Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Decoding transcriptional states in cancer

Jasper Wouters[1,2,*], Zeynep Kalender Atak[1,2,*] and Stein Aerts[1,2]

Gene regulatory networks determine cellular identity. In cancer, aberrations of gene networks are caused by driver mutations that often affect transcription factors and chromatin modifiers. Nevertheless, gene transcription in cancer follows the same *cis*-regulatory rules as normal cells, and cancer cells have served as convenient model systems to study transcriptional regulation. Tumours often show regulatory heterogeneity, with subpopulations of cells in different transcriptional states, which has important therapeutic implications. Here, we review recent experimental and computational techniques to reverse engineer cancer gene networks using transcriptome and epigenome data. New algorithms, data integration strategies, and increasing amounts of single cell genomics data provide exciting opportunities to model dynamic regulatory states at unprecedented resolution.

Addresses
[1] Laboratory of Computational Biology, VIB Center for Brain & Disease Research, Leuven, Belgium
[2] Department of Human Genetics, KU Leuven (University of Leuven), Leuven, Belgium

Corresponding author: Aerts, Stein (stein.aerts@kuleuven.vib.be)
[*] These authors contributed equally to the manuscript.

Current Opinion in Genetics & Development 2017, **43**:82–92

This review comes from a themed issue on **Genome architecture and expression**

Edited by **Bart Deplancke** and **Charles Sagerstrom**

http://dx.doi.org/10.1016/j.gde.2017.01.003

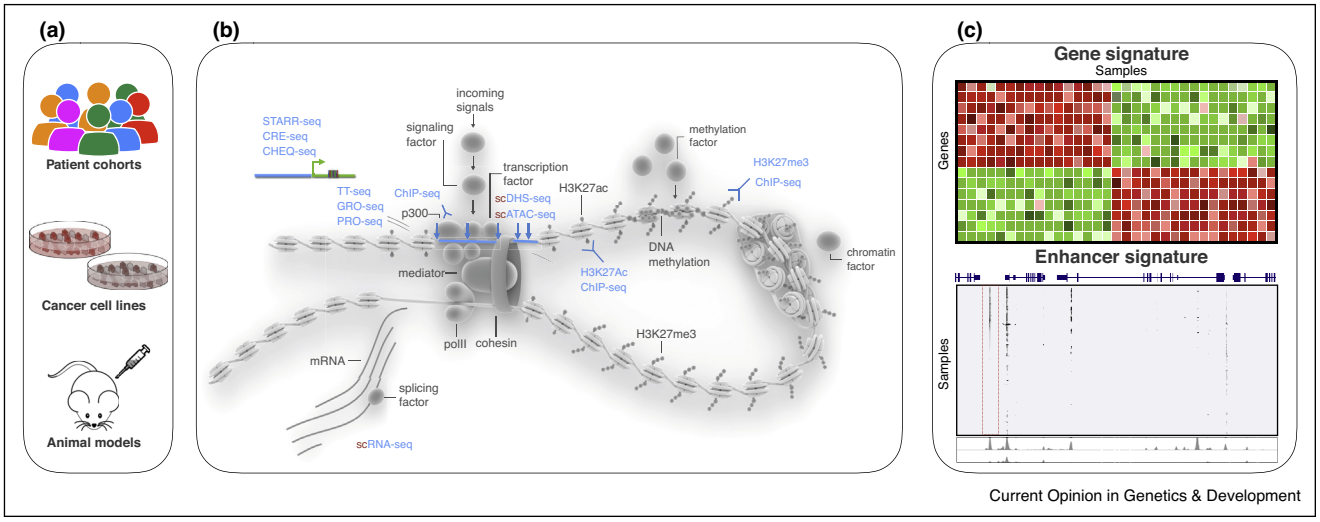
Cancer transcriptional states emerge from gene regulatory network perturbations

Genomic re-sequencing of tumour samples revealed that different patients often share one or two strong driver mutations (*e.g.* KRAS gain of function plus TP53 loss of function mutations are typical for pancreatic cancer [1]), together with a unique combination of less frequent driver mutations [2]. Transcriptome and epigenome profiling on the other hand, often result in defined clusters of reproducible regulatory subtypes. Thus, chaos in the cancer genome is often canalized into regulatory order; and cancer cells seem to follow the same rules for transcriptional regulation as normal cells, albeit with aberrant/ectopic combinations of transcription factors, co-factors, and genomic enhancers (Figure 1; [3]). In melanoma for example, three clusters of transcriptional states have been observed repeatedly (proliferative/pigmentation; invasive/MITF-low; and immune), but these do not show any obvious correlation to the three driver mutation groups (BRAF gain, NRAS gain, or NF1 loss) [4,5]. Also in many other cancers, transcriptomes are observed with a seemingly limited influence of the underlying genomic mutations, such as glioblastoma and colorectal cancer [6,7]. There exist a few exceptions though, where a driver mutation dominantly causes a specific transcriptomic state. For example, leukemic cells with a chromosomal translocation involving KMT2A (also known as MLL fusions), consistently yield a transcriptional state distinct from all other acute lymphoblastic leukaemia subtypes [8,9]. Likewise, sarcoma cells with the EWSR1 fusion oncogene have a specific transcriptome, unique amongst all subtypes of Ewing sarcoma [10]. Nonetheless, the vast majority of transcriptional states, and emerging phenotypic behaviour, seems to be a combination of both the initial genomic aberrations, likely as a consequence of mutations directly affecting the regulatory program (*i.e.* mutated transcription factors, co-factors, signalling molecules, and *cis*-regulatory regions; reviewed in Ref. [11]), and the influence of the tumour microenvironment.

