The Axis of Progression of Disease

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.4137/CIN.S17683</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:13454704">http://nrs.harvard.edu/urn-3:HUL.InstRepos:13454704</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>

(Article begins on next page)
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.\textsuperscript{1–4} 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 accommodation.\textsuperscript{5,6} 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,\textsuperscript{7} the saturation analysis of cancer genes across 21 tumor types has allowed identification of additional somatic mutations that are associated with cancers.\textsuperscript{8}

**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. 9–11 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 Oncotype Dx and Mammaprint gene expression signatures in breast cancer.12

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 functional 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.13–16 Moreover, the issue of polygenic causation itself,
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.\textsuperscript{17,18}

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.\textsuperscript{19–21} Other analytic tools also suggest that significant pathways or sets of genes work together.\textsuperscript{22–24}

**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.).\textsuperscript{25,26} 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.\textsuperscript{27}

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.\textsuperscript{28}

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.\textsuperscript{29} This strategy can be directly extended to investigation of model organisms with distinct genetic backgrounds, for example, different inbred strains of mice,\textsuperscript{10} and animals with engineered genomes.\textsuperscript{31}

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 explanation for why 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.\textsuperscript{33,34} 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. 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, 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. 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. Ongoing comparative exome sequencing of germline and tumor samples for cancer genomes will aid identification of further biomarkers for prediction of cancer risk.

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. 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.

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


