Reflexive Autocatalytic Sets as a Framework for Identifying Critical Disruption Targets in Cancer Transcriptional Regulatory Networks

A Network-Theoretic Approach to Oncogenic Circuit Collapse

Armando Vieira
Tartu University
[email]

May 7, 2026

Abstract

Cancer progression is increasingly understood as a systems-level phenomenon driven by self-sustaining transcriptional regulatory circuits rather than isolated gene mutations. Standard network analyses identify hub genes by connectivity or expression level, but fail to capture the *autocatalytic closure* that makes oncogenic circuits resistant to perturbation. Here we apply the theory of **Reflexive Autocatalytic and Food-generated sets (RAF)** to a genome-scale protein–protein interaction network of 138,479 regulatory edges spanning 994 transcription factors. We define three novel disruption metrics—*RAF Keystone Score*, *Bridge Edge Criticality*, and *Gatekeeper Feed Strength*—and introduce a composite *Disruption Score* that integrates structural and catalytic information. Applied to a 30% sample of the GRN, our method identifies a self-sustaining RAF core of 459 transcription factors, 294 critical bridge edges, and three high-priority disruption targets: **ESR1**, **FOXO1**, **RUNX2**, and **TP53**, whose removal collapses up to 1.09% of the RAF set per node. We argue that RAF-based disruption analysis provides a principled, biologically grounded framework for identifying cancer-specific therapeutic targets that are structurally irreplaceable rather than merely highly expressed.

Contents

1 Introduction **3**
1.1 Limitations of Current Approaches 3
1.2 Autocatalytic Sets in Biology 3
1.3 Our Contribution 3

2 Background and Related Work **4**
2.1 Transcriptional Regulatory Networks 5
2.2 RAF Theory 5
2.3 Related Network Medicine Approaches 6

3 Methods **6**
3.1 Data and Preprocessing 6
3.2 Catalyticity Score 6
3.3 RAF Set Detection 7
3.4 Novel Disruption Metrics 7
3.4.1 RAF Keystone Score 7
3.4.2 Bridge Edge Criticality 8
3.4.3 Gatekeeper Feed Strength 8
3.4.4 Composite Disruption Score 8

4 Experimental Setup **9**
4.1 Network Parameters 9
4.2 Computational Implementation 9

5 Results **9**
5.1 Keystone Analysis 10
5.2 Bridge Edge Analysis 10
5.3 Food Set Gatekeeper Analysis 11
5.4 High-Priority Disruption Targets 11

6 Discussion **11**
6.1 The RAF Collapse Hypothesis 12
6.2 Self-Loop Bridge Edges as Vulnerability Points 12
6.3 Limitations 12

7 Conclusion **13**

1 Introduction

Cancer is not simply a disease of overexpressed genes or broken proteins—it is a disease of *network topology*. Oncogenic transformation rewires the transcriptional regulatory network (TRN) of a cell, creating self-sustaining circuits that drive uncontrolled proliferation, immune evasion, and metastasis [Hanahan & Weinberg, 2011]. Understanding which genes maintain these circuits, and which edges are structurally irreplaceable, is a central challenge in translational oncology.

1.1 Limitations of Current Approaches

Existing computational approaches to cancer target identification fall into three broad categories:

1. **Differential expression analysis** identifies genes upregulated in cancer versus normal tissue [Love et al., 2014], but provides no network context and frequently highlights genes that are downstream consequences rather than upstream drivers.

2. **Hub gene analysis** targets highly connected nodes in protein–protein interaction (PPI) networks [Barabási & Oltvai, 2004]. While hubs are often essential, they are typically essential in *all* cell types, making them poor therapeutic targets due to systemic toxicity.

3. **Centrality-based methods** (betweenness, eigenvector, PageRank centrality) identify structural bottlenecks [Newman, 2010] but treat the network as static and ignore the *self-sustaining dynamics* of oncogenic circuits.

None of these approaches captures the key property of cancer networks: their ability to *catalyze their own maintenance*. A cancer circuit that can regenerate its own regulatory signals is fundamentally different from one that depends on external inputs, and requires a fundamentally different analytical framework.

1.2 Autocatalytic Sets in Biology

The concept of autocatalytic sets originates in the origin-of-life literature. Kauffman [1986] proposed that life emerged when a set of molecules became large enough that each molecule's formation was catalyzed by at least one other molecule in the set—a *collectively autocatalytic set*. Hordijk & Steel [2004] formalized this as the RAF (Reflexive Autocatalytic and Food-generated) framework, providing efficient algorithms for RAF detection in chemical reaction networks.

