White Paper

Transcriptome-informed Drug Design: A Tale of Six Compounds

HTG

Introduction

Failure of preclinical drug candidates during drug discovery and development is a well-known and expensive problem. One approach to de-risking compounds is the addition of transcriptomic profiling during target identification or selection,[1] compound design,[2] mechanism of action determination,[3] or as part of drug repurposing.[4] Transcriptomic data have two advantages: One, individual profiles reveal underlying biological effects of compound treatment beyond target engagement alone. Second, large-scale databases of ready-to-use profiles and associated analysis tools enable systems biology approaches to these problems at scale.[5,6] We envision enabling a conversation between chemical compounds and their transcriptomic effects that ultimately yields de-risked drug candidate molecules with greater potential for clinical development success, leading to positive patient outcomes.

In a previous study, we demonstrated that transcriptomic profiles generated using the proprietary HTG Transcriptome Panel (HTP) support a transcriptome-informed drug discovery paradigm.[7] Specifically, we demonstrated that closely related compounds were differentiated by their mRNA profiles, and that relevant biological information could be gleaned using the resulting expression data. This first proof of concept study used well-characterized compounds known to inhibit mTOR — itself a highly studied target — which allowed us to compare HTP data to those in public databases, then verify the biological findings using the literature to build confidence in the process. However, while the data compared favorably to previous studies, and indeed even uncovered novel biology, the scope of the first study was limited to transcriptome-based *characterization* of compounds, without a continuation to refinement of compound design based on the transcriptomic data.

In this white paper, we build further toward true transcriptome-informed drug design using the HTP, this time focusing on a small number of early-stage hits against our first therapeutic target of emphasis, here referred to as "HTGT1," an undisclosed RNA modifying protein (RMP). RMPs regulate the modification of cellular RNAs. Over 170 RNA modifications have been described; collectively, these modified RNAs are known as the epitranscriptome.[8,9] Modifications of RNA serve as an additional layer of control for regulation of location, stability, and fate. Numerous studies associate RMPs and/or the epitranscriptome with disease development and progression,[10,11] demonstrating the importance of exquisite control of cellular RNAs. One advantage of an RMP target is that a modified sample preparation process allows the use of the HTP for identification and quantitation of the RNA targets of HTGT1. Between the transcriptomic profiles, modification profiles, and *in vitro* inhibition (IC_{50}) data, each compound may thus be assessed for inhibition, mechanism of action, and downstream effects, prior to *in vitro* efficacy studies. The results can then be used to guide the refinement of the compound design.

The results presented below support our previous observation that changes in the structure of a small molecule meaningfully affect the transcriptomic profile of cells treated with that compound. More importantly, the profiles support an understanding of how these compounds affect the cell, and how the cellular background modulates those effects. Through the examples below, we demonstrate how such insights may be utilized to de-risk early-stage hit compounds.

Experimental Overview: Materials and Methods

Cell treatment with compounds

Six previously identified compounds known to inhibit HTGT1 (HTW0741, HTW0186, HTW0544, HTW0744, HTW0457, and HTW0902) were solubilized in DMSO and used to treat two cancer cell lines, referred to herein as "Cell Line 1" and "Cell Line 2," at 10 μM. DMSO alone was used as a control treatment. Following treatment for 24 hours, each well of cells was harvested and lysed in HTG lysis buffer.

RNA profiling

Transcriptome profiles were generated using HTG EdgeSeq™, a targeted RNA sequencing technology that couples a nuclease protection assay with next-generation sequencing for rapid and accurate RNA quantification. The HTG Transcriptome Panel (HTP) used in this study measures approximately 20,000 protein-coding genes (mRNAs) in the human transcriptome.[12] All samples were run in triplicate, using standard protocols.

QC and data quality

All samples passing quality control were used in downstream analyses. Repeatability was evaluated for sample replicates and inter-plate repeatability was assessed using the DMSO controls. Pairwise comparisons were performed for each replicate type and the Pearson correlation coefficients were calculated. The median correlation for all replicates was 0.98.

