

Decoding transcriptional states in cancer

Jasper Wouters^{*,1,2}, Zeynep Kalender Atak^{*,1,2}, Stein Aerts^{1,2}

¹ Laboratory of Computational Biology, VIB-KU Leuven Center for Brain & Disease Research, KU Leuven (University of Leuven), Leuven, Belgium

² Department of Human Genetics, KU Leuven (University of Leuven), Leuven, Belgium

* These authors contributed equally to the manuscript.

Abstract

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. Tumors 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.

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 [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 sub-networks 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]. 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 (GRNIinfer [33]), or machine learning (e.g. GENIE3 [34]); as reviewed by Liu [35]* and benchmarked in [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 [43]). Recent improvements made such methods more powerful by including cross-species comparisons, by using larger collections of candidate position weight matrices (*cis*BP contains around 6500 matrices, TRANSFAC approximately 5500 matrices, iRegulon uses more than 18.000 PWMs), and by incorporating enrichment of 'regulatory data tracks' from public resources (e.g., ENCODE and Roadmap Epigenomics [44,45]). Drawing a gene regulatory network using motif enrichment results (Figure 2d) can be performed using *ad-hoc* parsers; or by using tools like iRegulon that operate inside the Cytoscape framework and automatically convert motif-based predictions into a gene regulatory network [43]. One such network, in T-ALL, was constructed by Sanda *et al.*, using a combination of motif enrichment analyses, GSEA, TF ChIP-seq and experimental validations, and was centred around key regulators TAL1, GATA3 and RUNX1 [46]*.

To analyse more specific regulatory programs underlying certain cancer-related processes such as disease progression, metastasis, or therapy resistance, transcriptome profiling can be performed on *in vitro* cancer cell lines or animal models, measuring changes in response to genetic perturbations or drug treatments [20,47*]. Such model systems are also useful to experimentally validate gene regulatory predictions, for example by knock-down and/or overexpression of the potential master regulators using RNAi (repression) or CRISPRi (repression, activation or knockout). Recently, high-throughput perturbation screens have emerged, allowing independent assays to identify candidate regulators of cancer networks using genome-wide shRNA or CRISPR-Cas libraries [48–50]. For robust normalisation, minimising false positives/negatives and accounting for different screen libraries, specific computational methods have been developed, such as RSA or RIGER for RNAi screens [51,52] and MAGeCK or NEST for CRISPRi [53,54]. Interestingly, these screens can be performed in a very specific context, such as during the development of resistance to chemotherapeutics [50].

Although transcriptome-based studies are technically straightforward and can be highly effective, accurate inference of regulatory interactions from transcriptome profiling data often remain limited because of the large intergenic and intronic space around a gene's transcription start site, and the many-to-many relationship between regulatory regions and target genes. To overcome this limitation, epigenome-profiling is used more and more to directly assess deviations in the activity of promoters and enhancers, either alone or in combination with the transcriptome.

Epigenome profiling to reverse engineer cancer networks

Promoters and enhancers constitute the central processing nodes within a gene regulatory network, serving as docking stations for combinations of transcription factors, and connecting these inputs to a transcriptional output. Analogous to cancer-state gene signatures (derived from transcriptomics data, see above), epigenome profiling can be used to derive "enhancer signatures", which are sets of co-regulated enhancers in a particular state (Figure 2b). Taking advantage of the chromatin changes at active enhancers, multiple biochemical techniques have been developed to identify enhancer signatures underlying a certain cellular state, and enhancers are taking a central place in cancer research (Reviewed by Sur and Taipale [55]**). Some of the key epigenomic profiling methods that can be used to predict enhancers include DHS-seq and ATAC-seq to identify accessible chromatin; ChIP-seq against histone modifications (see Table 2 for an overview of histone modifications near regulatory elements; reviewed in [56] and [57]), such as H3K27ac to find active regulatory element, H3K4me1 for enhancers and H3K4me3 for promoters ([56,57]); bisulfite sequencing (e.g. WGBS [58], RRBS [59]), methylated DNA immunoprecipitation (MeDIP[60]) followed by sequencing, or array-based DNA methylation analyses (e.g. Infinium's BeadChip arrays) to identify hyper- (mostly

inactive) or hypomethylated (mostly active) regions (reviewed in [61]); and nascent RNA-seq (e.g., GRO-seq [62], PRO-seq [62], PRO-cap [63], TT-seq [64]) to find active enhancers by their bidirectional transcription (Figure 1). By combining DNA methylation, ChIP-, DHS- and ATAC-seq data of breast, prostate, and kidney tumor tissues, as compared to normal tissues, Rhie et al. identified more than 25,000 differentially-methylated enhancers ([65]). Using an in-house developed method, TENET, based on the inverse correlation of DNA methylation and gene expression, the authors identified putative master regulators in breast (GATA3, FOXA, and ESR1), prostate (HOXC6 and DLX1) and kidney (ZNF395) cancer. Denny *et al.* employed ATAC-seq to decipher regulatory changes associated with metastatic progression in small cell lung cancer [66]**. By comparing primary and metastatic tumours from genetically engineered mouse models they discovered that regions with increased chromatin accessibility at metastasis are enriched for NFIB motifs; and this finding lead to the experimentally validation of NFIB as master regulator of the metastatic state. Liu et al. used RNA-seq gene expression and chromatin ChIP-seq data (H3K27ac and H3K4me1) on glioblastoma cell lines, patient derived cultures and primary patient samples to characterize the regulatory cascade downstream of EGFR mutations, and identified SOX9 and FOXG1 as master regulators through motif discovery [47]. Rendeiro *et al.* applied ATAC-seq to a cohort of chronic lymphocytic leukaemia samples and identified subtype specific epigenome signatures [67]. Finally, in Verfaillie *et al.* [5], we used H3K27Ac ChIP-seq and FAIRE-seq across a melanoma cohort and identified two distinct epigenomic states in melanoma. Again, motif discovery in the enhancer signatures lead to the identification of master regulators for each state, and experimental follow-up confirmed a key role of TEAD in the invasive melanoma state [5]. Thus, decoding enhancer signatures often yields insight into the transcription factors, and indirectly (e.g., via protein-protein interactions with the TFs) also candidate co-factors and chromatin regulators. For the decoding step, a variety of bioinformatics techniques can be used, such as *de novo* motif discovery (e.g., using tools like RSAT or Homer [39,68]) and motif enrichment with PWM libraries (e.g., Homer or i-cisTarget). Furthermore, the relatively short sequence length (a typical enhancer is 300-600bp) allows using more advanced machine learning approaches for classification, such as Hidden Markov Models (chromHMM [69], segWay [70]), Support Vector Machines [71] and deep learning [72], which may be used for genome-wide predictions of *cis*-regulatory modules and for the assessment of *cis*-regulatory variation [73–75] (Figure 2c).

Although the above-mentioned high-throughput approaches allow accurate identification of candidate regulatory regions, not every accessible region is a functional enhancer. Often, complementary enhancer assays are required to discriminate between functional and non-functional regions, or to discriminate between *bona-fide* and phantom ChIP-peaks. One functional genomics approach that allows for medium to high-throughput follow-up of enhancer activity is the massively parallel enhancer reporter assay (MPRA; Figure 1; reviewed in [76,77]). The majority of MPRA use a

plasmid library of synthetic enhancer fragments followed by a specific 20-bp barcode sequence to measure the activity of each enhancer fragment simultaneously after transfection of the entire library into cells of interest [78]. Kwasnieski *et al.* for instance used such a technique, termed CRE-seq, in K562 erythroleukemia cells to demonstrate that not all ENCODE-predicted enhancers were truly functionally active cis-regulatory elements [79]. In Verfaillie *et al.* [80], we use a variant of MPRA, called CHEQ-seq, to capture and measure activity of 1526 candidate TP53 enhancers, simultaneously. We showed that TP53 employs an unsophisticated enhancer logic, whereby functional enhancers harbour a single TP53 canonical motif. Other MPRA variants, such as STARR-seq [81], use plasmid libraries of randomly sheared enhancer fragments downstream instead of upstream of a reporter gene. Hereby, functionally active enhancers transcribe themselves and hence can be identified without the use of barcodes. Interestingly, MPRA has also been modified to study the functionality of enhancers in a genomic instead of episomal context, for instance by using retroviral delivery and subsequent integration into the genome [82]. MPRAs in which the enhancer-reporter construct remains episomal have the advantage that the influence of the enhancer sequence itself can be analyzed, regardless of the potential impact of additional factors. MPRAs with genomic integration allow measuring the regulatory activity in a chromosomal context, even though this context might still differ between the location of the endogenous enhancer and the location of random insertion. Another promising technique within the field of functional genomics is the CRISPR-Cas9 system to either edit TF motifs in their endogeneous genomic context, to modulate enhancer activity, or to use in genetic enhancer screens [83] as reviewed recently [84].

