Epigenetics as a mediator of plasticity in cancer

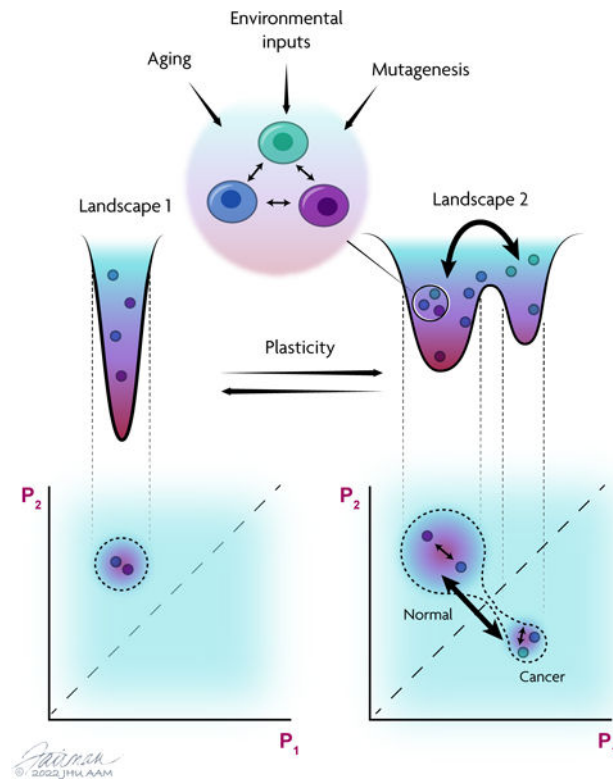**Andrew P. Feinberg**[1], **Andre Levchenko**[2]

[1]Center for Epigenetics, Johns Hopkins University Schools of Medicine, Biomedical Engineering, and Public Health, 855 N. Rangos St., Baltimore, MD 21205

[2]Yale Systems Biology Institute and Department of Biomedical Engineering, Yale University, 850 West Campus Dr., West Haven, CT 06516

Abstract

The concept of an epigenetic landscape describing potential cellular fates arising from pluripotent cells, first advanced by Conrad Waddington, has evolved in light of experiments showing non-deterministic outcomes of regulatory processes and mathematical methods for quantifying stochasticity. In this Review we discuss modern approaches to epigenetic and gene regulation landscapes and the associated ideas of entropy and attractor states, illustrating how their definitions are both more precise and relevant to understanding cancer etiology and the plasticity of cancerous states. We address the interplay between different types of regulatory landscapes, and how their changes underlie cancer progression. We also consider the roles of cellular aging and intrinsic and extrinsic stimuli in modulating cellular states, and how landscape alterations can be quantitatively mapped onto phenotypic outcomes, and thereby used in therapy development.

Graphical Abstract

afeinberg@jhu.edu .

Environmental inputs

Aging

Mutagenesis

Landscape 1

Landscape 2

Plasticity

P_2

P_2

Normal

Cancer

P_1

P_1

Summary sentence

A Review that discusses how epigenetic stochasticity can connect mutations and environmental perturbations to the phenotypic plasticity that drives tumor progression and therapeutic resistance.

Tumor cell heterogeneity and phenotypic plasticity have long been known to be principal drivers of invasion and metastasis (1). Complex and diverse behaviors displayed by cancer cells frequently involve coordinated expression of multiple genes, constituting programs that may be very distinct from those defining the phenotypes of the tissue of origin. Since oncogenic transformations are frequently accompanied by extensive mutagenesis, analysis of phenotypic heterogeneity displayed by cancer cell populations has typically been ascribed to the emergence of multiple genetically distinct clonal cell subpopulations within the same tumor (2). This genetic heterogeneity should be distinguished from phenotypic heterogeneity (i.e., emergence of stable phenotypic states in an isogenic population) and phenotypic plasticity (i.e., the ability of cells to adopt different phenotypic states transiently). Phenotypic plasticity is thought to underlie such complex phenotypic changes as epithelial-mesenchymal transition (EMT), drug resistance, and increased cell proliferation (3–6). Therefore, although genetic mutations may increase the probability of new phenotypic states, they may not completely define them and may not be strictly necessary for their emergence.

Elucidation of the interplay between epigenetic states and regulation of gene expression is crucial for our understanding of phenotypic plasticity. Multiple questions remain regarding coupling of these regulatory mechanisms. For example, since many regulators of gene

expression may also have epigenetic modification capabilities, e.g., by recruiting DNA methylases and histone acetyltransferases, it is not clear whether epigenetic control might be viewed as a constituent part of transcriptional regulation, or it has independent and essential roles not fully predictable by the analysis of gene networks alone.

In this Review, we argue that epigenetic mechanisms can have several key functions that are frequently not evident from traditional quantitative approaches involving the analysis of transcriptional regulation or RNA and protein stability. In particular, epigenetic regulation may have a critical role in regulating variability of gene expression, accounting for variance in the current cellular states and emergence of new ones. Recent methodological advances now allow us to explicitly define and measure epigenetic landscapes that, along with gene regulation landscapes, can be used to much more precisely understand the etiology of cancer, and better understand the mechanisms underlying phenotypic plasticity and heterogeneity.

Theoretical approaches to describe genetic and epigenetic landscapes

Regulatory networks impose complex inter-dependencies on gene expression, which are commonly modeled using ordinary differential equations (ODEs) (Box 1, Equation 1). A related concept of "landscape" introduced by Waddington and further elaborated mathematically (7–10) has been instructed for simpler and well characterized systems, but has faced important challenges. In particular, molecular interactions constituting regulatory networks are frequently unknown and may be restricted by epigenetic mechanisms, such as DNA methylation. Furthermore, these interactions and networks are subject to biomolecular 'noise' or variability of molecular concentrations or alterations of molecular states (11–13), which are crucial for non-deterministic many-fate behaviors of the system. Although the Waddington landscape is frequently interpreted as the epigenetic landscape, these concepts may no longer be fully synonymous. Indeed, whereas previously the term "epigenetic" referred to any mechanism outside traditional gene expression control, today, it more commonly refers to a specific set of regulatory processes, affecting in particular DNA methylaiton and chromatin modifications.

The probabilistic nature of biochemical processes has been historically addressed in several ways. Most simply, added noise might be described by extending the ODE description using the Langevin approach (Box 1, Equation 1). However, this approach can break down if the number of molecules becomes very small, making the functions defining the ODE system behavior discontinuous. This consideration drove a conceptual quantum leap, akin to transition from classical to quantum mechanics in physics. In this new view, instead of describing the biological variables, such as molecular concentrations, one operates with the probabilities that system-defining variables take on certain values. This probabilistic dynamics is captured by the Master Equation, based on the Markovian assumption about the system dynamics (Box 1, Equation 2). However, while appearing superficially simple, such equations require advanced computational resources to solve or estimate, although some success has been achieved for simpler cases (14–16). A fundamental challenge facing both the Langevin and Master Equation approaches is the lack of complete knowledge of the mechanisms governing interactions between system variables.

To date, relatively few biological networks have been fully characterized, and frequently this mechanistic characterization is only semi-quantitative. It is therefore difficult to theoretically or computationally predict the probability distributions of different molecular and cellular states. However, one can now directly experimentally estimate these probabilities (17–19) and thus attempt to solve an inverse problem of reconstructing the mechanisms underlying gene expression regulation and the corresponding genetic and epigenetic landscapes from experimental data.

The inferences based on experimentally obtained probability distributions can be put on a solid theoretical foundation that provide specific meanings to the notions of epigenetic and genetic landscapes, distinct from those implied by Waddington. Critical to this approach is the fundamental connection of potential energy and probability distributions provided by the Gibbs-Boltzmann approach (Box 1, Equations 3 and 4), which are particularly useful for definition and analysis of landscape attractors corresponding to specific coherent molecular, cellular, and phenotypic states. Specifically, this approach associates the probability of any outcome, e.g., gene expression or DNA methylation pattern and the corresponding phenotypic outcome, with a generically defined quasi-potential energy, a term adapted from physics to describe the origin of probability distributions of the possible cellular states (states with lower energy are occupied with higher probability; see Fig. 1A, and the definition in Box 1, Equation 3). This association allows one to define the corresponding genetic (gene expression) or epigenetic (e.g., DNA methylation markers) landscapes as the quasi-potential energy distributions. This landscape definition is consistent with the one implied by the Waddingtonian ODE approach, affording easy interpretation and visualization. The attractors within this landscape representation would be seen as the values of biological variables for which the local and global energy minima would be reached, thus constituting states with high relative probability values (Fig. 1A). The stability of the attractor is further defined by such landscape characteristics as the depth and width of the corresponding energy wells, and proximity, and thus accessibility, of other attractors (20). Increased variability or noise makes energy wells effectively more shallow, thus widening the distribution of biological variables (21). This allows cellular networks to explore multiple states across different attractors. The local energy (landscape) minima may be realized in different cells within a population or could only be accessible in specific developmental, physiological or pathological situations. Below we illustrate and discuss these and other important ideas, such as the entropy associated with the whole landscape and specific attrctor states, and the information associated with changes of this entropy (Box 1), particularly in the context of emergence and progression to cancer.

Landscape models of cancer

Gene regulatory landscapes can be evaluated from experimentally determined gene expression probability distributions, particularly those yielded by single cell RNA sequencing (scRNA-seq) data or from the analysis of multiple tissue samples (e.g., cancerous tissues from multiple patients) (Fig. 1A). The landscape attractors in these distributions correspond to groups of coordinately expressed genes and gene products, whose abundance and activity may be regulated by self-sustaining feedback interactions, endowing cells with specific phenotypic properties. Cancerous states may represent the

genetic landscape attractors that are inaccessible in untransformed tissues but can be unmasked by oncogenic mutations (20, 22). This is exemplified by the emergence of the mesenchymal state in EMT (3). The Langevin ODE approach has been used to analyze a tri-stable regulatory circuit assumed to control epithelial-mesenchymal plasticity (23). Additionally, four attractor states have been proposed to control EMT, including metastatic, anti-metastatic, and two intermediate states (24). These studies illustrated putative feedback-based mechanisms of emergence of well-defined attractors in cancer-associated gene regulation landscapes. However, they were based primarily on theoretical modeling constructs that require experimental validation, whereas mechanisms, such as epigenetic regulation, may not have been included. Alternatively, the properties of experimentally measured regulatory landscapes can be inferred without imposing hypothetical mechanisms.