Data analysis

Differential expression analysis was performed using the DESeq2 package in R;[13] data from each compound were compared to those of the corresponding DMSO control. An adjusted p-value of less than or equal to 0.05 was used as the significance cut-off for all differential expression analyses presented, including pathway analysis. When absolute $\log_2$ fold changes were used, the cutoff threshold is indicated. PCA plots were generated in JMP. CMap/LINCS data were queried using differential expression data restricted to the L1000 set of genes and using the R package *signatureSearch*.[14] PPathway analysis was performed using iPathwayGuide (Advaita Biosciences) and/or the R package ROntoTools.[15]

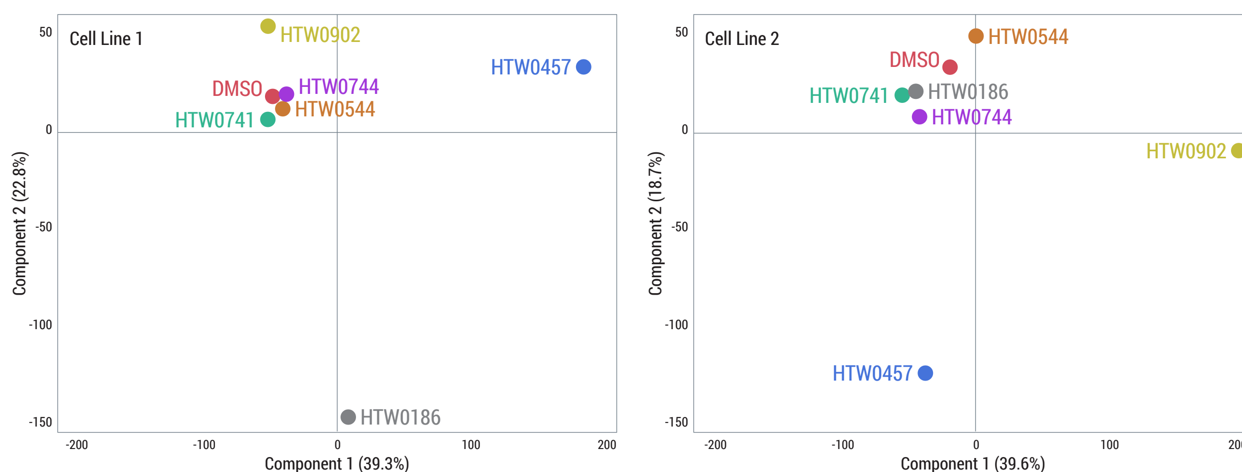


Figure 1. *Principal component analysis (PCA) of expression profiles generated by treatment of cell lines with the indicated compounds or DMSO control. Expression profiles are shown for Cell Line 1 (left) and Cell Line 2 (right).*

Transcriptomic and Modification Profiling

A foundational principle of transcriptome-informed drug design is that different compounds, even when closely related, may be differentiated by their transcriptomic profiles — which in turn represents differences in the resulting cellular perturbation following exposure. Those differences may then be used to uncover relevant biological responses to such perturbation, including specificity, mechanism of action, or adverse/undesirable characteristics of a given compound.

An overall look at the expression profiles using a principal component analysis (PCA) *(Figure 1)* shows that the effect of each compound is indeed distinct, although the magnitude of change differs (see for instance HTW0457 (blue dot) in the panel on the left compared to HTW0741 (green dot)). The relative distances of each compound from the DMSO control correlate well with the number of differentially expressed genes (not shown). A similar analysis was performed for the RNA modification profiles *(Figure 2)*, with a similar correlation between distance and the number of differentially modified genes observed (not shown). These results demonstrate that the downstream effects of the compound–HTGT1 interaction differ, despite each compound being known to inhibit HTGT1 as a primary pharmacologic mechanism. Interestingly, *Figures 1-2* also show that these compounds do not behave identically in the two cell lines tested. We will discuss below why these differences are important to understand.