RAF theory has since been applied to metabolic networks [Hordijk et al., 2012], ecosystem dynamics [Vasas et al., 2012], and cognitive systems [Steel, 2000]. Its application to transcriptional regulatory networks, and specifically to cancer biology, has not been systematically explored.

1.3 Our Contribution

We make the following contributions:

- We apply RAF theory to a genome-scale human TRN, identifying a self-sustaining transcriptional core of 459 genes.

- We introduce three novel disruption metrics: *Keystone Score*, *Bridge Edge Criticality*, and *Gatekeeper Feed Strength*.
- We define a composite *Disruption Score* and validate it against known cancer biology.
- We propose the *RAF Collapse Hypothesis*: that cancer cell viability depends on the integrity of a minimal autocatalytic transcriptional core, and that targeted disruption of RAF keystones can selectively collapse this core.

2 Background and Related Work

Cancer initiation is often described in terms of a small number of driver mutations. In many solid tumours, including breast cancer, transformation may begin with only a limited set of major driver alterations, such as mutations or amplifications affecting *TP53*, *MYC*, *PIK3CA*, *ESR1*, or related oncogenic regulators. However, these driver events should be understood as ignition points rather than as the complete functional architecture of the tumour. Once transformation is established, the cellular regulatory network is broadly rewired, recruiting hundreds to thousands of genes into a disease-maintaining transcriptional programme.

This distinction is central to the network-theoretic view of cancer. Driver genes act as high-level control points, but the malignant phenotype is executed by a much larger set of downstream regulatory genes, transcription factors, signalling intermediates, and differentially expressed genes. In breast cancer, transcriptomic studies commonly identify hundreds to more than one thousand differentially expressed genes between tumour and normal tissue, while genome-scale regulatory-network reconstructions may connect thousands of genes into subtype-specific regulatory programmes Mares-Quiñones et al. [2024]. Thus, the biologically relevant object is not merely the set of mutated genes, but the self-reinforcing regulatory system that emerges after those mutations have altered the cellular state.

Several well-known cancer genes illustrate this transition from driver mutation to network-wide regulation. *MYC*, for example, functions less like a conventional single-pathway switch and more like a global transcriptional amplifier, influencing a large fraction of the human transcriptome and increasing the output of growth, metabolism, ribosome biogenesis, and cell-cycle programmes Dang [2012]. Similarly, wild-type *TP53* regulates broad programmes involved in DNA repair, apoptosis, senescence, and cell-cycle arrest; when *TP53* is lost or mutated, this regulatory programme is disabled or redirected, allowing damaged cells to survive and proliferate. Other transcription factors, including *RUNX2* and *SOX2*, can act as pioneer or lineage-reprogramming factors, opening chromatin regions and enabling access to genes associated with stemness, invasion, and metastatic competence Smith et al. [2010].

Within these large regulatory systems, some genes function as network hubs even when they are not themselves recurrently mutated. These genes may stabilize oncogenic states by coordinating immune evasion, extracellular-matrix remodelling, proliferation, or metastatic behaviour. For example, transcription factors such as *ETS1* regulate matrix metalloproteinases and invasion-associated programmes, while interferon-related genes such as *IFIT3* have been identified as highly connected regulatory nodes in basal-like breast cancer networks **??**. Such genes are important because their contribution is structural and regulatory rather than purely mutational.

This motivates a multi-level view of gene involvement in cancer. A small number of driver mutations may initiate the disease, but a much larger dysregulated transcriptome maintains the malignant state, and an even larger regulatory network defines the space of possible feedback loops, compensatory pathways, and therapeutic escape routes. Table 1 summarizes this hierarchy.

Table 1: Approximate hierarchy of gene involvement in cancer regulatory networks.

Level	**Approximate number of genes**	**Functional role**
Driver mutations	3–10	Initiating oncogenic lesions or "spa
Major regulatory hubs	50–100	Control points for cancer traits su
Dysregulated transcriptome	600–1,000+	Differentially expressed genes exec
Total regulatory network	19,000+	Global map of regulatory interacti

The RAF framework is particularly suited to this multi-scale structure. Standard driver or hub-based analyses ask which genes are mutated, overexpressed, or highly connected. By contrast, RAF analysis asks which subset of regulatory interactions is capable of maintaining itself through autocatalytic closure. This distinction is important because the most therapeutically relevant target may not be the most mutated or most connected gene, but the gene whose removal causes the greatest collapse of the self-sustaining oncogenic circuit.

