

## Supplementary Issue: Network and Pathway Analysis of Cancer Susceptibility (B)

### The Axis of Progression of Disease

Alan M. Tartakoff<sup>1</sup> and Di Wu<sup>2</sup>

<sup>1</sup>Pathology Department and Cell Biology Program, Case Western Reserve University, Cleveland, OH, USA. <sup>2</sup>Statistics Department, Harvard University, Cambridge, MA, USA.

**ABSTRACT:** Starting with genetic or environmental perturbations, disease progression can involve a linear sequence of changes within individual cells. More often, however, a labyrinth of branching consequences emanates from the initial events. How can one repair an entity so fine and so complex that its organization and functions are only partially known? How, given the many redundancies of metabolic pathways, can interventions be effective before the last redundant element has been irreversibly damaged? Since progression ultimately proceeds beyond a point of no return, therapeutic goals must target earlier events. A key goal is therefore to identify early changes of functional importance. Moreover, when several distinct genetic or environmental causes converge on a terminal phenotype, therapeutic strategies that focus on the shared features seem unlikely to be useful – precisely because the shared events lie relatively downstream along the axis of progression. We therefore describe experimental strategies that could lead to identification of early events, both for cancer and for other diseases.

**KEYWORDS:** disease progression, cancer

**SUPPLEMENT:** Network and Pathway Analysis of Cancer Susceptibility (B)

**CITATION:** Tartakoff and Wu. The Axis of Progression of Disease. *Cancer Informatics* 2014;13(S6) 7–13 doi: 10.4137/CIN.S17683.

**RECEIVED:** June 4, 2014. **RESUBMITTED:** July 9, 2014. **ACCEPTED FOR PUBLICATION:** July 10, 2014.

**ACADEMIC EDITOR:** JT Efir, Editor in Chief

**TYPE:** Perspective

**FUNDING:** The authors thank the NIH for grant R01GM089872. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

**COMPETING INTERESTS:** Authors disclose no potential conflicts of interest.

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

**CORRESPONDENCE:** [amt10@case.edu](mailto:amt10@case.edu)

This paper was subject to independent, expert peer review by a minimum of two blind peer reviewers. All editorial decisions were made by the independent academic editor. All authors have provided signed confirmation of their compliance with ethical and legal obligations including (but not limited to) use of any copyrighted material, compliance with ICMJE authorship and competing interests disclosure guidelines and, where applicable, compliance with legal and ethical guidelines on human and animal research participants. Provenance: the authors were invited to submit this paper.

### Introduction

Diseases result from one or more forms of “stress.” In some cases, the stress is best described as environmental, while in others the instigator is genetic stress, that is, one or more mutations. It is commonplace for both forms of stress to contribute. Especially in the many cases for which the underlying cause is unknown, the identification of chinks in the armor of disease and selection of satisfactory therapeutic targets present a daunting challenge of broad significance. The following comments are generally relevant to cancers, as well as for other diseases.

Forms of cancer that show simple inheritance should be contrasted to those that appear to be of multigenic origin or to be sporadic. Unfortunately, only a minority of cases exhibit simple inheritance. These prototypes are instructive and important, but do not begin to account for the full scope of disease.

Although evolution has certainly contributed to mitigating severe forms of malignancy, the late onset and low incidence of most cancers place them in a chaotic realm that is largely outside of evolutionary improvement. Moreover, the fine-tuning that would seem desirable in order to limit expression of deleterious proteins is often not feasible: too many of the key players function in conjunction with multiple targets. Indeed, this issue lies at the heart of understanding the evolvability of organisms. If all control networks were separate from each other, specificity of regulation could be exquisite; however, the size of the corresponding genome or transcriptome would need to be vast.

### Progression Through States

It is plausible to conceive of the healthy cell as being in a dynamic “*status quo*,” for which many aspects of prevailing physiology



fluctuate. Examples of metabolic fluctuations are provided by studies in which fluorescent reporters allow cell-by-cell scrutiny of single transcripts or their products in real time. The causes of these fluctuations are often hard to pin-point; however, transcription is subject to stochastic variability of the concentration and localization of key regulatory factors.<sup>1-4</sup> Frequent adjustments of the levels of many metabolites and proteins are surely characteristic of all cells. Some of these adjustments may be homeostatic, while others may be destabilizing.

