Adopting tipping-point theory to transcriptome profiles unravels disease regulatory trajectory

Running title: transcriptomic network perturbation signature

Xinan H Yang[1,*], Zhezhen Wang[1], Dannie Griggs[1], Qier An[1,2], Fangming Tang[1], John M Cunningham[1]

[1] Department of Pediatrics, [2] Graham School of Continuing Liberal and Professional Studies, The University of Chicago, Chicago, IL 60637, USA

XHY: xyang2@uchicago.edu
ZW: zhezhen@uchicago.edu
DG: dgriggs@uchicago.edu
QA: ann72@uchicago.edu
FT: ftang@peds.bsd.uchicago.edu
JMC: jmcunningham@uchicago.edu

Key words: transcriptional regulation; network dynamics; translational genomics; critical transition; long noncoding RNA (lncRNA)

Highlights:

- We adopt 'tipping-point' theory to identify distribution-transition in disease regulatory states
- Critical transcriptional transition happens between low-risk and a high-risk neuroblastoma states
- A critical transition signal (CTS) based on coherent expression of genes and lncRNAs shows prognostic significance
- GWAS-scans with the CTS unveiled five overlooked genes and four lncRNAs with promising clinical implications
- We propose a CTS-amplifier model that unveils complex but mastering trans-regulation in disease

Abstract

Tipping-point models have had success identifying transcriptional critical-transition signal (**CTS**) in tumor phenotypes using time-course data. Can these mathematical models be adopted to cross-sectional transcriptome profiles? Furthermore, can the

CTS analysis that characterizes tumor progression be applied to lncRNA-expression patterns? This study introduces a novel network-perturbation signature (**NPS**) scoring scheme to model a phenotype-defined tumor regulatory system. Applying NPS to neuroblastoma transcriptome of two patient populations yielded two CTSs that reproducibly identified a critical system transition between the low-risk and a high-risk state. The coherent expression pattern of one specific CTS, consisting of mRNA and lncRNA components, showed prognostic significance. Associating GWAS-scans with the CTS unveiled four overlooked intergenic loci and five genes with promising clinical significance. Additionally, a new mechanism of 'CTS-amplifier' is proposed, modeling how CTS-transcript fluctuation response to complex master regulators such as c-*MYC* and *HNF4A* uniquely in the transition state. Overall, NPS is a powerful computational approach that provides a breakthrough in phenomenological analysis of collective regulatory trajectory by applying 'tipping-point' theory to '-omics' data.

Introduction

Transcriptional network oscillation induces critical transitions in biology ranging from single-cell yeast to looped human immune response systems (Blake, M et al., 2003, Bury-Mone & Sclavi, 2017, Lesterhuis, Bosco et al., 2017). Can we understand how these dichotomous phenotypes come about by modeling expressional oscillation over regulatory trajectory in disease? We aim to unravel transcriptional critical transition for cell fate determination and subsequent disease phenotypes.

Indeed, multiple gene products combine to produce a phenotype of complex disease. Therefore, it becomes necessary to elucidate gene regulatory network properties. Gene regulatory network is turned by temporal transcription with noise. By modelling dynamic network properties (or correlated gene-oscillation) along with disease states (grouped phenotypes), our group and others have identified transcriptomic signals for critical state transitions (Chen, Liu et al., 2012, Mojtahedi, Skupin et al., 2016, Yang, Li et al., 2018, Yang, Li et al., 2015). Some of these signals showed clinical significances, for instance epithelial-mesenchymal transition in cancer progression (Jolly, Ware et al., 2017, Tian, Zhou et al., 2018), metabolic transition between tumor and microenvironment (Kianercy, Veltri et al., 2014), and tumor-immune interaction (Lesterhuis et al., 2017).

The computational methods to identify these signals are based on a 'tipping point' theory (O'Regan & Drake, 2013, Scheffer, Bascompte et al., 2009, Scheffer, Carpenter et al., 2012). In this theory, complex gene regulatory systems exhibit abrupt shifts between two distinct stable states (**Figure 1a**). Each equilibrium or stable state is characterized by quick recovery rate on perturbations with large basin of attraction, which is maintained by specific attractors (e.g., steady gene interactions). Thus, between-sample autocorrelation is large in each stable state.

When a steady state is exited, there appears an increase in gene-expressions' fluctuation and autocorrelation until the critical transition (also called tipping point or regime shift) is passed and the next steady state is entered.

Fundamental of the tipping-point can be applied to transcriptomic profiles where highly connected transcriptional configurations serve as attractors (Scheffer et al., 2012). Although researchers have been able to identify critical transition signal (**CTS**) (impending tipping points) for complex biological systems including cancer (Mojtahedi et al., 2016, Yang et al., 2018), existing tipping-point models are designed for expensive longitudinal cohort studies and thus cannot be efficiently applied to cross-sectional transcriptomic data (Liu, Chen et al., 2015, O'Regan & Drake, 2013). These current methods present three main limitations.

1) Individuals in a phenotype-defined state hold genomic heterogeneities (noise) that aren't necessarily linked to disease-state changes.
2) Isolating the regulatory network impact of disease progression is hard without healthy controls in many studies of clinical samples.
3) Extracting the biological and clinical relevance from the identified CTS is unlikely and rarely discussed since the signature-identifications are data-driven (i.e., chaotic or driven by unknown factors).

Additionally from a systems biology viewpoint, lncRNAs modify the transcription of target genes either as noncoding enhancer-derived RNAs (**eRNAs**) or as regulatory RNA elements themselves (Kim, Xie et al., 2018). Therefore, a comprehensive description for transcriptome oscillation considers regulatory long noncoding RNAs (**lncRNAs**) (Engreitz, Haines et al., 2016, Pavlaki, Alammari et al., 2018, Washietl, Kellis et al., 2014).

Here, we develop a new network perturbation signature (**NPS**) framework to characterize transcription oscillation in a high probability space defined by phenotype, rather than an individual time-course space, that indicate disease regulatory 'tipping point' (**Figure 1a**). The NPS design is inspired by a previously 'tipping-point'-methylation study of multiple patients, each presenting an instance in the probability distribution of a phenotype-defined disease state (Teschendorff, Liu et al., 2014). The assumption is that slightly perturbations of a hidden process (i.e., expression of a master regulator) could trigger critical transition through a group of genes' expression changes. As proof of concept, NPS was applied to neuroblastoma (**NB**) transcriptomes and successfully identifies a critical regulatory transition between the low-risk and a high-risk state. Finally, previously uncharacterized CTS transcripts are analyzed for their role in biological pathways, demonstrating the broad utility for analysis of network perturbation for a wide range of tumors.

Result

NPS advances existing approaches to tipping-point modelling

NPS offers a new illustration of transcriptional dynamics which includes lncRNA expressional patterns. It also advances current computational pattern recognition in three ways outlined below (**Figure 1b** with details in **Appendix Table S1**).

NPS can be applied to transcriptome profiles without needing baseline or healthy control samples. Its scoring system calculates expressional oscillation for each state and then compares these state scores to one another. Its goal is to measure state-level trends not individual-level trends; in this was NPS minimizes the influence of individual patients' heterogeneity. Therefore, NPS overcomes two limitations of existing methods: needing normal control samples and an expressional distinction (fluctuation) between the case samples and the control samples (Chen et al., 2012, Teschendorff & Widschwendter, 2012). Compared to a recent latter-vs-former strategy (Zhang, Liu et al., 2019) that also overcomes the second limit, NPS is an ideal solution because it pick up a global maximum.

NPS constructs subnetworks (or modules) based on graph-structure, which is as efficient as conventional methods without requiring prior knowledge. Network partition optimization has been unfortunately disregarded in previous computational schemes for CTS identification. Hierarchical clustering (**HC**) requires prior knowledge to set either a partitioning threshold or a limit for the largest sub-cluster (Chen et al., 2012, Yang et al., 2018, Yang et al., 2015). Alternative strategies, including semi-supervised Partitioning Around Medoids (**PAM**), prove challenging when trying to optimize the parameter *K* (Teschendorff et al., 2014). A simple partitional clustering by splitting (Li, Zeng et al., 2014) is limited to finding two modules. However, random walk (**RW**) avoids these pitfalls while generally agreed with four other partition methods (**Appendix Figure S1a**). RW was able to construct relatively concentrated subnetworks, each with modest subnetwork size (**Appendix Figure S1b**).

NPS improves on a method called Dynamic Network Biomarker (**DNB**) that was calculated within each state (Chen et al., 2012). When studying NB, expressional patterns are often aggregates of multiple dynamic subsystems. Thus using DNB for NB tipping-point detection is not effective because the autocorrelation of non-DNB-modules at other states could attribute to an even higher DNB-score at a predicted 'tipping point' (**Appendix Figure S1c**). A new method is needed to allow for cross-state comparisons. By incorporating an index of critical state transition (**Ic-score**) (Mojtahedi et al., 2016), NPS overcomes DNB's pitfall by checking for drastic deviations across samples around the tipping point. The NPS scheme acts as a hybrid model to join the advantages of both DNB and Ic-scoring.

NPS is designed for a genome-wide CTS identification from transcriptome

To adopt tipping-point theory to widely available transcriptome data of complex disease, we designed a five-step analytical scheme (**Figures 1c, EV1**).

i) For data pre-procession, NPS introduces a concept of 'phenotype-defined state,' postulating that patients in the same state represent a distribution of expression-configuration effect (Chen et al., 2012, Honda, Kaneko et al., 2001, Teschendorff & Widschwendter, 2012). Samples were therefore categorized into phenotypic states based on the Children's Oncology Group (**COG**) risk group criteria (American-Cancer-Society) (Matthay, Maris et al., 2016). Given the lower expression levels of lncRNA relative to coding genes (Derrien, Johnson et al., 2012), a mild 'expression' cutoff was applied to leverage noise and informative transcription.

ii) Increased variance (transcript fluctuation) is a measurable marker of the critical slowing-down characteristic of tipping points (Scheffer et al., 2009). Without normal-control samples, NPS is still able to pre-select transcripts with high variance using a relative transcript fluctuation (**RTF**)-score. The RTF-score assesses the variance of transcripts t_r of patients in a state *r*, relative to its complement set t_{cr} of patients outside the state (ie, a dynamic 'control state'). This gives,

$$RTF(t)|r = \frac{\mathrm{sd}(\mathrm{t}_r)}{\mathrm{sd}(\mathrm{t}_{cr})} \qquad \textbf{(Formula 1)}$$

iii) Clustering pre-selected features into modules is necessary as these subnetworks will be the inputs of downstream scoring calculation. NPS conducts RW to find densely connected subnetworks robustly and effectively (Pons & Latapy, 2005) (**Appendix Figure S1 a-b**).

iv) NPS summaries the three most important indicators of an impending tipping point -- slower recovery from perturbations, increased transcript autocorrelation, and increased variance (Scheffer et al., 2009). It does so by employing the DNB score (Chen et al., 2012, Li et al., 2014, Liu, Chang et al., 2017, Teschendorff et al., 2014, Teschendorff & Relton, 2018, Yang et al., 2018, Yang et al., 2015) which is renamed as the scoring Module-Criticality Index (**MCI**). For a state *r* and its pre-selected transcripts, a module *m*, comparing transcript deviation and autocorrelation in this module t_m relative to its complement set t_{cm} gives,

$$MCI(m)|_r = \mathrm{Avg}(\mathrm{sd}(\mathrm{t}_{m.})) \times \frac{Avg(|\,PCC(\mathrm{t}_{mi},\mathrm{t}_{mj})\,|)}{Avg(PCC(\mathrm{t}_m,\mathrm{t}_{cm}))}|r \qquad \textbf{(Formula 2)}$$

Taking the absolute value (|.|) is meaningful because both positive and negative feedbacks are indicative in looped regulatory network. Then, the module with the highest MCI is termed the dominating module or "biomodule" using $m^{\wedge}|r = m|maximum(all\ MCI(m)|_r)$ within each state.

v) Finally, NPS employs the Ic-score to predict critical transition and an empirical simulation to make determination. Given a biomodule *m*^ defined from a state

r with samples $s_{r.}$, the Ic-score summarizes the decrease of autocorrelation between samples and concomitant increase of autocorrelation between transcripts, characteristic of a tipping point.

$$Ic(r)|_{m^\wedge} = \frac{Ave(|PCC(t_{m^\wedge i}, t_{m^\wedge j})|)}{Ave(PCC(s_{rk}, s_{rl}))} \qquad \textbf{(Formula 3)}$$

Following the induction of (Mojtahedi et al., 2016), the estimated tipping point r^ should be the state where Ic(r) achieves its maximum $r^\wedge|m^\wedge = r|\ maximum(all\ TC(r)|_{m^\wedge})$; and a successful biomodule would display an increased Ic-score around the tipping point while random gene combinations shouldn't (Mojtahedi et al., 2016). Therefore, NPS tests two null-hypotheses:

a) H_0: Given a m^, the value of $Ic(r^\wedge)|_{m^\wedge}$ is just an instance from those of random Ic*-scores calculated at r* but for any transcripts (of the same size of *m*^) from the 48.5k transcript background.
b) H_0: Given a r^, the value of $Ic(r^\wedge)|_{m^\wedge}$ is an instance from those of random Ic*-scores calculated for m* regardless of the phenotype-defined tumor states.

Lastly, an evaluation for both biological and clinical relevance of CTS transcripts is included in our NPS application (**Figure 1d**). In an 'omnigenic' model to link common genetic variants and disease, small trans effects can be amplified by co-regulated network (Liu, Li et al., 2019). Because CTS genes and lncRNAs exhibit increased autocorrelation at the tipping point, we expect NPS identification to shed light on novel predictions of master regulators at or near the top of the regulatory hierarchy including noncoding variants. These relevance analyses include (but are not limited to) prognosis analysis, DNA-binding enrichment analysis, and NB-susceptibility loci selection from previous genome-wide association study (**GWAS**) findings (**S Methods**). This is the first tipping-point model to cover lncRNA as a CTS component and will help surmise which lncRNA marks are the most possibly implicated in complex disease out of the thousands of possible candidates.

