noise-induced or unstable clusters can be particularly costly due to resource-intensive follow-up.

Notably, Untangled's performance depends on upstream preprocessing, dimensionality reduction and graph construction, and prioritization does not replace the need for orthogonal experimental validation. Taken together, these results support Untangled as an interpretive framework for large-scale biology screens applicable to various data modalities. To facilitate broad adoption, we provide Untangled as a user-friendly R package, offering a versatile clustering and prioritization workflow to guide actionable insights from any highly complex dataset with applications spanning drug discovery, target identification, and disease subtyping with potential for application in non-biological contexts.

Key Points

- Untangled aggregates pairwise co-assignment patterns across increasing clustering granularity to denoise relationships, then performs cluster number optimisation and robustness evaluation to prioritise the most stable, actionable groupings.
- Across simulated and benchmark datasets, Untangled performs comparably to existing methods for independent cluster recovery, while more accurately resolving nested or hierarchical substructure and capturing higher-purity clusters when meaningful variation exists within ground-truth classes.
- In ovarian cancer drug profiling screens, Untangled demonstrates high utility for target prioritisation by ranking robust morphology-defined drug clusters enriched for shared mechanisms and conserved across cell lines and orthogonal drug-response modalities.
- Untangled generalises across diverse data modalities and is provided as a user-friendly R package to support efficient clustering and prioritisation in complex, high-dimensional biological screens.

Acknowledgements

This work has been supported by grant funding from the NHMRC (MRFCDDM000033 to NP and 2007625 to NP) and the National Heart Foundation of Australia (106721 to NP). The Victorian Centre for Functional Genomics RRID: SCR_025582 (K.J.S.) is funded by the Australian Cancer Research Foundation (ACRF), Phenomics Australia, through funding from the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) program, the Peter MacCallum Cancer Centre Foundation, and the University of Melbourne Collaborative Research Infrastructure Program. KIP acknowledges funding from the Victorian Government (Victorian Cancer Agency Mid-Career Low Survival Cancer Philanthropic Research Fellowship, APP MCRF23014), the Department of Defense (Ovarian Cancer Research Program Pilot Award, APP OC220022), and The CASS Foundation. KJS, IGC, and DC acknowledge support from the Medical Research Future Fund (APP1200264) and Therapeutic Innovation Australia through the Pipeline Accelerator Scheme. DC further acknowledges funding from the Victorian Cancer Agency (MCRF19046) and the Ovarian Cancer Research Foundation (GA-2024-02). M.S.C and D.C would like to acknowledge additional support from the BC Cancer Foundation, Vancouver General Hospital/UBC Hospital Foundation, the Janet D. Cottrelle Foundation, Cure Our Ovarian Cancer, Ovarian Cancer Canada/OVCAN, and the Cancer Research Society. We gratefully acknowledge the patients and their families, including the MacKenzie, Lawler, MacRae, Ho, Luther, Ludemann, and Schmid families. We also thank Professor John Hooper (Mater Research) and the Australian Ovarian Cancer Study for generating and providing cell lines used in this study. We thank Jennii Luu and Robert Vary from the Victorian Centre for Functional Genomics for their screening support. We acknowledge Timothy Semple and Stuart Bencraig (Molecular Genomics Core, Peter MacCallum Cancer Centre) for assistance with MAC-Seq. We thank Compounds Australia for the provision of compounds and logistical support. We also thank Sophie Shen for her assistance with the schematic in Figs 1 and 2.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflicts of interest

No conflicts of interest are reported for any of the authors.

Funding

None declared.

Code and data availability

Untangled is implemented in an R package and is available in the following GitHub repository: https://github.com/palpant-comp/UnTANGLeD_R_Package. The raw imaging data from morphology drug screens can be downloaded from BioImage Archive (https://www.ebi.ac.uk/bioimage-archive/) with accession number S-BIAD1069. The matched multiplexed RNAseq datasets for a subset of compounds can be downloaded from GEO database (http://www.ncbi.nlm.nih.gov/gds) under accession number GSE313916.

References

1. Subramanian A, Narayan R, Corsello SM *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell.* 2017;**171**:1437–1452.e17. https://doi.org/10.1016/j.cell.2017.10.049
2. Regev A, Teichmann SA, Lander ES *et al.* The human cell atlas. *Elife.* 2017;**6**:e27041.
3. DL Donoho. High-dimensional data analysis: the curses and blessings of dimensionality. *AMS Math Challenges Lecture* **1**(2000), 32. 2000.
4. Monti S, Tamayo P, Mesirov J *et al.* Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach Learn* 2003;**52**:91–118. https://doi.org/10.1023/A:1023949509487