In the simplest model, progression of a healthy cell toward disease involves a linear sequence of intermediates, and culminates in changes that are responsible for overt symptomatology, which can coincide with entry into a terminal state, for example, complete lack of growth control or death (Fig. 1, upper rectangle).

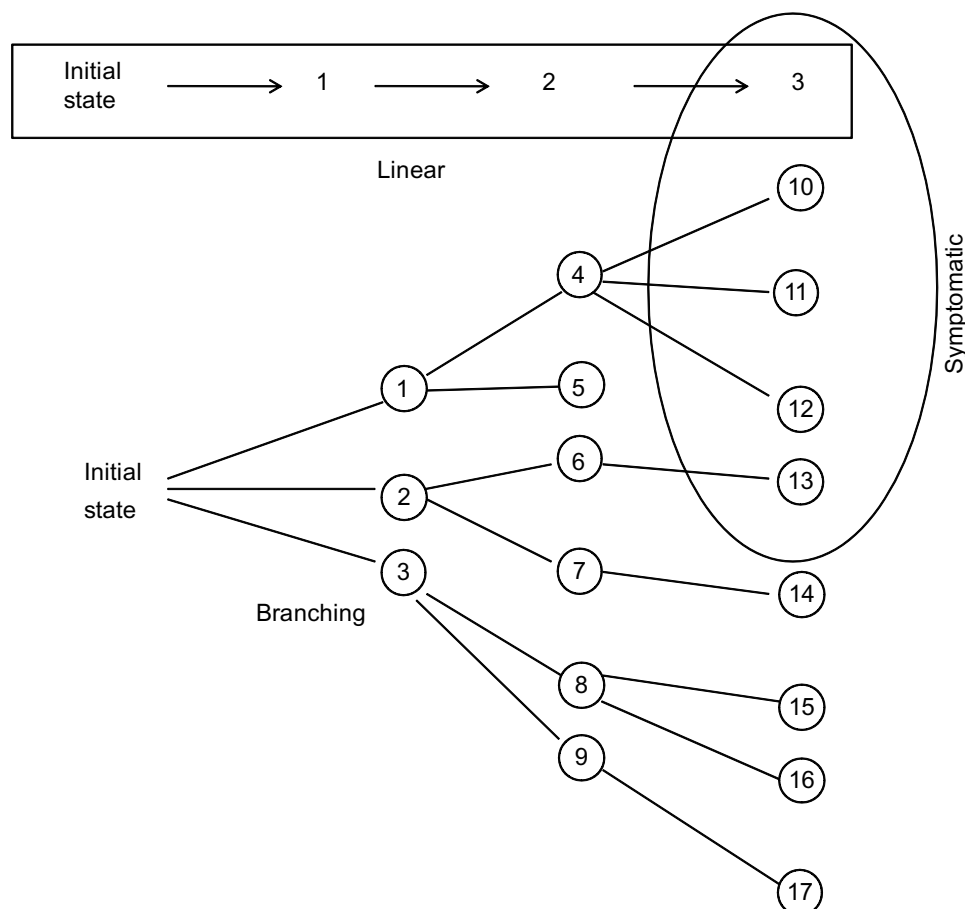
In reality, most primary molecular changes that are triggered in disease seem likely to have multiple downstream repercussions (Fig. 1, lower), reflecting widespread interdependence of the sort that is conspicuous in transcriptional profiling of cells in which a single gene has been silenced or overexpressed. The resulting branching cascades obviously become extremely complicated, especially if feed-forward

events and interactions between temporally separated events occur. Branching cascades define composite perturbed states for the cell. By including changes quite distinct from those that were first present, they can dramatically alter and amplify symptoms. They can readily be misleading with regard to identification of events of causal significance.