A key consequence of both the emergence of new attractors and elevated variability within the same attractors is the increase in the entropy of the underlying regulatory landscapes (Inputs 1 and 1' in Fig. 1B, see Box 1). Indeed, it is consistently observed that the entropy of gene expression increases in cancer versus the corresponding non-cancerous tissue of origin, as e.g., revealed by examining gene expression data alone (25), or by integrating these data with protein interaction networks (to account for molecular interactions stabilizing the attractors) (26). A similar approach using protein-interaction networks and RNA-seq for multiple tissue samples, rather than single cells within the same sample, suggested that tumor progression is associated with increasing entropy (27). This approach was further used to link the entropy of gene expression to the pluripotency of underlying cell population (28), and to demonstrate that specific cell differentiation, in contrast to cancer emergence, could be represented by emergent attractors, associated with progressively lower variability and entropy. Consistent with this analysis, it was shown that cancer stemness may be inferred from increased entropy of the attractor defining a specific cell sub-population (28). Interestingly, this analysis revealed that decreased entropy was associated not only with cell fate-specifying molecular circuits, but also with specific cancer-related phenotypic states such as increased cell proliferation (22), implying correlated gene expression and molecular interaction of the components of the cell cycle, representing another example of a well-defined attractor in the gene expression landscape. However, this high proliferation attractor may not be fully stable, with recent studies suggesting that cells may stochastically transition into and out of a cell proliferation state (6, 29). Other studies further pointed to a transient rather than stable nature of other key cancer-associated phenotypes: drug resistance and EMT, indicating that they may be a consequence of phenotypic plasticity within a broader attractor state, potentially allowing these phenotypes to be transiently achieved (30–33). This broader attractor may be associated with a higher entropy stemness state, whereas each of the transiently visited phenotypic states may have much lower associated entropy, i.e., they can be considered sub-attractors that can be occupied by the cells either stochastically or in response to environmental inputs (34) (Inputs 1 and 2 in Fig. 1B).

These studies reveal that reasons for higher entropy in gene expression networks in cancer cells can be complex, because the same increase in entropy could be due to the emergence of new stable attractors in the cell population (phenotypic diversity, Input 1' in Fig. 1B) or increased noisiness of the individual attractors that may already be present in the normal, pre-cancerous tissue or emerge due to cell de-differentiation (phenotypic plasticity, Input 1

in Fig. 1B). Furthermore, phenotypic plasticity can enable emergence of transiently stable, stochastically visited attractors that can nevertheless have important functional phenotypic characteristics, such as EMT, cell proliferation, and drug resistance (Input 2 in Fig. 1B). These phenotypes may be related to those occurring separately in normal development and tissue homeostasis, but might become mutually accessible in the context of a larger underlying attractor in the phenotypically plastic cancerous states (Fig. 1C). Thus, a more detailed landscape analysis is required to understand the mechanisms underlying the emergence and dynamics of cancerous cell populations. Furthermore, these studies leave the following questions open: what controls the availability of various transcriptional programs that enable diverse landscape attractors and what mechanisms regulate the transitions between different attractor states and thus the degree of plasticity of various phenotypes?

In contrast to gene regulation landscapes, in which variables, such as molecular concentrations, can be viewed as continuously adopting multiple values, epigenetic landscapes are essentially binary or "digital" in nature, e.g. the methylation patterns of specific CpG sites or the presence or absence of specific histone modifications. Nevertheless, the landscape analysis framework can also be used for these types of experimental data. For instance, epigenetic variation in the patterns of sets of four consecutive CpG sites display increasing stochasticity in aging fibroblasts and in cancer cells, within particular regions of the genome (35). The patterns of consecutive CpG sites, also known as epialleles, can thus be treated probabilistically, with their distributions within the same tumor estimated using a Bayesian model (36). The probability analysis lends itself to examination of entropy and information metrics (defined in Box 1), which was applied to several brain tumor and normal samples, revealing an increased methylation entropy of Alu repeat sequences (37). The epigenetic patterns of epialleles may display certain correlations, interpreted as specific attractor states in the underlying landscapes. Mechanistically, these correlated patterns may arise because DNA methylation at heritable (i.e. CpG) dinucleotides is affected by the state of the cell or location in the genome, as well as the nearest CpG neighbors, because DNA methyltransferase I, the enzyme that copies CpG from the parent to the daughter strand during DNA replication, resides at the replication fork, conferring a nonrandom methylation pattern to nearby CpG dinucleotides.

The underlying landscapes may be more formally defined recognizing the binary nature of the experimental data by exploiting the Ising model, introduced in the 1930's for one-dimensional magnetic dipoles, whose binary states are also influenced by external forces and the nearest dipole neighbors (38). It also allowed a highly accurate computation of the Shannon normalized methylation entropy (NME) for DNA methylation. Furthermore, the Ising model enabled a more sophisticated analysis, such as the calculation of the Jensen-Shannon distance to evaluate the mutual information between two epigenomes, and thus to rank genomic features by their relative contribution to epigenetic landscape change (39). This more rigorous treatment revealed substantial entropy increases in the epigenetic landscapes of a variety of cancer types compared to matched normal tissues of origin (38). Importantly, it was found that an increase in disordered methylation patterns predicts poor clinical outcomes in chronic lymphocytic leukemia (40) and acute myelogenous leukemia (41). This recent progress has been striking because cancer epigenetics was thought to be

computationally intractable due to being derived from "hundreds of landscapes" of normal tissue (42).

The increase in epigenetic landscape entropy observed in various cancers is accompanied by a greater variability in gene expression (43), thus linking epigenetic and gene regulation landscapes. A possible explanation for this linkage is that genetic programs accounting for new phenotypic properties can become accessible due to increased epigenetic variability, leading to the emergence of new gene regulation attractors. This hypothesis is supported by the observation that the increased variation in gene methylation and expression is often enriched in genomic regions that are larger than a single gene (43, 44), e.g., regions associated with EMT (43) or smaller domains, observed in pancreatic cancer (45), perhaps related to hypomethylated regions inferred for hematopoietic stem cells (46). Thus, epigenetic control of multiple genes needed to enable a specific phenotype could "loosen" and the corresponding genetic attractor may emerge and enable variable onset of a new phenotypic state in cell population. It was indeed found that the transition to the mesenchymal state depends on altered chromatin accessibility of the DNA regions incorporating EMT related genes (47). Furthermore, large heterochromatic domains related to developmentally regulated EMT (48) are topologically altered in colon cancer (49). These data suggest that a key cancer related phenotype is enabled by changes in both gene expression landscape (actualization of a cryptic attractor state) and epigenetic landscape (changes in DNA and histone markers that enable accessibility to the genetic information encoding components of this newly accessible genetic landscape attractor).

Together, these results suggest that a key condition for the emergence of a new attractor in the genetic landscape is the epigenetic re-organization of the corresponding genetic material, leading to increasing stochastic variability within the current attractor and increased probability of transition to a new state (21) (Input 1 in Fig. 1B). It is of interest therefore to explore the triggers of epigenetic and gene regulation landscape reorganization that occur during carcinogenesis, a task that is complicated by inter-dependencies of epigenetic and gene regulation mechanisms. Indeed, epigenetic modification, e.g., DNA methylation of gene regulatory sequences, can directly influence gene expression, and many transcription factors can recruit epigenetic regulators, suggesting substantial epigenetic and genetic landscape interdependence (Fig. 2). To provide more detailed insight into oncogenic processes that shape and are shaped by epigenetic and gene regulation landscapes, it will be instructive to examine the genetic and environmental inputs that can lead to precancerous and cancerous states.

Interplay between epigenetic and genetic landscapes

How can the large number of recurrent somatic mutations that drive carcinogenesis be connected to the pathways and genes responsible for cancer phenotypes (50)? Traditionally, this question has been addressed by analyzing the effects of various mutations on the magnitude of phenotypic changes, such as increased proliferation, in response to overexpression of diverse mitogenic regulators. What is currently less understood is how the increased variability (and thus entropy) of epigenetic and gene expression regulation is linked to cancer progression. Such analyses are now progressively enabled by the rank-

ordering of oncogenic mutations using metrics such as Jensen-Shannon distance (Box 1, Equation 7), which establish the effect size and its importance in epigenetic control of cancer etiology (39).

The roles of mutations affecting molecules that directly (so-called, epigenetic modifiers) and indirectly (epigenetic modulators) control epigenetic landscape through their enzymatic functions (51) can be elucidated in cancerous states (Fig. 2). For example, in acute myelogenous leukemia (AML), driver mutations can affect both epigenetic modulator molecules [such as transcription factors and signaling pathways that can recruit epigenetic modifiers to specific gene targets (52)] and modifier molecules [such as the DNA demethylation-related gene *TET2*, which is directly mutated in a subset of AML patients (53)]. Analysis of epiallele distribution across AML patients with different driver mutations indicated an increase in the entropy of epiallele distribution, consistent with other cancers. Importantly, this analysis revealed a more fine-grained epigenetic landscape structure underlying entropy increase. In particular, epialleles commonly showed specific patterns for a subset of modulator mutations, whereas the distribution was much broader for modifier mutations. This study further indicated that tighter and more specific epiallele distributions were corelated with better clinical outcomes, whereas broader distributions were generally associated with poor clinical prognoses. Furthermore, higher epiallele diversity was associated with gene expression hypervariability and, thus, likely higher phenotypic and developmental plasticity. This result was consistent with an independent analysis of AML involving genetic rearrangement of the the histone methyltransferase gene *KMT2A* (also known as *MLL*) which, when compared to untransformed controls, also demonstrated a greatly increased epigenetic entropy (more specifically, NME) and high expression variability of genes implicated in AML etiology (54). Importanlty, animal models of AML revealed that epigenetic variability was a precursor of transformation to AML, pointing to the causal rather than correlative nature of the greater entropy states (53). Overall, these data suggest that mutations affecting particular regulatory pathways may result in adoption of specific genetic and epigenetic states (narrow and stable landscape attractors). Conversely, dysregulation of epigenetic modifiers can lead to a much broader epiallele distribution, which may result in broader, more 'noisy' attractors, phenotypic plasticity and poorer prognosis (Input 1 in Fig. 1B and Fig. 2). Indeed, a greater phenotypic plasticity enabled by a more global epigenetic dysregulation may enhance escape of cancer cells from various treatments and other pressures, including immune surveillance.