PCA plots are an excellent complexity reduction tool and may be used to view the landscape of relative differences between points, or to group compounds for classification. To more fully describe the differences between closely related compounds based on structure, though, we use the underlying data, which are far richer and allow for more precise differentiation between similar compounds. We discuss two such similar compound pairs below.

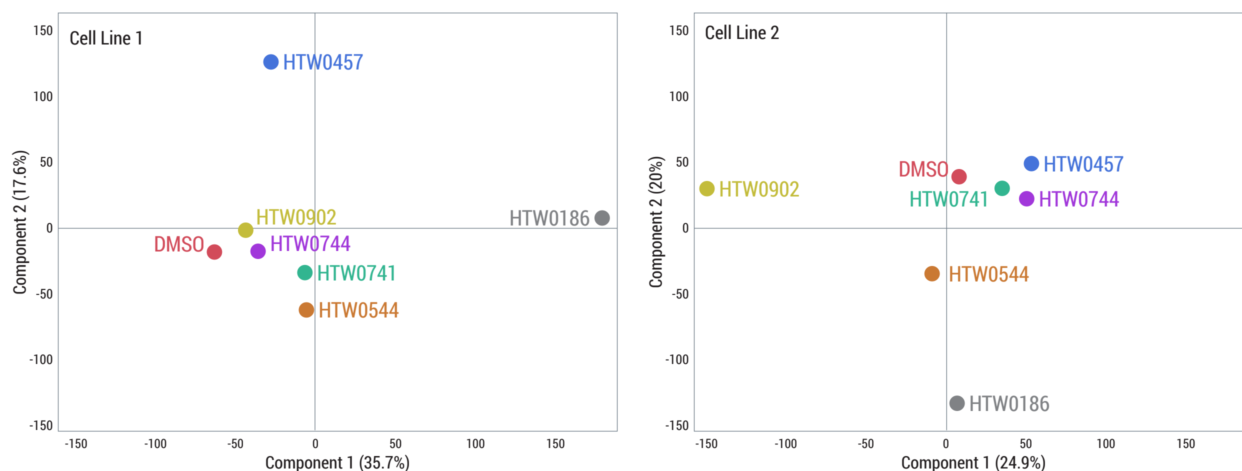


Figure 2. *Principal component analysis (PCA) of RNA modification profiles generated by treatment of cell lines with the indicated compounds or DMSO control. Modification profiles are shown for Cell Line 1 (left) and Cell Line 2 (right).*

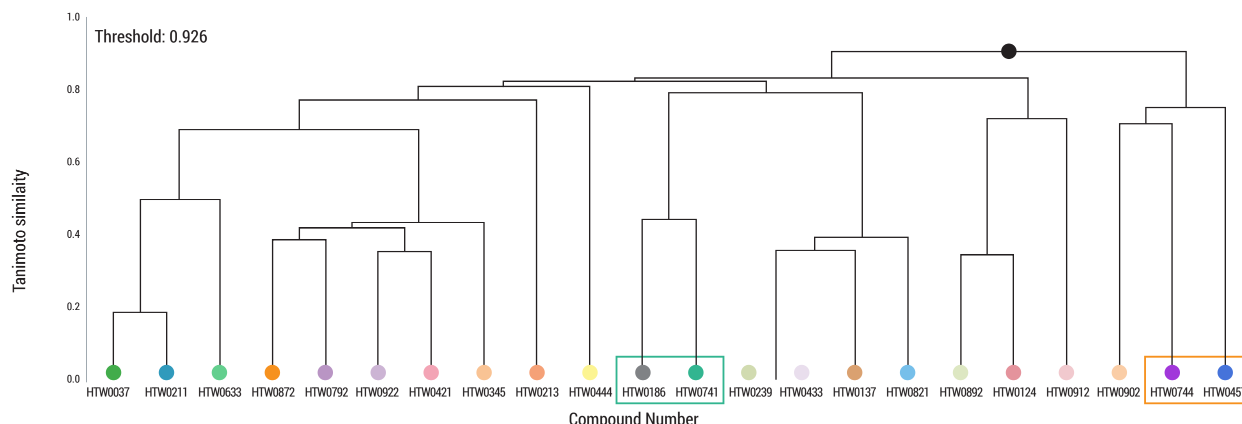


Figure 3. *Tanimoto scores from Morgan fingerprints were used to show the structural similarities of HTW0744/HTW0457 (orange box) and HTW0741/HTW0186 (green box) in comparison to a subset of tested compounds.*