A future challenge in the field will be to combine both enhancer and gene signatures into "enhancer-aware" gene regulatory networks; this is a challenging task because of the complexity of enhancer-to-gene associations. Indeed, enhancers have an extensive outreach, regulating very distal target genes (>1 Mb) and regulating multiple target genes at once. Studies using 3D chromatin interaction data can be incorporated to aid in this problem (3C, 4C, 5C, Hi-C, ChIA-PET [88–92]), but currently lack the required resolution in signal-to-noise ratio to accurately resolve this problem. Furthermore, some genes are regulated by arrays of enhancers, sometimes referred to as "super-enhancers", of which the combined output (e.g., additive [93] or synergistic [94]) is not always easy to disentangle. Nevertheless, such hyper-active enhancer clusters may hold potential to modulate cancer states as they are often found to be more sensitive to drugs that target chromatin modifiers, such as BRD4 [95] or CDK7 inhibitors [96].

Single-cell genomics to map dynamic cancer networks

Whole-genome transcriptomic and epigenomic analyses, as described above, have proven to be very useful for the identification and characterisation of bulk-level GRNs, transcriptional states, their master regulators, and target enhancers. However, when cancer cells transition between alternate states, cell populations will likely change asynchronously and may even use different trajectories. To avoid averaging-out these differences and to obtain high-resolution space and time trajectories, single cell techniques are crucial. In addition, single-cell studies will enable the discovery of novel, rare populations of cells and more precise characterisation of currently known cell types.

Both transcriptome and epigenome can be measured at single-cell level, using single-cell RNA-seq (scRNA-seq) and single-cell chromatin accessibility (scATAC-seq or scDHS-seq) respectively (Figure 1). To perform scRNA-seq various methodologies are available, with different degrees of sensitivity, efficiency, cost, and throughput (reviewed in [97]). For example, SMART-seq2 and CEL-seq2 achieve high recall in terms of measured genes per cell; while Drop-seq and In-Drop achieve higher throughput and lower cost per cell. Using SMART-seq, Patel *et al.* demonstrated considerable intratumoral transcriptional heterogeneity in human glioblastoma, showing that in principle these data can be used to map gene regulatory networks of cancer sub-populations [98]**. In a more recent study on human melanomas Tirosh *et al.* applied SMART-seq2 to mixed tumor and microenvironment populations. Bulk RNA-seq based classification of tumor samples revealed two distinct transcriptional states (MITF-high and AXL-high) and each tumor could be classified either as MITF-high or AXL-high. However, scRNA-seq analysis revealed subpopulations of melanoma cells with varying expression of MITF and AXL within single tumours, regardless of whether the tumour was classified as invasive (AXL-high) or proliferative (MITF-high), demonstrating the power of single cell sequencing [99]**. Advanced computational methods will be needed in the future to unravel gene regulatory programs from such data, as current methods are mainly limited to the statistical analysis [100–102] (reviewed by [103]) or pathway analysis [104] of single-cell RNA-seq data. It is also expected that cancer research will benefit from dynamical analysis of tumour progression, for example in model systems, so that single cancer cells can be dynamically ordered in pseudo-time, similarly to what has been done on cellular differentiation programs, such as myogenic differentiation [105] and adult neurogenesis [106].

Another layer of regulatory heterogeneity in cancer can be measured by single-cell epigenomic profiling. Although single-cell ChIP-seq [107], single-cell DNA methylation [108], and single-cell DNaseI hypersensitivity [109] have been shown to work in principle, the highest quality single-cell epigenomes can be derived by single-cell ATAC-seq [110,111]. Using single cell and bulk ATAC-seq in combination with RNA-seq, Corces *et al.* recently demonstrated how regulatory heterogeneity observed in bulk can be resolved with single-cell data. They showed that the regulatory heterogeneity

observed in AML is a combination of both intercellular and intracellular heterogeneity; the former representing clonal architecture of the tumor while the latter representing intermediate GRNs in tumors, that do not exist in normal haematopoiesis [112]*.

Using the increasing amounts and quality of single cell profiles, the development of novel experimental techniques such as the simultaneous epigenomic and transcriptomic profiling of a single cell, and novel computational methods for reverse-engineering GRNs, we will gain more fundamental insights into gene regulation and enhancer logic, and might also open up novel therapeutic strategies by uncovering specific oncogenic GRNs.

Acknowledgements

The authors would like to thank Dr. Sara Aibar for her help with the figures. This work is funded by ERC (Consolidator Grant to SA); the Research Foundation - Flanders (FWO, www.fwo.be) (grants G.0640.13 and G.0791.14 to SA), Special Research Fund (BOF) KU Leuven (<http://www.kuleuven.be/research/funding/bof/>) (grant PF/10/016 and OT/13/103 to SA), Foundation Against Cancer (<http://www.cancer.be>) (grant 2012-F2 to SA) and The Harry J. Lloyd Charitable Trust (<http://www.hjltrust.com>). JW and ZKA are funded by a postdoctoral grant from *Kom of tegen Kanker*. The funders had no role in the design of the manuscript.

Table 1**a. Computational resources and methods**

Description	Methods	Characteristics	Reference
(Cancer) gene signature databases	GeneSigDB	Contains literature-curated signatures. Release 4 contains 3515 signatures across more than 50 diseases and tissues	[26]
	MSigDB	Contains annotated (curated and computational) gene sets. MSigDB database v5.2 contains 18026 gene sets.	[27]
	OncoMine	A cancer-profiling database containing re-analyzed genomic data	[28]
Gene ontology enrichment analysis	WebGestalt	Web-based functional enrichment analysis tool that enables analysis not only on Gene Ontology terms but also on pathway, network module, gene–phenotype association, gene–disease association, gene–drug association categories	[29]
	HumanMine		[30]
	Ingenuity Pathway Analysis		
GRN construction from gene expression data	ARACNe	GRN construction with the use of mutual information. It can accommodate non-linear relationships between genes; requires large sample sizes to achieve a good performance	[31]
	BC3NET	GRN construction with Bayesian networks. It infers the structure of the network statistically; fails to accommodate feedback loops	[32]
	GRNIinfer	GRN construction with differential equation models. It is suited for small networks	[33]
	GENIE3	GRN construction with machine learning approaches. GENIE3 was best performer in two DREAM challenges; it's a computationally intensive approach	[34]
TFBS motif enrichment analysis	Homer	Available as a command-line tool. Search space can be defined by user. Enables use of TRANSFAC motif collection as well as user defined motifs.	
	RSAT	Implemented in a web-server. Search space can be defined by user. Enables use of several publicly available motif collections as well as user defined motifs.	[39]
	OPOSSUM	Implemented in a web-server. Search space can be up to 1kb up-and-downstream of TSS. Enables use of JASPAR motif collection.	[40]
	PSCAN	Implemented in a web-server. Search space is up to 1kb upstream and up to 50bp downstream of TSS. Enables use of JASPAR and TRANSFAC motif collections as well as user defined motifs.	[41]
	DIRE	Implemented in a web-server. Search space is pre-defined. Enables use of TRANSFAC motif collection.	[42]
	iRegulon	I Implemented as a Cytoscape plug-in and as a command-line tool. Search space can be up to 10 kb up-	[43]