Whereas mutational alterations of some modulators can lead to emergence of specific attractor states, mutations of other modulator proteins may exert their effect indirectly, by regulating the broad specificity epigenetic modifiers, thus also leading to extensive epigenetic alterations, and broader, more plastic attractor states. For example, inhibition of DNA methylation was highly effective in AML patients harboring *TP53* mutations (encoding the tumor suppressor p53), but had much lower effects in tumors with wild type *TP53* (55). These results were supported by lung cancer xenograft studies and *in vitro* experiments (56), together suggesting that p53 is an epigenetic modulator that is critically important for epigenetic stability. Indeed, in ESCs, p53 restricts the expression of modifiers *Dnmt3a* and *Dnmt3b* thus inhibiting methylation, as well as regulating *Tet1* and *Tet2*, making the loss of *Trp53* a major driver of epigenetic heterogeneity (56, 57).

Other genes commonly mutated in AML, *WT1*, *IDH1* and *IDH2* can also have more global effects on epigenetic landscapes, through inducing loss of TET2 function (58). These results suggest that, although epigenetic modifiers may not be mutated in specific tumors, in many cases the effect of modulator mutations can lead to changes in the methylation status of a large number of genes, similar to the effects of modifier mutations. The more specific, targeting effects of modifiers can be revealed through comparative gene expression analysis, e.g., by searching for enrichment of binding sites for specific transcription factors in the control regions of epigenetically variable genes. Enrichment of specific binding sites would suggest that a particular transcription factor can both recruit an epigenetic modifier to a particular gene and more directly control the expression of this gene, once it is epigenetically modified. This type of analysis yielded a number of putative modulator proteins, including GATA2, coupled to epiallele variation in many AML subtypes driven by distinct mutation events (53). Similar studies further suggested specific roles of *NFKB1* and *MYBL1* as modulator proteins, controlling the drug resistance phenotype in chronic lymphocytic leukemia (59).

Overall, these studies suggest the following refinement of our understanding of the interplay between genetic and epigenetic landscapes in the emergence of genetically driven cancers (Fig. 2). The effect of mutations in mediator and a subset of modulator genes may be broad, encompassing multiple genes across the genome and generating a high degree of epiallelic and genetic heterogeneity, which correlates with poor clinical outcome. The effect of mutations in another subset of modulators may be more specific, leading to emergence of less variable epiallele distributions and coordinated gene expression patterns, corresponding to a limited number of coupled attractors in genetic and epigenetic landscapes. In both cases, the entropy of epigenetic and genetic landscapes goes up, but for different reasons: in the former case, a more plastic state implies a noisy and relatively unstable attractor state, whereas in the latter case, increased entropy may correspond to the emergence and co-existence of new stable attractors in the genetic and epigenetic spaces (Fig 1B). This view implies that there may be different landscape paths to cancer, but the emergence of a more plastic, less defined attractor state may be particularly interesting to consider, because it may be related to both more pluripotent, stem cell-like states and increased plasticity of cancer cells.

Stabilization of attractor states by endogenous and exogenous inputs

Exogenous inputs.

Enhanced variability of both gene expression and epiallele abundance indicative of broad genetic and epigenetic attractors may permit cells to transiently achieve multiple alternative states that could be potentially further stabilized by additional inputs. For example, genetic perturbations of DNMT3A and DNMT3B dramatically increase genetic and epigenetic entropy in embryonic stem cells, whereas the effects on mean methylation were much more modest (60). Furthermore, several genes that are differentially controlled by this perturbation can promote cellular differentiation. Importantly, genetic perturbations of DNMT and TET genes revealed that the differentiated state is a separate attractor in the genetic landscape, and that this attractor exists dynamically, with cells reaching and leaving it according to the

relative activities of DNMT and TET enzymes (61). These results are consistent with the current understanding of the importance of epigenetic regulation of induced pluripotency by the Yamanaka factors, particularly the increased expression of TET enzymes to stabilize this state (62). Critically, this differentiation attractor can also be regulated by modulators in the form of signaling pathways, particularly dual inhibition of MAPK signaling and GSK3β (63–65), suggesting that external signaling inputs can stabilize the differentiation state.

The insights from epigenetic regulation of stemness can elucidate how conditional attractors can emerge and be stabilized e.g., by exogenous signaling inputs in cancer (Input 2 in Fig. 1B). As an example of signaling attractor stabilization, it was found in two independent studies that paracrine signaling involving IGF-1 receptor activation or NOTCH pathway signaling can lead to an altered chromatin structure that is mediated by the histone demethylases KDM5A and KDM6A/B respectively, which in turn control the drug resistance phenotype (66). The acquisition of this phenotype was, in both cases, transient and stochastic, suggesting continuous transitions between distinct states associated with two or more attractors in the underlying landscape. The dynamic acquisition can be analyzed in substational detail *in vivo*, as was the case for AML in a genetically engineered mouse model (67).

Metaplasia.

The landscape-based analysis implies that a loss of stabilizing inputs, exogenous or endogenous, may make transitions to other available attractors more likely, maintaining a plastic state and metaplastic tissue composition. A particularly interesting example of this is found in kidney differentiation. WT1, a tissue specifying transcription factor that is critical for embryonic kidney development, can be seen as a molecule stabilizing the corresponding differentiation program attractor. Loss-of-function mutation of *WT1* leading to a loss of DNA binding in the germline causes Denys-Drash syndrome with renal dysplasia (68). The resultant absence of *WT1* in somatic cells leads to metaplastic trans-differentiation of early pronephric cells into other mesodermal elements such as skeletal muscle, fat and cartilage and generation of pre-neoplastic "intralobar nephrogenic rests," relatively early in the developmental axis, increasing the risk of Wilms tumor development (69). WT1 also controls EMT, likely through interaction with the YAP transcriptional co-regulator in controlling E-cadherin expression (70), which may modulate the tissue organization and further influence the metaplastic state. Timed experimental activation of stem cell reprogramming factors in the developing mouse kidney can induce Wilms tumor in the absence of other perturbations, including absence of any *WT1* mutations (71). This same effect arises from loss of imprinting (LOI) of *IGF2* in Beckwith-Wiedemann syndrome, or sporadically in the general population, with hyperproliferation of developing nephrons also creating predisposition for Wilms tumor development. These results are consistent with losses of stabilizing control leading to a broad, variable and plastic landscape attractor, which can also decrease the probability of occupying the attractor corresponding to a stably differentiated nephron state (Figs. 1B and 3).

Metaplastic states do not need to involve processes that occur early in development or depend on reprogramming cues. Rather they may arise from stabilization of one of the

mutually accessible attractors even in the mature adult tissue (Input 2' in Fig. 1B). Acinar-ductal metaplasia (ADM; i.e., the process in which acinar cells in the pancreas differentiate into ductal cells) can occur in the untransformed adult pancreas as a part of the tissue repair process (72), but may also lead to pre-neoplastic lesions and, ultimately, progression to pancreatic ductal adenocarcinoma (PDAC) (73). In addition to ADM, which may occur as a normal adaptive process, de-stabilization of the attractor corresponding to the acinar cell type can lead to the emergence of pancreatic intra-epithelial neoplasms (PanINs). Furthermore, other cell types may also emerge in this process, e.g., tuft cells. This attractor destabilization may involve mutations of modulators *KRAS* and *DPC4* (also known as *SMAD4*) and upregulation of *KLF4* during inflammation or injury (74, 75), as well as changes in the epigenetic control of key regulators, such as AATK, as revealed by studies of a mouse model of ADM and pancreatic cancer (76).

Another example of an environmentally induced metaplastic process occurs in Barrett's esophagus. In this case, the environmental stimulus (modulator) is acid reflux and inflammatory cytokine enrichment, which promote trans-differentiation of normal squamous epithelium of the esophagus into intestine-like columnar epithelium (77). Animal and human data suggest that, in this setting, the attractor corresponding to the esophageal cell type is compromised due to induction of intestinal-specific factors CDX1 and CDX2, which might be induced by inflammatory inputs and can self-stabilize due to auto-regulation (78). This process is associated with a bimodality in the DNA methylation distribution reflective of co-existence of esophageal and intestinal attractors, characteristic of this type of metaplasia (79). Trans-differentiation in this setting is again a major risk factor of ensuing development of esophageal adenocarcinoma.

Together, these observations suggest that metaplasia or alternative differentiation can be an important characteristic of pre-malignant tissue states. Consequently, the new states explored by the cells may not be random but frequently belong to closely related lineages (80), in agreement with the expectation that cells explore adjacent attractors along the underlying epigenetic and genetic landscapes, due to an increased stochasticity and thus entropy of the regulatory networks (Fig. 1C). Intriguingly, upregulation of epigenetic modifiers, such as *ARID1A*, is commonly observed both in the metaplastic states and multiple cancers (81–83). Tissue disorganization related to metaplasia as well as the common dysregulation of the EMT-MET program can further perturb local cell-cell interactions leading to hyperplasia and the onset of neoplastic growth (84–86).