2.1 Transcriptional Regulatory Networks

A transcriptional regulatory network (TRN) is a directed graph $G = (V, E, w)$ where V is the set of transcription factors (TFs), $E \subseteq V \times V$ is the set of regulatory interactions, and $w : E \to \mathbb{R}^+$ assigns interaction confidence weights. An edge $(u, v, w_{uv}) \in E$ indicates that TF u regulates the expression of gene v with confidence w_{uv}.

Human TRNs are among the most complex biological networks, with estimates of $\sim$1,400 TFs regulating $\sim$20,000 genes through $\sim 10^6$ interactions [Lambert et al., 2018]. The ENCODE project [ENCODE Project Consortium, 2012] and subsequent efforts have produced genome-scale TRN maps, enabling computational analysis at unprecedented resolution.

2.2 RAF Theory

Definition 1 (Chemical Reaction System)**.** *A chemical reaction system (CRS) is a tuple* $\mathcal{Q} = (X, \mathcal{R}, C)$ *where* X *is a set of molecule types,* $\mathcal{R}$ *is a set of reactions, and* $C : \mathcal{R} \times X \to \{0, 1\}$ *is a catalysis function indicating which molecules catalyze which reactions.*

Definition 2 (RAF Set [Hordijk & Steel, 2004])**.** *Given a CRS* $\mathcal{Q}$ *and a food set* $F \subseteq X$*, a subset* $\mathcal{R}' \subseteq \mathcal{R}$ *is a* Reflexive Autocatalytic and Food-generated *(RAF) set if:*

1. ***Reflexivity****: Every reaction* $r \in \mathcal{R}'$ *is catalyzed by at least one molecule produced by* $\mathcal{R}' \cup F$*.*
2. ***Food-generation****: Every reactant of every* $r \in \mathcal{R}'$ *can be produced from* F *using reactions in* $\mathcal{R}'$*.*

The maximal RAF set (maxRAF) is unique and can be computed in polynomial time [Hordijk & Steel, 2004].

2.3 Related Network Medicine Approaches

Barabási et al. [2011] introduced the concept of the *disease module* — a connected subgraph of the interactome enriched for disease genes. Menche et al. [2015] showed that disease modules of different diseases overlap in predictable ways. Ghiassian et al. [2015] extended this to identify drug targets within disease modules.

Our approach differs fundamentally: rather than identifying *where* disease genes cluster, we identify *which genes maintain the self-sustaining dynamics* of the disease circuit. This distinction is critical for therapeutic targeting.

3 Methods

3.1 Data and Preprocessing

We use the Gencode p5 protein–protein interaction network comprising 138,479 regulatory edges across 994 unique transcription factors. Edge weights $w_{uv} \in [0, 1]$ represent interaction confidence scores derived from co-expression, ChIP-seq, and literature evidence.

Sampling. To manage computational complexity, we sample a fraction $\rho \in \{0.10, 0.30\}$ of edges uniformly at random:

$$E_\rho = \text{Sample}(E,\ \rho,\ \text{seed} = 42) \tag{1}$$

Thresholding. We retain only high-confidence edges:

$$E^* = \{(u, v, w) \in E_\rho \mid w \geq \tau\} \tag{2}$$

where $\tau = 0.90$ is the weight threshold. This yields a sparse, high-confidence subgraph $G^* = (V, E^*)$.

3.2 Catalyticity Score

We define the *catalyticity* of a node v as a weighted combination of its outgoing and incoming edge strengths:

$$\kappa(v) = \alpha \cdot S^{\text{out}}(v) + (1 - \alpha) \cdot S^{\text{in}}(v) \tag{3}$$

where $\alpha = 0.7$ and the out- and in-strengths are:

$$S^{\text{out}}(v) = \sum_{(v,u)\in E} w_{vu} \tag{4}$$

$$S^{\text{in}}(v) = \sum_{(u,v)\in E} w_{uv} \tag{5}$$

The weighting $\alpha = 0.7$ reflects the asymmetry of regulatory networks: a TF's ability to *catalyze* other genes (out-strength) is more relevant to RAF membership than its being regulated (in-strength).