In cancer genome research, it is useful to discriminate between mutations that “drive” the disease and genes that carry “passenger” mutations, which can result from secondary genetic accommodations.<sup>5,6</sup> Recent progress along these lines has been achieved by large-scale comparison of exome sequences of tumors and matched normal samples from the same individuals, for example, using samples from The Cancer Genome Atlas and International Cancer Genome Consortium. In addition to the published databases that have enumerated somatic mutations for different cancer types,<sup>7</sup> the saturation analysis of cancer genes across 21 tumor types has allowed identification of additional somatic mutations that are associated with cancers.<sup>8</sup>

### Commitment Points and the Point of No Return

The cumulative impact of initiating events and/or their combination with others can cause what began as an inconsequential



**Figure 1.** Axes of progression leading to pathogenesis.

**Notes:** The rectangle at the top encloses a linear pathway, with the axis of causal progression proceeding from left to right. The lower part of the figure diagrams branching paths in which many changes occur, some of which are ultimately symptomatic. A subset of these contributes to loss of growth control or to cell death. In this diagram, each of several early events (1, 2, and 3) is shown as being linked to a subset of downstream events.

or meta-stable perturbation to progress to a “commitment point,” signifying that the cell or organism can no longer readily return to its initial condition. At the organismal level, an example is that of cells that have already lost one functional allele of a tumor suppressor or – if already malignant – have entered the circulation and therefore gained wide access to the body. Other examples include those discussed in several recent overviews.<sup>9–11</sup> Once cells are “trapped” in such a state, they would be all the closer to a point that allows them to be pushed toward a terminal state. The stochastic nature of some such events, and their low probability, could critically account for much of the variability of the timing of symptoms.

At a later point in progression, it is useful to think of arrival at a “point of no return,” which leads to major incapacitation of the cell (Fig. 2). By definition, this second critical transition is also irreversible. Beyond the point of no return are terminal events that often furnish a characteristic metabolic or histologic signature of disease. This signature is likely to be far removed from the initiating circumstances.

It is often difficult to discriminate between the commitment point and the point of no return. Nevertheless, efforts to identify driver and passenger genes in cancer genome studies seek to target genes at these two stages, with the intent of using them as predictive biomarkers. Among these biomarkers for individual diagnosis are *KRAS* mutations in metastatic colorectal cancer, *EGFR* mutations in advanced non-small cell lung cancer, and *BRAF* mutations in metastatic malignant melanoma. Prognostic biomarkers of value after the point of no return can be identified through analysis of the recurrence risk stratification using the OncotypeDx and Mammaprint gene expression signatures in breast cancer.<sup>12</sup>

Diseases of monogenic causation provide a simplified prototype for reasoning. Yet the struggle against many diseases is fundamentally distinct from the game-like staged challenges of simplified experimental models. Only in exceptional cases do we know the initial provocateur and even in this situation,

there is every reason to expect that multiple genetic and/or environmental factors contribute to progression and outcome.

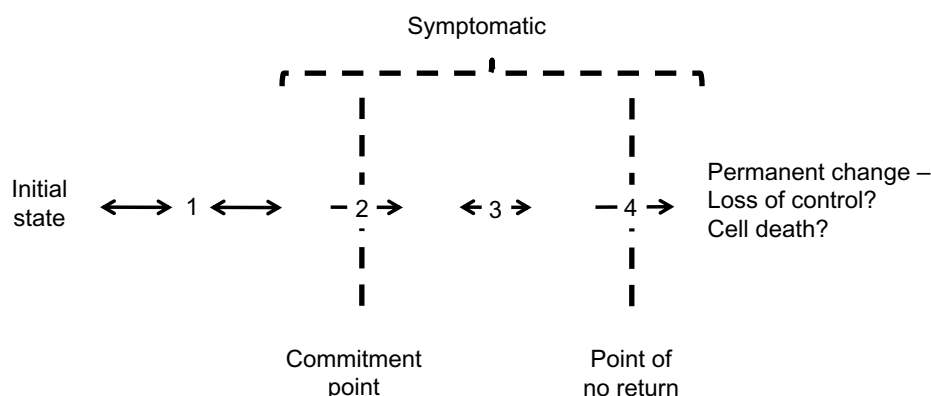
It is instructive to compare this situation to the notoriously high complexity of chess matches in which players start from fixed positions and are allowed access to only 64 positions. Even though the beginning of each match appears to be perfectly balanced, the winner can be different in successive matches between the same opponents. By comparison, in disease progression even the number of interacting elements and the equivalent of their initial positions are generally unknown.