CTS identification using NPS in neuroblastoma (NB)

NPS was applied to NB while considering three important aspects of this disease. Transcriptome profiles of NB have been collected not only from large populations (n=498, GSE49711, **Figure EV2, Appendix Table S2**) (Zhang, Yu et al., 2015) but have employed total RNA-seq library to cover non-polyadenylated transcripts, thus enabling the investigation of lncRNA transcripts. Additional '-omics' data detecting potential enhancers, disease susceptibility, and chromatin accessibility have already been the subject investigation for NB (Zeid, Lawlor et al., 2018). This allows for the evaluation of functional consequence for identified lncRNAs. Lastly, studying NB requires a data-driven model because there is limited knowledge on its

heterogeneous mechanisms (eg, MYCN-amplification and few common variants) (Bosse & Maris, 2016, Maris, Hogarty et al., 2007) .

NPS identified a CTS appearing at the Low-or-Intermediate-Risk (**LIR**) state in both event-free (**EF**) and event-occurring (**E**) populations. 48.5k (80%) transcripts were remained after applying a mild 'expression' cutoff (the cross-sample average expression level ≥1 on log2 scale). 20.6k (57.5%) of these transcripts are lncRNAs. A robust RTF-score was calculated by bootstraping 80% of samples (n=100) in each state to decrease the impact of individual heterogeneity. This bootstrapping pre-selected a variety of transcripts ranging from 72 to 274 at each state, and the LIR state showed increased and scattered transcript variance (**Figures 2a, Appendix S2**). After applying a network partition strategy, each pre-selected transcript set was further divided into 2-10 modules (**Figures 2b, Appendix S3**). Then, a dominating module herein the 'biomodule', was identified by ranking the MCI-scores per state (**Figures 2c, Appendix S4**). Again, the biomodules of the LIR stage (m_LIR.EF and m_LIR.E) showed the highest MCI-score, in both populations.

The critical transition occurring in the LIR state was confirmed by the pattern of Ic-scores -- the Ic-score of the LIR-derived biomodule peaked at the LIR-state. Additionally, the observed Ic-score was significantly higher than the scores (Ic*) of any randomly combined transcripts at the LIR-state (black line versus grey lines in **Figure 2d**). Sample-size controlled simulation analysis further confirmed a higher value for the observed Ic-score over the LIR patients than any randomly grouped patients (empirical P<0.001, **Figure 2e**).

Genes and lncRNAs coordinately co-expressed and fluctuated as a CTS

The identification of the LIR state as a tipping point was verified by an increased variability and transcript correlation among the CTS transcripts (**Figure 3a**). Theoretically, this fluctuation explains why the 'tipping point' is susceptible to disturbance. Although derived from independent cohorts, the two CTSs at LIR-state significantly overlapped, sharing 65 transcripts (**Figure 3b**, OR=825), thus indicating a common transcriptional configuration committing to phenotype-defined states in NB. Among these shared CTS transcripts, 17 were lncRNAs, posting a substantial significance from the 20.6k expressed lncRNA background (OR=1551). We conclude that lncRNA expression plays a role in CTS that determines systems dynamics.

Given the importance of mRNA-lncRNA interaction, the network-degrees were evaluated for all possible genes and lncRNAs in each CTS network. LncRNAs and genes exhibited a significantly higher proportion of connections than gene-gene connections did in both identified CTSs (P<5e-10, **Figure 3c, S Method**). Similarly, significantly higher connections existed between lncRNAs and genes than lncRNA-lncRNA connections in both identified CTSs (P<2e-7, **Figure 3d**). Note that over 90% of these co-expressed lncRNA-gene pairs reside on different chromosomes (**Figure 3e**).

This high proportion of trans lncRNA regulation agrees with that most heritability of gene expression is due to remote trans-regulation (Liu et al., 2019).

Additionally, both CTSs were significantly depleted with differentially expressed transcripts (**Figure 3f**). These results confirmed previous observations that a CTS can't be identified by conventional statistics on group difference in mean (Chen et al., 2012).

We conclude that NPS captures unique CTS based on significant correlations and deviations among transcriptional dynamics in the critical transition state, which are uncharacterized by traditional statistics.

Clinical relevance of the identified neuroblastoma CTSs

The prognostic implications for the identified CTS can be measured by a relative expression analysis with gene-set pairs (**RXA-GSP**) (Yang, Vasudevan et al., 2013), given critical transition is marked by concerted rather than individual changes of expression. RXA-GSP quantifies each patient's transcriptional ratio of poor-outcome and good-outcome components into a prognostic index and thus discerns the potential clinical utility of cancer-biomarkers (Yang et al., 2013, Yang et al., 2015, Yang, Tang et al., 2017a, Yang, Tang et al., 2017b). To apply RXA-GSP to the CTS in the EF population, its 137 transcripts were first grouped into 76 poor-outcome predictors (with higher expression levels in 56 High-Risk patients than in 168 Low-Risk patients) and 61 good-outcome predictors (**Appendix Table S3**). As intended, their transcriptional ratio showed significantly prognostic effects for event-free survival (log-rank P=2.6e-8, empirical P=0.007, **Figure 4a, S Method**) and overall survival (log-rank P=4.4e-11, empirical P=0.01). Notably, the same analysis of mRNA components was not prognostic, nor was the lncRNA analysis (both log-rank P>0.05), indicating the CTS has a collective prognostic effect of mRNAs and lncRNAs.

We asked if the identified CTS transcripts can indicate the critical transition between Low-Risk and High-Risk, using the Ic-score on independent transcriptome profiles of NB patients (Oberthuer, Juraeva et al., 2015, Oberthuer, Juraeva et al., 2010, Wang, Diskin et al., 2011). In all three datasets with over a hundred of samples, the observed Ic-score archived a summit at intermediate risk state (either LIR, LM, or HM) (**Figure Ev4a**). We conclude that there exists a critical system transition between the low-risk and a high-risk state in NB that can be detected from temporal transcription oscillation.

It is important to check whether CTS disclosed previously overlooked NB-associated variations. Seeing as GWAS rarely identifies single variant-trait associations but instead blocks of associated variants in linkage-disequilibrium (**LD**) blocks (Wainberg, Sinnott-Armstrong et al., 2019), we scanned 163 NB-associated linkage-disequilibrium (**LD**) blocks hosting significant or suggestive NB-susceptibility SNP(s) (GWAS P<5e-5, **S Methods**) (Wang, Cunningham et al., 2018). Because regulatory interactions are frequently observed within topologically associating domain (**TADs**) (Dixon, Selvaraj et

al., 2012, Jager, Migliorini et al., 2015), it was assumed that most NB-associated cis-acting happen in 151 TADs defined by the 163 LDs. Querying these 151 NB-associated TADs with both identified CTSs allowed for pinpointing nine TAD regions with susceptible SNPs and CTS transcripts. One notable finding was gene *DCDC2*, whose copy number gain presented in 6 out of 29 NB tumors studied (Costa & Seuanez, 2018) and lies adjacent to NB suppressive lincRNA pairs *CASC14/15* (Mondal, Juvvuna et al., 2018) (**Table 1**). Strong evidence also exists of DCDC2 interacting with both the CASC14/15 locus and the neuron-differentiate marker gene *NRSN1* which comes from Hi-C mapping in the NB cell (**Figure EV3a**). Overall, four lncRNAs and five genes (**Figure 4b**, green and blue dot, respectively) are CTS markers that have promising clinical significance for future investigation.

Note that eight suggestively susceptibility loci (8e-5<P<5e-5, **Table 1**) were overlooked in the previous GWAS (Wang et al., 2011), even all appear to have some connection with NB in literature (**Appendix Table S4**). Among those loci are two lncRNAs on chr5 q31.1 with likely cis-acting potential (*PPP2CA-AS* and *LINC01843)* given their overlap with the human novo lncRNA landscape (Iyer, Niknafs et al., 2015) and enhancer histone marks (**Figure 4c**)*.* Hi-C mapping revealed interaction with each other and with *SLC22A4* (that hosting a NB susceptible SNP rs11950562). Given the lower expression of *SLC22A4* in HR patients than non-HR patients (FC=0.45, FDR=2e-44), this interaction appears to be associated with the NB phenotype. Additionally, HNF4A was observed to interplay with this area through either DNA-binding at the *PPP2CA/CDKL3/PPP2CA-AS* shared promoter, or LIR-specific co-expression (PCC=0.68, FDR=1.1e-4). We conclude that suggestive variants could be significantly disease-associated (Natarajan, Peloso et al., 2018, Onengut-Gumuscu, Chen et al., 2015, van der Harst & Verweij, 2018) and NPS contributes to the hunting.

HNF4A presents as a master regulator of neuroblastoma CTS

In transcriptional regulatory network, little malfunction of master regulators may have enlarged consequence on downstream tissue- or condition-specific transcription. DNA-binding motif analysis of the identified two CTSs revealed 12 potential master regulators, including *ATF6, EHF, GPAM, MAX, NFE2L1, NR4A1, HNF4A, HNF4G, NFIX, PPP5C, KIF22*, and *HNF1B* (**Figure 5a-b**).

Hub gene often drives its gene-regulatory network thus be the optimal treatment target (Lesterhuis, Rinaldi et al., 2015). Using Ingenuity pathway analysis (IPA) (Kramer, Green et al., 2014), these upstream regulators were constructed into a disease-associated network in which most regulators intensively interacted with three nucleus hubs -- Max/Myc, Hdac, and NHF4A (**Figure 5c**). *HNF4A* gene encodes HNF4α, a nuclear transcription factor that binds DNA. Through protein-protein interaction (Rouillard, Gundersen et al., 2016), half of the 12 up-regulators (NFE2L1, KIF22, ATF6, MAX, HNF1B, and PPP5C) interacted with HNF4α (OR=7.5, P=0.0012, among 2581

HNF4α-interactors in the database). We further notify the importance of HNF4α by not only the observed network-hub but its NB tumorigenesis role *in vitro* and *in vivo* (Xiang, Zhao et al., 2015).

Next, the CTS-based gene regulatory network was evaluated to determine if its impact by *HNF4A* was significant rather than stochastic. For the LIR state, the patients' *HNF4A* expressions were divided into four levels to calculate the *HNF4A*-dependent fold-change of the 191 identified CTS transcripts (**Figure 3b**) by comparing the tumors within the upper-quarter of *HNF4A*-expression levels to the tumors within the lower-quarter levels. After conducting the same calculation on *MMP14*, positive regulatory correlation was found between *HNF4A*-dependence and *MMP14*-dependence (**Figure 5d**, PCC=0.65). This correlation was significant (empirical P<0.05) in both intermediate states and had the highest coefficient in the LIR, affirming a previously reported coherence between *HNF4A* and *MMP14* expression in NB tissues and cell lines (Xiang et al., 2015). These data suggest an uncharacterized enhanced correlation occurring before the high-risk state.

As a control, the inverse correlation between *MYCN*- and *MYC*-dependence *in vitro* (Huang, Cheung et al., 2011) was verified among all patient samples for the 191 CTS transcripts (**Figure 5d**, blue dots at the row *MYCN* and column *MYC*, PCC=-0.53, -0.81, -0.46, and -0.49 in all four states, respectively, but significant in the two intermediate states). *MYCN* and *MYC* serves as an oncogenic cause in distinct subsets of HR NB (Yang et al., 2017a, Zimmerman, Liu et al., 2018); both intriguingly interplay with HNF4A. In this study, *HNF4A*- and *MYCN*-dependence were positively correlated and empirically significant only in the LIR state; while *HNF4A*- and *MYC*-dependence were significantly negatively correlated in the two intermediate states. This validation provides further evidence that *HNF4A*-regulation of NB progression happens before the high-risk state.

We further evaluated the global transcription dependence on *HNF4A* relative to the dependence on *MYC* and *MYCN* genes. In all three datasets, we recaptured the consistent patterns of transcription-dependence: positive correlation between HNF4A- and MYCN-dependence, reverse correlation between HNF4A- and MYC-dependence, and reverse correlation between MYCN- and MYC-dependence (**Figure Ev4b**). These results posit an interplay between abnormal *HNF4A* expression and myc-pathway.

Given the importance of *HNF4A*, the identification of its antisense RNA as a CTS transcript promoted further investigation. *HNF4A-AS*, one of the 191 CTS transcripts, is located on the opposite strand of *HNF4A* and intriguingly showed an inverse correlation with *HNF4A* expression levels in all states but the Low-Risk state. Further findings included a significant inverse correlation between *HNF4A-AS* and *MMP14* in LIR and HIR states implying that *HNF4A*'s impact on NB ontogenesis is modified by the

predicted antisense RNA. These post the *HNF4A-AS* transcription to be a promising new marker of NB oncogenes.

Overall, the CTS's transcription changes are most strongly dependent on the evaluated four regulators (HNF4A, MMP14, MYCN, and MYC) in the LIR state, not only confirming LIR as a tipping point but also suggesting that these regulators are biologically significant.

A CTS-amplifier model

Transcriptional correlation of CTS will put associated master transcription factor (**TF**) in tune, which may interpret why *HNF4A* has two oncogenic potentials in NB. On one hand, HNF4α directly targets the MMP14 promoter to facilitate its transcription, promoting the invasion, metastasis and angiogenesis of NB cells *in vitro* and *in vivo*, suggesting that HNF4α has an oncogenic role (Xiang et al., 2015). On the other hand, this study observed occupancy of histone repressive mark (H3K27me3) *in vitro* indicative of a suppression of *HNF4A* in HR-NB (**Figure EV3b**). In agreement with the latter, HNF4α and other TF co-led differentiation of neural stem cell (Wang, Cheng et al., 2013). Furthermore, there is no significant differences in *HNF4A* expression between HR and non-HR patients, explaining why few literature describing HNF4A in NB. These suggest a complex if not contradictory role of HNF4α in differentiation that has recently emerged in cancer of multiple organs and embryonal cell (Chiba, Itoh et al., 2005, Lucas, Grigo et al., 2005, Sun, Xu et al., 2019, Tanaka, Jiang et al., 2006, Wei, Dai et al., 2017).