3.3 RAF Set Detection

We adapt the RAF detection algorithm of Hordijk & Steel [2004] to directed regulatory networks:

Definition 3 (Regulatory Food Set)**.** *The food set $F \subseteq V$ consists of nodes with zero in-degree in G^*:*

$$F = \{v \in V \mid \deg^-(v) = 0 \text{ in } G^*\} \tag{6}$$

These represent TFs that are constitutively active or driven by external signals not captured in the network.

Definition 4 (Regulatory RAF Set)**.** *A set $\mathcal{R} \subseteq V \setminus F$ is a RAF set if every node $v \in \mathcal{R}$ has at least one incoming edge from $\mathcal{R} \cup F$ in G^*.*

Algorithm 1 MaxRAF Detection for Regulatory Networks

Require: Graph $G^* = (V, E^*)$, food set F
Ensure: Maximal RAF set $\mathcal{R}^*$
$\mathcal{R} \leftarrow V \setminus F$
repeat
  $\mathcal{R}_{\text{prev}} \leftarrow \mathcal{R}$
  for each $v \in \mathcal{R}$ **do**
    $\text{catalysts}(v) \leftarrow \{u \mid (u, v) \in E^*,\ u \in \mathcal{R} \cup F\}$
    if $\text{catalysts}(v) = \emptyset$ **then**
      $\mathcal{R} \leftarrow \mathcal{R} \setminus \{v\}$
    end if
  end for
until $\mathcal{R} = \mathcal{R}_{\text{prev}}$
return $\mathcal{R}$

Algorithm 1 converges in $O(|V|^2)$ time in the worst case.

3.4 Novel Disruption Metrics

3.4.1 RAF Keystone Score

Definition 5 (RAF Keystone Score)**.** *For a node $v \in \mathcal{R}^*$, the keystone score measures the collapse of the RAF set upon removal of v:*

$$K(v) = |\mathcal{R}^*| - |\mathcal{R}^*_{-v}| - 1 \tag{7}$$

where $\mathcal{R}^*_{-v}$ is the maxRAF of $G^* \setminus \{v\}$. The term -1 corrects for the trivial loss of v itself.

The *keystone collapse fraction* normalizes by RAF set size:

$$\hat{K}(v) = \frac{K(v)}{|\mathcal{R}^*|} \tag{8}$$

3.4.2 Bridge Edge Criticality

Definition 6 (Bridge Edge Criticality)**.** *For an edge $(u, v) \in E^*$ with $u, v \in \mathcal{R}^*$, the bridge criticality measures RAF collapse upon edge removal:*

$$B(u, v) = |\mathcal{R}^*| - |\mathcal{R}^*_{-(u,v)}| \tag{9}$$

where $\mathcal{R}^*_{-(u,v)}$ is the maxRAF of $G^* \setminus \{(u, v)\}$.

An edge with $B(u, v) > 0$ is a *critical bridge*: its removal causes at least one gene to lose its only catalytic support within the RAF set.

3.4.3 Gatekeeper Feed Strength

Definition 7 (Gatekeeper Feed Strength)**.** *For a food set node $f \in F$, the gatekeeper feed strength measures its influence on the RAF core:*

$$\Gamma(f) = \sum_{v \in \mathcal{N}^+(f) \cap \mathcal{R}^*} w_{fv} \tag{10}$$

where $\mathcal{N}^+(f)$ is the out-neighbourhood of f.

The *gatekeeper breadth* counts the number of RAF members fed:

$$\gamma(f) = |\mathcal{N}^+(f) \cap \mathcal{R}^*| \tag{11}$$

3.4.4 Composite Disruption Score

We define a composite disruption score integrating all three metrics:

$$D(v) = \lambda_1 \cdot \tilde{K}(v) + \lambda_2 \cdot \tilde{\hat{K}}(v) + \lambda_3 \cdot \tilde{\kappa}(v) \tag{12}$$

where $\tilde{\cdot}$ denotes min-max normalization to $[0, 1]$, and $\lambda_1 = 0.5$, $\lambda_2 = 0.3$, $\lambda_3 = 0.2$ are weights reflecting the relative importance of absolute collapse, fractional collapse, and catalytic influence respectively.