Therefore, if these genetic or environmental inputs occur in tissues with a high potential for meta-plasticity, the occupancy of the attractors may change. The occupancy of the previously dominant attractors may become less likely, whereas pre-existing alternative attractors can be occupied with higher probabilities, leading to further metaplasia and potentially cancerous states (Fig. 1B). In another scenario, if epigenetic modulators are directly affected by genetic mutations or environmental inputs, even tissues with low initial plasticity can become more plastic and permit occupancy of previously inaccessible, cryptic attractors (Input 1' in Fig. 1B). These attractors can become stable due to either internal feedback or additional stabilizing mutations, or can remain transient and stochastically visited, depending on the presence of an external cue. The higher degree of epigenetic

plasticity and conditional, rather than stable, attractors can increase the corresponding phenotypic plasticity which can result in poor prognosis due to increased metastatic propensity, drug resistance, or immune escape. In this sense, stabilization of particular attractors, by reducing the plasticity of cell regulation, can make cancer progression more predictable and potentially increase the probability of success of personalized therapeutic interventions.

Attractor stability and aging

Stabilization of conditional attractors by an environmental cue enabled by dysregulated epigenetic control can occur through a variety of mechanisms, but particularly through activation of intracellular signaling networks. Indeed, new attractors, such as the mesenchymal cell state emerging in EMT, can be triggered by biochemical and biomechanical inputs from the microenvironment, and increased cell proliferation can be controlled by a progressive accumulation of growth factors and other ligands in the cellular milieu (87). Attractor stabilization may result in a decrease in entropy of gene expression and epiallele occurrence, which can be precisely quantified experimentally and used to evaluate the information carried by the signaling input (88) (see Box 1; information is defined as the decrease in entropy following an input). The entropy decrease is frequently modest, with only 1 bit of information about the signal amplitude commonly transferred in such events, on a cell-by-cell basis (89–96). Coordination of signaling between multiple cells can allow for an increase in this information transfer and thus further entropy decrease, although, since cellular communication is noisy, the effect may also not be very substantial (97). A further entropy decrease can occur if cells are engaged in response to the same input, but are not necessarily communicating (98).

One bit of information is sufficient to define a binary phenotypic outcome, such as cell death/survival, differentiation/self-renewal, or proliferation/migration. This binary outcome frequently reflects occupancy of alternative attractor states modulated by the signaling input (Inputs 2 and 2' in Fig. 1B). In the context of cancer progression, multiple external cues can trigger signaling events, leading to diversification of cell responses through conditional occupancy of different attractors, as a function of the combinations of different cues (and their persistence) in a specific tumor location or the cues received by cancer cells during metastatic spread. Similar quantitative analysis can be used to gauge the effects of different mutations and drug treatments that can perturb the entropy of different attractors more specifically or the entropy associated with the overall epigenetic or genetic state more generally (93).

In contrast to the entropy decrease due to signaling inputs, entropy of a particular attractor and thus the associated phenotypic plasticity can also naturally increase due to epigenetic drift (Input 1 in Fig. 1B). Epigenetic drift occurs both in tissue culture and *in vivo* with increasing age of cells and organisms (99–102). In the context of hyper-proliferation associated with cancer, epigenetic drift is due to epimutations accumulating with the increasing number of cell divisions, which can happen at a high frequency (103). This increasing entropy of the epigenetic landscape is distinct from the recurrent changes in DNA methylation serving as a basis for the epigenetic clocks (104). Unlike the age related DNA

hyper- or hypomethylation signatures at specific CpG sites, epigenetic drift is a reflection of increased DNA methylation variability which frequently occurs at loci related to malignancy and can be predictive of cancer emergence (105). One example of this site specificity is methylation at the locus encoding polycomb repressive complex-2 (PRC2, itself responsible for trimethylation of H3K27) (103). The entropy increase associated with epigenetic drift may also explain the emergence of a rare 5% subpopulation of bipotent-like breast cancer progenitor cells (106). Intra-tumor epigenetic drift may be dependent on the replication history of the individual clones, leading to a continually expanding spectrum of EMT identities in colorectal tumors (107). Overall, the landscape entropy is not static and can decrease and increase in a quantifiable fashion, even without genetic alterations, creating a more complex context for normal tissue function and progression to cancer.

Mapping regulatory landscapes onto phenotypic responses

Can quantitatively assessed changes in epigenetic and genetic landscapes inform selection of appropriate treatments? Since treatments are driven by disease specific phenotypes and alterations of lanscapes occur at the level of DNA modifications and molecular networks, it is important to consider in more detail how the specific phenotypes are controlled. In many cases, a phenotypic state is defined by a specific bio-molecular circuit. For instance, cell migration phenotypes are controlled by cytoskeleton reorganization, which in turn, is frequently controlled by a subset of Rho-family small GTPases (108, 109). Similarly, cell proliferation is phenotypically driven by the circuit involving regulation of cyclin-dependent kinases (110). It is therefore important to understand how the function of these specific, phenotype defining circuits may be influenced by the more global changes in cellular regulatory networks, as discussed below for the case of apoptosis.

Although the molecular circuits controlling apoptosis are complex, one can use the relative abundances of pro- and anti-apoptotic members of the BCL protein family controlling the phenotypic outcomes of either regulated cell death (relatively higher abundance of pro-apoptotic proteins, such as BAX) or cytoprotection (relatively higher abundance of anti-apoptotic proteins, such as Bcl2) (111, 112) (Fig. 3A). The equally matched abundance of pro- and anti-apoptotic proteins defines the boundary between these states within a phenotypic space (life versus death) onto which the epigenetic and genetic landscapes can be mapped (Fig. 3B). This mapping is based on biochemical input-output relatonships between the phenotype-defining molecules and the upstream genetic and epigenetic mediators, is a natural dimensionality reduction mechanism allowing the evaluation of effects of the underlying molecular network variability and regulation. In many circumstances, e.g., in the turnover of multiple epithelial tissues, the entropy associated with the underlying landscape and thus variability in the expression of apoptosis proteins can result in a subset of cells undergoing cell death even under homeostatic conditions. The distribution of cells across the epigenetic landscape can therefore define the fractions of cells displaying different phenotypes.

The mapping of signaling network activity onto this apoptosis-regulating circuit (Fig. 3B) can be established experimentally and visualized in 2D as the joint single cell distributions of BCL2-BAX pairs. The regulatory landscapes upstream of the phenotype-

defining molecular circuit control the relative abundances of the proteins in these pairs, which can alter in response to an epigenetic change, such as LOI of *IGF2* (113). The mis-regulation of the signaling inputs resulting from epigenetic alteration (particularly, the re-balancing of signaling through growth factor responsive ERK-MAPK and AKT pathways) is the key aspect of the newly emerging attractor state that translates into a new BCL2-BAX distribution, resulting in a higher propensity for cell death (Fig. 3C). Therefore, an epigenetic landscape change, in addition to promoting higher cell proliferation and tissue hyperplasia, creates a vulnerability (an increased propensity to undergo apoptosis) that can be exploited to selectively eliminate the cell population undergoing *IGF2* LOI. These landscape perturbations cause the LOI cells but not wild type cells to cross the 'death boundary' (Fig. 3C), thus predicting the therapeutic window relevant to this pre-cancerous state for specific drugs, such as receptor tyrosine kinase inhibitors (113). This analysis highlights the direct translation of landscape concepts to the development of more targeted therapeutic approaches, particularly if an epigenetic alteration is considered an important therapeutic target (114).

This phenotypic mapping analysis can be further extended within the recently introduced framework of cellular archetypes in tissues (32, 33). Archetypes are defined as extreme phenotypic states spanning the available optimized phenotypic space and thus reflecting the available degrees of freedom of the phenotypic plasticity. The analysis of Wilms tumor patient samples revealed three archetypal phenotypic states, suggesting a more complex landscape distribution *in vivo*. More advanced stages of the disease, accompanied by emergence of *IGF2* LOI, lead to a shift of cell population from a relatively uniform distribution across three archtypes: an epithelial and two mesenchymal states (one of which is also characterized by higher cell proliferation), towards a greater abundance of mesenchymal states versus the epithelial one (31). This result suggested a specific landscape change, potentially due to an increased expression of *IGF2* promoting more aggressive cell behaviors related to cell proliferation and EMT. In spite of this more complex landscape structure, we suggest that phenotypic mapping analysis can still be employed to target specific cell states.

In general, measuring and mapping genetic and epigenetic landscapes onto phenotypic spaces can be an effective strategy to both infer the corresponding phenotypic diversity and target specific subpopulations defined by landscape attractors. Importantly, it can also elucidate the effects of phenotypic plasticity and guide corresponding treatment strategies. If mutations, epigenetic changes or environmental inputs leading to the emergence of pre-cancerous and cancerous states also increase the plasticity within emerging attractors, cell populations may be substantially less targetable due to a larger overlap with wild type attractors and faster escape from a vulnerable area of the attractor (Fig. 3C). This consideration both provides further interpretation for poor prognostic correlates of the more plastic tumor attractors and suggests that cancer treatment should reduce plasticity. More specifically, the clinical intervention would be aimed at reshaping the underlying landscape to first stabilize the cancer specific attractors by perturbing the attractor-defining pathways and molecular circuits, and then eliminate cells occupying these attractors, while minimizing the effect on untransformed cells.

Conclusions and future directions

The studies reviewed here support the utility of the genetic and epigenetic landscape approaches in our understanding of cancer etiology, ultimately allowing us to arrive at a more general and conceptual understanding of carcinogenesis. Importantly, this approach presents a detailed view of the system, which may focus on a specific attractor, but also a combination of multiple attractors and dynamic transitions between them. Conversely, entropy is frequently a less resolved and more global metric of the overall diversity of cell states. Both approaches are useful and frequently complementary. For example, the landscape view may be better at capturing the difference between an increase in phenotypic plasticity, i.e., the ability of cells (and their molecular networks) to dynamically 'explore' diverse states and corresponding phenotypes, from phenotypic diversity, i.e., stable occupancy of different attractors by different cells in the population (Fig. 3C). In both cases, the entropy of the tumor tissue may be higher compared to the parental wild-type tissue state, but the landscapes would be distinct, which can be a crucial level of resolution to determine prognosis and possible interventions. Nevertheless, entropy can still be a convenient metric, particularly in assessing the effects of signaling inputs, mutational and epigenetic perturbations or medical interventions. In particular, changes in entropy [assessed e.g., through different information theory-based measures (Box 1)] can quantitatively characterize the effects of these diverse inputs, or evaluate how long-term processes, such as cellular aging, may affect cellular landscapes.