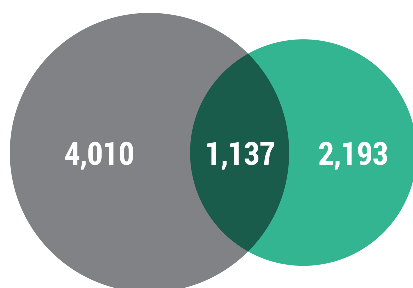
Compounds HTW0741 and HTW0186

Figure 3 displays a similarity analysis of the compounds HTW0741 and HTW0186 *(green box in Figure 3)* using Tanimoto similarity of calculated Morgan fingerprints[16] to compare structural similarities between a group of tested compounds. These two compounds share a core pharmacophore but have functional group substitutions. Both inhibit HTGT1; HTW0186 has the lower IC_{50}. HTW0186 also has the larger number of significant changes to transcriptomic and modification profiles, and the bulk level of modified RNA in samples treated with HTW0186 is significantly changed compared to HTW0741 or control. The number of and overlap between the genes that change in response to these two compounds is shown in *(Figure 4)*. This suggests that while both compounds inhibit HTGT1 as their primary pharmacologic mechanism, they may promote different effects downstream of this inhibition. This supports our thesis that differences in biological effects are represented by differences in transcriptomic profiles, that in turn are unique to differences in chemical structure, and that this effect is observed even when compounds share a common primary pharmacologic mechanism. Overall, the number of differences in profile and biological impact between two similar compounds is striking.

Given that HTW0186 has the lower IC_{50} and more robustly impacted modification levels, we focused on this compound in Cell Line 1 for further investigation. Pathway analysis of HTW0186 flagged the *Apoptosis* (upregulation) and *Cell cycle* (downregulation) pathways as significantly changed. This provides an indication that this compound causes growth arrest in transformed cells, as desired. However, another significantly perturbed pathway is *Regulation of actin cytoskeleton,* which was not observed for any other compound tested. While targeting the actin cytoskeleton has been a therapeutic ideal for many years, it also tends to have undesirable effects.[17] At this stage, the effect of this perturbation is unknown, but the transcriptome information helps guide the next stage of *in vitro* efficacy studies. These could include imaging of the cytoskeleton or compound treatment of primary or muscle cell lines to investigate and characterize this potential effect. The effect of dose and/or treatment time on this pathway could be determined using further transcriptomic profiling. Importantly, these early studies could then guide redesign of the compound structure to avoid an adverse effect; an example of early-stage de-risking via transcriptome-informed drug design.

Figure 4. *Euler diagram summarizing the number of significantly differentially expressed transcripts (compared to the DMSO control) resulting from treatment of Cell Line 1 with HTW0186 or HTW0741.*

● HTW0186
● HTW0741



In summary, the profiles of HTW0741 and HTW0186 support our two themes: structurally similar compounds show a marked difference in their transcriptomic and modification profiles, and pathway analysis of the data uncovered a potential mechanism of action via upregulation of apoptosis and downregulation of the cell cycle, as well as a potential adverse effect that can immediately be investigated further.