### Contributions from Neighboring Cells

Overt symptomatology at the level of the organism results from collective dysfunction of more than a critical number of cells. The multicellular nature of organs can buffer the physiologic consequences of changes in single cells. For example, neighboring cells in a given tissue can sustain their neighbors, both by providing extracellular nutrients and growth factors, and also *via* junctional complexes that allow exchange of low molecular weight constituents. Furthermore, extracellular factors can be critical, for example, immune and inflammatory mediators, metalloproteases that facilitate cell migration, etc. Therefore, any full understanding of disease progression will require cell-specific information for multiple cell types, and ultimately such information needs to be obtained in the intact organism.

### Differences Between Monogenic, Polygenic, and Environmentally Caused Disease

Linear and branching models of pathogenesis are relevant to the comparison of therapeutic options for diseases of monogenic, polygenic, or environmental origin. Adding to this complexity is the realization that different mutations in a single gene can sometimes lead to a broad range of seemingly distinct conditions.<sup>13–16</sup> Moreover, the issue of polygenic causation itself,



**Figure 2.** The commitment point and the point of no return.

**Notes:** As illustrated, the initial molecular changes are reversible without being symptomatic. Later changes eliminate the possibility of return to the initial state, for example (3) cannot revert to (2), thereby defining a commitment points. Still later changes are irreversible and lead to loss of growth control or to death.



although surely at least as complex as monogenic causation, defies generalization since in most cases polygenic causation is a hypothesis rather than an established fact.<sup>17,18</sup>

To understand how polygenic causation may work, bioinformatics/biostatistical tools have increasingly been focused on regulatory networks that make it possible to integrate multiple levels of genomic data from tumors.<sup>19–21</sup> Other analytic tools also suggest that significant pathways or sets of genes work together.<sup>22–24</sup>

### Therapeutic Target Priorities

What are the implications of these reflections for the choice of therapeutic targets? For diseases in which the key mutation is in an enzyme or receptor, rational design of active-site ligands can be enormously effective (Gleevec, Herceptin, Vemurafenib, etc.).<sup>25,26</sup> Furthermore, datasets based on high throughput drug screens of cancer lines, for example cMap, often can suggest which drugs or compounds will be most effective.<sup>27</sup>

Moreover, if the mutation is in an identified protein of unknown function, or if the normal protein is altogether unimportant, gene knock-out or RNAi-based strategies could ultimately be successful. If elimination of the normal protein is itself deleterious, it would, on the other hand, be necessary to replace the mutant copy with a normal copy or, perhaps, to silence only the mutant copy.<sup>28</sup>

For disorders of more complex causation, the value of attempting to correct any identified changes depends critically on their position along the axis of progression. Critical targets include those that include a feed-forward feature or those that control passage beyond a commitment point. Targets that perform the ultimate *coup de grace* in unleashing uncontrolled growth are less likely to be optimal.

### Natural Indicators of Therapeutic Options

Faced with the difficulty of identifying early events on a causal pathway that leads to pathogenesis, it could be valuable to focus on any candidate modifier genes (eg, identified through association with single nucleotide polymorphisms) that correlate with outcomes.<sup>29</sup> This strategy can be directly extended to investigation of model organisms with distinct genetic backgrounds, for example, different inbred strains of mice,<sup>30</sup> and animals with engineered genomes.<sup>31</sup>

A further important consideration is the cell type or tissue specificity of disease. For example, both upon transplantation and during metastasis, many cancers are known to flourish only at selected sites. In principal, these divergences provide an opportunity: Unaffected or less-affected tissues could express protective factors. Alternatively, affected tissues could express factors that sensitize them. Moreover, since cancer of any one tissue often comprises several distinct molecular cancer subtypes, distinct therapies may be required for different tumor subtypes, as in breast cancer and lung cancer.<sup>32</sup>

There is a central distinction between modifiers identified in populations and factors identified in varied cell types of

the same individual. Modifiers presumably are mostly allelic variants among naturally occurring polymorphisms. By contrast, factors that characterize varied cell types (or ages) largely reflect differences of expression of products of the same genes.

### Random Screens and Selections

Given the many molecular features that can distinguish normal cells from malignant cells, it is not obvious which aberrations could become therapeutic targets. Many such features could be entirely secondary, while others – although close to the axis of disease progression – could be so inextricably linked to other vital processes that their manipulation is fool-hardy.

