AutoHCS: AI-based scoring of high-content screens results in morphological clustering that predicts mechanisms of action

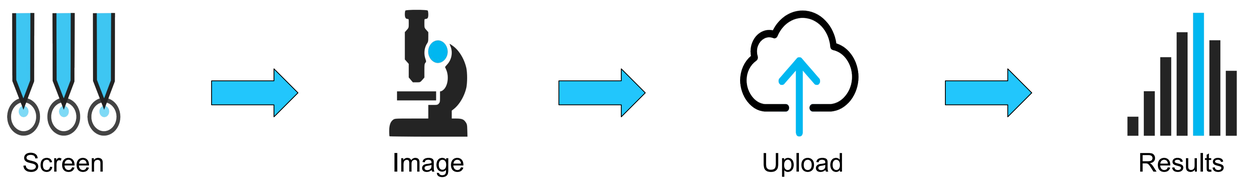
Teresa Findley[1], *Rebecca Winfree*[1], *John Delaney*[1], *Rupert Dodkins*[1], *Ibrahim Bilem*[2], *& Ilya Goldberg*[1]

[1]ViQi Inc., [2]Saguaro Technologies Inc.

ABSTRACT

Modern drug development increasingly depends on high-content compound screens where automation is the key to rapid, impactful discoveries. AutoHCS™ is a multifaceted AI-based analysis tool developed by ViQi Inc. that automatically detects and scores phenotypic responses to compounds in high-content screens. Because the system does not depend on segmentation, it works non-parametrically with multichannel fluorescence, a combination of fluorescence and brightfield, or brightfield alone. The only inputs to the analysis are images from any automated plate imager and a plate map specifying concentrations, replicates, and controls. A few core AutoHCS analytical tools are: 1) comparing compounds of interest against negative and positive controls or target phenotypes 2) evaluating the dose response of compounds of interest and 3) computing morphological clusters across many different compounds of interest. Importantly, AutoHCS AIs can conduct each of these analyses independently or in combination. For example, comparing the dose response of a compound of interest against positive controls will determine which dose, if any, is most similar to a known target phenotype. Whereas, investigating dose-dependent responses independently of controls permits the discovery of novel phenotypes. AutoHCS can also be used to evaluate the ability of compounds to reverse a background phenotype, such as in screens for neuroprotectant drugs, anti-inflammatory drugs, and antivirals. AutoHCS entirely determines its training parameters using the experimental controls rather than user input, which eliminates subjective criteria selection that may bias phenotype scoring. It also allows multiplex scoring of screens both for a positive target phenotype and against negative phenotypes such as cellular toxicity. AutoHCS is cloud-based, so there is no software or specialized computing hardware to install locally. Accordingly, AutoHCS is scalable to millions of images and works regardless of contrast method, cell type, or cellular responses generated.
A key function of AutoHCS is its ability to morphologically cluster compounds according to their induced phenotype. To further investigate the significance of these clusters, we used AutoHCS to analyze a subset of the JUMP dataset, a public HCS dataset with many compounds and replicates. We converted our morphological clusters into gene lists using standard databases and conducted pathway analysis on these gene lists. We found that gene lists from our morphological clusters resulted in a high degree of overlap in database-queried mechanisms of action when compared with randomly generated clusters. As we build our knowledge base using tools like the JUMP dataset, this bioinformatics-based approach to cluster validation will not only allow us to make predictions about novel compounds, but may provide deeper insight into other AutoHCS analyses. With all its capabilities, AutoHCS harnesses the pattern recognition abilities of modern AIs to precisely score and phenotypically hc screens in an entirely automated, objective manner.