Compounds HTW0744 and HTW0457

Compounds HTW0744 and HTW0457 are based on a different pharmacophore than are HTW0741 and HTW0186, but are similarly close to one another, as seen by the fingerprint analysis *(orange box on Figure 3)*. Expression and modification results for both cell lines reveal that HTW0744 has significantly less differential expression and differential modification compared to

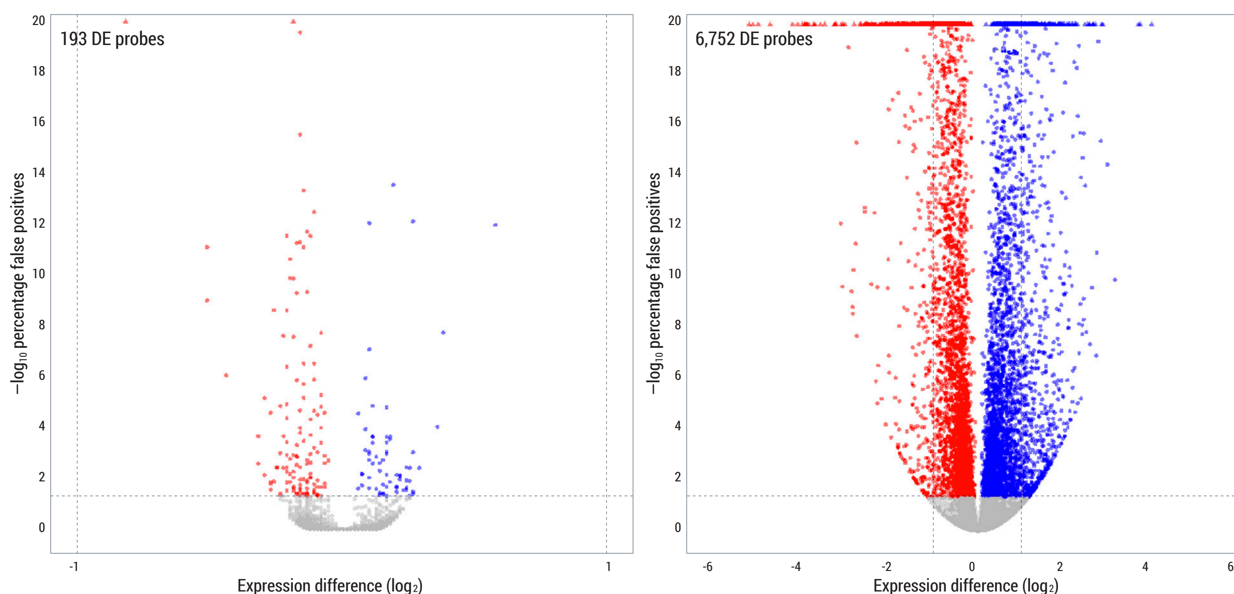


Figure 5. *Volcano plots summarizing the differential expression resulting from treatment of Cell Line 1 by HTW0744 (left) or HTW0457 (right) compared to DMSO control treatment.*

HTW0457. *Figure 5* summarizes the differential expression resulting from treatment of Cell Line 1 with these two compounds using volcano plots, which display the $\log_2$ fold change on the x-axis and the significance of the change on the y-axis; each dot represents a gene that is significantly differentially expressed.

Pathway analysis of HTW0457 indicates *Cell cycle, Apoptosis, Aminoacyl-tRNA biosynthesis*, and *Biosynthesis of amino acids* are impacted in Cell Line 1. The first two are similar to those shown by HTW0186. However, HTW0457 has a larger effect on the cell cycle. Compare, for instance, the two impact analyses of the *Cell cycle* pathway shown in *Figure 6*. Impact analysis uses both the measured differential expression of genes in a pathway (y-axis) and the calculated perturbation of each gene, computed using the pathway topology (x-axis).[18] In the analysis of HTW0457, the key genes *MYC, CCND2, CCND1*, and *CDK4* are downregulated and accumulate negative perturbation. HTW0186 shows downregulation and accumulation only for *MYC*. Additionally, HTW0457 causes strong downregulation of histone gene expression levels, an indication of cell cycle arrest at the G1/S boundary.[7] Therefore, HTW0457 appears to be a potent inhibitor of cell growth in cancer cell lines. Conversely, the investigation of HTW0744 was hampered by the small number of significantly differentially regulated genes. Applying an absolute $\log_2$ fold change threshold of 0.6 returned two (Cell Line 1) or one (Cell Line 2) significantly changed genes. *DELE1* was on both lists, but a single gene gives little upon which to build. Neither of these two compounds had a predicted effect on the actin cytoskeleton, as was seen with HTW0186 (discussed above).