		and-downstream of TSS. Enables use of several publicly available motif collections (more than 18.000 PWMs in total). motif2TF procedure links motifs to candidate human TFs enabling generation of an actual gene regulatory network in Cytoscape environment.	
	i-cisTarget	Implemented as a web-server and as a command-line tool. Search space is 10 kb up-and-downstream of TSS. Enables use of several publicly available motif collections (more than 18.000 PWMs in total).	[113]
Chromatin state identification and analysis	ChromHMM	Chromatin state identification with Hidden Markov Models. It can be trained on the entire genome since it has a lower genomic resolution (200bp)	[69]
	SegWay	Chromatin state identification with dynamic Bayesian Networks. More suitable for subset of the genome since it has single base-pair genomic resolution	[70]
RNAi screen analysis	RSA	Provides a statistical score indicating the probability of a gene being hit by integrating information about multiple RNAi reagents per gene	[51]
	RIGER	Provides an GSEA methodology based enrichment score for each shRNA and rank ordered list of genes targeted by it	[52]
CRISPRi screen analysis	MAGECK	Statistical framework for identifying significant hits using negative binomial model	[53]
	NEST	Enables interpretation of CRISPR screens using protein-protein interaction networks	[54]

b. Experimental methods

Description	Methods	Characteristics	Reference
Perturbation of transcriptional states	genome-wide RNAi libraries	Established method/protocols; Only inhibition possible	[114]
	genome-wide CRISPR libraries	Inhibition, activation, knockout and mutagenesis possible; less off-target effects than shRNA	[49,50]
Accessible chromatin	DHS-seq	Established method/protocols; "footprinting" of DNA-binding proteins possible; High-input needed; laborious method	[115]
	ATAC-seq	Fast and low-input; nucleosome positioning and "footprinting" of DNA-binding proteins possible; contamination with mitochondrial reads	[116]
Histone modifications	ChIP-seq	Established method/protocols; High-input needed; highly dependent on quality of antibody	[117]
	ChIPmentation	Fast and low-input; Highly dependent on quality of antibody	[118]
MPRAs	CRE-seq	Quantitative assessment of enhancer functionality; using barcodes; episomal	[79]
	CHEQ-seq	Quantitative assessment of enhancer functionality; using barcodes; episomal	[80]
	STARR-seq	Quantitative assessment of enhancer functionality; episomal	[119]
	TRIP	Chromosomal; quantitative; low resolution	[84]
	FIREWACH	Chromosomal; FACS-based; not quantitative	[82]

Nascent RNA sequencing	SIF-seq	Chromosomal; FACS-based; not quantitative	[83]
	LV-MPRA	Chromosomal; quantitative	[85]
	GRO-seq	High sensitivity and robustness; enhancer RNA detectable; complicated and laborious method	[62]
	PRO-seq	Base-pair resolution; enhancer RNA detectable; RNA polymerases very close to the TSS might be missed	[63]05/01/ 2017 21:23:00
	PRO-cap	Base-pair resolution; identification of transcription start sites; enhancer RNA detectable; RNA polymerases very close to the TSS might be missed	[120]
3D chromatin interaction	TT-seq	Enhancer RNA detectable; most straight-forward nascent RNA sequencing method	[64]
	3C	Crosslinking of interacting chromatin segments; PCR primers interrogate specific interaction	[88]
	4C	Application of 3C to an entire locus by sequencing; PCR primers define the "viewpoint"	[89]
	5C	Combination of 3C and multiplex ligation-mediated amplification	[90]
	Hi-C	Genome-wide variant of 3C	[91]
DNA methylation	ChIA-PET	Interactions of chromatin that is bound by a specific protein, using ChIP	[92]
	WGBS	All CpG analysed; single base resolution; Very expensive; no distinction between 5mC and 5hmC	[58]
	RRBS	Single base resolution; high sensitivity; No distinction between 5mC and 5hmC	[59]
	Infinium	Lower cost; low-input; Only CpGs with a probe on the array can be analyzed	[121]
Single cell sequencing	MeDIPseq	Distinction between 5mC and 5hmC possible; Low resolution (~100bp)	[122]
	FACSorting- and well plate-based single-cell sequencing (e.g. SMART-seq2 and CEL-seq2)	High sensitivity	[123,124]
	Droplet microfluidics based single-cell sequencing (e.g. Drop-seq and In-Drop)	High-throughput	[125,126]

c. Markers for ChIP-seq and ChIPmentation

Type of regulatory element	Histone modifications
Active and poised promoters	H3K4me2, H3K4me3, H3K27ac (for active) and H3K4me3, H3K27me3 (for poised)
Active enhancers	H3K4me1, H3K4me2, H3K27ac
Inactive promoters	H3K9me3, H3K27me3
Inactive enhancers	H3K9me2, H3K9me3
Active gene bodies	H3K36me3, H3K79me2
Inactive gene bodies	H3K9me2, H3K9me3

Figure legends

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 profiling 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.

Figure 2. Computational methods for deciphering gene regulatory networks. **A.** Transcriptome data can be analysed with network inference techniques that predict relationships between genes based on (linear and/or non-linear) expression similarities. **B.** *cis*-regulatory sequence analysis is applicable to both gene and enhancer signatures. In the case of gene signatures, a pre-defined regulatory space (usually >10kb) around transcription start site is scanned for enrichment of transcription factor motifs (indicated as orange and blue rectangles), while in the case of enhancer signatures only the enhancer regions themselves (usually < 1kb) (indicated as grey peaks) are subjected to *de novo* motif discovery or motif/PWM enrichment analysis. **C.** Enhancer signatures are also suitable for training enhancer models using machine-learning approaches. Each line in the figure represents one enhancer region with rectangles of different colours representing predicted motif occurrences or other features such as dinucleotide repeat motifs or DNA shape preferences. **D.** Gene regulatory networks as obtained from methods described in A and B can be represented in a hierarchical network where nodes indicate genes (transcription factors and target genes) and edges represent direct TF-target regulatory interactions. **E.** High-throughput single-cell assays measuring transcriptome or epigenome activity can be used to identify different transcriptomic/epigenomic states and decipher gene regulatory networks governing them. The scatter plot above represents 2D-plot after dimensionality reduction, where each dot represents a single cell and spatial distribution reflects transcriptome similarities between cells. Attractor states are inferred from this distribution and gene regulatory networks are then predicted from state specific signatures.

References

1. Bardeesy N, DePinho RA: **Pancreatic cancer biology and genetics.** *Nat. Rev. Cancer* 2002, **2**:897–909.
2. Sandhu V, Wedge DC, Bowitz Lothe IM, Labori KJ, Dentro SC, Buanes T, Skrede ML, Dalsgaard AM, Munthe E, Myklebost O, et al.: **The Genomic Landscape of Pancreatic and Periampullary Adenocarcinoma.** *Cancer Res.* 2016, doi:10.1158/0008-5472.CAN-16-0658.
3. Atkins M, Potier D, Romanelli L, Jacobs J, Mach J, Hamaratoglu F, Aerts S, Halder G: **An Ectopic Network of Transcription Factors Regulated by Hippo Signaling Drives Growth and Invasion of a Malignant Tumor Model.** *Curr. Biol. CB* 2016, **26**:2101–2113.
4. Cancer Genome Atlas Network: **Genomic Classification of Cutaneous Melanoma.** *Cell* 2015, **161**:1681–1696.
5. Verfaillie A, Imrichova H, Atak ZK, Dewaele M, Rambow F, Hulselmans G, Christiaens V, Svetlichnyy D, Luciani F, Van den Mooter L, et al.: **Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state.** *Nat. Commun.* 2015, **6**:6683.
6. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, et al.: **Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.** *Cancer Cell* 2010, **17**:98–110.
7. Cancer Genome Atlas Network: **Comprehensive molecular characterization of human colon and rectal cancer.** *Nature* 2012, **487**:330–337.
8. Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR, Korsmeyer SJ: **MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia.** *Nat. Genet.* 2002, **30**:41–47.
9. Yeoh E-J, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A, et al.: **Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling.** *Cancer Cell* 2002, **1**:133–143.
10. Brohl AS, Solomon DA, Chang W, Wang J, Song Y, Sindiri S, Patidar R, Hurd L, Chen L, Shern JF, et al.: **The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation.** *PLoS Genet.* 2014, **10**:e1004475.
11. Lee TI, Young RA: **Transcriptional regulation and its misregulation in disease.** *Cell* 2013, **152**:1237–1251.
12. Huang S, Ernberg I, Kauffman S: **Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective.** *Semin. Cell Dev. Biol.* 2009, **20**:869–876.
13. Huang S, Kauffman S: **How to escape the cancer attractor: rationale and limitations of multi-target drugs.** *Semin. Cancer Biol.* 2013, **23**:270–278.
14. Li Q, Wennborg A, Aurell E, Dekel E, Zou J-Z, Xu Y, Huang S, Ernberg I: **Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape.** *Proc. Natl. Acad. Sci. U. S. A.* 2016, **113**:2672–2677.