A CTS-amplifier model (**Figure 5e**) considers pathological implications for a gene on two levels – first, the direct mechanisms for how that gene might contribute to a disease pathway; second, the dynamic/indirect effects a gene can have through influencing regulatory networks. For example, HNF4α blocks cell proliferation could be controlled by a completion between Myc and HNF4α for control of the CDKN1A (p21) promoter (Hanse, Mashek et al., 2012, Sladek, 2012), evident by the reverse correlation in expression between *HNF4A* and *MYC* in NB patients (**Fig 4d**) and in liver cancer (Sladek, 2012). Most notable among many liver-specifically *HNF4A* down-regulated genes (Bonzo, Ferry et al., 2012) were *Ccnd1*, *Cdk1*, and *Myc* that were previously identified to be important for NB tumor proliferation and progression (Yang et al., 2017a). In mouse induced pluripotent stem cell models, *HNF4A* and *MYC* were both predicted as upstream regulators of genes involved in neuronal differentiation (Ando, Kato et al., 2015). Additionally, if validated in NB that HNF4α binds with both USP36 and SOX9 (Odom, Zizlsperger et al., 2004, Rouillard et al., 2016), HNF4α may have a direct impact on the core CASC14/15-USP36/CHD7-SOX9 pathway of NB suppression (Mondal et al., 2018). In this amplifier model, the weak but essential perturbations caused by suggestive genic variants or abnormal TF activity can be amplified by CTS and configure the complex and dynamic HNF4α-regulation. Different with the amplifier

in a ‘omnigenic’ model (Liu et al., 2019), gene regulatory network dynamics is introduced for the first time.

The identified genetic variants (**Table 1**) may trigger a system transition through an enlarged impact on core regulatory pathways through its target gene(s). For example, the SNP rs3828112 on chr1 was identified as a NB-risk loci by optimal FDR control of GWAS-findings (Xie, Cai et al., 2011). Significance of rs3828112 can be further informed by its local genes: Its host gene NR5A2 has been of interest in oncogenic processes, chemoresistance, and therapeutic target (Fernandez-Marcos, Auwerx et al., 2011, Lin, Aihara et al., 2014, Wang, Zou et al., 2018, Xiao, Wang et al., 2018); while its adjacent gene *KIF14* is HR up-regulated (FC=2.03 FDR<2e-16) in NB. Therefore, the CTS-amplifier model could facilitate the discovery of cis-acting loci, providing future direction for investigation.

We suggest, the CTS-amplifier model has the potential to shed light on understanding suggestive variants and TFs that largely practice in looped pathway to cause disease phenotypes. Because small-scale variants are able to escape natural selection’s strong constraints against large-effect variants (Chen, Wu et al., 2017, Simons, Bullaughey et al., 2018), grouping individuals into phenotype-defined groups allows NPS to unveil the ‘distribution transition’ that is seemly trivial in individual but disease-causal in distribution. This amplifier model also explains why over 90% of trans-regulation in the identified CTS (**Figure 3e**) as well most of the heritability of gene expression (~70%) are trans that seem to have tiny effects on individual (Liu et al., 2019). Since current static statistics have been unable to address these issues, the CTS-amplifier concept provides versatile perspectives for future cancer biomarker research and the development of cancer treatment.

Discussion

In this study, we proposed a test of ‘probability distribution-transition’ rather than individual ‘state-transition’ to detect the tipping point and CTS from cross-sectional ‘omics’ data. Genetic and epigenetic factors are the biologically significant contributors triggering gene regulatory network transitions but unfortunately random noise can also influence this (Scheffer et al., 2012). Previous studies of noisy time-course data have used ‘moment-variables’ (generated using sliding-windows of consecutive time points) to estimate the ‘probability distribution-transition’ (Liu et al., 2015). Instead, we assembled samples within a phenotype-defined state to estimate the distribution-transition of disease regulatory trajectory. Therefore, we filled this fundamental translational gap between a personal unit (or time-course) and a regulatory unit (phenotypic window) by introducing the computational framework, NPS, and demonstrating its ability to unveil disease causal regulators.

Applying pharmacological intervention during the window of transcriptional fluctuation has the most potential for halting disease-state progression (Lesterhuis et al., 2017, Zhang et al., 2019), given that critical state transitions (or tipping points) are fragile, reversible, and can be identified by statistical tests (Scheffer et al., 2012, Trefois, Antony et al., 2015). Our identification of the LIR state as the NB tipping point is affirmed by the fact that intermediate-risk tumors have been shown to present high cure rates and are susceptible to chemotherapy (reversible) (Baker, Schmidt et al., 2010). Thus, this study serves as proof-of-concept that molecular signatures can be indicators of arrival at a reversible state in disease progression and may provide new treatment options. The jointly identified CTS transcripts from the two independent cohorts (Event and Event-Free patients) seem to be involved in pushing the system over the tipping point towards a cellular differentiated, drug sensitive, and non-malignant state. In fact, critical transition state reflects vulnerable regime of the system that could be reversible (Trefois et al., 2015).

This work presents the adoption of the fundamental of the tipping-point theory into an advanced computational framework, NPS, which successfully identifies the transcriptional transition signals from phenotype-defined ‘longitudinal’ cancer states. NPS focuses on interactions between SNP loci, TFs, and their target genes in the CTS, providing particular use in its ability to identify previously uncharacterized signals that were previously missed because of the suggestive GWAS p-value of variant or insignificant expression change (by testing group-mean difference) of the TF itself. NPS provides five major contributions towards translation bioinformatics workflow.

1) NPS introduces the concept of ‘distribution-transition’ of disease trajectory into re-analysis of massively available transcriptome profiles of patients. Distribution-transition pinpoints phenotype-contributing transcription changes whose individual effects are insignificant in static statistics, while their coherent oscillation makes a dynamic signature.
2) There is a wealth of information available on transcriptional interaction where lncRNAs are embed in noise, being regulated by upstream regulators and are in turn affecting gene expression (Huarte, 2015). In this way, NPS provides immense value by filtering through noise to identify master regulators that mark disease regulatory trajectory.
3) The small effects of trans-regulation are hard to detect but dominate the transcriptomic architecture of complex traits (Liu et al., 2019). The CTS transcripts propose a theoretical path forward for linking suggestive loci to trans-regulatory network to disease phenotypes.
4) Lastly but not least, the CTS-amplifier model fits well with an assumption of correlated genes amplified the expressional oscillation of weak but essential

regulators (Liu et al., 2019). Moreover, the CTS-amplifier model encompasses phenotype-associated transcription factor activities and polygenicity.

Certain limitations of the NPS approach must be addressed. The model's underpinning of disease regulatory trajectory relies on sample-grouping which is not always practicable. Instead of comparing the data to normal controls (which do not exist), NPS uses a 'relative control' by comparing one state against the complement states in the same disease. This requires a large sample size and enrollment of most states of a disease. It remains unclear if this pre-selection strategy can correctly identify multiple critical-transition states that are postulated in some complex biological systems (Majdandzic, Braunstein et al., 2016, Scheffer et al., 2009). In fact, the findings of this study suggest another possible tipping point at the High-Risk state where the Ic-scores of a CTS subset exhibited a second peak (**Figures Ev4a**).

Based on master regulator analyses of the CTS transcripts in NB, a novel CTS-amplifier model is proposed. This model interprets seemingly controversial reports as caused by feedback loops between core networks and the CTS-based regulatory network. We reported HNF4α to be a master regulator of the transcriptional regulatory hierarchy which may engage repressive histone modification (**Figure EV3b**) and be affected by other genes' expression, such as CTS member *HNF4A-AS*, known NB core pathway members *MYC* and *MYCN*, and its co-regulator *MMP14* as an instance of dynamic impact on NB (**Figure 5d**). Although more experiments are required to verify the impact of *HNF4A* expression levels on NB, we predict the existence of a CTS-amplifier of *HNF4A*-regulation in NB system. This model goes beyond the previous 'omnigenic' hypotheses (Liu et al., 2019) by considering gene regulatory network dynamics.

Epigenetic regulation plays a role in the CTS. The presence of histone deacetylases (HDACs, **Figure 5c**) is of particular interest given that the HNF4α1 (isoform 1) recruits HDAC activity to target promoters and thus mediates multiple interactions (Torres-Padilla, Sladek et al., 2002). Previous studies suggest HDAC inhibitor targeting of HNF-4α serves as an effective treatment for advanced carcinomas. In colon carcinoma cells, the HDAC inhibitor 'butyrate' significantly repressed the transcription of HNF4α and its target genes associated with proliferation (Algamas-Dimantov, Yehuda-Shnaidman et al., 2013). In human hepatoma cells, the recruitment of HDAC activity played a partial role in p53-mediated repression of the HNF4α P1 promoter (Maeda, Hwang-Verslues et al., 2006). In multiple cancer cell lines, an HDAC inhibitor, TSA (trichostatin A), has been shown to partially recover p53-mediated repression of the HNF4α promoter modifications of histones, thus playing a critical role in regulating gene expression (Maeda, Seidel et al., 2002). Furthermore, HDACs are involved in controlling MYCN function and are upregulated in chemotherapy-resistant NB cell. Treatment with pan-HDAC inhibitors caused cell cycle arrest, differentiation, apoptosis, and inhibition of clonogenic growth of NB cells and restored susceptibility to chemotherapy treatment

(Witt, Deubzer et al., 2009). Given these previous findings (Ghandi, Huang et al., 2019), functional interaction seems likely among the key genetic regulator HNF4A, epigenetic regulator HDAC, and the identified transcriptional CTS in NB tumorigenesis.

In conclusion, use of the NPS method in this investigation affirms that studying the perturbed connectivity in a transcriptional network between two equilibrium states can identify important CTS biomarkers (Grechkin, Logsdon et al., 2016, Mukherjee, Carignano et al., 2018). The identified CTS and their up-stream regulators pinpoint novel treatment targets surrounding a reversible state in disease progression. Although more wet-lab experiments are required to better understand the predicted CTS-amplifier model, the concept has broad utility for computational analysis of gene-regulation networks in a wide range of disease systems.

Methods

RNA-seq dataset

To study NB transcription systematically, we collected normalized RNA-Seq profiling of 498 primary patients (tumor > 60%, sequencing depth around 100 M, GSE49711) (Su, Fang et al., 2014, Zhang et al., 2015). Besides being defined into four phenotype-based risk-groups (LR, LIR, IR, and HR), patients were also disaggregated into two populations according to whether or not an event (the occurrence of progression, relapse or death) was observed in the following-up records (**Figure EV2**, **Appendix Table S2**).

Differential expression analysis (Figure 3f)

For the 48.5k expressed transcripts, differential expression was conducted by comparing all 176 high-risk (HR) tumors with 322 non-HR tumors in the dataset. We used Bioconductor tool *edgeR* (Robinson, McCarthy et al., 2010) with significance considered at (FC>2, FDR<0.05). This analysis identified 5,043 up-regulated and 3,490 down-regulated transcripts in HR vs non-HR tumors, among which 1,881 and 2,626 were lncRNAs.

Motif-enrichment analysis (Figure 5a)

A promoter was defined as the 2k-nt window ([-1,500, +500]) around each transcription-start site (**TSS**). Enrichment was evaluated for 1,571 known DNA binding motifs within a 2,000nt-window ([-1,500, +500]) surrounding TSS of the NPS transcripts (**S Method**). The significance was considered with the criteria: FDR <0.05, average binding-affinity score >1; and the breadth of enrichment using a 5% ranking threshold >15%.

Acknowledgements

This work was supported by NIH grant 5R21LM012619 (XY, ZW, FT, JC), the University of Chicago, Biological Sciences Collegiate Division Quantitative Biology Fellowship (DG), and Graham School of Continuing Liberal and Professional Studies, Master of Science in Biomedical Informatics Capstone Project (QA). We thank Drs. Charles Y Lin and Brian J. Abraham for providing global ATAC signature and H3K27ac signature, respectively, in NB cells. We thank Jorge Andrade for supporting the Ingenuity IPA analysis and high performance computing services. We thank Susan Cohn for her insightful discussion when initiating the project.

Author contributions

XHY designed the study, helped data analysis, interpreted the results, and wrote the manuscript, with comments from JMC. ZW collected data, performed data analysis, and contributed to result interpretation. DA wrote the Appendix Table S4 and edited the manuscript. QA contributed to Figure 2a and Appendix Table S1. FT contributed to the biological discussion and model evaluation. JMC directed the project design and result interpretation, and revised the abstract.

Completing interests

The authors declare no competing interests.

