Navigating the Cancer Transcriptome by Decoding Divergent Oncogenic States

Samirkumar B. Amin[1] and Roel G.W. Verhaak[1,*]
[1]The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06032, USA
*Correspondence: roel.verhaak@jax.org
http://dx.doi.org/10.1016/j.cels.2017.08.006

A new approach decomposes aberrant signaling mediated by an oncogenic mutation into underlying core cellular states that may be more permissive to available therapeutic options.

Genomic alterations such as KRAS mutation deregulate a transcriptional network rather than one signaling cascade, making it difficult to compare these effects across cellular states and lineages. In this issue of *Cell Systems*, Kim et al. (Kim, 2017) present Onco-GPS, an analytical approach that decomposes the effects of a single oncogenic mutation into multiple transcriptional components. Using semi-supervised clustering, these components are grouped into divergent and potentially functional states. These states are associated with specific signaling pathways, allowing the oncogenic effects of the seed mutation to be evaluated across different cell types. Onco-GPS may also be used to infer potentially relevant genetic and pharmacological perturbations.

Recent advances in high-throughput functional genomics and drug discovery have resulted in improved and expanded numbers of targeted therapies aimed at recurrent driver mutations, such as the BRAF V600E mutation in melanoma and EGFR or ALK abnormalities in lung cancer. However, tumor cells can adapt to such targeted therapies, posing an obstacle to the implementation of precision medicine. For example, patients harboring cells with the BRAF V600E mutation show variable response to monotherapy with BRAF inhibitor Vemurafenib (Hyman et al., 2015). Progressive loss of sensitivity is attributed to the selection of subclonal cell populations due to both primary and acquired modes of resistance, which in turn promotes maintenance of downstream oncogenic pathways independent of mutant BRAF activity (Rizos et al., 2014). Identifying the maintenance pathways or cellular states under a selective upstream driver mutation is of critical value in understanding how to overcome therapeutic resistance to targeted therapies.

The work of Kim et al. offers a promising integrative analytical approach to decompose the transcriptional circuitry downstream of an oncogenic genomic alteration into its transcriptional principal components, equivalent to oncogenic states. These predicted oncogenic states, such as WNT or NF-kB activation, provide insights into potential vulnerabilities for targeting common driver mutants that are otherwise "difficult to drug" (Dang et al., 2017), such as KRAS and MYC mutations.

Kim et al.'s method is aptly named Onco-GPS, for oncogenic positioning system. It associates tumor samples, or model-system-derived specimens, with a shared upstream driver event (GPS receiver) near the location of an underlying dominant oncogenic state (satellite). Such placement is based on overlapping enrichment of that state within the transcriptional profile of the sample of interest (Figure 1). Onco-GPS uses a three step approach. Steps 1 and 2 create a "map" of cellular states affected by a particular mutation. Then in Step 3 this map is used to "position" new samples harboring the mutation.

In step 1, isogenic cell lines are engineered with the mutation using CRISPR-Cas9 and then analyzed by gene expression profiling. Analyses by the authors of these cell lines carrying oncogenic BRAF and KRAS mutations revealed a non-linear relationship between upstream driver events and multiple downstream pathways, indicative of heterogeneous and context-dependent driver mutation transcriptional programs.

In step 2, Kim et al. model the effects of transcriptional deregulation by analyzing the top 1,000 most differentially expressed genes in the isogenic cell lines. They extract a matrix of the expression values of these 1,000 genes in 750 cancer cell lines in the Cancer Cell Line Encyclopedia (CCLE) dataset (Barretina et al., 2012). Then, this matrix is decomposed using non-negative matrix factorization, which groups tightly correlated genes into several sparse components. As the Cancer Cell Line Encyclopedia cohort is comprised of cell lines from 25 solid tumor types, these components represent not only the mutant-specific core signature but also relate to the lineage-specific context and underlying activity of signaling pathway(s). Next, the components are clustered and annotated with their potential cellular or functional states based on genomic features derived from multiple molecular characterization databases, e.g., MSigDB, copy number, mutation, and protein expression data from the Cancer Cell Line Encyclopedia, RNAi-based genetic dependency from Project Achilles, and drug sensitivity data from Cancer Cell Line Encyclopedia and the Cancer Therapeutics Response Portal (Seashore-Ludlow et al., 2015). The resulting states represent likely activation of divergent pathways, for example, activation of epithelial mesenchymal transition with MYC versus NF-kB.

In step 3, new samples carrying the specific mutation, such as those derived from a patient's tumor biopsy, are assigned to the most relevant cellular state, on the basis of their expression profile. The resulting positioning is visualized in a "ternary" graphical layout, which projects one or more mutant samples onto a reference map based on the Cancer Cell Line Encyclopedia database (Figure 1). Besides its visual appeal, the ternary plot allows samples sharing common transcriptional states to be identified.