15. Cheng W-Y, Ou Yang T-H, Anastassiou D: **Biomolecular events in cancer revealed by attractor metagenes.** *PLoS Comput. Biol.* 2013, **9**:e1002920.
16. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MDM, Niu B, McLellan MD, Uzunangelov V, et al.: **Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin.** *Cell* 2014, **158**:929–944.
17. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, et al.: **The prognostic landscape of genes and infiltrating immune cells across human cancers.** *Nat. Med.* 2015, **21**:938–945.
18. Peng L, Bian XW, Li DK, Xu C, Wang GM, Xia QY, Xiong Q: **Large-scale RNA-Seq Transcriptome Analysis of 4043 Cancers and 548 Normal Tissue Controls across 12 TCGA Cancer Types.** *Sci. Rep.* 2015, **5**:13413.
19. O'Connell MP, Marchbank K, Webster MR, Valiga AA, Kaur A, Vultur A, Li L, Herlyn M, Villanueva J, Liu Q, et al.: **Hypoxia induces phenotypic plasticity and therapy resistance in melanoma via the tyrosine kinase receptors ROR1 and ROR2.** *Cancer Discov.* 2013, **3**:1378–1393.
20. Landsberg J, Kohlmeyer J, Renn M, Bald T, Rogava M, Cron M, Fatho M, Lennerz V, Wölfel T, Hözel M, et al.: **Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation.** *Nature* 2012, **490**:412–416.
21. Sun C, Wang L, Huang S, Heynen GJJE, Prahallas A, Robert C, Haanen J, Blank C, Wesseling J, Willems SM, et al.: **Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma.** *Nature* 2014, **508**:118–122.
22. De Craene B, Berx G: **Regulatory networks defining EMT during cancer initiation and progression.** *Nat. Rev. Cancer* 2013, **13**:97–110.
23. Hoek KS, Goding CR: **Cancer stem cells versus phenotype-switching in melanoma.** *Pigment Cell Melanoma Res.* 2010, **23**:746–759.
24. Yoo CB, Jones PA: **Epigenetic therapy of cancer: past, present and future.** *Nat. Rev. Drug Discov.* 2006, **5**:37–50.
25. Sáez-Ayala M, Montenegro MF, Sánchez-Del-Campo L, Fernández-Pérez MP, Chazarra S, Freter R, Middleton M, Piñero-Madrona A, Cabezas-Herrera J, Goding CR, et al.: **Directed phenotype switching as an effective antimelanoma strategy.** *Cancer Cell* 2013, **24**:105–119.
26. Culhane AC, Schwarzl T, Sultana R, Picard KC, Picard SC, Lu TH, Franklin KR, French SJ, Papenhausen G, Correll M, et al.: **GeneSigDB--a curated database of gene expression signatures.** *Nucleic Acids Res.* 2010, **38**:D716–725.
27. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P: **The Molecular Signatures Database (MSigDB) hallmark gene set collection.** *Cell Syst.* 2015, **1**:417–425.
28. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincaid-Beal C, Kulkarni P, et al.: **Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles.** *Neoplasia N. Y. N* 2007, **9**:166–180.
29. Wang J, Duncan D, Shi Z, Zhang B: **WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013.** *Nucleic Acids Res.* 2013, **41**:W77–83.

30. Smith RN, Aleksić J, Butano D, Carr A, Contrino S, Hu F, Lyne M, Lyne R, Kalderimis A, Rutherford K, et al.: **InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data.** *Bioinforma. Oxf. Engl.* 2012, **28**:3163–3165.
31. Margolin AA, Nemenman I, Bassi K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A: **ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context.** *BMC Bioinformatics* 2006, **7 Suppl 1**:S7.
32. Simoes R de M, Emmert-Streib F: **Bagging Statistical Network Inference from Large-Scale Gene Expression Data.** *PLOS ONE* 2012, **7**:e33624.
33. Wang Y, Joshi T, Zhang X-S, Xu D, Chen L: **Inferring gene regulatory networks from multiple microarray datasets.** *Bioinformatics* 2006, **22**:2413–2420.
34. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P: **Inferring Regulatory Networks from Expression Data Using Tree-Based Methods.** *PLOS ONE* 2010, **5**:e12776.
35. Liu Z-P: **Reverse Engineering of Genome-wide Gene Regulatory Networks from Gene Expression Data.** *Curr. Genomics* 2015, **16**:3–22.
This is an excellent review on reverse-engineering methods from gene expression data.
36. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, DREAM5 Consortium, Kellis M, Collins JJ, et al.: **Wisdom of crowds for robust gene network inference.** *Nat. Methods* 2012, **9**:796–804.
37. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, et al.: **The transcriptional network for mesenchymal transformation of brain tumours.** *Nature* 2010, **463**:318–325.
This study used a popular reverse engineering method, ARACNe, to infer master regulators of mesenchymal state in high grade gliomas.
38. Della Gatta G, Palomero T, Perez-Garcia A, Ambesi-Impiombato A, Bansal M, Carpenter ZW, De Keersmaecker K, Sole X, Xu L, Paietta E, et al.: **Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL.** *Nat. Med.* 2012, **18**:436–440.
This study used gene expression and transcription factor ChIP-seq data in combination with reverse engineering method ARACNe to decipher regulatory interactions in TLX-connected oncogenic networks in T-ALL.
39. Medina-Rivera A, Defrance M, Sand O, Herrmann C, Castro-Mondragon JA, Delerce J, Jaeger S, Blanchet C, Vincens P, Caron C, et al.: **RSAT 2015: Regulatory Sequence Analysis Tools.** *Nucleic Acids Res.* 2015, **43**:W50–56.
40. Kwon AT, Arenillas DJ, Hunt RW, Wasserman WW: **oPOSSUM-3: Advanced Analysis of Regulatory Motif Over-Representation Across Genes or ChIP-Seq Datasets.** *G3 GenesGenomesGenetics* 2012, **2**:987–1002.
41. Zambelli F, Pesole G, Pavesi G: **Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes.** *Nucleic Acids Res.* 2009, **37**:W247–W252.

42. Gotea V, Visel A, Westlund JM, Nobrega MA, Pennacchio LA, Ovcharenko I: **Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers.** *Genome Res.* 2010, **20**:565–577.
43. Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D, et al.: **iRegulon: from a gene list to a gene regulatory network using large motif and track collections.** *PLoS Comput. Biol.* 2014, **10**:e1003731.
44. Chadwick LH: **The NIH Roadmap Epigenomics Program data resource.** *Epigenomics* 2012, **4**:317–324.
45. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, et al.: **ENCODE data in the UCSC Genome Browser: year 5 update.** *Nucleic Acids Res.* 2013, **41**:D56–63.
46. Sanda T, Lawton LN, Barrasa MI, Fan ZP, Kohlhammer H, Gutierrez A, Ma W, Tatarek J, Ahn Y, Kelliher MA, et al.: **Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia.** *Cancer Cell* 2012, **22**:209–221.
This study identified master regulators of TAL1-connected networks in T-ALL using gene expression and transcription factor ChIP-seq data.
47. Liu F, Hon GC, Villa GR, Turner KM, Ikegami S, Yang H, Ye Z, Li B, Kuan S, Lee AY, et al.: **EGFR Mutation Promotes Glioblastoma through Epigenome and Transcription Factor Network Remodeling.** *Mol. Cell* 2015, **60**:307–318.
This study identified SOX9 and FOXG1 as master regulators of a subset of glioblastomas (EGFR mutated) using RNA-seq and chromatin ChIP-seq on different glioblastoma model systems.
48. FANTOM Consortium, Suzuki H, Forrest ARR, van Nimwegen E, Daub CO, Balwierz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, et al.: **The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line.** *Nat. Genet.* 2009, **41**:553–562.
49. Wang T, Wei JJ, Sabatini DM, Lander ES: **Genetic screens in human cells using the CRISPR-Cas9 system.** *Science* 2014, **343**:80–84.
50. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, et al.: **Genome-scale CRISPR-Cas9 knockout screening in human cells.** *Science* 2014, **343**:84–87.
51. König R, Chiang C, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, et al.: **A probability-based approach for the analysis of large-scale RNAi screens.** *Nat. Methods* 2007, **4**:847–849.
52. Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, Hinkle G, Boehm JS, Beroukhim R, Weir BA, et al.: **Highly parallel identification of essential genes in cancer cells.** *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**:20380–20385.
53. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS: **MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.** *Genome Biol.* 2014, **15**:554.