Nodes are classified into priority tiers:

$$\text{Priority}(v) = \begin{cases} \text{High} & D(v) > 0.66 \\ \text{Medium} & 0.33 < D(v) \leq 0.66 \\ \text{Low} & D(v) \leq 0.33 \end{cases} \tag{13}$$

4 Experimental Setup

4.1 Network Parameters

Table 2: Network statistics for the 30% sample used in the main analysis.

Parameter	10% Sample	30% Sample
Total edges (full)	138,479	138,479
Edges sampled	13,848	41,544
Unique nodes	958	994
Weight threshold τ	0.90	0.90
Food set size $\lvert F\rvert$	53	32
RAF set size $\lvert\mathcal{R}^*\rvert$	211	459
RAF fraction	22.0%	46.2%
Connected components	6	8
Largest component	953	987

4.2 Computational Implementation

All analyses were implemented in Python 3.10 using NetworkX 3.1 for graph operations and NumPy/Pandas for numerical computation. Keystone analysis requires $\lvert\mathcal{R}^*\rvert$ RAF recomputations; for $\lvert\mathcal{R}^*\rvert = 459$ this is computationally tractable ($\sim$5 minutes on a standard workstation). Bridge edge analysis requires $\lvert E^*_{\mathcal{R}}\rvert$ recomputations where $E^*_{\mathcal{R}}$ is the set of intra-RAF edges above threshold.

5 Results

The maxRAF algorithm converged to a stable core of **459 transcription factors** (46.2% of all nodes) supported by a food set of 32 constitutively active TFs. The RAF set is not random: it is enriched for known master regulators and contains all major transcriptional regulatory families (Table 3).

Table 3: Transcription factor families represented in the RAF core.

Family	Key Members	Biological Role
Pluripotency	SOX2, NANOG, POU5F1	Stem cell identity
Oncogenes	MYC, STAT3, RELA	Proliferation, survival
Tumour suppressors	TP53, SMAD3, SMAD5	Growth control
Immune	IRF1/3/4/5/6/7/8/9, STAT1/2	Interferon response
Cell cycle	E2F1/2/3/4/6, TFDP1	Proliferation control
Differentiation	MYOD1, MYOG, GATA1/4	Lineage commitment
Circadian	ARNTL, NR1D1/2, NPAS2	Clock regulation
HOX cluster	HOXA9/10/11, HOXD4/11	Developmental patterning
Nuclear receptors	RXRA/B/G, RARA/B, PPARG	Hormone signalling

5.1 Keystone Analysis

Table 4 presents the top 20 RAF keystones ranked by collapse impact $K(v)$.

Table 4: Top 20 RAF keystones ranked by RAF collapse $K(v)$.

Gene	$K(v)$	$\hat{K}(v)$ (%)	$\kappa(v)$	k^+	k^-	$D(v)$
ESR1	5	1.09	42.46	112	117	0.931
RUNX2	4	0.87	36.49	102	97	0.751
TP53	3	0.65	63.17	148	151	0.675
MYC	2	0.44	64.51	159	130	0.xxx
NFATC2	2	0.44	16.39	46	46	0.xxx
SMAD3	2	0.44	37.53	93	107	0.xxx
GATA2	2	0.44	44.92	131	155	0.xxx
PPARG	2	0.44	32.08	96	82	0.xxx
RBPJ	2	0.44	36.89	131	116	0.xxx
RXRA	1	0.22	24.33	78	71	0.xxx
ATF4	1	0.22	17.60	50	48	0.xxx
RUNX1	1	0.22	16.79	60	58	0.xxx
MYOG	1	0.22	19.44	65	70	0.xxx
HES1	1	0.22	18.03	65	41	0.xxx
MAFK	1	0.22	14.49	42	48	0.xxx
HEY1	1	0.22	22.19	69	63	0.xxx
NKX2-5	1	0.22	34.70	118	104	0.xxx
FOXA2	1	0.22	42.95	133	134	0.xxx
BHLHE41	1	0.22	9.40	32	45	0.xxx
SOX10	1	0.22	34.64	103	110	0.xxx

k^+: out-degree, k^-: in-degree, $D(v)$: disruption score

Key observation. ESR1 and RUNX2 have *lower catalyticity* than MYC and TP53, yet cause *greater RAF collapse*. This demonstrates that RAF keystones are **not** simply the most connected or most catalytic genes—they occupy structurally irreplaceable positions in the autocatalytic closure.