Reference

Algamas-Dimantov A, Yehuda-Shnaidman E, Peri I, Schwartz B (2013) Epigenetic control of HNF-4alpha in colon carcinoma cells affects MUC4 expression and malignancy. Cell Oncol (Dordr) 36: 155-67

American-Cancer-Society The Children’s Oncology Group (COG) risk group criteria. In

Ando T, Kato R, Honda H (2015) Differential variability and correlation of gene expression identifies key genes involved in neuronal differentiation. BMC Syst Biol 9: 82

Baker DL, Schmidt ML, Cohn SL, Maris JM, London WB, Buxton A, Stram D, Castleberry RP, Shimada H, Sandler A, Shamberger RC, Look AT, Reynolds CP, Seeger RC, Matthay KK, Children's Oncology G (2010) Outcome after reduced chemotherapy for intermediate-risk neuroblastoma. N Engl J Med 363: 1313-23

Blake WJ, M KA, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. Nature 422: 633-7

Bonzo JA, Ferry CH, Matsubara T, Kim JH, Gonzalez FJ (2012) Suppression of hepatocyte proliferation by hepatocyte nuclear factor 4alpha in adult mice. J Biol Chem 287: 7345-56

Bosse KR, Maris JM (2016) Advances in the translational genomics of neuroblastoma: From improving risk stratification and revealing novel biology to identifying actionable genomic alterations. Cancer 122: 20-33

Bury-Mone S, Sclavi B (2017) Stochasticity of gene expression as a motor of epigenetics in bacteria: from individual to collective behaviors. Res Microbiol 168: 503-514

Chen H, Wu CI, He X (2017) The genotype-phenotype relationships in the light of natural selection. Mol Biol Evol

Chen L, Liu R, Liu ZP, Li M, Aihara K (2012) Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci Rep 2: 342

Chiba H, Itoh T, Satohisa S, Sakai N, Noguchi H, Osanai M, Kojima T, Sawada N (2005) Activation of p21CIP1/WAF1 gene expression and inhibition of cell proliferation by overexpression of hepatocyte nuclear factor-4alpha. Exp Cell Res 302: 11-21

Costa RA, Seuanez HN (2018) Investigation of major genetic alterations in neuroblastoma. Mol Biol Rep 45: 287-295

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M et al. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22: 1775-89

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485: 376-80

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature 539: 452-455

Fernandez-Marcos PJ, Auwerx J, Schoonjans K (2011) Emerging actions of the nuclear receptor LRH-1 in the gut. Biochim Biophys Acta 1812: 947-55

Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER, 3rd, Barretina J, Gelfand ET, Bielski CM, Li H, Hu K, Andreev-Drakhlin AY, Kim J, Hess JM, Haas BJ, Aguet F, Weir BA, Rothberg MV, Paolella BR, Lawrence MS et al. (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature 569: 503-508

Grechkin M, Logsdon BA, Gentles AJ, Lee SI (2016) Identifying Network Perturbation in Cancer. PLoS Comput Biol 12: e1004888

Hanse EA, Mashek DG, Becker JR, Solmonson AD, Mullany LK, Mashek MT, Towle HC, Chau AT, Albrecht JH (2012) Cyclin D1 inhibits hepatic lipogenesis via

repression of carbohydrate response element binding protein and hepatocyte nuclear factor 4alpha. Cell Cycle 11: 2681-90

Honda M, Kaneko S, Kawai H, Shirota Y, Kobayashi K (2001) Differential gene expression between chronic hepatitis B and C hepatic lesion. Gastroenterology 120: 955-66

Huang R, Cheung NK, Vider J, Cheung IY, Gerald WL, Tickoo SK, Holland EC, Blasberg RG (2011) MYCN and MYC regulate tumor proliferation and tumorigenesis directly through BMI1 in human neuroblastomas. FASEB J 25: 4138-49

Huarte M (2015) The emerging role of lncRNAs in cancer. Nat Med 21: 1253-61

Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, Feng FY, Iyer HK, Chinnaiyan AM (2015) The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 47: 199-208

Jager R, Migliorini G, Henrion M, Kandaswamy R, Speedy HE, Heindl A, Whiffin N, Carnicer MJ, Broome L, Dryden N, Nagano T, Schoenfelder S, Enge M, Yuan Y, Taipale J, Fraser P, Fletcher O, Houlston RS (2015) Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. Nat Commun 6: 6178

Jolly MK, Ware KE, Gilja S, Somarelli JA, Levine H (2017) EMT and MET: necessary or permissive for metastasis? Mol Oncol 11: 755-769

Kianercy A, Veltri R, Pienta KJ (2014) Critical transitions in a game theoretic model of tumour metabolism. Interface Focus 4: 20140014

Kim YJ, Xie P, Cao L, Zhang MQ, Kim TH (2018) Global transcriptional activity dynamics reveal functional enhancer RNAs. Genome Res 28: 1799-1811

Kramer A, Green J, Pollard J, Jr., Tugendreich S (2014) Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 30: 523-30

Lesterhuis WJ, Bosco A, Millward MJ, Small M, Nowak AK, Lake RA (2017) Dynamic versus static biomarkers in cancer immune checkpoint blockade: unravelling complexity. Nat Rev Drug Discov 16: 264-272

Lesterhuis WJ, Rinaldi C, Jones A, Rozali EN, Dick IM, Khong A, Boon L, Robinson BW, Nowak AK, Bosco A, Lake RA (2015) Network analysis of immunotherapy-induced regressing tumours identifies novel synergistic drug combinations. Sci Rep 5: 12298

Li M, Zeng T, Liu R, Chen L (2014) Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: study of type 2 diabetes by cross-tissue analysis. Brief Bioinform 15: 229-43

Lin Q, Aihara A, Chung W, Li Y, Chen X, Huang Z, Weng S, Carlson RI, Nadolny C, Wands JR, Dong X (2014) LRH1 promotes pancreatic cancer metastasis. Cancer Lett 350: 15-24

Liu R, Chen P, Aihara K, Chen L (2015) Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. Sci Rep 5: 17501

Liu X, Chang X, Liu R, Yu X, Chen L, Aihara K (2017) Quantifying critical states of complex diseases using single-sample dynamic network biomarkers. PLoS Comput Biol 13: e1005633

Liu XY, Li YI, Pritchard JK (2019) Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell 177: 1022-1034

Lucas B, Grigo K, Erdmann S, Lausen J, Klein-Hitpass L, Ryffel GU (2005) HNF4alpha reduces proliferation of kidney cells and affects genes deregulated in renal cell carcinoma. Oncogene 24: 6418-31

Maeda Y, Hwang-Verslues WW, Wei G, Fukazawa T, Durbin ML, Owen LB, Liu X, Sladek FM (2006) Tumour suppressor p53 down-regulates the expression of the human hepatocyte nuclear factor 4alpha (HNF4alpha) gene. Biochem J 400: 303-13

Maeda Y, Seidel SD, Wei G, Liu X, Sladek FM (2002) Repression of hepatocyte nuclear factor 4alpha tumor suppressor p53: involvement of the ligand-binding domain and histone deacetylase activity. Mol Endocrinol 16: 402-10

Majdandzic A, Braunstein LA, Curme C, Vodenska I, Levy-Carciente S, Stanley HE, Havlin S (2016) Multiple tipping points and optimal repairing in interacting networks. Nat Commun 7: 10850

Maris JM, Hogarty MD, Bagatell R, Cohn SL (2007) Neuroblastoma. Lancet 369: 2106-20

Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, Weiss WA (2016) Neuroblastoma. Nat Rev Dis Primers 2: 16078

Mojtahedi M, Skupin A, Zhou J, Castano IG, Leong-Quong RY, Chang H, Trachana K, Giuliani A, Huang S (2016) Cell Fate Decision as High-Dimensional Critical State Transition. PLoS Biol 14: e2000640

Mondal T, Juvvuna PK, Kirkeby A, Mitra S, Kosalai ST, Traxler L, Hertwig F, Wernig-Zorc S, Miranda C, Deland L, Volland R, Bartenhagen C, Bartsch D, Bandaru S, Engesser A, Subhash S, Martinsson T, Caren H, Akyurek LM, Kurian L et al. (2018) Sense-Antisense lncRNA Pair Encoded by Locus 6p22.3 Determines Neuroblastoma Susceptibility via the USP36-CHD7-SOX9 Regulatory Axis. Cancer Cell 33: 417-434 e7

Mukherjee S, Carignano A, Seelig G, Lee SI (2018) Identifying progressive gene network perturbation from single-cell RNA-seq data. Conf Proc IEEE Eng Med Biol Soc 2018: 5034-5040

Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, Khera AV, Zhou W, Bloom JM, Engreitz JM, Ernst J, O'Connell JR, Ruotsalainen SE, Alver M, Manichaikul A, Johnson WC, Perry JA, Poterba T, Seed C, Surakka IL et al. (2018) Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. Nat Commun 9: 3391

O'Regan SM, Drake JM (2013) Theory of early warning signals of disease emergenceand leading indicators of elimination. Theor Ecol-Neth 6: 333-357

Oberthuer A, Juraeva D, Hero B, Volland R, Sterz C, Schmidt R, Faldum A, Kahlert Y, Engesser A, Asgharzadeh S, Seeger R, Ohira M, Nakagawara A, Scaruffi P, Tonini GP, Janoueix-Lerosey I, Delattre O, Schleiermacher G, Vandesompele J, Speleman F et al. (2015) Revised risk estimation and treatment stratification of low- and intermediate-risk neuroblastoma patients by integrating clinical and molecular prognostic markers. Clin Cancer Res 21: 1904-15

Oberthuer A, Juraeva D, Li L, Kahlert Y, Westermann F, Eils R, Berthold F, Shi L, Wolfinger RD, Fischer M, Brors B (2010) Comparison of performance of one-color and two-color gene-expression analyses in predicting clinical endpoints of neuroblastoma patients. Pharmacogenomics J 10: 258-66

Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, Fraenkel E, Bell GI, Young RA (2004) Control of pancreas and liver gene expression by HNF transcription factors. Science 303: 1378-81

Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, Achuthan P, Guo H, Fortune MD, Stevens H, Walker NM, Ward LD, Kundaje A, Kellis M, Daly MJ, Barrett JC et al. (2015) Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet 47: 381-6

Pavlaki I, Alammari F, Sun B, Clark N, Sirey T, Lee S, Woodcock DJ, Ponting CP, Szele FG, Vance KW (2018) The long non-coding RNA Paupar promotes KAP1-dependent chromatin changes and regulates olfactory bulb neurogenesis. EMBO J 37

Pons P, Latapy M (2005) Computing communities in large networks using random walks. Lect Notes Comput Sc 3733: 284-293

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-40

Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) 2016

Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, van Nes EH, Rietkerk M, Sugihara G (2009) Early-warning signals for critical transitions. Nature 461: 53-9

Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, van de Koppel J, van de Leemput IA, Levin SA, van Nes EH, Pascual M, Vandermeer J (2012) Anticipating critical transitions. Science 338: 344-8

Simons YB, Bullaughey K, Hudson RR, Sella G (2018) A population genetic interpretation of GWAS findings for human quantitative traits. PLoS Biol 16: e2002985

Sladek FM (2012) The yin and yang of proliferation and differentiation: cyclin D1 inhibits differentiation factors ChREBP and HNF4alpha. Cell Cycle 11: 3156-7

Su Z, Fang H, Hong H, Shi L, Zhang W, Zhang W, Zhang Y, Dong Z, Lancashire LJ, Bessarabova M, Yang X, Ning B, Gong B, Meehan J, Xu J, Ge W, Perkins R, Fischer M, Tong W (2014) An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era. Genome Biol 15: 523

Sun Q, Xu W, Ji S, Qin Y, Liu W, Hu Q, Zhang Z, Liu M, Yu X, Xu X (2019) Role of hepatocyte nuclear factor 4 alpha in cell proliferation and gemcitabine resistance in pancreatic adenocarcinoma. Cancer Cell Int 19: 49

Tanaka T, Jiang S, Hotta H, Takano K, Iwanari H, Sumi K, Daigo K, Ohashi R, Sugai M, Ikegame C, Umezu H, Hirayama Y, Midorikawa Y, Hippo Y, Watanabe A, Uchiyama Y, Hasegawa G, Reid P, Aburatani H, Hamakubo T et al. (2006) Dysregulated expression of P1 and P2 promoter-driven hepatocyte nuclear factor-4alpha in the pathogenesis of human cancer. J Pathol 208: 662-72

Teschendorff AE, Liu X, Caren H, Pollard SM, Beck S, Widschwendter M, Chen L (2014) The dynamics of DNA methylation covariation patterns in carcinogenesis. PLoS Comput Biol 10: e1003709

Teschendorff AE, Relton CL (2018) Statistical and integrative system-level analysis of DNA methylation data. Nat Rev Genet 19: 129-147

Teschendorff AE, Widschwendter M (2012) Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. Bioinformatics 28: 1487-94

Tian X, Zhou D, Chen L, Tian Y, Zhong B, Cao Y, Dong Q, Zhou M, Yan J, Wang Y, Qiu Y, Zhang L, Li Z, Wang H, Wang D, Ying G, Zhao Q (2018) Polo-like kinase 4 mediates epithelial-mesenchymal transition in neuroblastoma via PI3K/Akt signaling pathway. Cell Death Dis 9: 54

Torres-Padilla ME, Sladek FM, Weiss MC (2002) Developmentally regulated N-terminal variants of the nuclear receptor hepatocyte nuclear factor 4alpha mediate multiple interactions through coactivator and corepressor-histone deacetylase complexes. J Biol Chem 277: 44677-87

Trefois C, Antony PM, Goncalves J, Skupin A, Balling R (2015) Critical transitions in chronic disease: transferring concepts from ecology to systems medicine. Curr Opin Biotechnol 34: 48-55

van der Harst P, Verweij N (2018) Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ Res 122: 433-443

Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K, Bjorkegren JLM, Im HK, Pasaniuc B, Rivas MA, Kundaje A (2019) Opportunities and challenges for transcriptome-wide association studies. Nat Genet 51: 592-599

Wang J, Cheng H, Li X, Lu W, Wang K, Wen T (2013) Regulation of neural stem cell differentiation by transcription factors HNF4-1 and MAZ-1. Mol Neurobiol 47: 228-40

Wang K, Diskin SJ, Zhang H, Attiyeh EF, Winter C, Hou C, Schnepp RW, Diamond M, Bosse K, Mayes PA, Glessner J, Kim C, Frackelton E, Garris M, Wang Q, Glaberson W, Chiavacci R, Nguyen L, Jagannathan J, Saeki N et al. (2011) Integrative genomics identifies LMO1 as a neuroblastoma oncogene. Nature 469: 216-20

Wang S, Zou Z, Luo X, Mi Y, Chang H, Xing D (2018) LRH1 enhances cell resistance to chemotherapy by transcriptionally activating MDC1 expression and attenuating DNA damage in human breast cancer. Oncogene 37: 3243-3259

Wang ZZ, Cunningham JM, Yang XH (2018) CisPi: a transcriptomic score for disclosing cis-acting disease-associated lincRNAs. Bioinformatics 34: 664-670

Washietl S, Kellis M, Garber M (2014) Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Res 24: 616-28

Wei L, Dai Y, Zhou Y, He Z, Yao J, Zhao L, Guo Q, Yang L (2017) Oroxylin A activates PKM1/HNF4 alpha to induce hepatoma differentiation and block cancer progression. Cell Death Dis 8: e2944