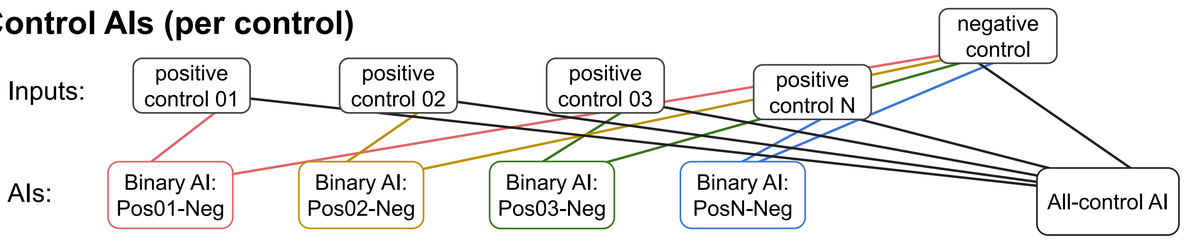
AUTO-HCS USER WORKFLOW



Using AutoHCS is simple and user-friendly. 1) Plate cells configured to your experiment. These can include a target phenotypes, dose-response, background conditions, and negative controls. 2) Capture images using one of the many high-throughput microplate imaging devices that can automatically image plates at high resolution. 3) Upload the images to ViQi Inc. servers along with a plate map. 4) Receive a complete analysis report and quantitative assay readout for each well.

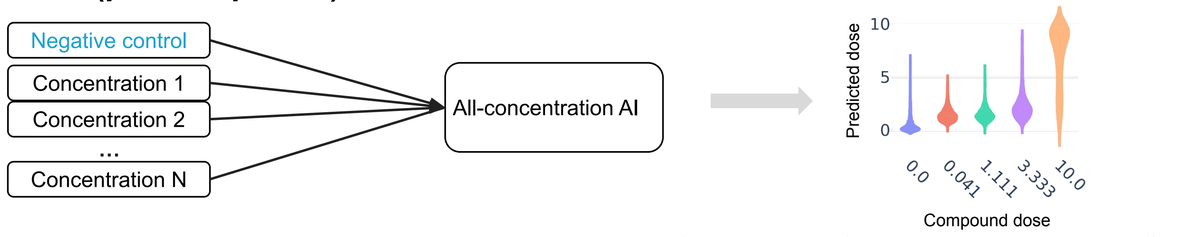
THE BASICS OF AI TRAINING

Control AIs (per control)



For each positive control, we train a binary AI to distinguish that control from the negative control resulting in a simple binary prediction of positive or negative. We also train an AI on all controls that will distinguish phenotypic distance between all controls.

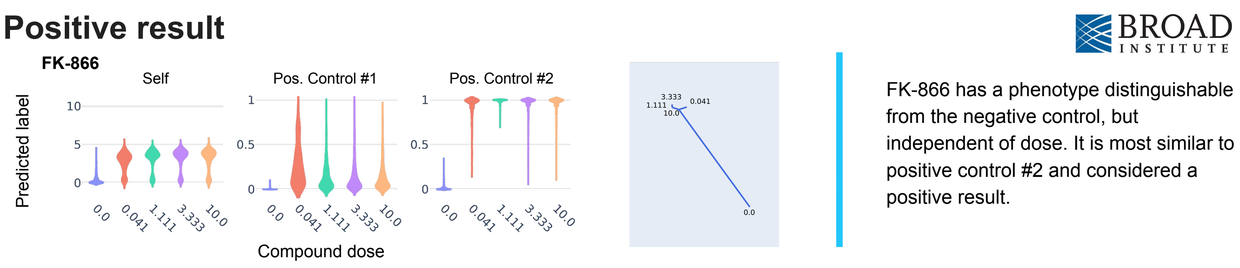
Dose AIs (per compound)



For each compound, an AI is trained on all compound concentrations independently of positive controls. Scoring concentration-dependent effects independently of positive controls allows for the discovery of new phenotypes.