The observation that multiple varying genetic alterations can lead to similar phenotypes is related to the concept of 'cancer attractor states', which represent stable lower-energy valleys within a Waddingtonian landscape of all potential GRN configurations [12–14]. Some of these attractors are shared across cancer types, such as the mesenchymal transition attractor, the mitotic chromosomal instability attractor, and the lymphocyte-specific attractor [15]. Not unexpectedly, regulatory subnetworks controlling cell proliferation and cell cycle, DNA damage response, and immune response are consistently found in pan-cancer transcriptome analyses [16–18].

Importantly, cancer cells are not necessarily fixated in a 'stable' state, but they may switch dynamically between alternate states, under the influence of the microenvironment such as hypoxia [19], or induced by drugs [20,21]. One of the best known state transitions in cancer is the epithelial-to-mesenchymal transition, causing cancer cells from epithelial origin to transition into a migratory, drug-resistant state [22]; and a comparable 'phenotype switching' that occurs in non-epithelial tumours such as melanoma and glioblastoma [23]. Such state transitions underlie tumour regulatory heterogeneity and understanding these regulatory programs can be highly important to invent effective therapeutic strategies [13,24].

Figure 1



Decoding cancer transcriptional states: biological systems, experimental methods and data. **(a)** Different types of samples and model systems can be used for constructing gene regulatory networks in cancer including patient cohorts, cell lines and animal models. **(b)** These systems are used to profile transcriptome and epigenome with the methods indicated in blue. These methods provide information on chromatin accessibility (DHS-seq, scDHS-seq, ATAC-seq and scATAC-seq), gene expression (RNA-seq and scRNA-seq), histone modifications (H3K27ac, H3K27me3, TF ChIP-seq), enhancer reporter activity (STARR-seq, CRE-seq, CHEQ-seq), and enhancer RNA (TT-seq, GRO-seq, PRO-seq). **(c)** Depending on the experimental method used, findings can be summarized in gene signatures or enhancer signatures.

Indeed, interfering with cancer gene networks using 'network drugs' could be an interesting avenue, for example by 'pushing' cells towards a transcriptional state that is vulnerable to a particular drug [25].

Here, we will review high-throughput and computational approaches to study cancer regulatory genomics; some studies are inspired by clinically relevant features, such as escape from apoptosis, DNA damage, drug resistance, invasive behaviour, or immune escape; while many other interesting studies use cancer cells as a convenient model system to study human transcription and chromatin.

Transcriptome profiling to reverse engineer cancer networks

A commonly used approach to infer cancer gene regulatory networks from high-throughput transcriptomic data starts by clustering samples according to sample-wise correlations, followed by the definition of subtype-specific gene signatures using statistical tests for differential gene expression. GeneSigDB, MSigDB, and OncoMine [26–28] contain thousands of cancer-related gene signatures, curated from the literature and online databases. The consequent downstream analysis of a cancer gene signature can involve pathway and Gene Ontology enrichment analysis (*e.g.* WebGestalt, HumanMine, Ingenuity Pathway Analysis [29,30]). Next, to infer master regulators and to predict their candidate target genes, two types of bioinformatics approaches are commonly used. Firstly, co-expression networks can be inferred by gene–gene co-expression correlations, and can be further structured into TF-target hierarchies using a variety of approaches that infer dependencies between TFs and candidate target genes (Figure 2a). Examples of methods that use gene expression correlations (*e.g.* ARACNe [31]), Boolean or Bayesian networks (BC3NET [32]), differential equations (GRNInfer [33]), or machine learning (*e.g.* GENIE3 [34]); as reviewed by Liu [35•] and benchmarked in Ref. [36]. Hallmark studies include Carro *et al.*, who applied ARACNe on high-grade gliomas with increased mesenchymal gene expression [37•], inferring a network controlled by 53 potentially important TFs, including bHLH-B2, C/EBP, FOSL2, RUNX1 and STAT3, some of which were experimental validated in mouse models. In similar research, Gatta *et al.* investigated TLX-connected oncogenic transcriptional networks in T-ALL, and used the ARACNe algorithm and GSEA to identify RUNX1 as an important tumour suppressor, which is often mutated in T-ALL [38•].

A complementary bioinformatics strategy to analyse the regulatory underpinnings of a cancer gene signature is based on *cis*-regulatory sequence analysis (Figure 2b). Overall these methods exploit the fact that master regulator TFs regulate their target genes, or *regulons*, by binding sequence-specific *cis*-regulatory elements near them. Methods like RSAT [39], OPOSSUM [40], PSCAN [41], DIRE [42], and iRegulon [43] predict enriched motifs across the upstream regions of all genes in a gene set (these algorithms are benchmarked in Ref. [43]). Recent improvements made such methods more powerful by including cross-species comparisons, by using larger collections of