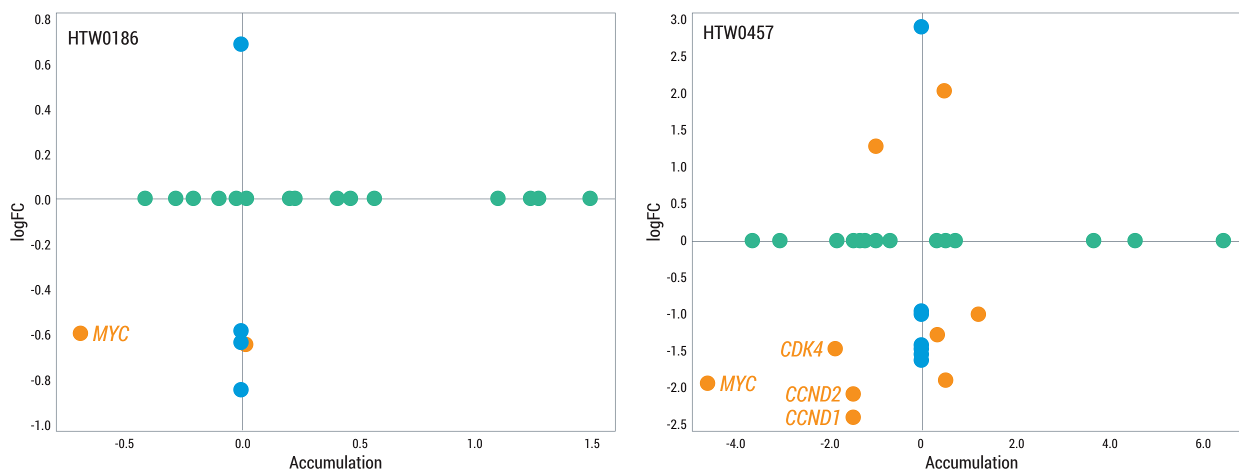


Figure 6. *Perturbation analysis comparing the relative impacts of HTW0186 (left) and HTW0457 (right) to the Cell cycle pathway in Cell Line 1. MYC is indicated on both two-way plots; CDK4, CCND2, and CCND1 are additionally indicated for HTW0457 (right).*

As illustrated with HTW0741 and HTW0186, small changes to a pharmacophore can clearly generate significant differences in expression profiles, and indeed the biological response of the treated cells. The path forward for these four compounds likely will follow an iterative process in pharmacophore modification guided by our profiling model in order to strike a balance between undesirable and desirable characteristics of the candidate molecules. In the case of HTW0744, which had few differentially expressed genes, further characterization could include dose or time responses. HTW0457 will reflex to *in vitro* efficacy studies to determine the potency of its impact on cell growth, and whether this effect is selective for transformed cells (or specific indications). There is also the question of specificity. Specificity for HTGT1 could be approached by the standard method of *in vitro* binding assay using panels of similar targets, but the transcriptome-informed approach is complementary and elegant: compare the changes between a wild-type and HTGT1-knockout cell line.

The Importance of Indication

Previous work using large-scale data sets has shown that the transcriptomic changes caused by a perturbagen in one cell line will not necessarily match those of another,[5] likely due to genetic and mutational backgrounds of immortalized cell lines. De-risking of compounds at the early stage should thus take into account that indication or cell line background may greatly affect the downstream biological changes. An example of such a difference is provided by HTW0902.