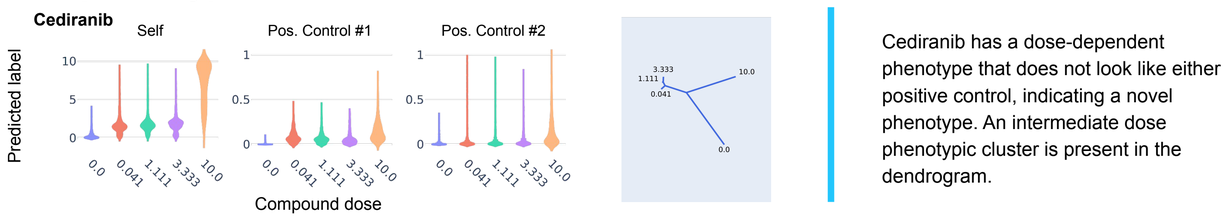
DOSE-RESPONSE ASSAYS: SCORING COMPOUNDS

Positive result



FK-866 has a phenotype distinguishable from the negative control, but independent of dose. It is most similar to positive control #2 and considered a positive result.

Novel phenotype

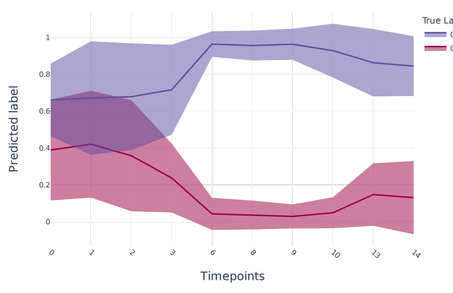


Cediranib has a dose-dependent phenotype that does not look like either positive control, indicating a novel phenotype. An intermediate dose phenotypic cluster is present in the dendrogram.

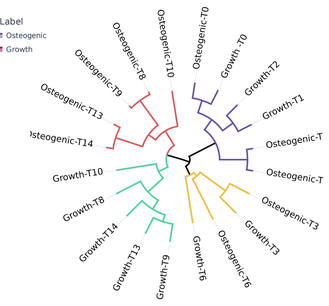
We identified 52 compounds in a JUMP consortium dose response dataset (cpg0004) and their pilot dataset (cpg0000) that included CRISPR (gene knockouts) and ORF (overexpression) edits of each compound's corresponding gene targets. The above examples of our dose-response screen come from the dose-response dataset from the JUMP consortium. Each compound is scored using the AI trained on its compound concentrations (Self) and using each binary control AI (Binaries). By presenting scores using violin plots (left), we help users quickly identify phenotypic profile for each compound across dosages, dose-dependent intermediate phenotypes, and to which control (positive or negative) the phenotypic responses to compounds are more similar. A dendrogram of each dose per compound (right) demonstrates phenotypic clustering across doses. This clearly demonstrates intermediate phenotypes that occur at lower/intermediate doses.