As a complement to classical genetic studies of animals or random mutagenesis, available libraries of drugs, cDNAs, or shRNAs/siRNAs make it possible to explore the impact of near-random groups of single agents on cell-culture-based models of disease. These strategies can either test single candidates separately, or – for the nucleic acid-based strategies – pool thousands of candidates and then recognize and pursue the phenotypic consequences of those that are shown to be effective.<sup>33,34</sup> In the simplest case in which a single, well-defined molecular target exists, one might expect all effective drugs or DNAs/RNAs to be recognizably related to each other. Alternatively, they could appear unrelated yet (a) perturb distinct sites on the same molecular target or (b) perturb components that function upstream or downstream of that target. As a first approximation, the possibility of their affecting the same target can be assessed by inquiring whether the simultaneous use of more than one agent increases efficacy.

Given the often incomplete specificity of corrective agents and their association with secondary effects, it seems reasonable to anticipate that effective molecular therapies will require combinatorial approaches. One strategy to identify pairs of agents could begin with a candidate that is helpful and use it as an “anchor.” Secondary screens or selections can then be conducted with the first agent already in place. Combinatorial options for which no experimental procedure presently exists are those for which the single agents do not by themselves affect phenotype. Examples of such effective combinations likely exist among the genetic background effects that are characteristic of outbred populations.

### Diseases with Fractional Genetic Linkage

Diseases are initially classified according to phenotype, emphasizing terminal characteristics. With the realization that many diseases with a characteristic terminal phenotype do not show uniform genetic linkage, their analysis becomes highly complex, poses therapeutic difficulties, and raises problems of nomenclature. In diseases for which no more than a fraction of cases share a given genetic linkage, it is reasonable to suppose that distinct events can be initiators and that their effects ultimately converge on similar outcomes (Fig. 3). A good example is that of amyotrophic lateral sclerosis. Here, mutations of multiple distinct genes – even

though they seem quite unrelated to each other (TDP-43, FUS/TLS, SOD1) – can account for the same ultimate phenotype.<sup>35,36</sup> Fractional linkage is also characteristic of Alzheimer's Disease, for which only a small minority of cases are inherited.

The most valuable therapeutic targets are those that lie relatively early along the axis of progression; however, in cases of fractional linkage – since distinct events initiate progression – early events surely differ from one example to the next. Since later events are increasing likely to lie downstream of a point of no return, one can only hope that the ultimate intersection of physiologic changes is not limited to late events.

The implications of fractional linkage for therapy development are sobering in the context of the development of genetically based animal models of disease. If an animal model is based on phenotypic similarity rather than on an orthologous underlying mutation, understanding of the phenocopy seems unlikely to be sufficient.

### Progression Signatures and the Axis of Time

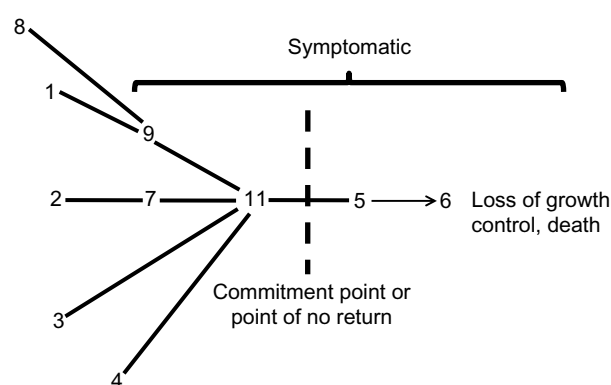
#### Prospective

To identify predictive biomarkers, one interrogates selected tissues, cell types, or fluids biochemically,<sup>29,37</sup> both from individuals who will remain healthy and from those who later will exhibit a disease characteristic. One then looks empirically for single parameters or conjunctions of parameters that correlate with outcome. For example, scrutiny of transcriptional profiles can allow subclassification of cancers, prediction of their progression, and response to therapeutic regimens.<sup>38</sup> Classical biomarkers are collected at a single time point; however, in principal, they could define a chronology of change at a succession of time points. Biomarkers in general are not causal precursors of the outcome.