54. Jiang P, Wang H, Li W, Zang C, Li B, Wong YJ, Meyer C, Liu JS, Aster JC, Liu XS: **Network analysis of gene essentiality in functional genomics experiments.** *Genome Biol.* 2015, **16**:239.
55. Sur I, Taipale J: **The role of enhancers in cancer.** *Nat. Rev. Cancer* 2016, **16**:483–493.
This is an excellent review on role of enhancers in tumorigenesis and their therapeutic potential.
56. Zhou VW, Goren A, Bernstein BE: **Charting histone modifications and the functional organization of mammalian genomes.** *Nat. Rev. Genet.* 2011, **12**:7–18.
57. ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57–74.
58. Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR: **MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing.** *Nat. Protoc.* 2015, **10**:475–483.
59. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, et al.: **Genome-scale DNA methylation maps of pluripotent and differentiated cells.** *Nature* 2008, **454**:766–770.
60. Zhao M-T, Whyte JJ, Hopkins GM, Kirk MD, Prather RS: **Methylated DNA immunoprecipitation and high-throughput sequencing (MeDIP-seq) using low amounts of genomic DNA.** *Cell. Reprogramming* 2014, **16**:175–184.
61. Yong W-S, Hsu F-M, Chen P-Y: **Profiling genome-wide DNA methylation.** *Epigenetics Chromatin* 2016, **9**:26.
62. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters.** *Science* 2008, **322**:1845–1848.
63. Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT: **Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq).** *Nat. Protoc.* 2016, **11**:1455–1476.
64. Schwalb B, Michel M, Zacher B, Frühauf K, Demel C, Tresch A, Gagneur J, Cramer P: **TT-seq maps the human transient transcriptome.** *Science* 2016, **352**:1225–1228.
65. Rhee SK, Guo Y, Tak YG, Yao L, Shen H, Coetzee GA, Laird PW, Farnham PJ: **Identification of activated enhancers and linked transcription factors in breast, prostate, and kidney tumors by tracing enhancer networks using epigenetic traits.** *Epigenetics Chromatin* 2016, **9**:50.
66. Denny SK, Yang D, Chuang C-H, Brady JJ, Lim JS, Grüner BM, Chiou S-H, Schep AN, Baral J, Hamard C, et al.: **Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility.** *Cell* 2016, **166**:328–342.
This study applied ATAC-seq on genetically engineered mouse models of small cell lung cancer and identified NFIB transcription factor as master regulator of the metastatic state.
67. Rendeiro AF, Schmidl C, Strefford JC, Walewska R, Davis Z, Farlik M, Oscier D, Bock C: **Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks.** *Nat. Commun.* 2016, **7**:11938.