LIVE CELL IMAGING: SCORING TIMECOURSES

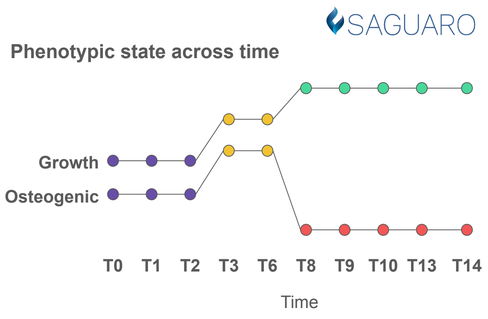
Binary AI trained per timepoint



AI trained on all timepoints

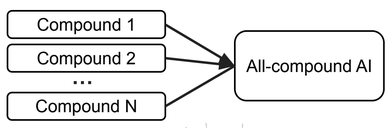


Phenotypic state across time

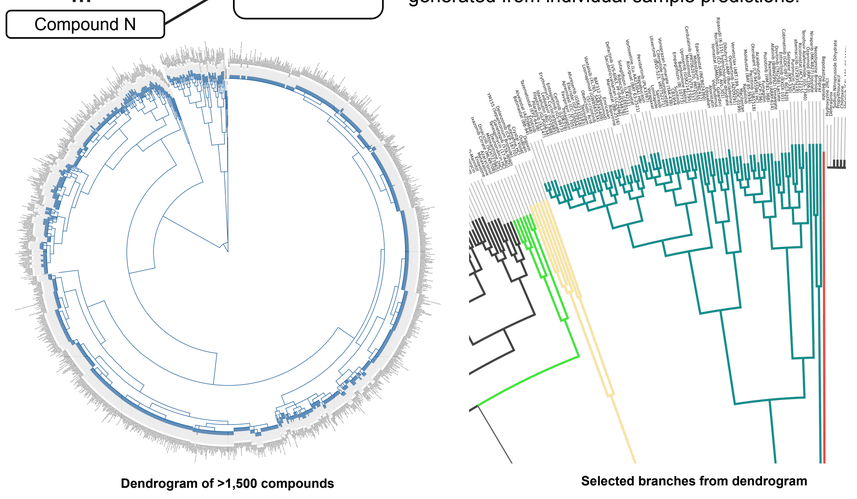


hMSCs stained with ChromaLive, a non-toxic dye, were grown in either growth medium ("Growth") or osteogenic medium ("Osteogeic") to investigate stem cell differentiation. When binary AIs (one per timepoint) are trained on osteogenic vs growth conditions, stem cell commitment towards the osteoblastic lineage is detected at day 6. We can also capture changes in phenotypic state across time by training an AI on all timepoints at each medium condition. By clustering the results into four phenotypic states, we can see clear differentiation in state by day 8.

MORPHOLOGICAL CLUSTERING ACROSS COMPOUNDS

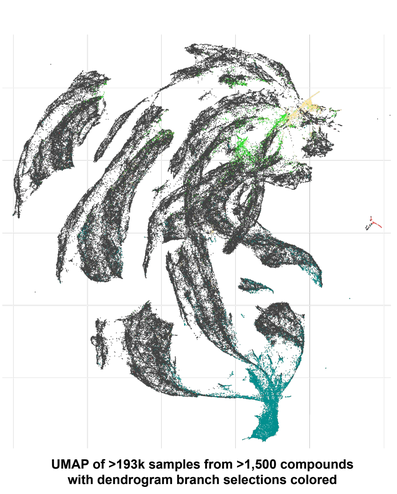


A single AI is trained across compounds. Prediction probabilities determine phenotype similarity, and dendrograms are constructed from the pairwise phenotype distance matrix. UMAPs are generated from individual sample predictions.