5.2 Bridge Edge Analysis

We identified **294 critical bridge edges** whose removal causes at least one RAF member to lose catalytic support. The top bridge edges include:

- **Self-loops** (e.g., ZNF17→ZNF17, FOXP1→FOXP1, ZNF445→ZNF445): These represent autoregulatory TFs whose only catalytic support within the RAF set is themselves. Disrupting their autoregulation immediately removes them from the RAF.
- **Cross-family bridges** (e.g., SMAD3→TFAP2B, SMAD3→DLX1): These connect distinct regulatory modules; their removal isolates downstream genes from catalytic support.

The prevalence of self-loops as critical bridges reveals an important biological insight: **autoregulatory TFs are maximally vulnerable** because they have no redundant catalytic support within the RAF set.

5.3 Food Set Gatekeeper Analysis

Table 5 presents the top food set gatekeepers.

Table 5: Top 10 food set gatekeepers ranked by RAF members fed $\gamma(f)$.

Gene	$\gamma(f)$	$\Gamma(f)$	$K_F(f)$	k^+
NR2F2	80	24.33	1	144
PAX5	57	19.45	0	108
FOXM1	56	18.99	0	96
HOXA5	56	16.58	0	91
TCF7L1	55	14.83	0	97
SNAI1	52	18.57	1	96
REL	52	18.78	1	72
NKX6-1	46	15.14	0	100
FOXC2	46	13.20	0	78
OLIG2	45	16.58	0	85

$K_F(f)$: RAF loss upon food node removal

NR2F2 feeds 80 RAF members with a total feed strength of 24.33, making it the most influential external input to the RAF core. In cancer, NR2F2 (COUP-TFII) is known to regulate angiogenesis and metastasis [Qin et al., 2010], suggesting that blocking its input to the RAF core could have broad anti-tumour effects.

5.4 High-Priority Disruption Targets

Three genes achieved High priority disruption scores (Table 6):

Table 6: High-priority RAF disruption targets.

Gene	$K(v)$	$\hat{K}(v)(\%)$	$\kappa(v)$	$D(v)$	**Cancer relevance**
ESR1	5	1.09	42.46	0.931	Breast cancer driver; endocrine resistance
RUNX2	4	0.87	36.49	0.751	Bone metastasis; osteosarcoma
FOXO1	4	0.67	67.17	0.695	regulator of metabolism
TP53	3	0.65	63.17	0.675	Pan-cancer tumour suppressor

6 Discussion

The most striking finding is the *dissociation* between catalyticity and keystone impact. MYC has the highest catalyticity score (64.51) yet causes only 2-node RAF collapse. ESR1, with catalyticity 42.46, causes 5-node collapse. This demonstrates that **RAF keystones are not hub genes**: they are genes that occupy structurally irreplaceable positions in the autocatalytic closure, regardless of their overall connectivity.

This has profound therapeutic implications. MYC has been extensively targeted in cancer with limited success, partly due to its essential role in normal cell proliferation [Dang, 2012]. ESR1 and RUNX2, as RAF keystones, may represent more selective targets whose disruption specifically collapses the cancer-specific autocatalytic circuit.

FOXO1 as a context-dependent regulatory switch. FOXO1 illustrates the ambiguity of assigning fixed oncogenic or tumour-suppressive labels to transcription factors. Canonically, FOXO1 acts as a tumour suppressor by promoting cell-cycle arrest, apoptosis, oxidative-stress response, and metabolic homeostasis. In breast cancer, reduced FOXO1 expression, promoter methylation, and PI3K/AKT-mediated nuclear exclusion have been associated with tumour progression and poorer clinical features. However, FOXO1 can also support tumour survival under specific stress conditions. In triple-negative breast cancer models, FOXO1-mediated autophagy can protect cells from paclitaxel-induced apoptosis, and FOXO1 activation has been linked to proliferation, invasion, EMT, autophagy, and therapy resistance. Similarly, in diffuse large B-cell lymphoma, mutant FOXO1 maintains enhancer accessibility and supports an oncogenic transcriptional program.

Thus, FOXO1 should not be interpreted as intrinsically tumour-suppressive or oncogenic. Its functional role depends on localization, phosphorylation state, mutation status, p53 competence, tumour subtype, stress exposure, and network embedding. From a RAF perspective, FOXO1 is therefore best understood as a conditional regulatory switch: it may suppress oncogenic closure in one context while stabilizing a self-sustaining cancer circuit in another.