Kim et al. demonstrate the utility of Onco-GPS by using it to infer cellular

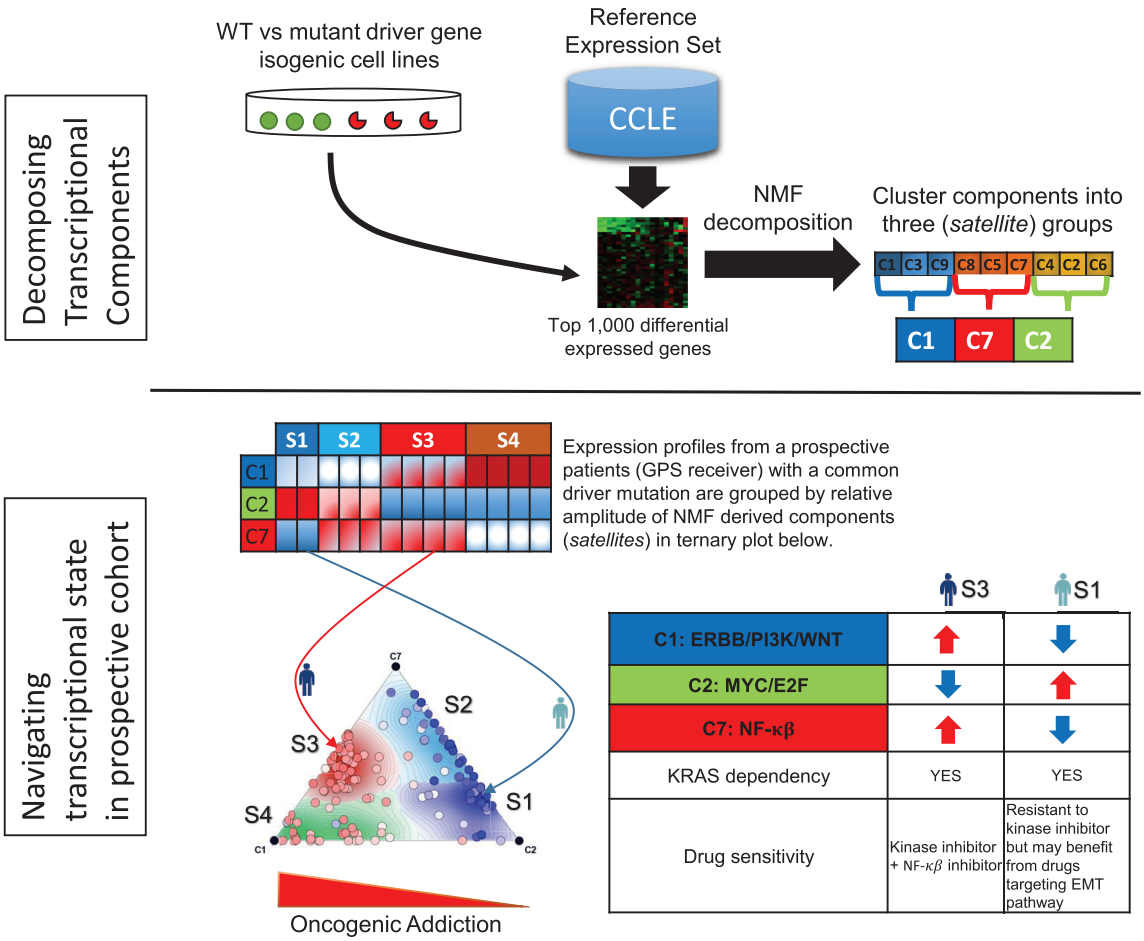

Figure 1. Onco-GPS Workflow and Potential Therapeutic Applications

(Top) The Onco-GPS approach starts with the top 1,000 genes differentially expressed between wild-type and mutant oncogene expressing isogenic cell line cultures. These genes are projected on a Cancer Cell Line Encyclopedia reference expression map to derive stable transcriptional components, each representing a core signaling pathway based on gene set enrichment across several databases for molecular characterization. These components are then subjected to semi-supervised clustering to derive three core groups (analogous to satellites in GPS), each reflecting predicted underlying functional or cellular states with one or more dominant active pathways.

(Bottom) Transcriptomic profiles from biospecimens with a shared oncogenic mutation (e.g., KRAS G12V mutation) can be projected in the four-state model, thereby homing the sample to its closest underlying cellular state. The ternary plot in the bottom panel shows four colored zones reflecting different cellular states, each with activation of different pathways and varying sensitivity (oncogene addiction) to inhibitor of upstream driver event.

states downstream of BRAF V600E mutations and KRAS G12V mutations. They identify three components which are involved in core RAS signaling, MAPK, and cytokine signaling (Figure 1, components C3, C6, and C7, respectively). This analysis highlights the importance of extracting oncogenic states defined by a somatic driver mutation, in addition knowing the mutation status prior to starting targeted monotherapy or combination regimes. Further, the authors propose that mutually exclusive components, such as C3 and C6, represent transcription programs related to oncogene addiction. Cancer cells may escape from such addiction by transitioning to an alternate expression program, thus rendering resistance to targeted therapies. Although further validation is needed, tools like Onco-GPS represent valuable resources enabling the identification of such transient states, thereby ultimately optimizing the efficacy of targeted, oncogene-inhibiting therapies.