Dendrogram of >1,500 compounds

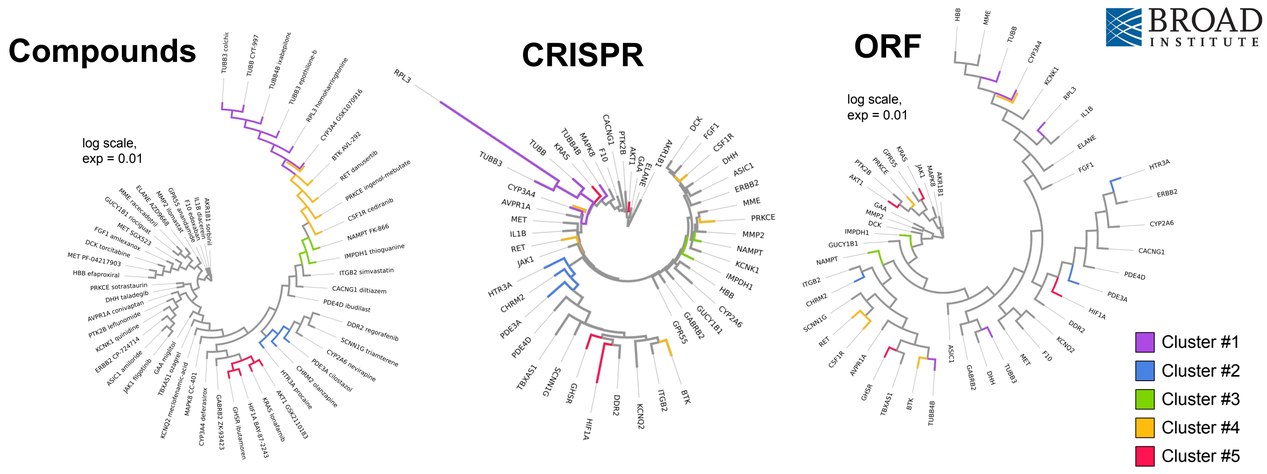
Selected branches from dendrogram



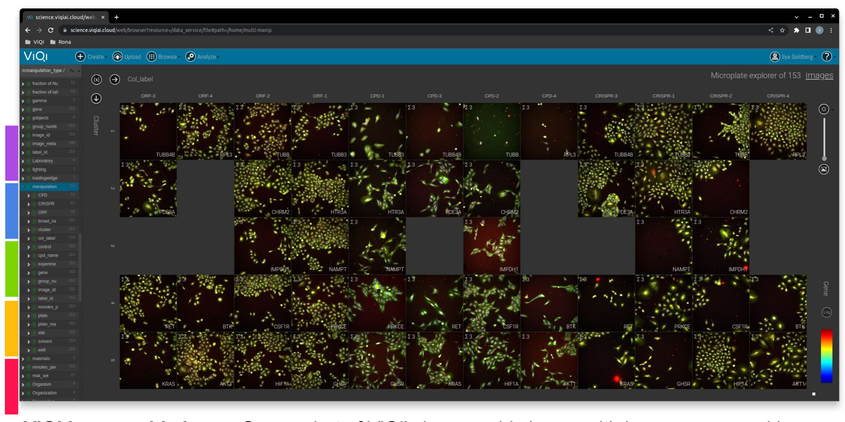
UMAP of >193k samples from >1,500 compounds with dendrogram branch selections colored

A >1,500 compound screen on human primary patient cells conducted by the Sunnybrook Research Institute is morphologically clustered by AutoHCS. Branches can be selected in the dendrogram, highlighted in color and displayed in the UMAP.

JUMP: IDENTIFYING COMPOUND/GENE TARGET CLUSTERS



Dendrograms of 52 compounds (and their gene targets) from the JUMP pilot screen. Left: 52 compound dendrograms labeled with gene target and common compound name. Dendrogram below is scaled exponentially for easier viewing (exp=0.01). Middle: dendrogram for CRISPR edits (gene knockouts) of compound targets. Right: dendrogram for ORF (overexpression) of compound gene targets. Dendrogram below is scaled exponentially for easier viewing (exp=0.01).



ViQi image grid viewer. Screenshot of ViQi's image grid viewer, with images arranged by cluster number (rows) vs. manipulation type (Compounds: CPD, CRISPR, ORF) and target gene. Clicking on image thumbnails launches ViQi's fully-featured browser-based image viewer.

Cluster #1-3: these clusters are conserved between compound and CRISPR dendrograms. Analyzed for mechanism of action (MoA) below.

Cluster #4: contains several tyrosine kinase inhibitors.

- When overexpressed (ORF), *CSF1R* and *RET* are in the same cluster and close to *BTK*.
- When deleted (CRISPR), these genes do not produce similar phenotypes as detected by AI.

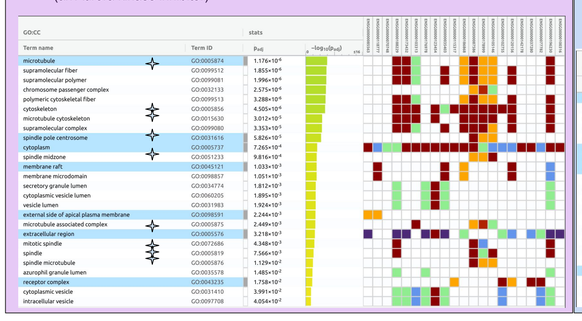
Cluster #5:

- *HIF1A* inhibitors have anti-tumor properties, while ibutamoren is an agonist of a growth factor receptor (*GHSR*).
- Despite expecting opposing proliferation effects, these drugs have similar phenotypes on treated cells (Compound) and their target gene deletions (CRISPR) produce similar phenotypes.
- Genes *AKT1* and *KRAS* are part of the same signaling pathway (tumor cell survival) and cluster when overexpressed (ORF).