For diseases that are known to be of simple causation, a directed experimental strategy could be used to search for biomarkers (Fig. 4, upper). Thus, one could activate a single oncogene using cells in culture or a model organism and then monitor the successive appearance of biochemical or transcriptional changes (a, b, c, etc. in Fig. 4, upper). If the simulation generates a sufficiently distinctive “progression signature,” single or composite early changes that are characteristic of the condition under study should provide useful biomarkers.

As an extension of this strategy, one could ask whether any potential biomarker lies along the causal axis of pathogenesis, as opposed to being irrelevant bystanders. This would involve opposing individual changes (a, b, c, etc.) – so long as indirect consequences are tolerable – and then inquiring whether progression still occurs. The search for such markers could be conducted with model organisms which had been engineered to express the oncogene in question.

Since BRCA1/2 mutation carriers tend to develop especially aggressive breast tumors, BRCA1 is often considered a prospective biomarker.<sup>39</sup> Ongoing comparative exome sequencing of germline and tumor samples for cancer genomes will aid identification of further biomarkers for prediction of cancer risk.



**Figure 3.** Phenotypically similar diseases can result from multiple causes.

**Notes:** Distinct causes (1, 2, 3, 4, and 8) lead to shared events (5, 6, and 11). If the shared events (eg, 5, 6) lie beyond a commitment point or point of no return, therapies that target them will not reverse earlier events.

#### Retrospective

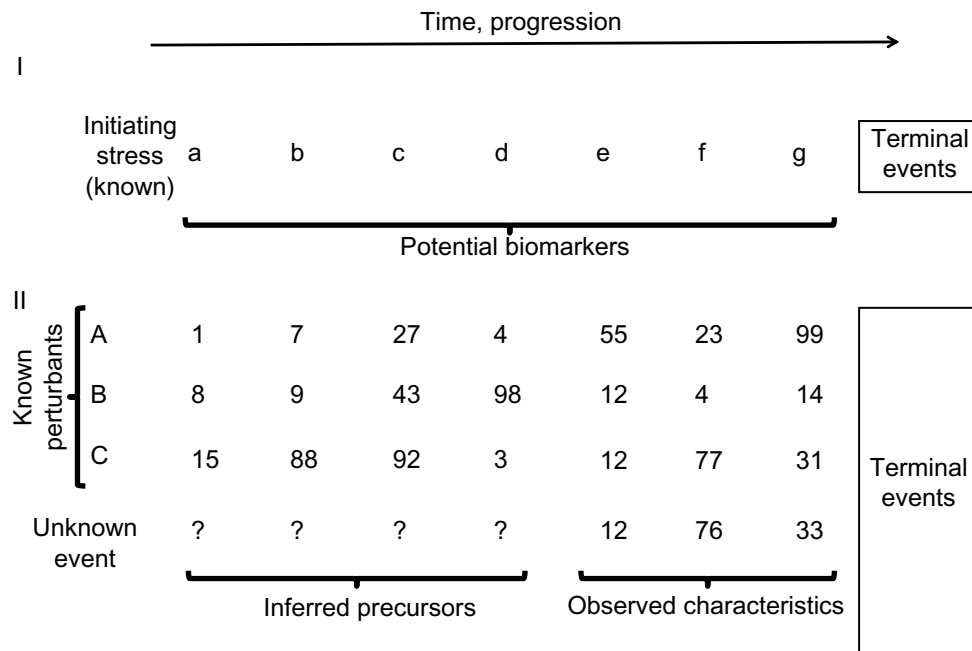
When confronted with a recurrent condition of unknown etiology, one must learn how to combat both precursor events and progression. We suggest that an interpolation strategy could be used to identify early targets of functional significance.

In interpolation strategies, one compares the state of an unknown condition to a reference dataset (eg, transcriptional profiles) obtained after treating the same cell type with panels of drugs, shRNAs, etc., or expressing pathogenic proteins.<sup>40,41</sup> The discriminatory power of such reference datasets depends on the density of their information content. Such reference sets could be extended to progression signatures, that is, following the chronology of changes of transcript levels through time (Fig. 4, lower). In the present context, the central idea is to identify progressive changes that occur either in cell culture or in tissues of an intact organism – comparing the unknown to a set of experimental variants imposed on normal cells or organisms.