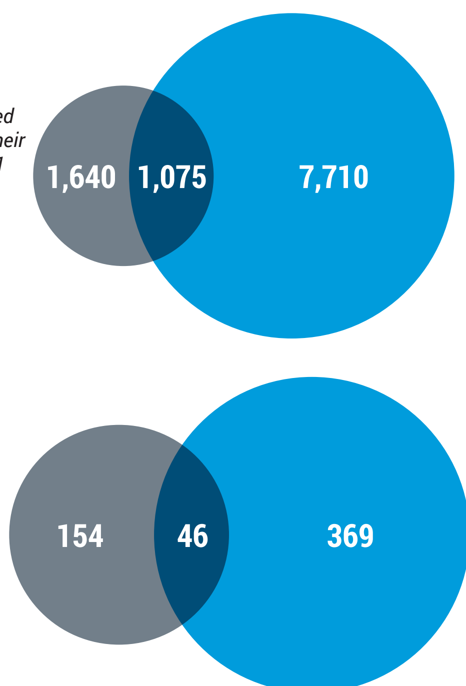


Figure 7. *Euler diagrams summarize the number of significantly differentially expressed (top) and modified (bottom) transcripts (and their intersections) for Cell Line 1 and Cell Line 2 treated with HTW0902.*

- Cell Line 1
- Cell Line 2

Figure 7 represents the differentially expressed and differentially modified genes measured in both cell lines following treatment with HTW0902. Both profile comparisons are similar: Cell Line 2 shows the larger overall number of perturbations and the larger number of specific perturbations. Cell Line 2 additionally has a significant difference in the *bulk* of modified RNA measured compared to the control, an effect not observed in Cell Line 1 (not shown).

While the absolute number of differentially expressed or modified genes is not necessarily an indication of specificity or efficacy, in this case, the same compound and dose are generating quite different magnitudes of effect. Modified transcripts were observed in both cases, which suggests that HTGT1 is indeed affected, as it affects modifications, and therefore no RNA modification changes would otherwise be expected. A likely scenario is that the downstream (or off-target) effects of inhibition of HTGT1 thus differ in these two cell lines. What can transcriptome profiling tell us about these differences and what they mean?

We performed pathway analysis of the two cell lines treated with HTW0902, starting with Cell Line 2. The most significant pathway enriched is *AMPK signaling pathway*. Significantly impacted genes within this pathway include the RNAs that encode subunits of AMPK itself, as well as *RPTOR, ULK1,* and *FOXO*. Indeed, the related *FOXO signaling pathway* is the second pathway on the list, followed by *Cell cycle,* which shows a likelihood of cell cycle arrest occurring via downregulation of *CDK1*. To determine if the cell cycle perturbation was likely to be a primary effect, we used the differential expression profile from Cell Line 2 treated with HTW0902 to query the CMap/L1000 database for similar profiles. The top matching result was SB-218078, a CHK1 inhibitor; Chk1 prevents entry into mitosis through inhibition of Cdk1 activity.[19] A second top result was aloisine, which is a CDK1 inhibitor. Both the pathway analysis and CMap similarities thus support that cell cycle regulation, and the AMPK pathway are likely to be significantly impacted by treatment with HTW0902 in Cell Line 2.

However, in Cell Line 1, treatment with HTW0902 did not result in significant perturbation of the AMPK signaling pathway, nor was an impact to the cell cycle predicted by pathway analysis. This surprising result is explained by the known properties of the cell line: AMPK is already constitutively upregulated/misregulated. Additional upregulation of the AMPK pathway by HTW0902, therefore, makes no impact. The results, while initially surprising, are consistent with the underlying biology.

What is the path forward for HTW0902? An important question is the expected use for this compound. Could the indication-specific results be exploited to treat a particular cancer or disease? The same consideration of specificity to HTGT1 (as discussed above) also applies. But overall, the HTW0902 data are an indication that

transcriptomic profiling of multiple cell lines, either between or within a given indication, has two potential advantages: Overlapping changes may indicate more "universal" changes caused specifically by the compound, and non-overlapping changes may indicate a useful cell line- or indication-specific effect.