Therapy resistance frequently emerges following treatment with targeted agents, showing that optimal personalized therapies cannot be based on driver mutations alone but also require an understanding of the tumor context, for example, the underlying transcriptional or epigenetic state and tumor microenvironment. Methods such as Onco-GPS, or Driver-gene Inference by Genetical-Genomic Information Theory (DIGGIT) (Chen et al., 2014), that account for the underlying mode of oncogen activation and downstream pathways can potentially facilitate discovery of treatments with increased efficacy, ultimately resulting in improved cancer-free survival.

Several key issues need to be addressed to improve the reproduciblity and utility of analytical tools such as Onco-GPS. For instance, it would be valuable to have community efforts to improve predicted cellular states and drug sensitivities for new samples, using orthogonal functional and drug screening data (Iorio et al., 2016). In addition, a web portal

showing both descriptive (perturbed pathways) and performance metrics (classifier accuracy) for frequent driver mutations from the Cancer Genome Atlas dataset would be a useful resource in target identification and drug discovery processes. Finally, inferring intricate oncogenic transcriptional states could benefit from an integrative framework that accounts for (1) epigenetic changes that result in transcriptional plasticity and drug resistance (Flavahan et al., 2017), and (2) that harnesses single-cell data to facilitate the measurement of clonal diversity and crosstalk between cancer cells and tumor microenvironment (Datlinger et al., 2017). Nevertheless, the ability of Onco-GPS to delineate the transcriptional effects of tumor-driving mutations provides an important step forward in defining the precision of our medicines.

REFERENCES

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). Nature *483*, 603–607.

Chen, J.C., Alvarez, M.J., Talos, F., Dhruv, H., Rieckhof, G.E., Iyer, A., Diefes, K.L., Aldape, K., Berens, M., Shen, M.M., and Califano, A. (2014). Cell *159*, 402–414.

Dang, C.V., Reddy, E.P., Shokat, K.M., and Soucek, L. (2017). Nat. Rev. Cancer *17*, 502–508.

Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Nat. Methods *14*, 297–301.

Flavahan, W.A., Gaskell, E., and Bernstein, B.E. (2017). Science *357*, 357.

Hyman, D.M., Puzanov, I., Subbiah, V., Faris, J.E., Chau, I., Blay, J.-Y., Wolf, J., Raje, N.S., Diamond, E.L., Hollebecque, A., et al. (2015). N. Engl. J. Med. *373*, 726–736.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). Cell *166*, 740–754.

Kim, E. (2017). Cell Syst. *5*, this issue, 105–118.

Rizos, H., Menzies, A.M., Pupo, G.M., Carlino, M.S., Fung, C., Hyman, J., Haydu, L.E., Mijatov, B., Becker, T.M., Boyd, S.C., et al. (2014). Clin. Cancer Res. *20*, 1965–1977.

Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Cancer Discov. *5*, 1210–1223.

When Seeing Isn't Believing: How Math Can Guide Our Interpretation of Measurements and Experiments

Paul Macklin[1,*]
[1]Intelligent Systems Engineering, Indiana University, Bloomington, IN 47408, USA
*Correspondence: Paul.Macklin@MathCancer.org
http://dx.doi.org/10.1016/j.cels.2017.08.005

Mathematical thought experiments probe the meaning and pitfalls of experimental measurements and demonstrate that caution is in order when measuring heterogeneity.

As powerful new technologies allow us to probe single-cell dynamics in high throughput, are we at risk of reaching the wrong conclusions? An intriguing paper by Schumacher, Maini and Baker in this issue of *Cell Systems* uses a mathematical model of collective cell migration—a 21st-century computational thought experiment—to examine the potential pitfalls of measuring heterogeneity (Schumacher et al., 2017). In the process, they highlight the potential for mathematical modeling to help us better understand the meaning and limits of experimental measurements.

Cellular heterogeneity has emerged as an important factor in studying clonal dynamics in cancer, predicting therapeutic response, and assessing the limitations of molecular profiling (e.g., Fisher et al. (2013)). An accurate appreciation of heterogeneity is clinically important: undersampling a heterogeneous tumor, for example, runs the risk of making treatment decisions from unrepresentative subclones or missing aggressive clones altogether. Cellular heterogeneity allows a cell population to explore the space of possibilities, increasing the odds of finding a strategy that circumvents a treatment. Hence, there is an increasing interest in quantitating cellular heterogeneity at a variety of levels, including genomic, metabolic, motility, and other characterizations.

Single-cell molecular analyses and single-cell tracking in time course experiments have opened fascinating new possibilities in examining cell-to-cell variability, but we need to develop quantitative measures of heterogeneity to make full use of these data and to relate heterogeneity to disease phenotypes and clinical outcomes. Essentially, this is a challenge of relating molecular-level heterogeneity within the genome or proteome to observable variations in cellular behavior. One behavior that has received much attention is cell motility because it is clinically important for understanding metastasis and can be directly quantified. For example, the peak delay time—the time difference at which two cells' velocities are most aligned—has emerged as a metric to quantitate collective cell migration; measure this across a migrating collective of cells, and we can quantitate heterogeneity in collective cell migration (Sharma et al., 2015).