New experimental approaches may also allow further analysis of causality versus correlation in the relationship between landscape alterations and the emergence of cancer states. For instance, new CRISPR methods can allow recruitment of specific epigenetic modifiers to a targeted locus within the genome and direct control of epigenetic and gene expression variability and attractor properties (30). Studies such as this, combined with further analysis of pre-cancerous states in patients and animal models involving specific perturbations of genetic and epigenetic mechanisms can provide further data to refine our understanding of cancer-specific lanscapes.

Within oncology applications, key advantages of the landscape and entropy concepts lie in an improvement of our understanding of cancer etiology and development of approaches to targeted and personalized cancer treatments. Indeed, more plastic attractors can allow cells to potentially rapidly adjust to both pharmacological and immune therapeutic interventions, as well as enable cells to dynamically switch between diverse phenotypic states required, e.g., for metastatic spread. Therefore, arguably, further approaches may focus on a combination of treatments that would not only attempt to eliminate specific cell sub-populations but also reduce the phenotypic plasticity of tumor cells. In developing these approaches, characterization of the underlying landscapes will be of particular importance. In these efforts, it will be important to refine our understanding of how the landscapes determined on the epigenetic and gene regulation levels can be quantitatively translated into the corresponding descriptions of the activity of the molecular circuits controlling specific phenotypes. Such phenotypic mapping can direct more specific drug development efforts and quantitatively establish the therapeutic windows of specific compounds.

In this Review we primarily focused on DNA methylation as a key epigenetic mechanism whose landscape and entropic analysis has already been performed in diverse biological systems. However, epigenetic control can occur on other levels, including through histone modifications. As rich data accounting for further levels of epigenetic control are gathered, an important challenge will be to understand whether extensive, landscape-based description of all levels of epigenetic and gene regulatory controls is necessary, or whether this complexity and data dimensionality can be reduced to a simpler measurement set. In these efforts, a key role may be played by various machine learning techniques to reveal mutual interdependencies of various molecular control mechanisms and their influences on phenotypic plasticity.

ACKNOWLEGMENTS:

References

1. Fidler IJ, Tumor Heterogeneity and the Biology of Cancer Invasion and Metastasis1. Cancer Research 38, 2651–2660 (1978). [PubMed: 354778]

2. Nowell PC, The Clonal Evolution of Tumor Cell Populations: Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. Science 194, 23–28 (1976). [PubMed: 959840]

3. Jia D, Jolly MK, Kulkarni P, Levine H, Phenotypic Plasticity and Cell Fate Decisions in Cancer: Insights from Dynamical Systems Theory. Cancers (Basel) 9, (2017).

4. Sacchetti A et al. , Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. Elife 10, (2021).

5. Sharma SV et al. , A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell 141, 69–80 (2010). [PubMed: 20371346]

6. Chang CA et al. , Ontogeny and Vulnerabilities of Drug-Tolerant Persisters in HER2+ Breast Cancer. Cancer Discov 12, 1022–1045 (2022). [PubMed: 34911733]

7. Jia D, Jolly MK, Kulkarni P, Levine H, Phenotypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. Cancers (Basel) 9, 70 (2017). [PubMed: 28640191]

8. Zhou JX, Aliyu M, Aurell E, Huang S, Quasi-potential landscape in complex multi-stable systems. Journal of the Royal Society Interface 9, 3539–3553 (2012). [PubMed: 22933187]

9. Wang J, Zhang K, Xu L, Wang E, Quantifying the Waddington landscape and biological paths for development and differentiation. Proceedings of the National Academy of Sciences 108, 8257–8262 (2011).

10. Li C, Wang J, Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. Proceedings of the National Academy of Sciences 111, 14130–14135 (2014).

11. Balázsi G, van Oudenaarden A, Collins JJ, Cellular decision making and biological noise: from microbes to mammals. Cell 144, 910–925 (2011). [PubMed: 21414483]

12. Eling N, Morgan MD, Marioni JC, Challenges in measuring and understanding biological noise. Nature Reviews Genetics 20, 536–548 (2019).

13. Levchenko A, Nemenman I, Cellular noise and information transmission. Current opinion in biotechnology 28, 156–164 (2014). [PubMed: 24922112]

14. Golightly A, Gillespie CS, in In Silico Systems Biology (Springer, 2013), pp. 169–187.

15. Gillespie DT, Hellander A, Petzold LR, Perspective: Stochastic algorithms for chemical kinetics. The Journal of chemical physics 138, 05B201_201 (2013).

16. Ullah M, Wolkenhauer O, Stochastic approaches in systems biology. Wiley Interdisciplinary Reviews: Systems Biology and Medicine 2, 385–397 (2010). [PubMed: 20836037]

17. Grün D, van Oudenaarden A, Design and analysis of single-cell sequencing experiments. Cell 163, 799–810 (2015). [PubMed: 26544934]

18. Kelsey G, Stegle O, Reik W, Single-cell epigenomics: Recording the past and predicting the future. Science 358, 69–75 (2017). [PubMed: 28983045]

19. Shema E, Bernstein BE, Buenrostro JD, Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. Nature genetics 51, 19–25 (2019). [PubMed: 30559489]

20. Li Q et al. , Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. Proc Natl Acad Sci U S A 113, 2672–2677 (2016). [PubMed: 26929366]

21. Pujadas E, Feinberg AP, Regulated noise in the epigenetic landscape of development and disease. Cell 148, 1123–1131 (2012). [PubMed: 22424224]

22. Taherian Fard A, Ragan MA, Modeling the Attractor Landscape of Disease Progression: a Network-Based Approach. Front Genet 8, 48 (2017). [PubMed: 28458684]

23. Tripathi S, Chakraborty P, Levine H, Jolly MK, A mechanism for epithelial-mesenchymal heterogeneity in a population of cancer cells. PLoS Comput Biol 16, e1007619 (2020). [PubMed: 32040502]

24. Cheng Y et al. , Role of metastasis-associated lung adenocarcinoma transcript-1 (MALAT-1) in pancreatic cancer. PloS one 13, e0192264 (2018). [PubMed: 29389953]

25. van Wieringen WN, van der Vaart AW, Statistical analysis of the cancer cell's molecular entropy using high-throughput data. Bioinformatics 27, 556–563 (2011). [PubMed: 21172912]

26. West J, Bianconi G, Severini S, Teschendorff AE, Differential network entropy reveals cancer system hallmarks. Sci Rep 2, 802 (2012). [PubMed: 23150773]

27. Cheng F, Liu C, Shen B, Zhao Z, Investigating cellular network heterogeneity and modularity in cancer: a network entropy and unbalanced motif approach. BMC systems biology 10, 301–311 (2016).

28. Teschendorff AE, Enver T, Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. Nature communications 8, 1–15 (2017).

29. Chaligne R et al. , Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. Nat Genet 53, 1469–1479 (2021). [PubMed: 34594037]

30. Lu A et al. , Reprogrammable CRISPR/dCas9-based recruitment of DNMT1 for site-specific DNA demethylation and gene regulation. Cell Discov 5, 22 (2019). [PubMed: 31016028]

31. Trink A et al. , Geometry of Gene Expression Space of Wilms' Tumors From Human Patients. Neoplasia 20, 871–881 (2018). [PubMed: 30029183]

32. Hausser J, Alon U, Tumour heterogeneity and the evolutionary trade-offs of cancer. Nature Reviews Cancer, 1–11 (2020).

33. Hausser J et al. , Tumor diversity and the trade-off between universal cancer tasks. Nature communications 10, 1–13 (2019).

34. Gupta PB et al. , Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 146, 633–644 (2011). [PubMed: 21854987]

35. Landan G et al. , Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat Genet 44, 1207–1214 (2012). [PubMed: 23064413]

36. Barrett JE et al. , Quantification of tumour evolution and heterogeneity via Bayesian epiallele detection. BMC bioinformatics 18, 1–10 (2017). [PubMed: 28049414]

37. Xie H et al. , Genome-wide quantitative assessment of variation in DNA methylation patterns. Nucleic acids research 39, 4099–4108 (2011). [PubMed: 21278160]

38. Jenkinson G, Pujadas E, Goutsias J, Feinberg AP, Potential energy landscapes identify the information-theoretic nature of the epigenome. Nat Genet 49, 719–729 (2017). [PubMed: 28346445]

39. Jenkinson G, Abante J, Koldobskiy MA, Feinberg AP, Goutsias J, Ranking genomic features using an information-theoretic measure of epigenetic discordance. BMC Bioinformatics 20, 175 (2019). [PubMed: 30961526]

40. Landau DA et al. , Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer cell 26, 813–825 (2014). [PubMed: 25490447]

41. Li S et al. , Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nature medicine 22, 792–799 (2016).

42. Jones PA et al. , Moving AHEAD with an international human epigenome project. Nature 454, 711 (2008). [PubMed: 18685699]

43. Hansen KD et al. , Increased methylation variation in epigenetic domains across cancer types. Nat Genet 43, 768–775 (2011). [PubMed: 21706001]

44. Berman BP et al. , Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina–associated domains. Nature genetics 44, 40 (2012).

45. McDonald OG et al. , Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. Nat Genet 49, 367–376 (2017). [PubMed: 28092686]

46. Jeong M et al. , Large conserved domains of low DNA methylation maintained by Dnmt3a. Nat Genet 46, 17–23 (2014). [PubMed: 24270360]

47. Johnson KS et al. , Gene expression and chromatin accessibility during progressive EMT and MET linked to dynamic CTCF engagement. bioRxiv, (2020).