Witt O, Deubzer HE, Lodrini M, Milde T, Oehme I (2009) Targeting histone deacetylases in neuroblastoma. Curr Pharm Des 15: 436-47

Xiang X, Zhao X, Qu H, Li D, Yang D, Pu J, Mei H, Zhao J, Huang K, Zheng L, Tong Q (2015) Hepatocyte nuclear factor 4 alpha promotes the invasion, metastasis and angiogenesis of neuroblastoma cells via targeting matrix metalloproteinase 14. Cancer Lett 359: 187-97

Xiao L, Wang Y, Xu K, Hu H, Xu Z, Wu D, Wang Z, You W, Ng CF, Yu S, Chan FL (2018) Nuclear Receptor LRH-1 Functions to Promote Castration-Resistant Growth of Prostate Cancer via Its Promotion of Intratumoral Androgen Biosynthesis. Cancer Res 78: 2205-2218

Xie J, Cai TT, Maris J, Li H (2011) Optimal False Discovery Rate Control for Dependent Data. Stat Interface 4: 417-430

Yang B, Li M, Tang W, Liu W, Zhang S, Chen L, Xia J (2018) Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. Nat Commun 9: 678

Yang X, Vasudevan P, Parekh V, Penev A, Cunningham JM (2013) Bridging cancer biology with the clinic: relative expression of a GRHL2-mediated gene-set pair predicts breast cancer metastasis. PLoS One 8: e56195

Yang XH, Li M, Wang B, Zhu W, Desgardin A, Onel K, de Jong J, Chen J, Chen L, Cunningham JM (2015) Systematic computation with functional gene-sets among leukemic and hematopoietic stem cells reveals a favorable prognostic signature for acute myeloid leukemia. BMC Bioinformatics 16: 97

Yang XH, Tang F, Shin J, Cunningham JM (2017a) A c-Myc-regulated stem cell-like signature in high-risk neuroblastoma: A systematic discovery (Target neuroblastoma ESC-like signature). Sci Rep 7: 41

Yang XH, Tang F, Shin J, Cunningham JM (2017b) Incorporating genomic, transcriptomic and clinical data: a prognostic and stem cell-like MYC and PRC imbalance in high-risk neuroblastoma. BMC Syst Biol 11: 92

Zeid R, Lawlor MA, Poon E, Reyes JM, Fulciniti M, Lopez MA, Scott TG, Nabet B, Erb MA, Winter GE, Jacobson Z, Polaski DR, Karlin KL, Hirsch RA, Munshi NP, Westbrook TF, Chesler L, Lin CY, Bradner JE (2018) Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. Nat Genet 50: 515-523

Zhang F, Liu X, Zhang A, Jiang Z, Chen L, Zhang X (2019) Genome-wide dynamic network analysis reveals a critical transition state of flower development in Arabidopsis. BMC Plant Biol 19: 11

Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, Wang J, Furlanello C, Devanarayan V, Cheng J, Deng Y, Hero B, Hong H, Jia M, Li L, Lin SM, Nikolsky Y, Oberthuer A, Qing T, Su Z et al. (2015) Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. Genome Biol 16: 133

Zimmerman MW, Liu Y, He S, Durbin AD, Abraham BJ, Easton J, Shao Y, Xu B, Zhu S, Zhang X, Li Z, Weichert-Leahey N, Young RA, Zhang J, Look AT (2018) MYC Drives a Subset of High-Risk Pediatric Neuroblastomas and Is Activated through Mechanisms Including Enhancer Hijacking and Focal Enhancer Amplification. Cancer Discov 8: 320-335

Figure Legends

Figure 1. Tipping-point theory and its adoption to transcriptome

a, Schematic illustration of a ‘tipping point” (red) in a complex system where a tiny perturbation can produce a large transition from one stable state to another (attractor, green). The x-axis is the disease-development trajectory, which is continuous in nature, however we project this by assembling non-time course individual samples into phenotypic-defined states (z-axis). The y-axis is the recovery rate which tells how strong the attractor state is. The dispersion of the dots shows the strength of the attractor in that state. At the transitional point, the system has a decreased recovery rate and is very sensitive to characteristic fluctuations which can be measured by of increased variance and autocorrelation of specific characteristic changes.

b, Comparison of NPS to seven existing alternative approaches. These prior studies exhibited three different CTS identification models: differential variance (orange), DNB model (blue), and ratio of transcript autocorrelation versus sample autocorrelation (purple). More details are given in **Table S1**.

The legend for the symbols used in the table are as follows. 1: Feature selection required a control group; 2: Network partitioning required prior knowledge on gene-gene interaction; 3: Bootstrapped of samples to empirically evaluate the critical state transition; 4: Bootstrapped genes to empirically evaluate biomarkers; #: Used the previous-time point samples to verify the critical transition. *: Used a scaled-down simulation to conduct the mathematical induction of ‘distribution transition;’ **: Analysis

conducted on coding-gene methylation profiles. PAM: Partitioning Around Medoids; PCC: Pearson Correlation Coefficient; HC: Hierarchical clustering.
c, Five-step analytics pipeline of the NPS method to identify critical transition and the corresponding biomarkers in a system. Note that in complex transcriptomic systems, a characteristic change is described by multiple transcripts. More details are given in **Figure EV1**.
d, Clinical and biological relevance analyses of the NPS transcripts identified as neuroblastoma CTS.

Figure 2. Applying NPS on transcriptome profiles of two NB populations
a, Boxplot compares the standard deviation (s.d.) of the preselected transcripts, with the transcript numbers at the top. In both populations, the highest variance presents at the LIR state.
b, A graph view of network modulation determined by the random-walk approach for the preselected transcripts at the LIR state. Background colors represent different clusters. The module size (number of transcripts) is listed in parentheses.
c, Bar plots illustrate the MCI scores calculated for transcript clusters per state. The dominating cluster (biomodule) is in grey with module size in parentheses. In both populations, the MCI score peaked in the LIR state.
d, Line plot compares the observed Ic-score (black) to 100 random results (grey). The observed score was calculated for the LIR-state specific biomodule but at all four states; the random results were calculated in the same way used randomly-chosen transcripts (listed in parentheses) from the 48.5k-transcript background. In both populations, the observed Ic score peaked in the LIR state.
e, Density histogram compares each observed Ic-score to the random results obtained from 1000 runs of bootstrapping the same number of samples.

Each panel is divided into two subpanels: one with the Event-Free patient group and one with Event patient group.

Figure 3. Genomic features and biological relevance of two identified CTSs, both at the LIR state
a, Scaled expression levels of the CTS transcripts respectively the EF and E populations (lncRNAs in red and mRNAs in black). Vertical lines separate the samples into four phenotype-defined states. The yellow triangles pinpoint the tipping-point indicated by the NPS method. The number of transcripts is listed on the top, and the sample size per risk-group is listed at the bottom.
b, Venn diagram showing the shared and different CTS transcripts identified between the EF and E populations. Out of the 65 shared transcripts between the populations, 48

are mRNAs and 17 are lncRNAs. Enrichment was assessed using the Fisher's exact test against its corresponding background and is given on the bottom.
c-d, Network-degree proportion for the mRNA (**c**) and the lncRNA (**d**) components. The proportion of all gene's network-degrees (yellow or red) was compared to the proportion of lncRNA network-degrees (blue or orange). The Wilcoxon rank sum tested P values are displayed on the top.
e, Stacked bar comparing all co-expressed (FDR<0.05) lncRNA-mRNA pairs for each CTS whose edge-to-edge distance fall within a certain nucleotide range of each other. The cross-chromosomal pairs take the largest fraction in both CTSs.
f, The identified CTS transcripts compared to those differentially expressed transcripts between HR and non-HR patients (FC>2, FDR<0.05) in the same dataset. The number of overlap, fisher's exact tested Odds Ratios (OR), and P-values are indicated. ****: P<2e-16. Green color indicates depletion.

Figure 4. Clinical relevance of the CTS detected from the event-free LIR-risk tumors
a, Kaplan-Meier curve of the event-free survival for neuroblastoma patients (n=498, top. The number of patients divided by the RXA-GSP index (I) is listed in parenthesis. Density plot comparing 1000 empirical log-ranked p-values to the observed p-value in a vertical line (bottom). For each run of simulation, 76 and 61 transcripts were randomly chosen from the 48.5k-transcript background. Then RXA-GSP stimulated a sample stratification.
b, Manhattan plot showing the 379 NB-susceptibility SNPs extracted from the GRASP database with a cutoff of P<5e-5. The dashed red line represents the conventional significance threshold of P<1e-8. SNPs are plotted relative to their position on each chromosome (alternating black and gray). The nine CTS-identified loci are denoted in color (green for lncRNA and blue for mRNA).
c. Visualization of Hi-C interactome at chr5 q31.1 in a NB cell (SKNDZ, ENCODE3, hg19, using the Hi-C data Browser at Pennsylvania State University) (top). Bottom, Zooming-in on the region between a suggestive NB-susceptibility LD (consisting rs11950562 at *SLC22A4*) and two NPS lincRNAs (*PPP2CA-AS* and *LINC01843* marked by pink vertical bars). At the transcription start sites of *SLC22A4* as well the two lncRNAs, we see occupancy of active enhancer markers (H3K27ac, H3K4me3), Myc-binding (from ChIP-seq), open chromatin (ATAC-seq signature) but not repressive histone marker (H3K27me3), suggesting the enhancer activity of these loci in NB. Data of NB cells were downloaded from GEO (**Table S5**).

Figure 5. Biological relevance of two LIR-specific CTS transcripts

a, DNA-binding motifs overrepresented at the CTS transcription promoters in the EF population (left) and the E population (right).

b, Venn-diagram compares the significantly overrepresented DNA-binding motifs between two populations. The significance of overlap was evaluated using Fisher's Exact Test.

c, Protein-protein interaction network view of 12 reproducibly overrepresented DNA-binding motifs (red nodes), generated by Ingenuity IPA tool.

d, Visualization of the transcription correlations between five regulators on which the CTS transcripts are dependent. Circle size and color denote the strength of the correlation with larger, more saturated colors signify stronger correlations versus subdued, smaller circles. The filled circle indicates an empirical significance (PCC: Pearson correlation coefficient, *p*-value < 0.05, 100 bootstraps of transcripts).

e, Model of CTS-amplifier. CTS triggers the impact of weak disruption (by essential TF dysregulation or DNA variation) on core gene regulatory networks that drives distinct disease phenotypes. In the exampled looped pathways, HFN4A interactions are in grey and HFN4A-repressed genes are in red (extracted from prior literature).

Tables and their legends

Table 1. Nine loci highlighted by both CTS and neuroblastoma-associated SNPs (PubMed ID: 21124317, P<5e-5 in GWAS of 8343 samples).

Transcript	CTS	GRASP records			lincRNA ci-acting		
transcript coordinate (hg19)		SNP Id	GWAS P (<5e-5) smallest one	InGene	distance to LD	distance to gene	lincRNA adjacent gene / mRNA symbol
chr1:200036596-200060673	1,2	rs3828112	3.80E-05	NR5A2	320k	adjacent	ZNF281, KIF14
chr5:133562096-133563520:+; chr5:133842243-133844920:+	1 2	rs11950562	3.90E-05	LOC553103, SLC22A4	1.9m	antisense 2.7k	CDKL3, PPP2CA LINC01843 (right to U6)
chr20:42988905-43031510:-	1,2	rs6073330	3.90E-05		250k	antisense	HNF4A
chr21:30554342-30560119:-	1	rs2268225	5.60E-06	TIAM1	2.1m	2m	MAP3K7CL, BACH1
chr6:24169843-24383546:-	1,2	rs4712653	1.69E-15	CASC15	31.6k		DCDC2
		rs9295536	5.29E-16	CASC15			
		rs6939340	9.33E-15	CASC15, CASC14			
		rs885769	6.55E-06	CASC15			

		rs196048	5.05E-06	CASC15		
		rs9393227	3.85E-05	CASC15		
		rs12210008	0.000047	NRSN1		
chr16:72088492-72094960:+	2	rs11648088	1.77E-05		1.4m	HP (Haptoglobin)
chr14:100789680-100796715:+	1,2	rs8017950	7.22E-06	LINC00523	327k	SLC25A47
chr6:15255659-15256283:+	1	rs16876210	2.80E-05	JARID2	44.3k	JARID2
chr2:230084754-230086779:-	1	rs4438469	3.55E-05	SPHKAP	1.2m	PID1
		rs6760875	2.73E-05	SPHKAP		

1: m_LIR.EF; 2: m_LIR.E

Aliases for CASC15: LINC00340, Lnc-SOX4-1

Aliases for CASC14: LOC729177

up/dnDE: significantly up- or dn-regulated in 176 HR relative to 322 non-HR patients uisng 17k genes with normalized value> -5 of GSE62564 samples (FC>2, FDR<0.05).

Expanded View (Supplementary Information)

Figure EV1. Network-perturbation signature workflow.
The five essential steps are outlined in the grey boxes on the left margin.

Figure EV2. Clinical information of four phenotypic states in two populations
The box and pie plots compare the clinical features of the four phenotype-defined NB risk-stages in the EF and E patient populations. These groups were assigned according to the Children's Oncology Group (COG) risk group criteria (https://www.cancer.org/cancer/neuroblastoma/detection-diagnosis-staging/risk-groups.html). Sample sizes are listed at the top of each sub-panel.

Figure EV3. Cis-interacting potential at CTS transcript and suggestive variant in a TAD
a, Visualization of the Hi-C interactome at chr6 in a NB cell (hg19) using the Hi-C data Browser (Pennsylvania State University). Vertical line locates suggestive SNP. The strongest Hi-C value in this area indicates an interaction between the DCDC2 and CASC14/15 loci. Zooming-in on the region between DCDC2 and CASC14/15 reveals the occupancy of active enhancer markers (H3K27ac, H3K4me3), Myc-binding (from ChIP-seq), open chromatin (ATAC-seq signature) but not the repressive histone marker (H3K27me3).

b, The identified WES transcript *HNF4A-AS* is marked in pink box which overlaps with ENSEMBL-annotation *HNF4A-AS1*. The suggestive NB SNP (red vertical line) is only 250k nt away from *HNF4A-AS* (pink box). The occupancy of repressive histone marker H3K27me and lack of enhancer marks at both loci suggests their inactivity in NB.