Conclusions

In our initial white paper, we investigated two parallel themes, asking whether small changes in the structure of compounds designed against the same target resulted in changes in transcriptomic profiles, and whether these profiles could be used to generate biological insights.[7] The study and data described above build upon this work, showing that small changes in a pharmacophore can have significant and measurable effects on the transcriptomic and modification profiles; that studying those changes can reveal underlying biology; and that inclusion of these data supports and enhances early-stage drug design and development. Our vision is to expand this approach to larger numbers of structurally diverse compounds and introduce a novel platform approach to establish feedback between pharmacophore iteration/optimization and transcriptomic and epitranscriptomic dynamics.

Overall, these data show the advantages of utilizing a transcriptome-informed approach in the early stages of drug design and development. The results of such experiments are expected to enhance current best practices, help with guidance and interpretation of early *in vitro* studies, and inform compound redesign, to ultimately de-risk early-stage drug candidates.

References

1: Baillif B, Wichard J, Méndez-Lucio O, Rouquié D. (2020). Exploring the Use of Compound-Induced Transcriptomic Data Generated From Cell Lines to Predict Compound Activity Toward Molecular Targets. *Front. Chem.* 8, 296.

2: Méndez-Lucio O, *et al*. (2020). De Novo Generation of Hit-like Molecules from Gene Expression Signatures Using Artificial Intelligence. *Nat. Commun.* 11, 10.

3: Ravindranath AC, *et al*. (2015). Connecting Gene Expression Data from Connectivity Map and in Silico Target Predictions for Small Molecule Mechanism-of-Action Analysis. *Mol. Biosyst.* 11, 86-96.

4: Iorio F, Rittman T, Ge H, Menden M, Saez-Rodriguez J. (2013). Transcriptional data: a new gateway to drug repositioning? *Drug Discov.* 18, 350-357.

5: Subramanian A, *et al.* (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 171, 1437-1452.e17.

6: Koleti A, *et al.* (2018). Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res.* 46, D558-D566.

7: Therapeutics WP1, HTG. *(https://www.htgmolecular.com/assets/htg/publications/HTG_Therapeutics_White_Paper_HTP_Proof_of_Concept_C-122-03.28.R0_.pdf).*

8: Van Nostrand EL, *et al*. (2020). A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins. *Nature* 583, 711-719.

9: Schaefer MR. (2021). The Regulation of RNA Modification Systems: The Next Frontier in Epitranscriptomics? *Genes* 12, 345.

10: Jonkhout N, *et al*. (2017). The RNA Modification Landscape in Human Disease. *RNA* 23, 1754-1769.

11: Frye M, Harada BT, Behm M, He C. (2018). RNA Modifications Modulate Gene Expression during Development. *Science* 361, 1346-1349.

12: HTG white papers *(https://www.htgmolecular.com/assays/htp)*.

13: Love MI, Huber W, Anders S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biol.* 15, 550.

14: Duan Y, *et al.* (2020). signatureSearch: environment for gene expression signature searching and functional interpretation. *Nucleic Acids Res.* 48, e124-e124.

15: Voichita C, Ansari S, Draghici S. (2022). ROntoTools: R Onto-Tools suite. R package version 2.24.0.

16: Bajusz D, Rácz A, Héberger K. (2015). Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminformatics* 7, 20.

17: Stehn JR, *et al.* (2013). A Novel Class of Anticancer Compounds Targets the Actin Cytoskeleton in Tumor Cells. *Cancer Res.* 73: 5169-5182.

18: Draghici S, *et al.* (2007). A Systems Biology Approach for Pathway Level Analysis. *Genome Res.* 17, 1537-1545.

19: Enomoto M, *et al*. (2009). Novel Positive Feedback Loop between Cdk1 and Chk1 in the Nucleus during G2/M Transition. *J. Biol. Chem.* 284, 34223-30.

HTG Molecular Diagnostics, Inc.
Call 1-877-289-2615
www.htgmolecular.com

For Research Use Only. Not for use in diagnostic procedures. HTG EdgeSeq and VERI/O are trademarks of HTG Molecular Diagnostics, Inc. Any other trademarks or trade names used herein are the intellectual property of their respective owners. ©2022 All Rights Reserved.