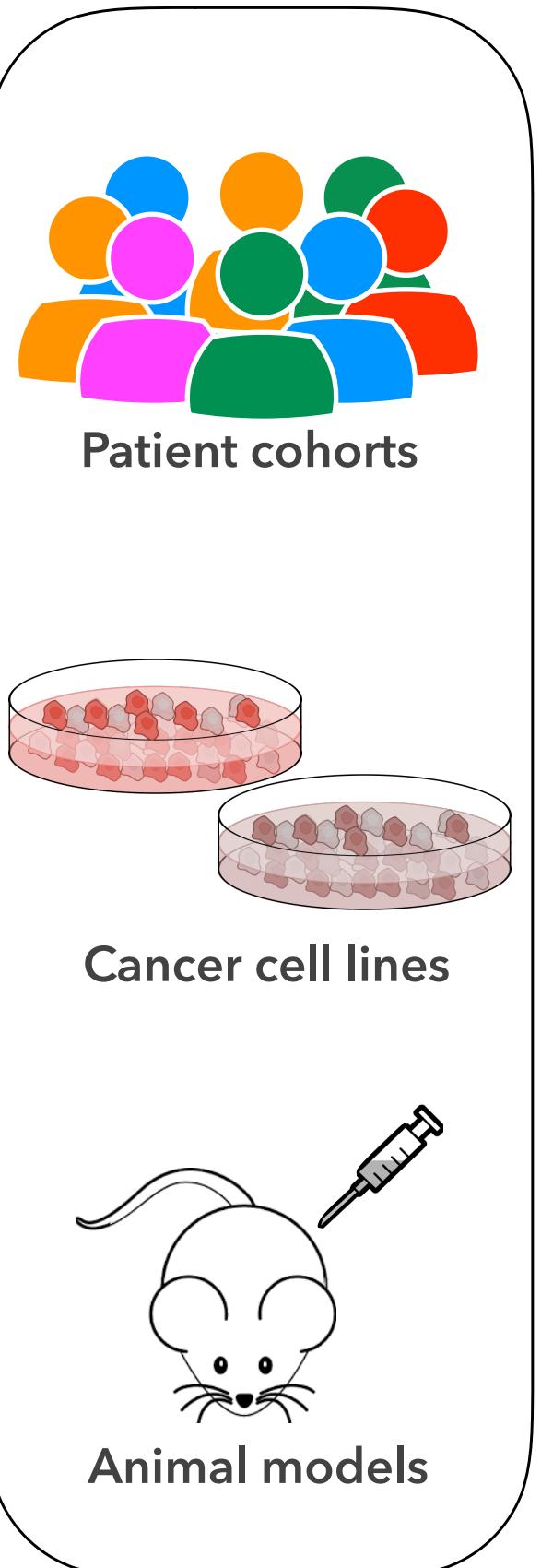
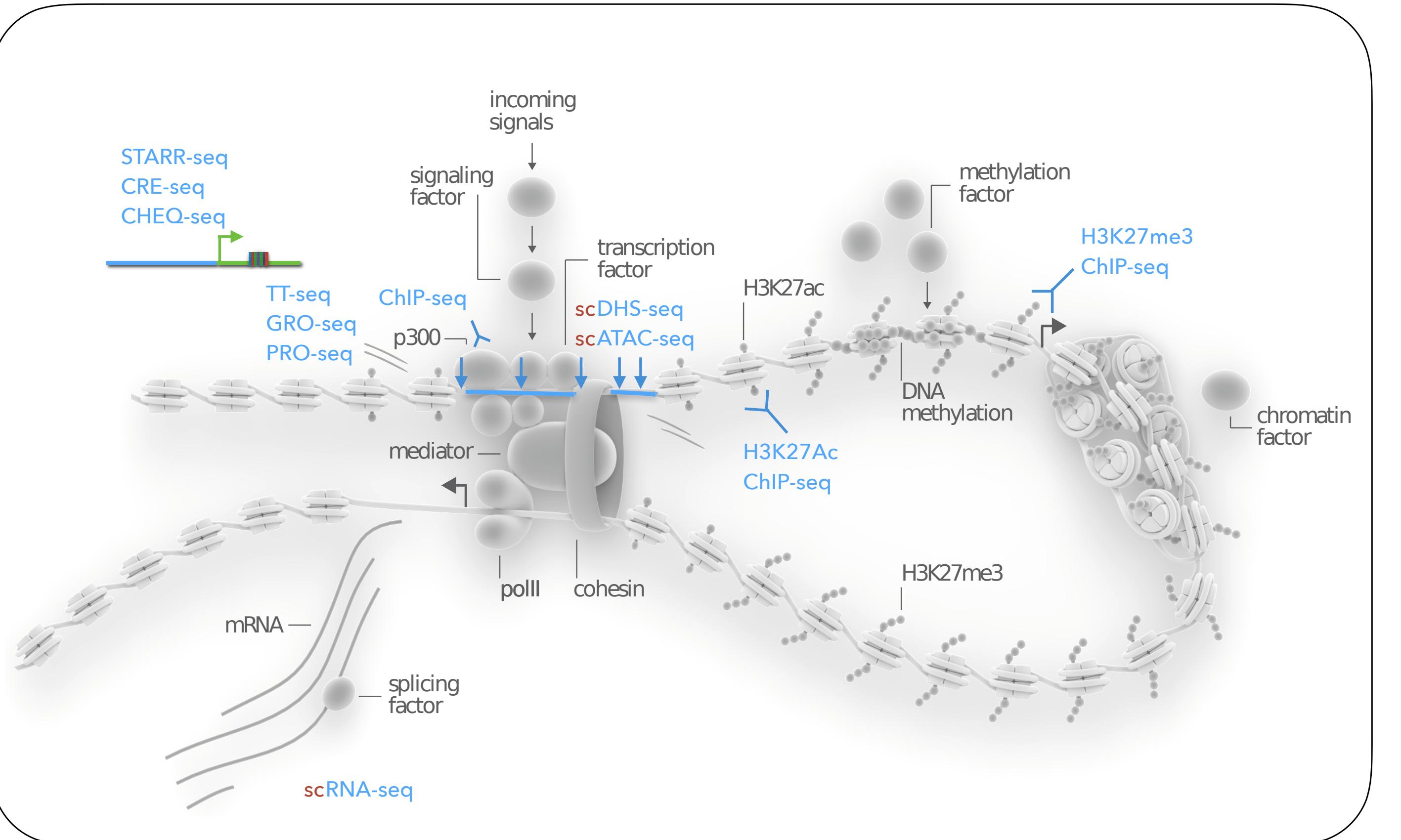
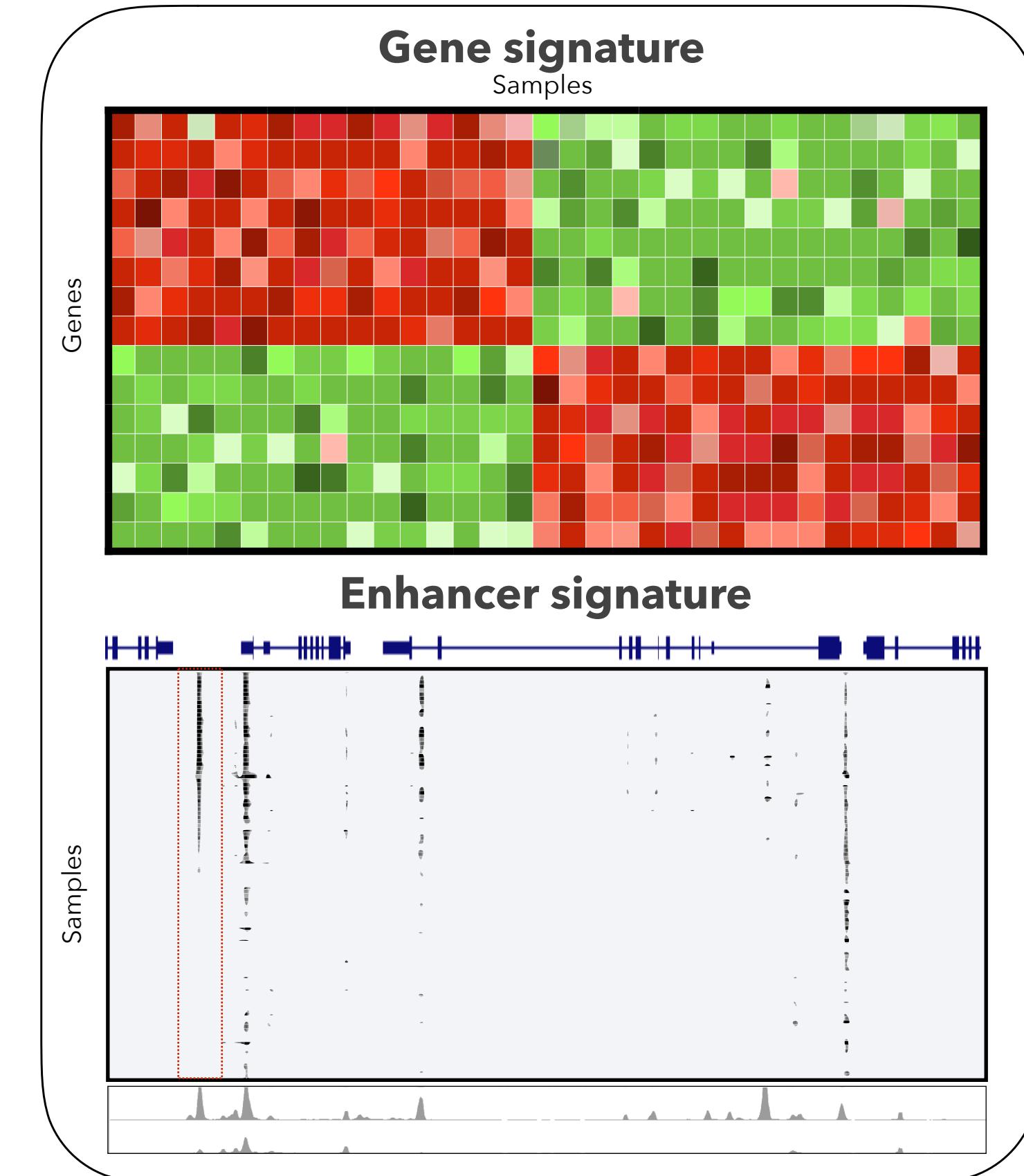
68. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.** *Mol. Cell* 2010, **38**:576–589.
69. Ernst J, Kellis M: **ChromHMM: automating chromatin-state discovery and characterization.** *Nat. Methods* 2012, **9**:215–216.
70. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS: **Unsupervised pattern discovery in human chromatin structure through genomic segmentation.** *Nat. Methods* 2012, **9**:473–476.
71. Fletez-Brant C, Lee D, McCallion AS, Beer MA: **kmer-SVM: a web server for identifying predictive regulatory sequence features in genomic data sets.** *Nucleic Acids Res.* 2013, **41**:W544–556.
72. Kelley DR, Snoek J, Rinn J: **Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks.** *Genome Res.* 2016, doi:10.1101/gr.200535.115.
73. Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, Beer MA: **A method to predict the impact of regulatory variants from DNA sequence.** *Nat. Genet.* 2015, **47**:955–961.
74. Zhou J, Troyanskaya OG: **Predicting effects of noncoding variants with deep learning-based sequence model.** *Nat. Methods* 2015, **12**:931–934.
75. Svetlichnyy D, Imrichova H, Fiers M, Atak ZK, Aerts S: **Identification of High-Impact cis - Regulatory Mutations Using Transcription Factor Specific Random Forest Models.** *PLOS Comput Biol* 2015, **11**:e1004590.
76. Inoue F, Ahituv N: **Decoding enhancers using massively parallel reporter assays.** *Genomics* 2015, **106**:159–164.
77. White MA: **Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences.** *Genomics* 2015, **106**:165–170.
78. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J: **High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis.** *Nat. Biotechnol.* 2009, **27**:1173–1175.
79. Kwasnieski JC, Fiore C, Chaudhari HG, Cohen BA: **High-throughput functional testing of ENCODE segmentation predictions.** *Genome Res.* 2014, **24**:1595–1602.
80. Verfaillie A, Svetlichnyy D, Imrichova H, Davie K, Fiers M, Kalender Atak Z, Hulselmans G, Christiaens V, Aerts S: **Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic.** *Genome Res.* 2016, **26**:882–895.
81. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A: **Genome-wide quantitative enhancer activity maps identified by STARR-seq.** *Science* 2013, **339**:1074–1077.
82. Murtha M, Tokcaer-Keskin Z, Tang Z, Strino F, Chen X, Wang Y, Xi X, Basilico C, Brown S, Bonneau R, et al.: **FIREWACH: high-throughput functional detection of transcriptional regulatory modules in mammalian cells.** *Nat. Methods* 2014, **11**:559–565.