Figure EV4. Independent transcriptomic evaluation

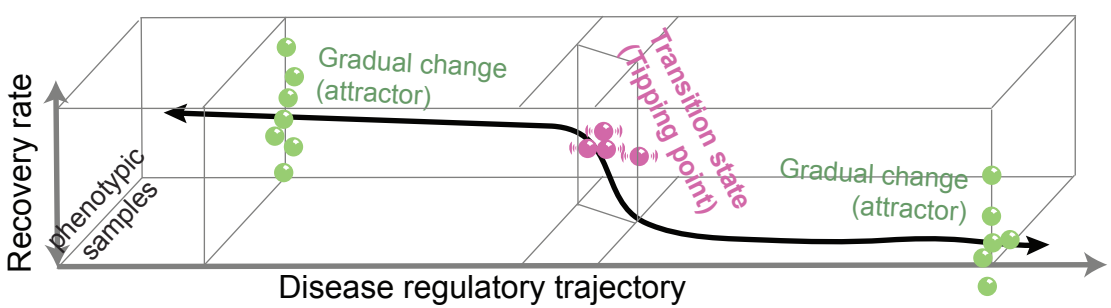
a, Line plot comparing the observed Ic-score (red) to 1000 random results (grey) in each independent dataset of NB (subpanel). Ic-score was calculated for the subset of the 140 coding genes among the identified 191 CTS transcripts, with the mapped number of transcripts given in square bracket. The sample size and state size are given in parenthesis. The observed Ic-score was significantly higher than its simulated scores in only one dataset. This could indicate the existence of multiple sets of CTS transcripts; another reason could be the subset of our identified CTS losses its power to indicate the tipping point in a system.

b, Visualization of the correlation among expressed transcripts [numbers in the square bracket] in three datasets, set to display only significant correlations (Pearson correlation p-value < 0.05). Circle size and color denote the strength of the correlation with larger, more saturated colors signify stronger correlations versus subdued, smaller circles. A question mark denotes insignificant PCC. When multiple probes measuring the same gene, each probe was independently calculated and shown.

Appendix

Supplemental Materials include Abbreviation, Supplementary Methods, four Supplementary Figures, and five Supplementary Tables and can be found with this article online.

a



b

Features \ methods	NPS	Teschendorff 2012 (Variance)	Chen 2012 (DNB)	Teschendorff 2014 (DNB)	Richard 2016 (DNB)	Mojtahedi 2016 (DNB)	Yang 2018 (DNB)	Zhang 2019 (DNB)
Applying to phenotypic series (cross-sectional omics data)	✓	✓	✓	*✓	x	x	x	x
Control-free fluctuated features	✓	1	1	✓	1	NA	1	#✓
Parameter-free network partition	✓	NA	HC (k<30)	PAM (k)	NA	NA	top 1% |PCC|	2
Identifying biomodule	✓	x	✓	✓	✓	x	✓	✓
Biological function relevance	✓	✓	✓	✓	✓	x	✓	✓
Phenotypic (clinical) relevance	✓	✓	x	✓	x	x	✓	✓
Cohort-independence	✓	✓	x	✓	x	✓	x	x
Incl. lncRNA (mRNA-lncRNA pair)	✓	x	x	x**	x	x	x	x
Determining "tipping-point"	✓	✓	✓	✓	✓	✓	✓	✓
Robustness	3,4✓	✓	✓	4✓	x	3,4✓	x	4✓

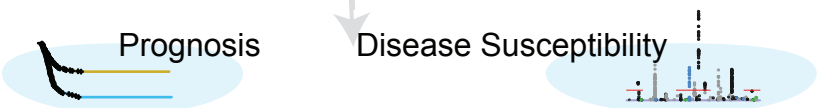
c

NPS

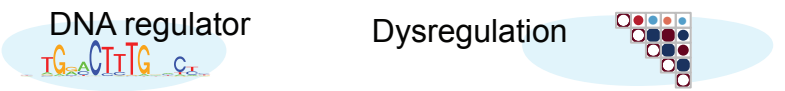
1) Gene & lncRNA expression (phenotype-defined states)
2) Pre-selecting top-flucturated transcripts (w/wo healthy controls)
3) Partition transcritp-interaction network into clusters (modules)
4) Identifying biomarkers (dominating module for each state)
5) Tipping-point determination (Empirically local and global maximum)

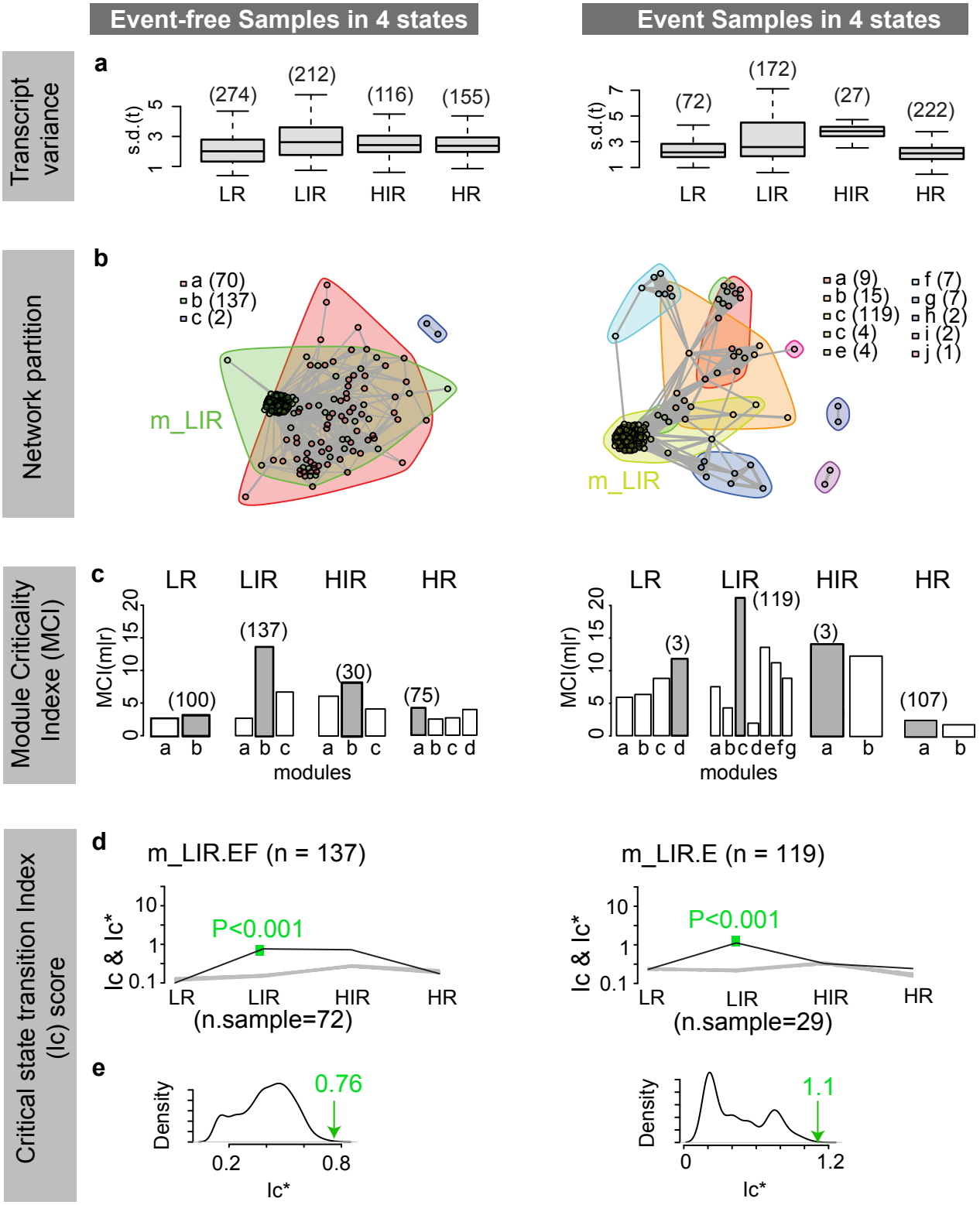
d

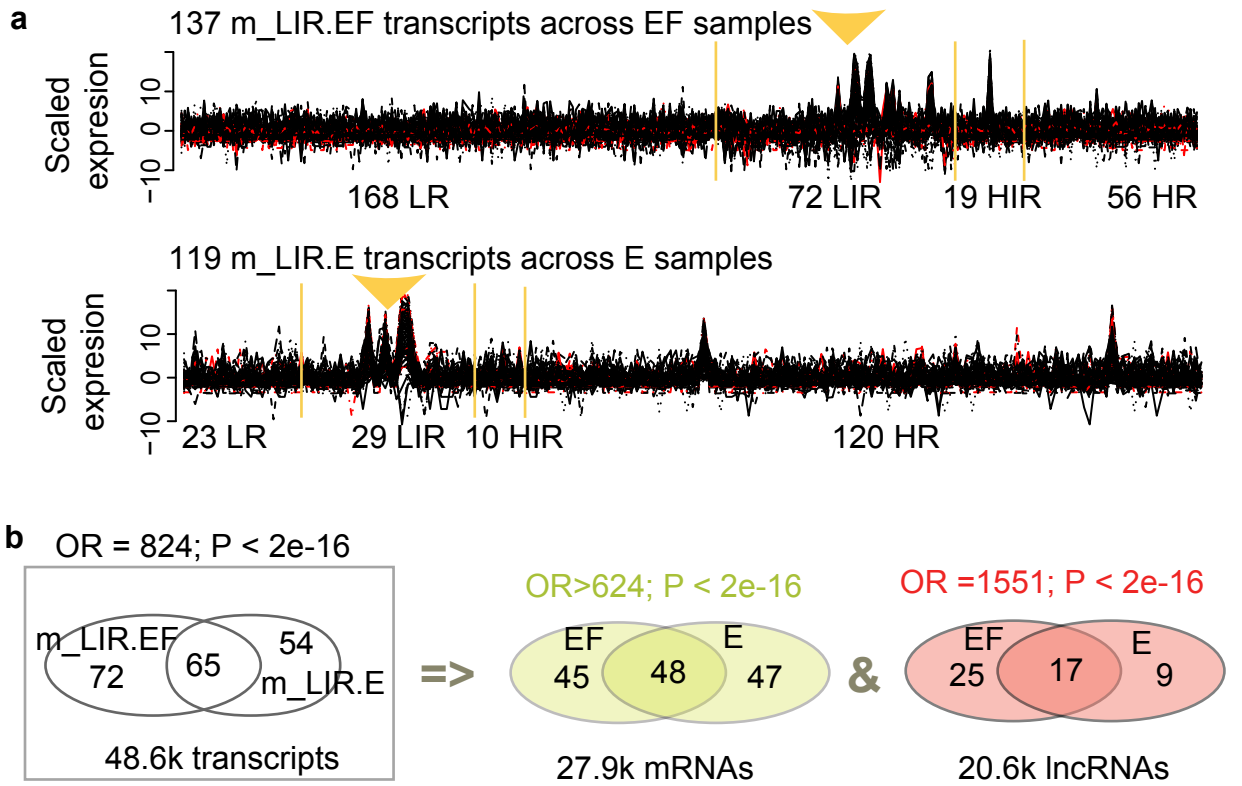
Clinical relevance

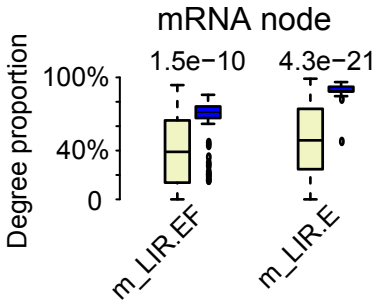


Biological relevance

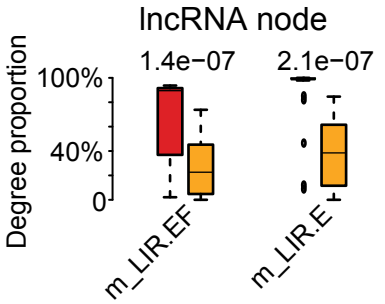



Event-free Samples in 4 states
Event Samples in 4 states
Transcript variance
a
s.d.(t)
(274)
(212)
(116)
(155)
LR
LIR
HIR
HR
(72)
(172)
(27)
(222)
Network partition
b
a (70)
b (137)
c (2)
m_LIR
a (9)
b (15)
c (119)
c (4)
e (4)
f (7)
g (7)
h (2)
i (2)
j (1)
Module Criticality Indexe (MCI)
c
MCI(m|r)
(100)
(137)
(30)
(75)
modules
(3)
(119)
(107)
Critical state transition Index (Ic) score
d
m_LIR.EF (n = 137)
Ic & Ic*
P<0.001
(n.sample=72)
m_LIR.E (n = 119)
(n.sample=29)
e
Density
0.76
Ic*
1.1

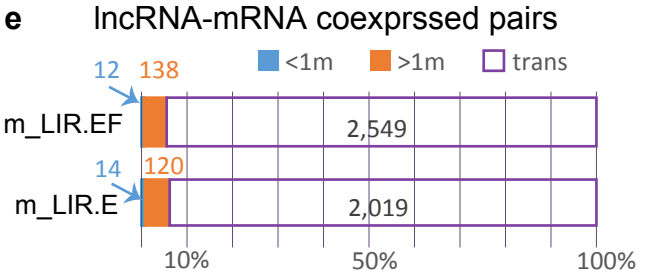




#(mRNA-mRNA)/#(mRNAs)
#(mRNA-lncRNA)/#(lncRNAs)