Once the progression signature of the unknown has been defined (eg, 12, 76, 33), if a sufficiently close match can be found among the reference sets (eg, 12, 77, 31 in Fig. 4), the earlier states of that entry in the reference set (15, 88, 92, 3) could approximate the circuitry that led to the downstream observable characteristics for the unknown. This inferential strategy thus could provide a way to read time backwards and, therefore, to identify corresponding early therapeutic targets. Even when many cells have already undergone irreversible changes, identification of such molecular targets should make it possible to rescue cells that had not yet been irreversibly affected.

In the *upper panel*, cells are exposed to a known insult and the goal is to identify relatively early events (potential biomarkers) that precede the terminal change. The biomarkers might be collected at a single time point or correspond to a sequence of characteristics.

The *lower panel* schematizes the sequential consequences of experimental perturbations (A, B, C, etc.) that have been imposed on a cell of interest. Once reference panels of changes are available



**Figure 4.** Repertoires of response can allow inference of later and earlier states.

**Notes:** In these schemes, a given cell has been subjected to either known or unknown perturbations. In the upper panel (I), the perturbant is known to initiate disease. In the lower panel (II), specific experimental drugs, shRNAs, etc., are used, as well as an unknown. In each situation, there is a sequence of changes which are indicated by letters or numbers. There is no reason to expect these changes to be limited to single proteins or transcripts. Each one could include an ensemble of characteristics. Among the latest events, there will be a “terminal” group that inevitably leads to loss of growth control or to death.

(1, 7, 27, etc.), when confronted with an unknown, one would ask whether it exhibits characteristics (static or progressive) that match one of the prototypes (A, B, C, etc.). When a satisfactory match is found (eg, for the observed characteristics of perturbant C), one would then read backwards in time for the closest match to infer the precursor events. Those which are causally linked to progression could become targets for intervention, unless they already lie far along the axis of progression. The complexity and resolution of these strategies depend on the extent to which changes that occur with one perturbant overlap with others.

## Acknowledgments

We thank the Visconsi family for their support.

## Author Contributions

Wrote the first draft of the manuscript: AMT. Contributed to the writing of the manuscript: DW. Agree with manuscript results and conclusions: AMT, DW. Made critical revisions and approved final version: AMT, DW. Both authors reviewed and approved of the final manuscript.

## REFERENCES

- Maheshri N, O’Shea EK. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct.* 2007;36:413–34.
- Miyawaki A. Visualization of the spatial and temporal dynamics of intracellular signaling. *Dev Cell.* Mar 2003;4(3):295–305.
- Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? *Cell.* May 14;141(4):559–63.
- Cattaneo E, Zuccato C, Tartari M. Normal huntingtin function: an alternative approach to Huntington’s disease. *Nat Rev Neurosci.* Dec 2005;6(12):919–30.
- Teng X, Dayhoff-Brannigan M, Cheng WC, et al. Genome-wide consequences of deleting any single gene. *Mol Cell.* November 21, 2013;52(4):485–94.
- Tamborero D, Gonzalez-Perez A, Perez-Llamas C, et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Scientific reports.* 2013;3:2650.
- Bamford S, Dawson E, Forbes S, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer.* July 19, 2004;91(2):355–8.
- Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* January 23, 2014;505(7484):495–501.
- Floor SL, Dumont JE, Maenhaut C, Raspe E. Hallmarks of cancer: of all cancer cells, all the time? *Trends Mol Med.* Sep 2012;18(9):509–15.
- Sonnenschein C, Soto AM. The aging of the 2000 and 2011 Hallmarks of Cancer reviews: a critique. *J Biosci.* Sep 2013;38(3):651–63.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* March 4, 2011;144(5):646–74.
- Gonzalez de Castro D, Clarke PA, Al-Lazikani B, Workman P. Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance. *Clinical pharmacology and therapeutics.* Mar 2013;93(3):252–9.
- Worman HJ. Nuclear lamins and laminopathies. *J Pathol.* Jan 2012;226(2):316–25.
- Shimi T, Pflieger K, Kojima S, et al. The A- and B-type nuclear lamin networks: microdomains involved in chromatin organization and transcription. *Genes Dev.* December 15, 2008;22(24):3409–21.
- Fullston T, Finnis M, Hackett A, et al. Screening and cell-based assessment of mutations in the Aristaless-related homeobox (ARX) gene. *Clin Genet.* Dec 2011;80(6):510–22.
- Suri M. The phenotypic spectrum of ARX mutations. *Dev Med Child Neurol.* Feb 2005;47(2):133–7.
- McClellan J, King MC. Genetic heterogeneity in human disease. *Cell.* April 16, 2010;141(2):210–7.
- Manolio TA, Collins FS. The HapMap and genome-wide association studies in diagnosis and therapy. *Annu Rev Med.* 2009;60:443–56.
- Glass K, Huttenhower C, Quackenbush J, Yuan GC. Passing messages between biological networks to refine predicted interactions. *PLoS One.* 2013;8(5):e64832.
- Wang W, Baladandayuthapani V, Morris JS, Broom BM, Manyam G, Do KA. iBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics data. *Bioinformatics.* January 15, 2013;29(2):149–59.