6.1 The RAF Collapse Hypothesis

We propose the following testable hypothesis:

Theorem 1 (RAF Collapse Hypothesis)**.** *Cancer cell viability is maintained by the integrity of a minimal autocatalytic transcriptional core $\mathcal{R}^*$. Targeted disruption of RAF keystones v with $K(v) > K_{\min}$ causes cascade collapse of $\mathcal{R}^*$, selectively impairing cancer cell transcriptional programs while sparing normal cells whose RAF sets have greater redundancy.*

This hypothesis is testable via CRISPR knockout screens targeting the identified keystones, with readout of transcriptional program collapse (RNA-seq) and cell viability.

6.2 Self-Loop Bridge Edges as Vulnerability Points

The dominance of self-loops among critical bridge edges (e.g., ZNF17, FOXP1, ZNF445) reveals a previously underappreciated vulnerability: TFs that maintain their own expression through autoregulation are maximally dependent on that single edge for RAF membership. Disrupting autoregulation (e.g., through promoter-targeting approaches or degrader technologies) could selectively remove these TFs from the RAF set.

6.3 Limitations

1. **Sampling variability**: The 30% sample may not capture all regulatory interactions. Stability analysis across multiple seeds is needed.

2. **Static network**: The RAF framework as applied here treats the network as static. Dynamic RAF analysis incorporating temporal gene expression data would be more biologically realistic.

3. **Tissue specificity**: The PPI network is not tissue-specific. Cancer-type-specific RAF sets may differ substantially.

4. **Validation**: Computational predictions require experimental validation via CRISPR screens and transcriptomic readouts.

7 Conclusion

We have demonstrated that RAF theory provides a principled framework for identifying structurally irreplaceable genes in cancer transcriptional regulatory networks. Our three novel metrics—Keystone Score, Bridge Edge Criticality, and Gatekeeper Feed Strength—capture complementary aspects of network vulnerability that are invisible to standard hub or centrality analyses.

The identification of ESR1, RUNX2, and TP53 as high-priority disruption targets, and the discovery that RAF keystones are not simply hub genes, opens new directions for computational cancer target identification. Future work will extend this framework to tissue-specific networks, dynamic RAF analysis, and experimental validation.

References

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1), 56–68.

Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2), 101–113.

Dang, C. V. (2012). MYC on the path to cancer. *Cell*, 149(1), 22–35.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74.

Ghiassian, S. D., Menche, J., & Barabási, A.-L. (2015). A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLOS Computational Biology*, 11(4), e1004120.

Emmert-Streib, F., de Matos Simoes, R., Mullan, P., Haibe-Kains, B., and Dehmer, M. (2014). The gene regulatory network for breast cancer: integrated regulatory landscape of cancer hallmarks. *Frontiers in Genetics*, 5, 15. doi:10.3389/fgene.2014.00015

Smith, K. N., Singh, A. M., and Dalton, S. (2010). MYC transcription factors: key regulators behind establishment and maintenance of pluripotency. *Regenerative Medicine*, 5(6), 947–959. doi:10.2217/rme.10.79

Fischer, M. (2017). Census and evaluation of p53 target genes. *Oncogene*, 36, 3943–3956. doi:10.1038/onc.2016.502

Mares-Quiñones, M. D., et al. (2024). Identification of modules and key genes associated with breast cancer subtypes through network analysis. *Scientific Reports*, 14, 11855. doi:10.1038/s41598-024-61908-4

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646–674.

Hordijk, W., & Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4), 451–461.

Hordijk, W., Hein, J., & Steel, M. (2012). Autocatalytic sets and the origin of life. *Entropy*, 12(7), 1733–1742.

Kauffman, S. A. (1986). Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1), 1–24.

Lambert, S. A., et al. (2018). The human transcription factors. *Cell*, 172(4), 650–665.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.

Menche, J., et al. (2015). Uncovering disease-disease relationships through the incomplete interactome. *Science*, 347(6224), 1257601.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.

Qin, J., et al. (2010). COUP-TFII inhibits TGF-β-induced growth barrier to promote prostate tumorigenesis. *Nature*, 493(7431), 236–240.

Steel, M. (2000). The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters*, 13(3), 91–95.

Vasas, V., Szathmáry, E., & Santos, M. (2012). Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. *Proceedings of the National Academy of Sciences*, 107(4), 1470–1475.