5. Kiselev VY, Kirschner K, Schaub MT *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* 2017;**14**:483–6. https://doi.org/10.1038/nmeth.4236
6. Strehl A, Ghosh J. Cluster ensembles a knowledge reuse framework for combining multiple partitions. *J Mach Learn Res* 2002;**3**:583–617.
7. Fred AL, Jain AK. Combining multiple clusterings using evidence accumulation. *IEEE Trans Pattern Anal Mach Intell* 2005;**27**:835–50. https://doi.org/10.1109/TPAMI.2005.113
8. Risso D, Purvis L, Fletcher RB *et al.* clusterExperiment and RSEC: a bioconductor package and framework for clustering of single-cell and other large gene expression datasets. *PLoS Comput Biol* 2018;**14**:e1006378. https://doi.org/10.1371/journal.pcbi.1006378
9. Huh R, Yang Y, Jiang Y *et al.* SAME-clustering: single-cell aggregated clustering via mixture model ensemble. *Nucleic Acids Res* 2020;**48**:86–95. https://doi.org/10.1093/nar/gkz959
10. Yang Y, Huh R, Culpepper HW *et al.* SAFE-clustering: single-cell aggregated (from ensemble) clustering for single-cell RNA-seq data. *Bioinformatics.* 2019;**35**:1269–77. https://doi.org/10.1093/bioinformatics/bty793
11. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010;**26**:1572–3. https://doi.org/10.1093/bioinformatics/btq170
12. Ravasz E, Somera AL, Mongru DA *et al.* Hierarchical organization of modularity in metabolic networks. *Science.* 2002;**297**:1551–5. https://doi.org/10.1126/science.1073374
13. Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* 2009;**10**:141–8. https://doi.org/10.1038/nrg2499
14. Hartwell LH, Hopfield JJ, Leibler S *et al.* From molecular to modular cell biology. *Nature.* 1999;**402**:C47–52. https://doi.org/10.1038/35011540
15. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience.* 2018;**7**:giy083. https://doi.org/10.1093/gigascience/giy083
16. Peyvandipour A, Shafi A, Saberian N *et al.* Identification of cell types from single cell data using stable clustering. *Sci Rep* 2020;**10**:12349. https://doi.org/10.1038/s41598-020-66848-3
17. Sonpatki P, Shah N. Recursive consensus clustering for novel subtype discovery from transcriptome data. *Sci Rep* 2020;**10**:11005. https://doi.org/10.1038/s41598-020-67016-3
18. Mizikovsky D, Sanchez MN, Nefzger CM *et al.* Organization of gene programs revealed by unsupervised analysis of diverse gene-trait associations. *Nucleic Acids Res* 2022;**50**:e87. https://doi.org/10.1093/nar/gkac413
19. LeCun Y, Cortes C, Burges C. *MNIST Handwritten Digit Database*. New York, NY: ATT Labs [Online] Available: http://yann.lecun.com/exdb/mnist, 2010.
20. Abdelaal T, de Raadt P, Lelieveldt BPF *et al.* SCHNEL: scalable clustering of high dimensional single-cell data. *Bioinformatics.* 2020;**36**:i849–56. https://doi.org/10.1093/bioinformatics/btaa816
21. Kopf A, Fortuin V, Somnath VR *et al.* Mixture-of-experts variational autoencoder for clustering and generating from similarity-based representations on single cell data. *PLoS Comput Biol* 2021;**17**:e1009086. https://doi.org/10.1371/journal.pcbi.1009086
22. Duo A, Robinson MD, Soneson C. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res.* 2018;**7**:1141. https://doi.org/10.12688/f1000research.15666.2
23. Ashburner M, Ball CA, Blake JA *et al.* Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium Nat Genet* 2000;**25**:25–9. https://doi.org/10.1038/75556
24. Pishas KI, Cowley KJ, Llaurado-Fernandez M *et al.* High-throughput drug screening identifies novel therapeutics for low grade serous ovarian carcinoma. *Sci Data* 2024;**11**:1024. https://doi.org/10.1038/s41597-024-03869-x
25. Nichols RJ, Sen S, Choo YJ *et al.* Phenotypic landscape of a bacterial cell. *Cell.* 2011;**144**:143–56. https://doi.org/10.1016/j.cell.2010.11.052
26. Shrestha R, Llaurado Fernandez M, Dawson A *et al.* Multiomics characterization of low-grade serous ovarian carcinoma identifies potential biomarkers of MEK inhibitor sensitivity and therapeutic vulnerability. *Cancer Res* 2021;**81**:1681–94. https://doi.org/10.1158/0008-5472.CAN-20-2222
27. Etemadmoghadam D, Azar WJ, Lei Y *et al.* EIF1AX and NRAS mutations Co-occur and cooperate in low-grade serous ovarian carcinomas. *Cancer Res* 2017;**77**:4268–78. https://doi.org/10.1158/0008-5472.CAN-16-2224
28. Mabuchi S, Kuroda H, Takahashi R *et al.* The PI3K/AKT/mTOR pathway as a therapeutic target in ovarian cancer. *Gynecol Oncol* 2015;**137**:173–9. https://doi.org/10.1016/j.ygyno.2015.02.003
29. Huang TT, Lampert EJ, Coots C *et al.* Targeting the PI3K pathway and DNA damage response as a therapeutic strategy in ovarian cancer. *Cancer Treat Rev* 2020;**86**:102021. https://doi.org/10.1016/j.ctrv.2020.102021
30. Grisham RN, Slomovitz BM, Andrews N *et al.* Low-grade serous ovarian cancer: expert consensus report on the state of the science. *Int J Gynecol Cancer* 2023;**33**:1331–44. https://doi.org/10.1136/ijgc-2023-004610
31. Cagnol S, Chambard JC. ERK and cell death: mechanisms of ERK-induced cell death - apoptosis, autophagy and senescence. *FEBS J* 2010;**277**:2–21. https://doi.org/10.1111/j.1742-4658.2009.07366.x
32. Li XM, Yoannidis D, Ramm S *et al.* MAC-Seq: coupling low-cost, high-throughput RNA-Seq with image-based phenotypic screening in 2D and 3D cell models. *Methods Mol Biol* 2023;**2691**:279–325. https://doi.org/10.1007/978-1-0716-3331-1_22

Briefings in Bioinformatics, 2026, **27**, bbag242
https://doi.org/10.1093/bib/bbag242
Problem Solving Protocol