EVALUATING CLUSTERS BY MECHANISM OF ACTION

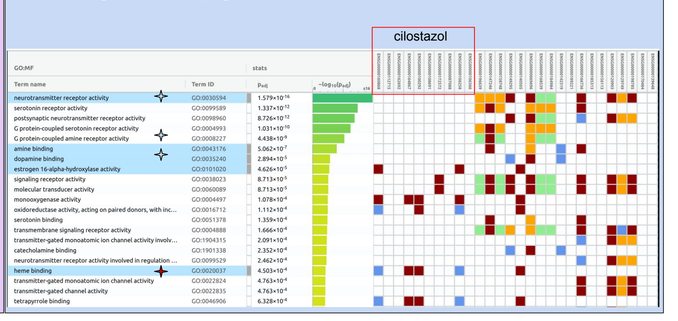
Cluster #1 - Functional enrichment of cellular components (GO:CC) for a conserved cluster of tubulin-targeting compounds (purple).

- Colchicine, CYT-997, ixabepilone, and epothilone-b disrupt tubulin and thus microtubule dynamics.
- Additional compounds in morphological cluster: homoharringtonine (a protein synthesis inhibitor) and GSK-1070916 (an Aurora kinase inhibitor)
- This suggests that AutoHCS identifies diverse mechanisms of action that may contribute to similar phenotypic cellular profiles.

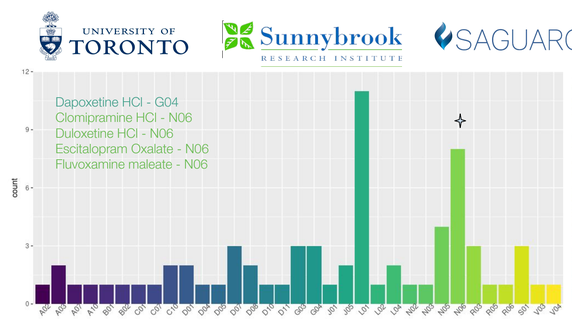


Cluster #2. Functional enrichment of molecular functions (GO:MF) for a conserved cluster of central nervous system-active compounds (blue).

- Olanzapine and procaine have been studied heavily in the context of neurotransmission.
- Olanzapine is an atypical antipsychotic acting on dopamine and serotonin pathways.
- Procaine, while commonly used as a local anesthetic, blocks sodium channels on nerve endings therefore disrupting neurotransmission.
- In contrast, cilostazol is a vasodilator and anti-platelet compound which is primarily known for treatment of peripheral claudication.
- g:Profiler results highlight this contrast in known mechanism of action between the 3 compounds.



A subcluster of 5 serotonin reuptake inhibitors (SSRIs) of interest in a cluster of 98 compounds are found to be enriched amongst a group of antineoplastic compounds using a large > 1,500 compound dataset of primary patient prostate cancer cells.



Summary:

- AutoHCS elucidates the relationships between cell phenotype and compound pharmacology.
- Within a similar phenotypic space, drug groups of interest are often enriched amongst others compounds of similar and sometimes diverging pharmacological classes.
- AutoHCS enables rapid discovery of divergent compound mechanisms of action converging on a cellular phenotype as well as potential compound off-target effects.

CONCLUSIONS

- AutoHCS can rapidly score compounds in an automated and objective manner for any target phenotype in any cell line.
- AI-based morphological clustering is validated by functional enrichment analysis

Learn More at www.viqiai.com/autohcs