48. McDonald OG, Wu H, Timp W, Doi A, Feinberg AP, Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. Nat Struct Mol Biol 18, 867–874 (2011). [PubMed: 21725293]

49. Johnstone SE et al. , Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. Cell 182, 1474–1489 e1423 (2020). [PubMed: 32841603]

50. Garraway LA, Lander ES, Lessons from the cancer genome. Cell 153, 17–37 (2013). [PubMed: 23540688]

51. Feinberg AP, Koldobskiy MA, Gondor A, Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. Nat Rev Genet 17, 284–299 (2016). [PubMed: 26972587]

52. Zhen T et al. , RUNX1 and CBFβ-SMMHC transactivate target genes together in abnormal myeloid progenitors for leukemia development. Blood 136, 2373–2385 (2020). [PubMed: 32929473]

53. Li S et al. , Somatic Mutations Drive Specific, but Reversible, Epigenetic Heterogeneity States in AML. Cancer Discovery 10, 1934–1949 (2020). [PubMed: 32938585]

54. Koldobskiy MA et al. , A Dysregulated DNA Methylation Landscape Linked to Gene Expression in MLL-Rearranged AML. Epigenetics 15, 841–858 (2020). [PubMed: 32114880]

55. Levine AJ, The p53 protein plays a central role in the mechanism of action of epigenetic drugs that alter the methylation of cytosine residues in DNA. Oncotarget 8, 7228–7230 (2017). [PubMed: 28129641]

56. Filipczak PT et al. , p53-Suppressed Oncogene TET1 Prevents Cellular Aging in Lung Cancer. Cancer Res 79, 1758–1768 (2019). [PubMed: 30622117]

57. Tovy A et al. , p53 is essential for DNA methylation homeostasis in naive embryonic stem cells, and its loss promotes clonal heterogeneity. Genes & development 31, 959–972 (2017). [PubMed: 28607180]

58. Rampal R et al. , DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. Cell Rep 9, 1841–1855 (2014). [PubMed: 25482556]

59. Marusyk A, Janiszewska M, Polyak K, Intratumor heterogeneity: The rosetta stone of therapy resistance. Cancer cell 37, 471–484 (2020). [PubMed: 32289271]

60. Tsankov AM et al. , Loss of DNA methyltransferase activity in primed human ES cells triggers increased cell-cell variability and transcriptional repression. Development 146, (2019).

61. Singer ZS et al. , Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol Cell 55, 319–331 (2014). [PubMed: 25038413]

62. Costa Y et al. , NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 495, 370–374 (2013). [PubMed: 23395962]

63. Doege CA et al. , Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature 488, 652–655 (2012). [PubMed: 22902501]

64. Gao Y et al. , Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell stem cell 12, 453–469 (2013). [PubMed: 23499384]

65. Silva J et al. , Nanog is the gateway to the pluripotent ground state. Cell 138, 722–737 (2009). [PubMed: 19703398]

66. Sharma SV et al. , A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell 141, 69–80 (2010). [PubMed: 20371346]

67. Rockne RC et al. , State-Transition Analysis of Time-Sequential Gene Expression Identifies Critical Points That Predict Development of Acute Myeloid Leukemia. Cancer Res 80, 3157–3169 (2020). [PubMed: 32414754]

68. Pelletier J et al. , Germline mutations in the Wilms' tumor suppressor gene are associated with abnormal urogenital development in Denys-Drash syndrome. Cell 67, 437–447 (1991). [PubMed: 1655284]

69. Schedl A, Hastie N, Multiple roles for the Wilms' tumour suppressor gene, WT1 in genitourinary development. Mol Cell Endocrinol 140, 65–69 (1998). [PubMed: 9722170]

70. Park J et al. , Switch-like enhancement of epithelial-mesenchymal transition by YAP through feedback regulation of WT1 and Rho-family GTPases. Nature communications 10, 1–15 (2019).

71. Ohnishi K et al. , Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. Cell 156, 663–677 (2014). [PubMed: 24529372]

72. Reichert M, Rustgi AK, Pancreatic ductal cells in development, regeneration, and neoplasia. J Clin Invest 121, 4572–4578 (2011). [PubMed: 22133881]

73. Storz P, Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. Nature reviews Gastroenterology & hepatology 14, 296–304 (2017). [PubMed: 28270694]

74. Wei D et al. , KLF4 is essential for induction of cellular identity change and acinar-to-ductal reprogramming during early pancreatic carcinogenesis. Cancer cell 29, 324–338 (2016). [PubMed: 26977883]

75. Giroux V, Rustgi AK, Metaplasia: tissue injury adaptation and a precursor to the dysplasia–cancer sequence. Nature Reviews Cancer 17, 594–604 (2017). [PubMed: 28860646]

76. Ding LY et al. , Epigenetic silencing of AATK in acinar to ductal metaplasia in murine model of pancreatic cancer. Clin Epigenetics 12, 87 (2020). [PubMed: 32552862]

77. Spechler SJ, Souza RF, Barrett's esophagus. N Engl J Med 371, 836–845 (2014). [PubMed: 25162890]

78. Barros R et al. , CDX2 autoregulation in human intestinal metaplasia of the stomach: impact on the stability of the phenotype. Gut 60, 290–298 (2011). [PubMed: 21148572]

79. Luebeck EG et al. , Identification of a key role of widespread epigenetic drift in Barrett's esophagus and esophageal adenocarcinoma. Clin Epigenetics 9, 113 (2017). [PubMed: 29046735]

80. Wu J et al. , A single-cell survey of cellular hierarchy in acute myeloid leukemia. J Hematol Oncol 13, 128 (2020). [PubMed: 32977829]

81. Streppel M et al. , Next-generation sequencing of endoscopic biopsies identifies ARID1A as a tumor-suppressor gene in Barrett's esophagus. Oncogene 33, 347–357 (2014). [PubMed: 23318448]

82. Livshits G et al. , Arid1a restrains Kras-dependent changes in acinar cell identity. Elife 7, e35216 (2018). [PubMed: 30014851]

83. Treger TD, Chowdhury T, Pritchard-Jones K, Behjati S, The genetic changes of Wilms tumour. Nature Reviews Nephrology 15, 240–251 (2019). [PubMed: 30705419]

84. Conteduca V et al. , Barrett's esophagus and esophageal cancer: an overview. International journal of oncology 41, 414–424 (2012). [PubMed: 22615011]

85. Taube JH et al. , Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proceedings of the National Academy of Sciences 107, 15449–15454 (2010).

86. Rhim AD et al. , EMT and dissemination precede pancreatic tumor formation. Cell 148, 349–361 (2012). [PubMed: 22265420]

87. Park J et al. , Switch-like enhancement of epithelial-mesenchymal transition by YAP through feedback regulation of WT1 and Rho-family GTPases. Nat Commun 10, 2797 (2019). [PubMed: 31243273]

88. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A, Information transduction capacity of noisy biochemical signaling networks. science 334, 354–358 (2011). [PubMed: 21921160]

89. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A, Information transduction capacity of noisy biochemical signaling networks. Science 334, 354–358 (2011). [PubMed: 21921160]

90. Rhee A, Cheong R, Levchenko A, The application of information theory to biochemical signaling systems. Phys Biol 9, 045011 (2012). [PubMed: 22872091]

91. Brennan MD, Cheong R, Levchenko A, Systems biology. How information theory handles cell signaling and uncertainty. Science 338, 334–335 (2012). [PubMed: 23087235]

92. Uda S et al. , Robustness and compensation of information transmission of signaling pathways. Science 341, 558–561 (2013). [PubMed: 23908238]

93. Levchenko A, Nemenman I, Cellular noise and information transmission. Curr. Opin. Biotechnol 28, 156–164 (2014). [PubMed: 24922112]

94. Selimkhanov J et al. , Systems biology. Accurate information transmission through dynamic biochemical signaling networks. Science 346, 1370–1373 (2014). [PubMed: 25504722]

95. Hansen AS, O'Shea EK, Limits on information transduction through amplitude and frequency regulation of transcription factor activity. Elife 4, (2015).

96. Suderman R, Bachman JA, Smith A, Sorger PK, Deeds EJ, Fundamental trade-offs between information flow in single cells and cellular populations. Proc Natl Acad Sci U S A 114, 5755–5760 (2017). [PubMed: 28500273]

97. Ellison D et al. , Cell–cell communication enhances the capacity of cell ensembles to sense shallow gradients during morphogenesis. Proceedings of the National Academy of Sciences 113, E679–E688 (2016).

98. Suderman R, Bachman JA, Smith A, Sorger PK, Deeds EJ, Fundamental trade-offs between information flow in single cells and cellular populations. Proceedings of the National Academy of Sciences 114, 5755–5760 (2017).

99. Mendelsohn AR, Larrick JW, Epigenetic drift is a determinant of mammalian lifespan. Rejuvenation research 20, 430–436 (2017). [PubMed: 28942711]

100. Maegawa S et al. , Widespread and tissue specific age-related DNA methylation changes in mice. Genome research 20, 332–340 (2010). [PubMed: 20107151]

101. Teschendorff AE, West J, Beck S, Age-associated epigenetic drift: implications, and a case of epigenetic thrift? Human molecular genetics 22, R7–R15 (2013). [PubMed: 23918660]

102. Takeshima H, Ushijima T, Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. NPJ Precision Oncology 3, 1–8 (2019). [PubMed: 30623031]

103. Zheng SC, Widschwendter M, Teschendorff AE, Epigenetic drift, epigenetic clocks and cancer risk. Epigenomics 8, 705–719 (2016). [PubMed: 27104983]

104. Horvath S, DNA methylation age of human tissues and cell types. Genome biology 14, 3156 (2013).

105. Teschendorff AE, Jones A, Widschwendter M, Stochastic epigenetic outliers can define field defects in cancer. BMC bioinformatics 17, 178 (2016). [PubMed: 27103033]

106. Teschendorff AE, Morabito SJ, Kessenbrock K, Meyer K, Integrated single-cell potency and expression landscape in mammary epithelium reveals novel bipotent-like cells associated with breast cancer risk. bioRxiv, 496471 (2018).