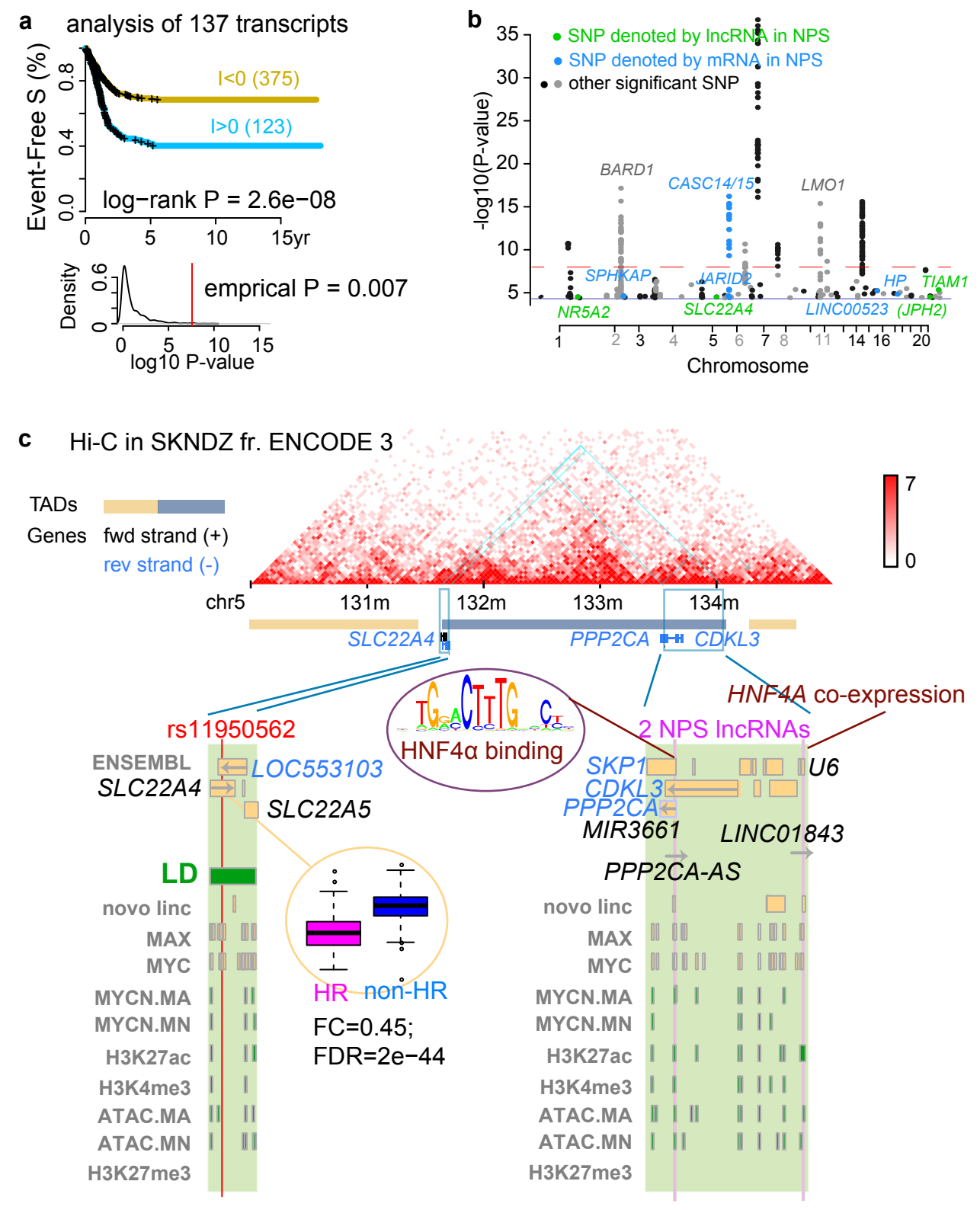


#(lncRNA-mRNA)/#(mRNAs)
#(lncRNA-lncRNA)/#(lncRNAs)



f

	all transcript (48,539)	m_LIR.EF (137)	m_LIR.E (119)
HR.dn (5,043)		9 NS	6 OR=0.05 ****
HR.up (3,490)		1 OR<0.1 ****	2 OR<0.1 ****


a
analysis of 137 transcripts
Event-Free S (%)
0.0
0.4
0.8
I<0 (375)
I>0 (123)
log−rank P = 2.6e−08
0
5
10
15yr
Density
0
0.6
emprical P = 0.007
0
5
10
15
log10 P-value
b
SNP denoted by lncRNA in NPS
SNP denoted by mRNA in NPS
other significant SNP
-log10(P-value)
35
30
25
20
15
10
5
BARD1
CASC14/15
LMO1
SPHKAP
JARID2
HP
TIAM1
NR5A2
SLC22A4
LINC00523
(JPH2)
1
2
3
4
5
6
7
8
11
14
16
20
Chromosome
c
Hi-C in SKNDZ fr. ENCODE 3
TADs
Genes
fwd strand (+)
rev strand (-)
7
0
chr5
131m
132m
133m
134m
SLC22A4
PPP2CA
CDKL3
HNF4α binding
HNF4A co-expression
rs11950562
2 NPS lncRNAs
ENSEMBL
SLC22A4
LOC553103
SLC22A5
SKP1
CDKL3
PPP2CA
U6
MIR3661
LINC01843
PPP2CA-AS
LD
novo linc
MAX
MYC
MYCN.MA
MYCN.MN
H3K27ac
H3K4me3
ATAC.MA
ATAC.MN
H3K27me3
HR
non-HR
FC=0.45;
FDR=2e−44

a

Top 15 of 23 enriched motifs

Target	PWM	Score	P-value	top %
OLIG1		1.3	2.2e−07	19
NFE2L1		2.6	2.7e−06	22
HNF4A		4.8	4.0e−06	16
KIF22		1.2	7.4e−06	18
HNF4G		4.1	9.8e−06	18
EHF		2.8	1.0e−05	26
MYF6		1.4	1.5e−05	16
PICK1		1.2	3.6e−05	23
FIP1L1		1.1	5.4e−05	19
GPAM		1.1	7.2e−05	21
HNF1B		7.0	0.00013	15
NR4A1		1.0	0.00013	16
SNRPB2		1.5	0.00019	23
MAX		1.1	0.00022	17
CNOT6		1.5	0.00024	19

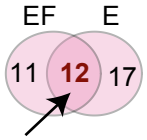
Top 15 of 29 enriched motifs

Target	PWM	Score	P-value	top %
YY1		1.1	6.7e−10	20
HNF4A		8.5	9.1e−10	24
HNF4G		6.6	2.5e−09	23
NR2F1		5.6	8.7e−09	21
ATF6		1.4	1.6e−06	24
DUS3L		1.3	2.4e−06	25
RXRA		2.7	1.3e−05	19
PTCD1		1.2	1.7e−05	17
NR2F6		5.5	2.7e−05	17
EHF		2.8	2.7e−05	24
NFIX		1.3	4.7e−05	20
GPAM		1.1	4.9e−05	21
NELFB		1.3	5.9e−05	19
NR2C2		4.8	6.4e−05	18
PPP5C		1.1	6.5e−05	20

b

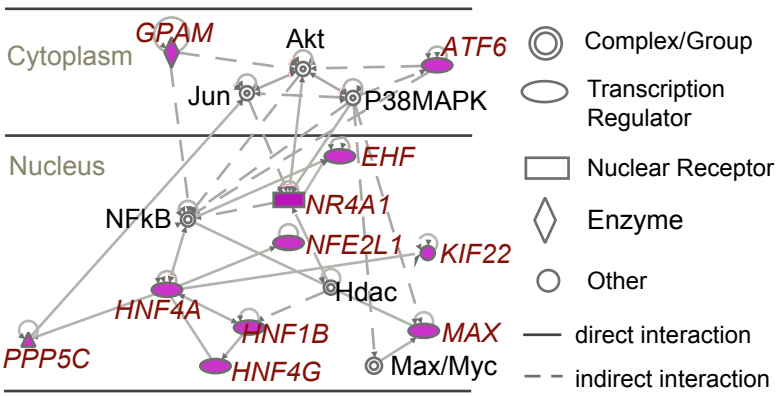
fr. 866 unique DNA_binding motifs

OR=52; P=1.7e-13



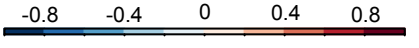
ATF6, EHF, GPAM, MAX, NFE2L1, NR4A1, HNF4A, HNF4G, NFIX, PPP5C, KIF22, HNF1B

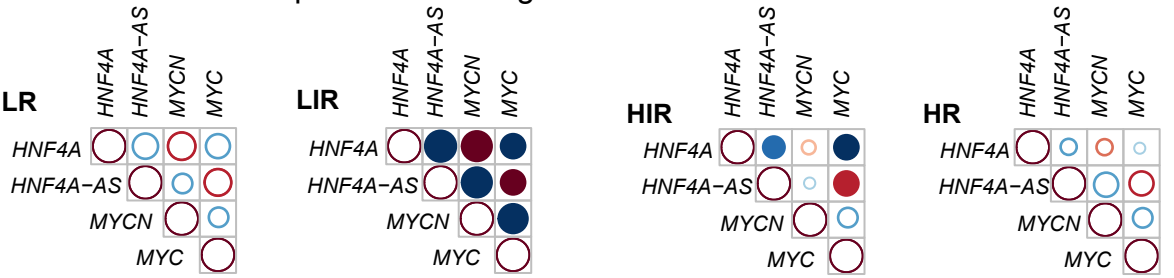
c



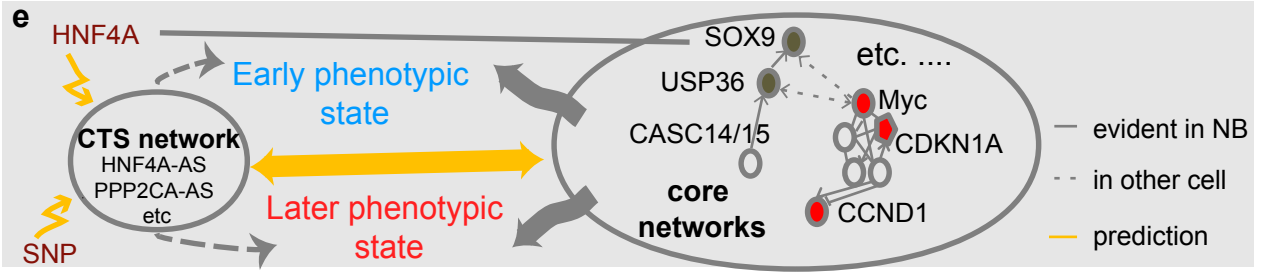
d

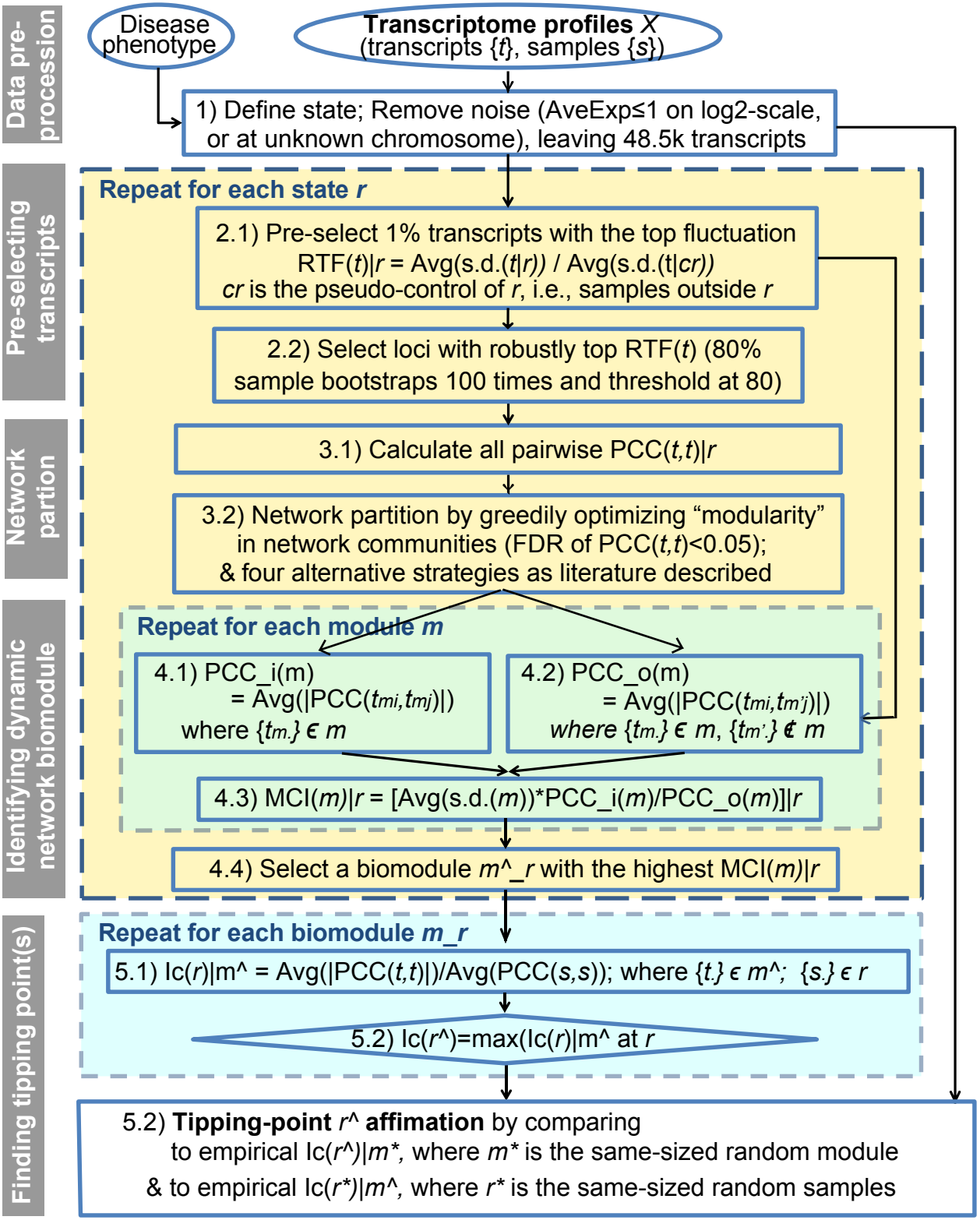
PCC of the NPS's dependence on regulator