21. Wang W, Baladandayuthapani V, Holmes CC, Do KA. Integrative network-based Bayesian analysis of diverse genomics data. *BMC Bioinformatics*. 2013;14 Suppl 13:S8.
22. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. October 25, 2005;102(43):15545–50.
23. Wu D, Lim E, Vaillant F, Asselin-Labat ML, Visvader JE, Smyth GK. ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics*. September 1, 2010;26(17):2176–82.
24. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res*. September 1, 2012;40(17):e133.
25. Jang S, Atkins MB. Treatment of BRAF-mutant melanoma: the role of vemurafenib and other therapies. *Clinical pharmacology and therapeutics*. Jan 2014;95(1):24–31.
26. Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer*. Jan 2009;9(1):28–39.
27. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. September 29, 2006;313(5795):1929–35.
28. Carroll JB, Warby SC, Southwell AL, et al. Potent and selective antisense oligonucleotides targeting single-nucleotide polymorphisms in the huntington disease gene / allele-specific silencing of mutant huntingtin. *Mol Ther*. Dec 2011; 19(12):2178–85.
29. Okada Y, Diogo D, Greenberg JD, et al. Integration of sequence data from a Con-sanguineous family with genetic data from an outbred population identifies PLB1 as a candidate rheumatoid arthritis risk gene. *PLoS One*. 2014;9(2):e87645.
30. Singer JB, Hill AE, Burrage LC, et al. Genetic dissection of complex traits with chromosome substitution strains of mice. *Science*. April 16, 2004;304(5669):445–8.
31. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. Nov 2013;8(11):2281–308.
32. Wu D, Pang Y, Wilkerson MD, Wang D, Hammerman PS, Liu JS. Gene-expression data integration to squamous cell lung cancer subtypes reveals drug sensitivity. *Br J Cancer*. September 17, 2013;109(6):1599–608.
33. Bilen J, Bonini NM. Genome-wide screen for modifiers of ataxin-3 neurodegeneration in *Drosophila*. *PLoS Genet*. Oct 2007;3(10):1950–64.
34. Fernandez-Funez P, Nino-Rosales ML, de Gouyon B, et al. Identification of genes that modify ataxin-1-induced neurodegeneration. *Nature*. November 2, 2000;408(6808):101–6.
35. Da Cruz S, Cleveland DW. Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. *Curr Opin Neurobiol*. Dec 2011;21(6):904–19.
36. Lagier-Tourenne C, Cleveland DW. Neurodegeneration: An expansion in ALS genetics. *Nature*. August 26, 2010;466(7310):1052–3.
37. Taguchi A, Hanash S, Rundle A, et al. Circulating pro-surfactant protein B as a risk biomarker for lung cancer. *Cancer epidemiology, biomarkers and prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. Oct 2013;22(10):1756–61.
38. Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med*. January 18, 2007;356(3):217–26.
39. Arun B, Bayraktar S, Liu DD, et al. Response to neoadjuvant systemic therapy for breast cancer in BRCA mutation carriers and noncarriers: a single-institution experience. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. October 1, 2011;29(28):3739–46.
40. Hughes TR, Marton MJ, Jones AR, et al. Functional discovery via a compendium of expression profiles. *Cell*. July 7, 2000;102(1):109–26.
41. Parsons AB, Lopez A, Givoni IE, et al. Exploring the mode-of-action of bio-active compounds by chemical-genetic profiling in yeast. *Cell*. August 11, 2006;126(3):611–25.