107. Meir Z, Mukamel Z, Chomsky E, Lifshitz A, Tanay A, Single-cell analysis of clonal maintenance of transcriptional and epigenetic states in cancer cells. Nature Genetics 52, 709–718 (2020). [PubMed: 32601473]

108. Lawson CD, Ridley AJ, Rho GTPase signaling complexes in cell migration and invasion. J Cell Biol 217, 447–457 (2018). [PubMed: 29233866]

109. Holmes WR, Park J, Levchenko A, Edelstein-Keshet L, A mathematical model coupling polarity signaling to cell adhesion explains diverse cell migration patterns. PLoS Comput Biol 13, e1005524 (2017). [PubMed: 28472054]

110. Kapuy O et al. , System-level feedbacks control cell cycle progression. FEBS Lett 583, 3992–3998 (2009). [PubMed: 19703449]

111. Walensky L, BCL-2 in the crosshairs: tipping the balance of life and death. Cell Death & Differentiation 13, 1339–1350 (2006). [PubMed: 16763614]

112. Youle RJ, Strasser A, The BCL-2 protein family: opposing activities that mediate cell death. Nature reviews Molecular cell biology 9, 47–59 (2008). [PubMed: 18097445]

113. Vanaja KG, Timp W, Feinberg AP, Levchenko A, A loss of epigenetic control can promote cell death through reversing the balance of pathways in a signaling network. Molecular cell 72, 60–70. e63 (2018). [PubMed: 30244832]

114. Zhong H et al. , Overproduction of IGF-2 drives a subset of colorectal cancer cells, which specifically respond to an anti-IGF therapeutic antibody and combination therapies. Oncogene 36, 797–806 (2017). [PubMed: 27399333]

Box 1.

Mathematical analysis of stochastic biological systems

Biological systems are governed by underlying interactions that could be described mathematically. These interactions can be stochastic due to the variability of the participating elements (molecular modifications, molecular and cellular numbers, etc.). The mathematical approach frequently used to capture stochastic interactions is the Langevin approximation based on the Ordinary Differential Equations (ODEs) describing temporal dynamics of variables x_i, using the function $f(x)$ *of the vector x of all variables* x_j, with the added noise terms $\eta(x_i)$:

$$\frac{dx_i}{dt} = f_i(x) + \eta_i(x_i) \tag{1}$$

An alternative but equivalent description that is particularly useful for systems with very small values of x_i that are expressed as whole numbers is to consider the dynamics of probabilities of x_i, rather than the dynamics of x_i themselves. This is done using the so-called Master Equation based on the assumption that the underlying processes are Markovian (memoryless):

$$\frac{dP(x_i)}{dt} = \sum_j k_{ij} P(x_j) \tag{2}$$

Unfortunately, other than for a few relatively simple systems, the mechanistic information that would define the functions $f(x)$ and k_{ij} is largely incomplete for most biological systems. Therefore, it is frequently hard to use these models for real biological systems to characterize their dynamic behavior and thus landscapes. Instead, it has proven easier to experimentally measure the distributions $P(x_i)$. Therefore, the problem can be inverted, i.e., reconstruct the underlying regulatory landscape from the knowledge of $P(x_i)$. The recipe for doing so is provided by statistical mechanics (Boltzmann-Gibbs distribution), and in particular the relationship between the energy $U(x_i)$ associated with a specific x_i and the probability of this value of the variable (the energy here is quasi-potential, corresponding to measuring the energy in units of $k_B T$, i.e., the product of the boltzmann constant and temperature):

$$U(x_i) = -\ln P(x_i) - \ln\left(\sum_j e^{-U(x_j)}\right) \tag{3}$$

If the distribution of different values that can be taken by variables x_I is constant (i.e., the landscape does not change) then the second term in (3) is constant and, since the energy $U(x_i)$ is only determined up to a constant, it can be omitted for the analysis of a constant landscape:

$$U(x_i) \simeq -\ln P(x_i) \tag{4}$$

The relationship (3) can be used to directly define the energy (quasi-potential) landscape corresponding to the gene expression and epigenetic landscapes and to relate it to the entropy of the system, again up to a constant:

$$H(x) = -\sum_i P(x_i) \log P(x_i) \simeq -\sum_i P(x_i) \, \ln P(x_i) \simeq \\ < U(x_i) > \tag{5}$$

This relationship, linking the entropy and average energy $< U >$ for a system, emphasizes the 'global' nature of the entropy measure, involving summation over multiple states. Thus, as illustrated in Fig.1, different landscapes may have the same entropy and thus cannot be uniquely distinguished using this function of the state. Conversely, the distributions captured by $P(x_i)$ and $U(x_i)$, can be more informative.

Finally, the mutual information is defined as the change of entropy (and thus the underlying landscape) of a state, $\Delta H(x)$, given a regulatory event or a signal. If the entropy decreases the mutual information is positive, which provides a measure of comparison between two distributions. Another important measure of dissimilarity between two distributions $P(x_i)$ and $Q(x_i)$ is the frequently used Kullback-Leibler (KL) divergence:

$$D_{KL}(P\|Q) = \sum_i P(x_i) \log\left(\frac{P(x_i)}{Q(x_i)}\right) \tag{6}$$

Mutual information is thus the KL divergence between the joint distribution of the variable values prior and following the event (or signal) and the product of their marginal distributions. A symmetrized version of the KL divergence of particular value in comparing epigenetic landscapes is the Jensen-Shannon distance, defined as:

$$\mathrm{JSD}(P \parallel Q) = \frac{1}{2} \, D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(Q \parallel M) \text{ where } M = \frac{1}{2} \\ (P + Q) \tag{7}$$
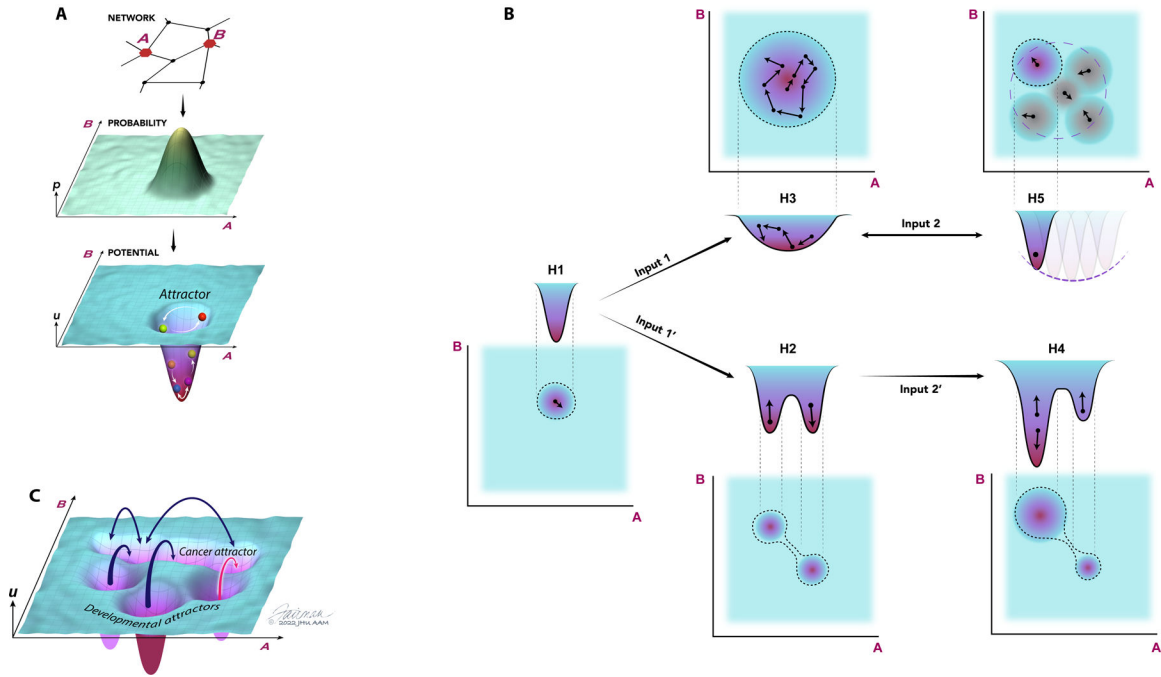
Figure 1. Gene expression and epigenetic landscapes control normal and cancer cell functions.
(**A**) Gene regulatory networks and availability of genes for expression can define the probabilistic distributions of proteins expressed within a cell population. In the example here, the network of interacting proteins that includes the molecules A and B (top) and the underlying epigenetic control determining the availability of the corresponding genes for expression define the distribution of expression of A and B (middle). This probability distribution can be experimentally measured and converted into a gene expression landscape by calculating the corresponding quasi-potential distribution (bottom, see Box). The epigenetic landscape can be similarly determined by experimentally measuring the probabilistic distributions of epi-alleles, DNA methylation marks at specific loci or by performing other measurements of epigenetic regulation across populations of cells and tissues, and then also converting these probability distributions into corresponding underlying quasi-potential landscapes. The landscape analysis allows conceptual accounting for abundance and dynamics of molecular species, shown here as a trajectory of particle inside a quasi-potential well, with the particle position defined by the current concentrations of A and B that can change probabilistically in time, *with the quasi-potential wells interpreted as the landscape attractors*. (**B**) Various scenarios of landscape alterations and the corresponding changes in the molecular distributions, shown as joint distributions of the molecules A and B and the corresponding entropies H1–5. Implementation of these scenarios in the context of carcinogens is extensively illustrated and discussed in the text. Oncogenic mutations of epigenetic modifiers and modulators or environmental inputs can lead to formation of new stable attractors with the overall entropy H2 greater than the original entropy H1 (H2>H1) generating phenotypic *heterogeneity* (Input 1') or, alternatively, enlarge the existing attractor with the new entropy H3>H1, generating a more plastic state (phenotypic *plasticity*), with cells capable of stochastically and dynamically 'exploring' this attractor and thus transiently adopting different phenotypes. Note that in