83. Dickel DE, Zhu Y, Nord AS, Wylie JN, Akiyama JA, Afzal V, Plajzer-Frick I, Kirkpatrick A, Göttgens B, Bruneau BG, et al.: **Function-based identification of mammalian enhancers using site-specific integration.** *Nat. Methods* 2014, **11**:566–571.
84. Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LFA, van Lohuizen M, van Steensel B: **Chromatin position effects assayed by thousands of reporters integrated in parallel.** *Cell* 2013, **154**:914–927.
85. Maricque BB, Dougherty JD, Cohen BA: **A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of cis-regulatory activity in neural cells.** *Nucleic Acids Res.* 2016, doi:10.1093/nar/gkw942.
86. Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han R, Myacheva K, Zwart W, Elkorn R, Agami R: **Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9.** *Nat. Biotechnol.* 2016, **34**:192–198.
87. Lopes R, Korkmaz G, Agami R: **Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers.** *Nat. Rev. Mol. Cell Biol.* 2016, **17**:597–604.
88. Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forné T: **Quantitative analysis of chromosome conformation capture assays (3C-qPCR).** *Nat. Protoc.* 2007, **2**:1722–1733.
89. Simonis M, Klous P, Splinter E, Moshkin Y, Willemse R, de Wit E, van Steensel B, de Laat W: **Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).** *Nat. Genet.* 2006, **38**:1348–1354.
90. Dostie J, Dekker J: **Mapping networks of physical interactions between genomic elements using 5C technology.** *Nat. Protoc.* 2007, **2**:988–1002.
91. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al.: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science* 2009, **326**:289–293.
92. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, et al.: **An oestrogen-receptor-alpha-bound human chromatin interactome.** *Nature* 2009, **462**:58–64.
93. Hay D, Hughes JR, Babbs C, Davies JOJ, Graham BJ, Hanssen LLP, Kassouf MT, Oudelaar AM, Sharpe JA, Suciu MC, et al.: **Genetic dissection of the α -globin super-enhancer in vivo.** *Nat. Genet.* 2016, **48**:895–903.
94. Shin HY, Willi M, Yoo KH, Zeng X, Wang C, Metser G, Hennighausen L: **Hierarchy within the mammary STAT5-driven Wap super-enhancer.** *Nat. Genet.* 2016, **48**:904–911.
95. Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA: **Selective inhibition of tumor oncogenes by disruption of super-enhancers.** *Cell* 2013, **153**:320–334.
96. Chipumuro E, Marco E, Christensen CL, Kwiatkowski N, Zhang T, Hatheway CM, Abraham BJ, Sharma B, Yeung C, Altabef A, et al.: **CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer.** *Cell* 2014, **159**:1126–1139.

97. Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, Mburu FM, Mantalas GL, Sim S, Clarke MF, et al.: **Quantitative assessment of single-cell RNA-sequencing methods.** *Nat. Methods* 2014, **11**:41–46.
98. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, et al.: **Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.** *Science* 2014, **344**:1396–1401.
This study demonstrated the existence of different transcriptional states within single glioblastoma tumors using single-cell transcriptomics.
99. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, et al.: **Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.** *Science* 2016, **352**:189–196.
This study applied SMART-seq2 to primary melanoma samples and demonstrated that both invasive and proliferative subtypes can co-exist.
100. Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, van Oudenaarden A: **Single-cell messenger RNA sequencing reveals rare intestinal cell types.** *Nature* 2015, **525**:251–255.
101. Kharchenko PV, Silberstein L, Scadden DT: **Bayesian approach to single-cell differential expression analysis.** *Nat. Methods* 2014, **11**:740–742.
102. Bacher R, Kendziora C: **Design and computational analysis of single-cell RNA-sequencing experiments.** *Genome Biol.* 2016, **17**:63.
103. Stegle O, Teichmann SA, Marioni JC: **Computational and analytical challenges in single-cell transcriptomics.** *Nat. Rev. Genet.* 2015, **16**:133–145.
104. Fan J, Salathia N, Liu R, Kaeser GE, Yung YC, Herman JL, Kaper F, Fan J-B, Zhang K, Chun J, et al.: **Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis.** *Nat. Methods* 2016, **13**:241–244.
105. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.** *Nat. Biotechnol.* 2014, **32**:381–386.
106. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, Enikolopov G, Nauen DW, Christian KM, Ming G, et al.: **Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis.** *Cell Stem Cell* 2015, **17**:360–372.
107. Rotem A, Ram O, Shores N, Sperling RA, Goren A, Weitz DA, Bernstein BE: **Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state.** *Nat. Biotechnol.* 2015, **33**:1165–1172.
108. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F: **Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing.** *Genome Res.* 2013, **23**:2126–2135.
109. Jin W, Tang Q, Wan M, Cui K, Zhang Y, Ren G, Ni B, Sklar J, Przytycka TM, Childs R, et al.: **Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples.** *Nature* 2015, **528**:142–146.

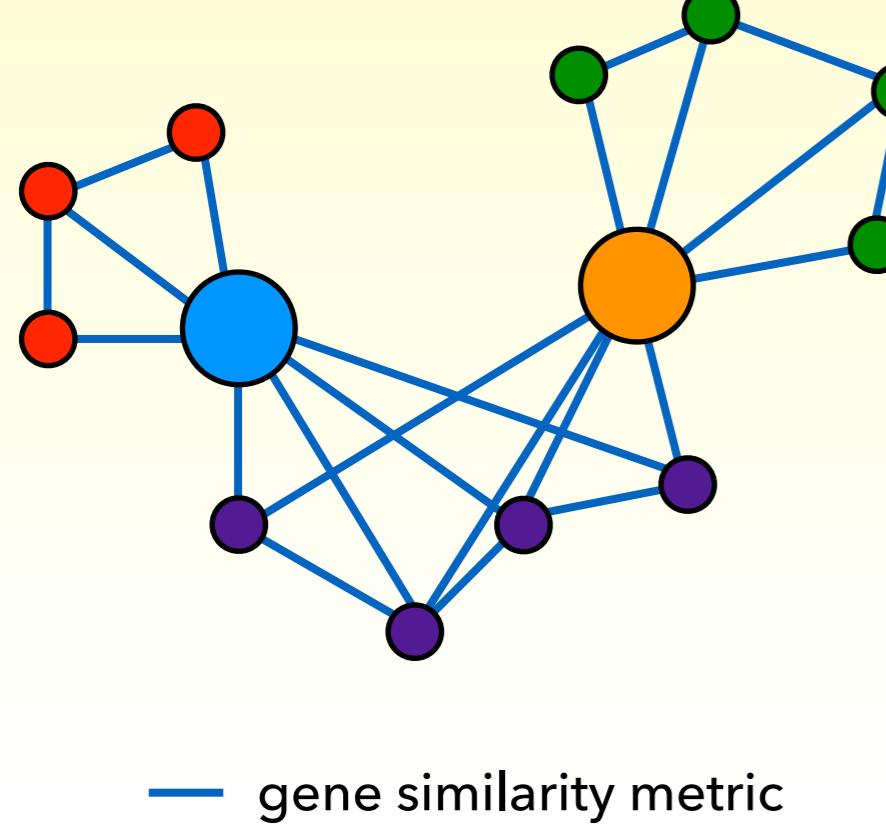
110. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ: **Single-cell chromatin accessibility reveals principles of regulatory variation.** *Nature* 2015, **523**:486–490.
111. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C, Shendure J: **Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing.** *Science* 2015, **348**:910–914.
112. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, et al.: **Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution.** *Nat. Genet.* 2016, doi:10.1038/ng.3646.
113. Imrichová H, Hulselmans G, Kalender Atak Z, Potier D, Aerts S: **i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly.** *Nucleic Acids Res.* 2015, doi:10.1093/nar/gkv395.
114. Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, et al.: **A large-scale RNAi screen in human cells identifies new components of the p53 pathway.** *Nature* 2004, **428**:431–437.
115. Song L, Crawford GE: **DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells.** *Cold Spring Harb. Protoc.* 2010, **2010**:pdb.prot5384.
116. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.** *Nat. Methods* 2013, **10**:1213–1218.
117. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129**:823–837.
118. Schmidl C, Rendeiro AF, Sheffield NC, Bock C: **ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors.** *Nat. Methods* 2015, **12**:963–965.
119. Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, Lau NC, Stark A: **Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution.** *Nat. Genet.* 2014, **46**:685–692.
120. Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT: **Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq).** *Nat. Protoc.* 2016, **11**:1455–1476.
121. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al.: **High density DNA methylation array with single CpG site resolution.** *Genomics* 2011, **98**:288–295.
122. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S, Butcher LM: **Methylome analysis using MeDIP-seq with low DNA concentrations.** *Nat. Protoc.* 2012, **7**:617–636.
123. Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R: **Full-length RNA-seq from single cells using Smart-seq2.** *Nat. Protoc.* 2014, **9**:171–181.

124. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, et al.: **CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.** *Genome Biol.* 2016, **17**:77.
125. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW: **Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.** *Cell* 2015, **161**:1187–1201.
126. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al.: **Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.** *Cell* 2015, **161**:1202–1214.