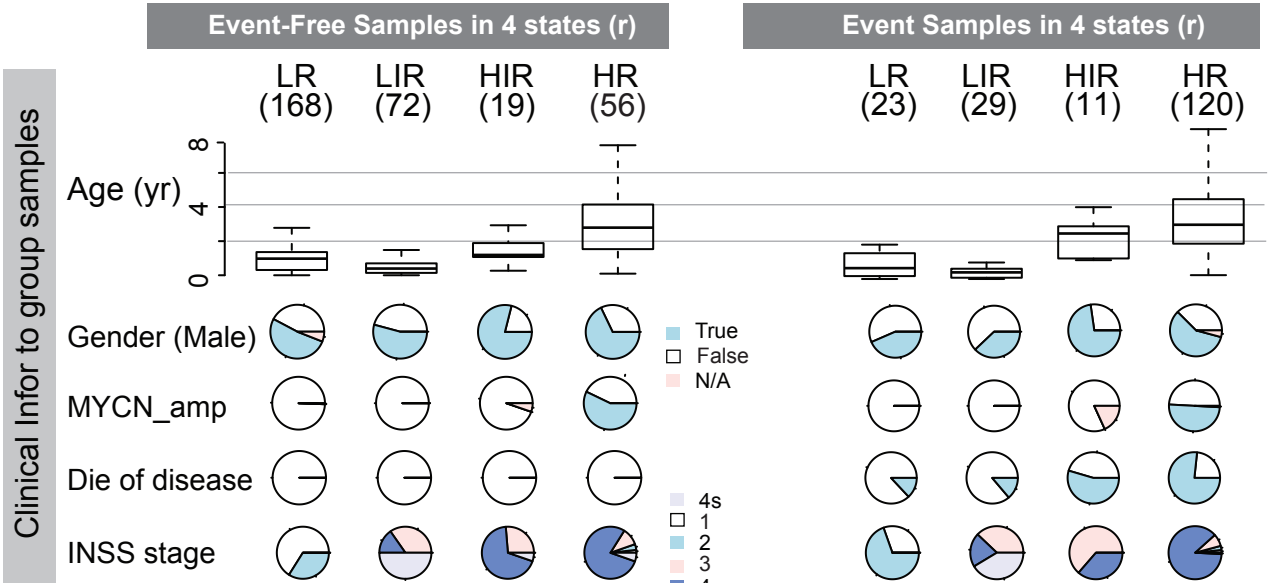


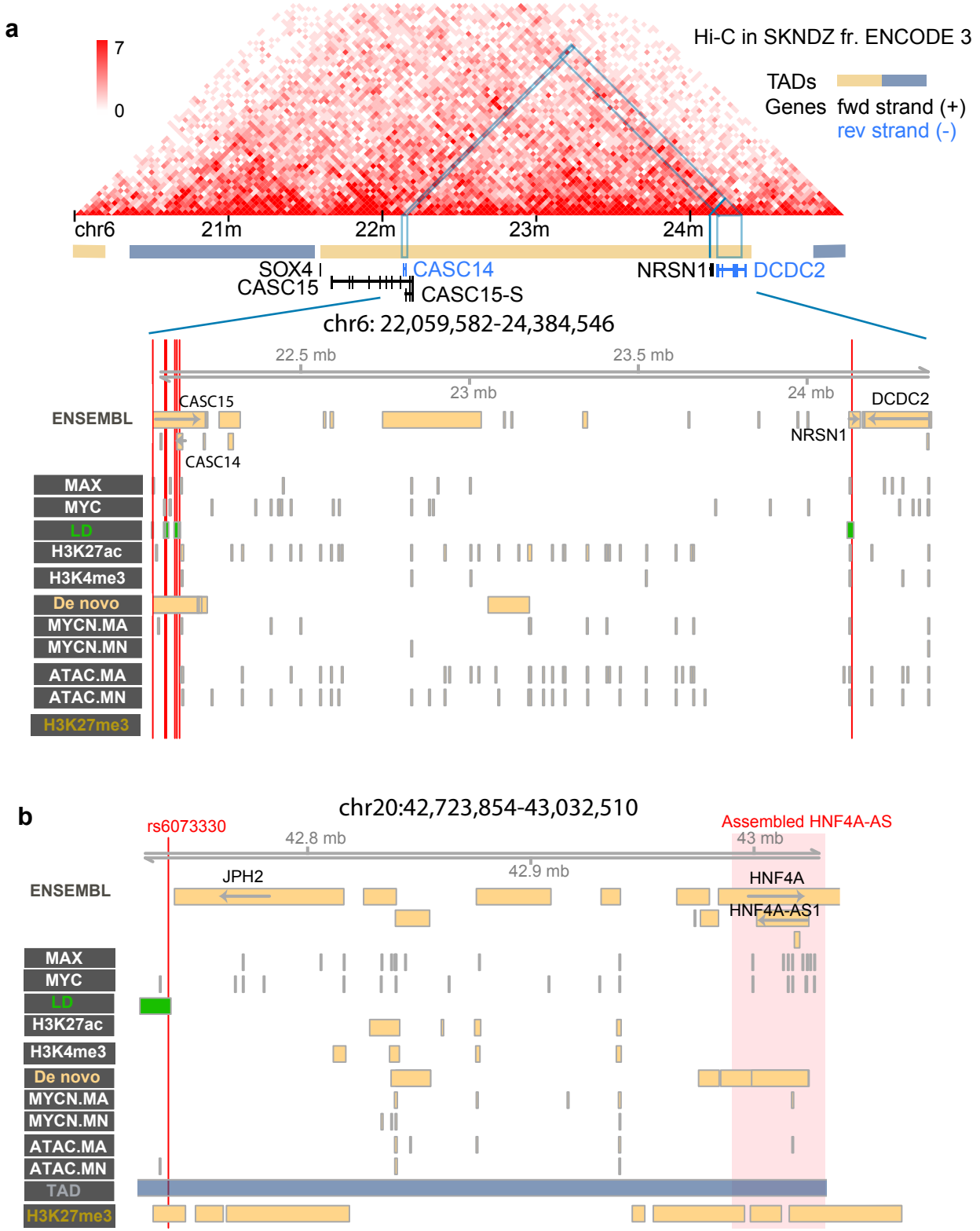


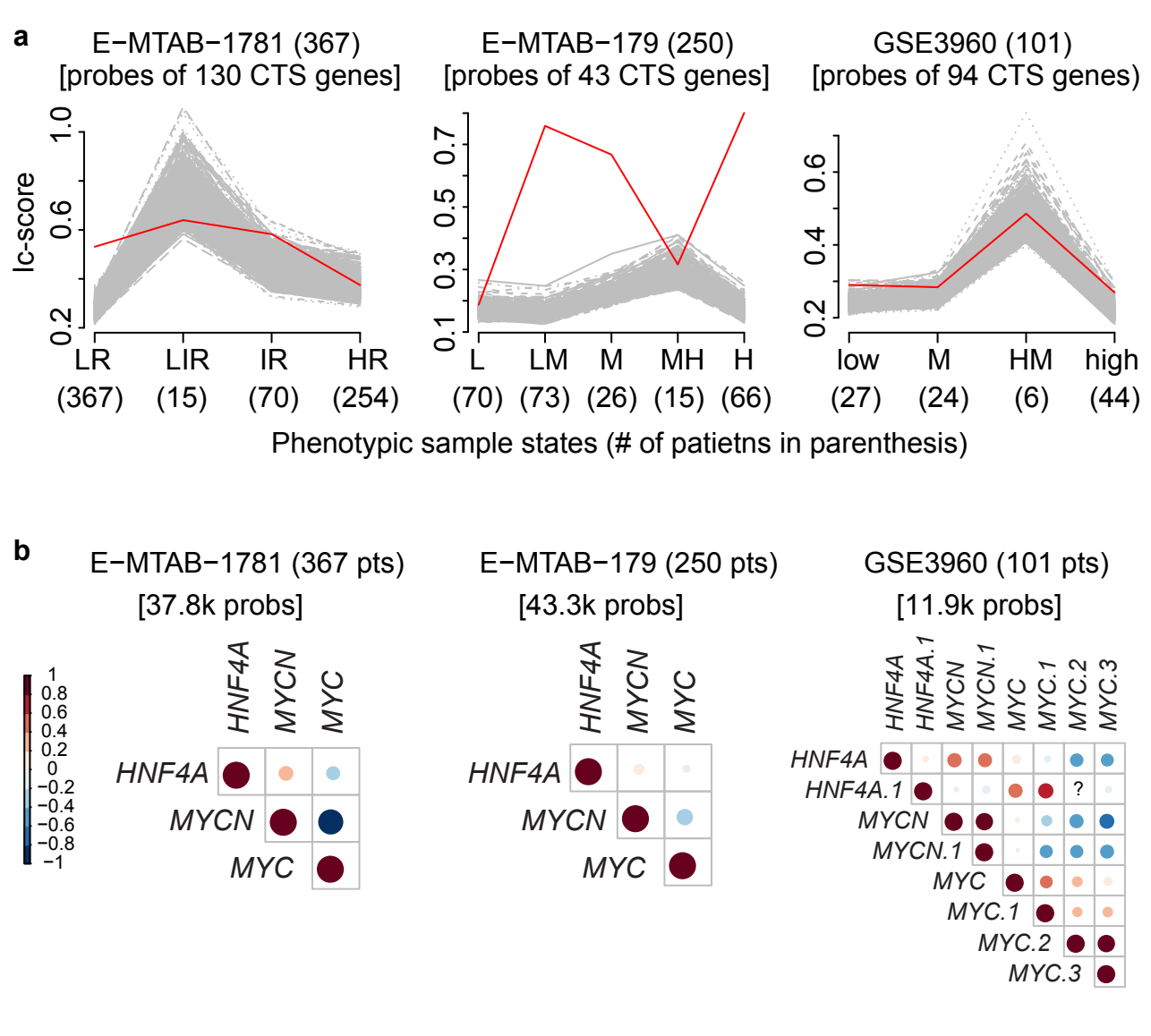
e




Data pre-procession
Disease phenotype
Transcriptome profiles X (transcripts {t}, samples {s})
1) Define state; Remove noise (AveExp≤1 on log2-scale, or at unknown chromosome), leaving 48.5k transcripts
Pre-selecting transcripts
Repeat for each state r
2.1) Pre-select 1% transcripts with the top fluctuation RTF(t)|r = Avg(s.d.(t|r)) / Avg(s.d.(t|cr))
cr is the pseudo-control of r, i.e., samples outside r
2.2) Select loci with robustly top RTF(t) (80% sample bootstraps 100 times and threshold at 80)
Network partion
3.1) Calculate all pairwise PCC(t,t)|r
3.2) Network partition by greedily optimizing "modularity" in network communities (FDR of PCC(t,t)<0.05); & four alternative strategies as literature described
Identifying dynamic network biomodule
Repeat for each module m
4.1) PCC_i(m) = Avg(|PCC(tmi,tmj)|) where {tm.} ϵ m
4.2) PCC_o(m) = Avg(|PCC(tmi,tm'j)|) where {tm.} ϵ m, {tm'.} ∉ m
4.3) MCI(m)|r = [Avg(s.d.(m))*PCC_i(m)/PCC_o(m)]|r
4.4) Select a biomodule m^_r with the highest MCI(m)|r
Finding tipping point(s)
Repeat for each biomodule m_r
5.1) Ic(r)|m^ = Avg(|PCC(t,t)|)/Avg(PCC(s,s)); where {t.} ϵ m^; {s.} ϵ r
5.2) Ic(r^)=max(Ic(r)|m^ at r
5.2) Tipping-point r^ affimation by comparing to empirical Ic(r^)|m*, where m* is the same-sized random module & to empirical Ic(r*)|m^, where r* is the same-sized random samples


Event-Free Samples in 4 states (r)
Event Samples in 4 states (r)
LR (168)
LIR (72)
HIR (19)
HR (56)
LR (23)
LIR (29)
HIR (11)
HR (120)
Clinical Infor to group samples
Age (yr)
0
4
8
Gender (Male)
MYCN_amp
Die of disease
INSS stage
True
False
N/A
4s
1
2
3
4


a
7
0
Hi-C in SKNDZ fr. ENCODE 3
TADs
Genes fwd strand (+)
rev strand (-)
chr6
21m
22m
23m
24m
SOX4
CASC15
CASC14
CASC15-S
NRSN1
DCDC2
chr6: 22,059,582-24,384,546
22.5 mb
23.5 mb
23 mb
24 mb
ENSEMBL
CASC15
CASC14
NRSN1
DCDC2
MAX
MYC
LD
H3K27ac
H3K4me3
De novo
MYCN.MA
MYCN.MN
ATAC.MA
ATAC.MN
H3K27me3
b
chr20:42,723,854-43,032,510
rs6073330
Assembled HNF4A-AS
42.8 mb
43 mb
42.9 mb
ENSEMBL
JPH2
HNF4A
HNF4A-AS1
MAX
MYC
LD
H3K27ac
H3K4me3
De novo
MYCN.MA
MYCN.MN
ATAC.MA
ATAC.MN
TAD
H3K27me3


a
E−MTAB−1781 (367)
[probes of 130 CTS genes]
E−MTAB−179 (250)
[probes of 43 CTS genes]
GSE3960 (101)
[probes of 94 CTS genes)
lc-score
LR (367)
LIR (15)
IR (70)
HR (254)
L (70)
LM (73)
M (26)
MH (15)
H (66)
low (27)
M (24)
HM (6)
high (44)
Phenotypic sample states (# of patietns in parenthesis)
b
E−MTAB−1781 (367 pts)
[37.8k probs]
E−MTAB−179 (250 pts)
[43.3k probs]
GSE3960 (101 pts)
[11.9k probs]
HNF4A
MYCN
MYC
HNF4A.1
MYCN.1
MYC.1
MYC.2
MYC.3
?