both cases entropy increases vs H1, and it is possible that H2=H3, thus making entropy less discriminating than the full landscape picture in the analysis of cell states. These new landscapes can be further altered by oncogenic and environmental inputs, so that one of the attractors becomes dominant (Inout 2'), associated with a lower entropy value (H4<H2), or, alternatively, with the narrowing of the wider (and more plastic) attractor (Inout 2, H5<H3). Again, it is possible that H4=H5, requiring the landscape analysis rather than entropy analysis alone for full characterization. The narrowing of the wide attractor due to either environmental or intrinsic inputs (Input 2) is frequently reversible and context dependent, further elaborating the more plastic overall state (transient nature of Input 2 described by a bidirectional arrow). The transiently occupied attractors can be simultaneously occupied by discerns cells in the population. Small arrows correspond to stochastic fluctuations of molecular concentrations within individual attractors. **(C)** Gene regulation and epigenetic landscapes of cancer cells can be complex and have multiple attractors, corresponding to distinct and stable cell states and phenotypes, which may be reshaped by oncogenic mutations, cell aging, environmental inputs and other perturbations, leading to mutual accessibility of the attractors, more plastic cell states and an increase in the phenotypic plasticity.
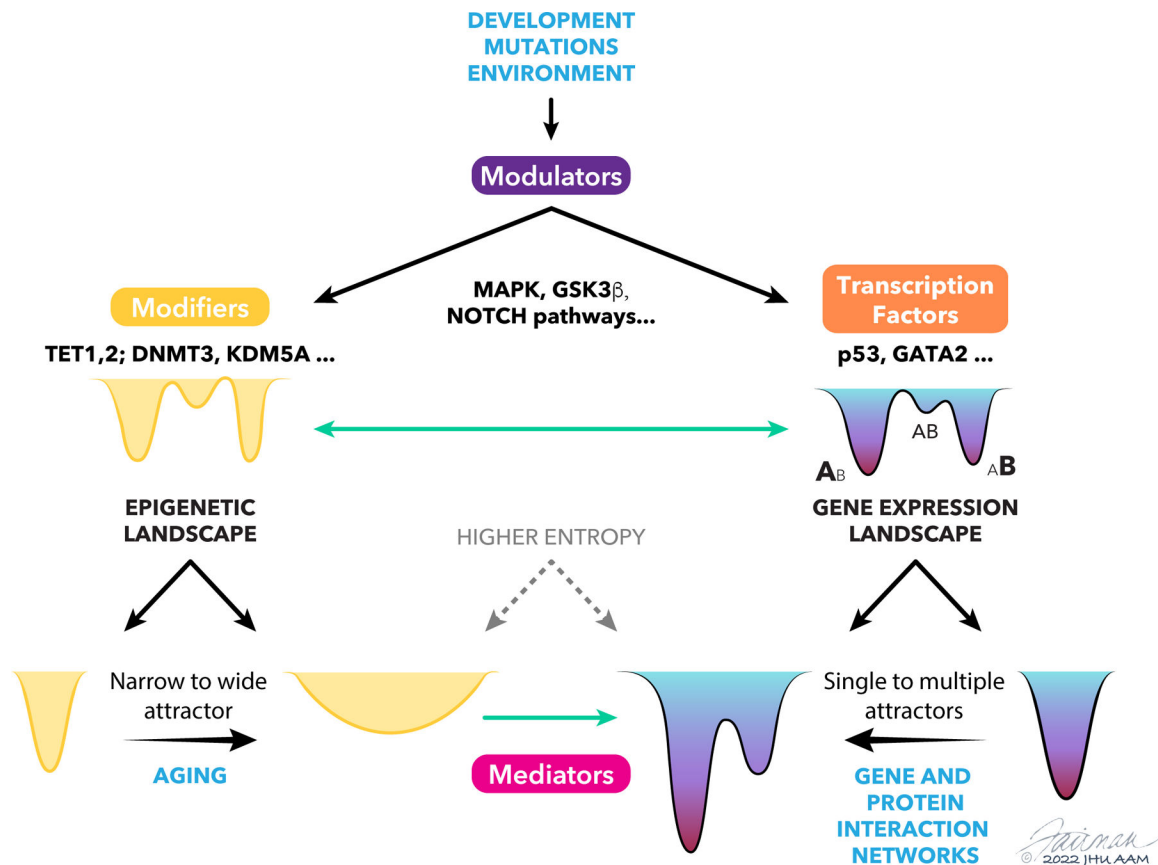
Figure 2. Interplay between epigenetic and gene expression landscapes.
Developmental and environmental factors and genetic mutations can impact diverse
modulators of epigenetic control and gene expression, such as signaling and cell
communication networks, frequently leading to diversification of cell states. These
modulators may directly impact modulators of epigenetic states, such as DNA demethylases,
and gene expression, such as transcription factors, which also can directly interact with each
other. Examples of these molecular regulators discussed in the text are shown here. The
result is alterations of the epigenetic and gene regulation landscapes that are tightly coupled,
e.g., through the action of mediators of epigenetic control, influencing accessibility of genes
for regulation, and the magnitude and variability of gene expression. Certain additional
inputs may be more specific to each of the landscapes, such as the epigenetic drift with
cell aging primarily leading to a widening of the landscape attractors, higher plasticity and
higher entropy of the state, and protein-protein interaction and gene regulatory networks,
stabilizing various attractors and serving to decrease the plasticity and entropy.
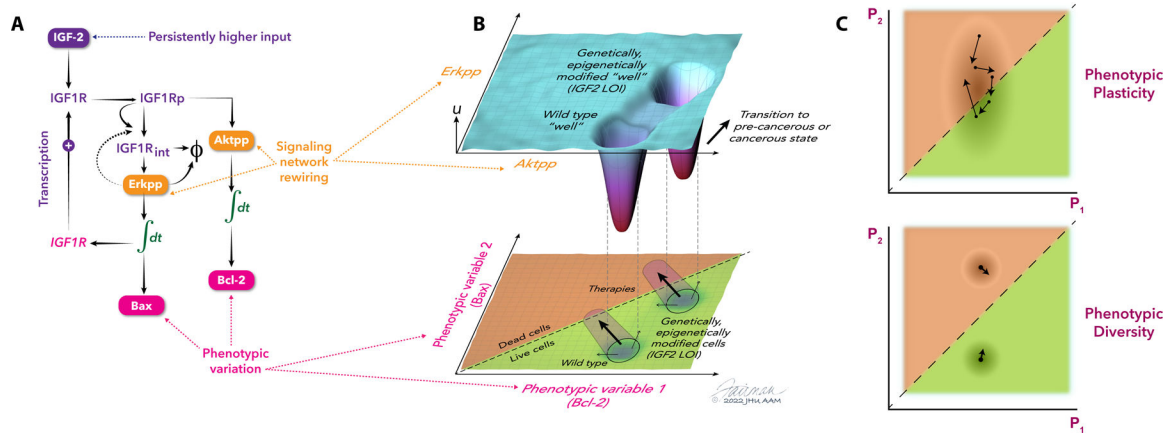
Figure 3. Connection between an epigenetic landscape and variable phenotypic outcomes.
(A) An epigenetic alteration – loss of imprinting (LOI) of the insulin-like growth factor 2 gene (*IGF2*), implicated in Wilms tumor, doubling of the signaling input, can lead to rewiring of the signaling network activated by the IGF2 receptor IGF1R (depicted as IGF1Rp) through altered receptor trafficking (IGF1Rint), degradation (ϕ) and altered balance of activation of the downstream signaling pathways activating Erk (Erkpp) and Akt (Aktpp) kinases. Rebalancing of Erk and Akt activities translates into transcriptional upregulation of IGF1R, a higher proliferation rate but also rebalancing of pro- and anti-apoptotic protein abundances (BAX versus Bcl-2, respectively) leading to an increased propensity for cell death (113). The integral signs represent integration over time of signaling activities. **(B)** The landscape alterations that correspond to a change in phenotype (upper panel) is the altered expression and activity of signaling pathway molecules (and thus gene regulatory landscapes in the lower panel) in response to alteration of epigenetic landscapes (IGF2 loss of imprinting, LOI). This leads to emergence of a new attractor in addition to the wild type attractor, resulting in a mosaic wild type/LOI cell distribution in the tissue. This landscape alteration can be 'mapped' onto, for example, the apoptosis phenotype-defining network by a quantitative analysis of the dependence of the BCL family protein distributions on the signaling inputs, thus enabling a direct translation of the landscape alterations into phenotype distributions. In this example, the mapping can be visualized as wild type and LOI cell distributions mapped with respect to the areas of cell survival and death on the (BAX, Bcl-2) phenotypic plane, suggesting how treatments targeting LOI cells may be developed to spare the wild type cells. Arrows in the lower panel represent the effect of drugs, such as IGF1R inhibitors shifting the landscape and phenotypic distribution towards the boundary separating survival and death, with the red areas depicting the effect on the wild type and LOI cell populations. **(C)** A more general view of landscape mapping onto the apoptosis phenotypic plane. By analogy with Fig. 1B, one can contrast mapping of a large attractor versus two more limited attractors, representing the difference between a plastic and stochastic state (phenotypic plasticity) versus a state with two alternative stable attractors (phenotypic diversity). The more plastic state can allow cells to escape from the death area to the survival area even in the presence of a treatment (such as in (B)), by stochastically 'exploring' the available attractor, whereas a combination of more stable attractors (with the same overall entropy as the more plastic state) can allow for selective targeting by one but not the other attractor. Therefore, the treatment strategy

suggested in (B) may benefit from the initial intervention stabilizing smaller attractors within a larger one and thus decreasing the plasticity of the state, particularly through epigenetic perturbations.