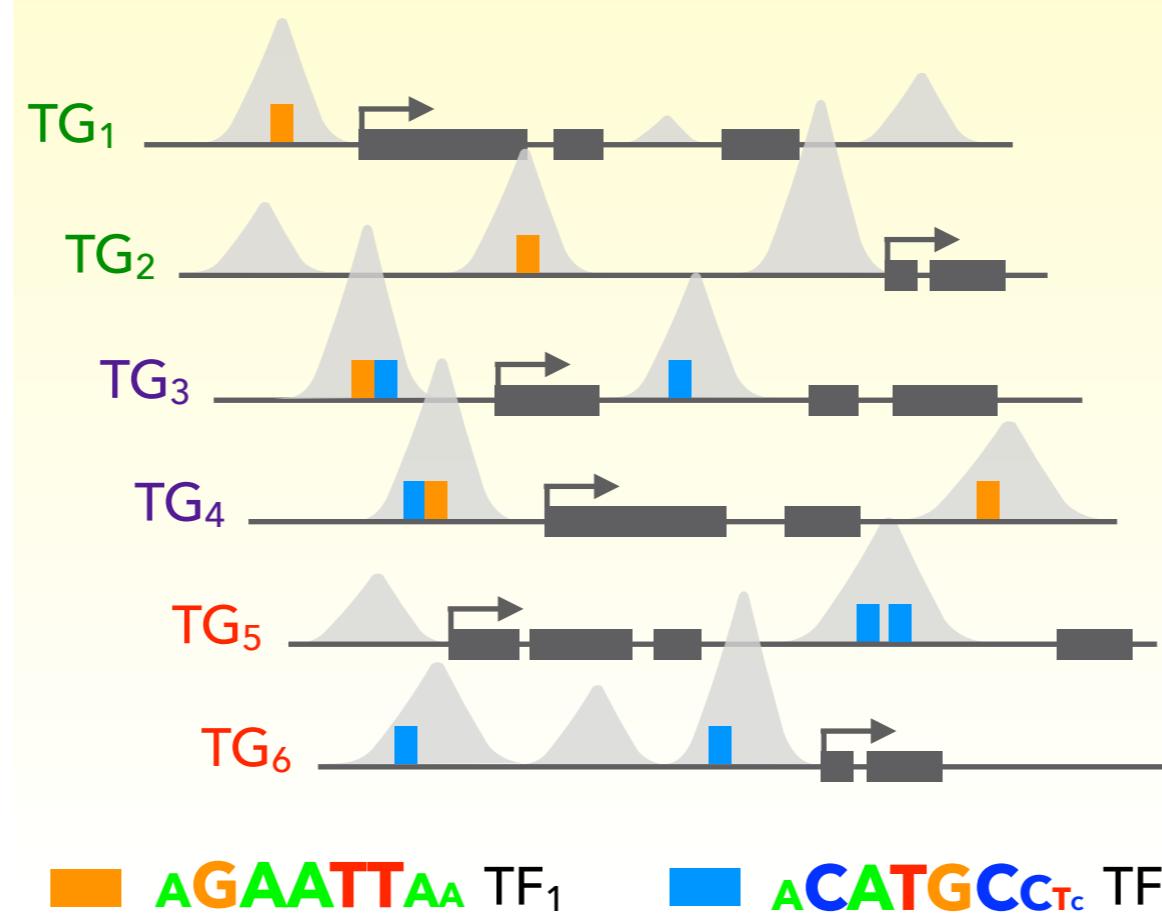
a**b****c**

a

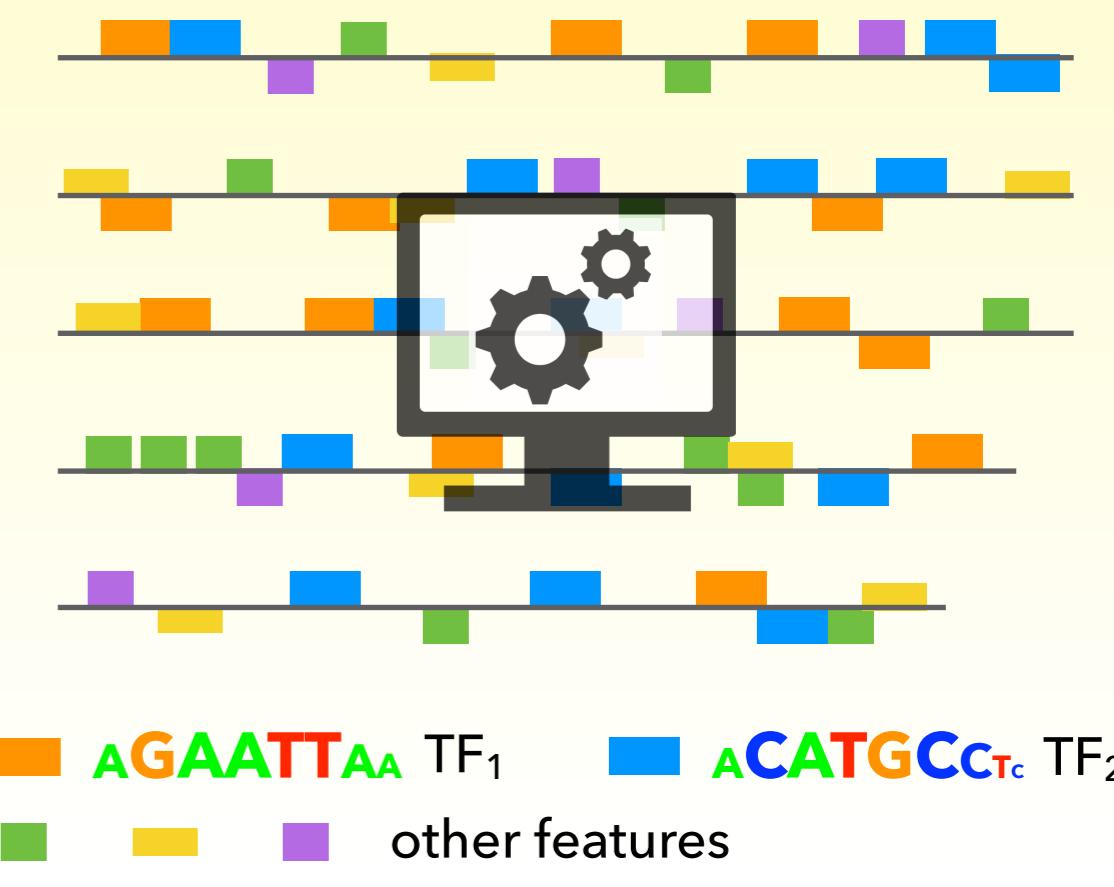
Reverse engineering gene signatures

**b**

cis-regulatory sequence analysis on gene and enhancer signatures

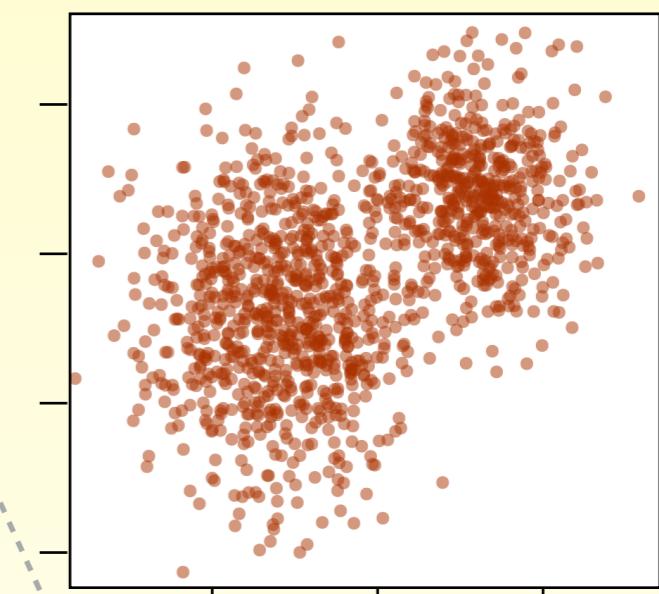
**c**

Enhancer models via machine learning

**d****e**

Network inference from high-throughput single cell assays

Single cell t-SNE



Attractor states

