POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

A network medicine approach for drug repurposing in arrhythmogenic cardiomyopathy

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Author: **Aurora Vido**

Student ID: 252189
Advisor: Prof. Paolo Zunino
Co-advisors: Dr. Cristina Banfi, Ph.D., Dr. Elena Sommariva Ph.D.
Academic Year: 2024-25

Abstract

Arrhythmogenic cardiomyopathy (ACM) is an inherited heart muscle disease characterized by progressive fibrofatty replacement and electrical instability, for which current pharmacological treatments remain palliative rather than curative. This thesis presents a network medicine approach to identify FDA-approved or experimental drugs capable of modifying the progression of the underlying pathology by treating ACM as a "network failure" disease. The study employs a human interactome model and applies the Random Walk with Restart (RWR) algorithm to quantify the functional relevance of drugs with respect to the ACM disease module. To address the topological hub bias inherent in scale-free networks, the methodology integrates a degree-preserving permutation test with a phenotype-based context score. This dual-filter approach effectively prioritizes candidates that are both topologically significant and phenotypically relevant to the cardiac context. The framework demonstrated high predictive performance, achieving an Area Under the Curve (AUC) of 0.861 in the retrieval of drugs with known clinical relevance or use in treating ACM, and the topological characterization revealed that effective target genes are not generic hubs, but rather network "bridges" with high betweenness and eigenvector centrality. Furthermore, transcriptomic validation using the Connectivity Map (CMap) on five independent datasets identified compounds capable of reversing the pathogenic gene expression signatures. The analysis highlighted *Isradipine* and *Fostamatinib* as promising new candidates, along with the rediscovery of established therapies such as *Amiodarone* and *Flecainide*. This work provides an open-source and adaptable computational pipeline, as well as a prioritized list of drugs that can be further validated *in vitro*.

Keywords: Network Medicine, Drug Repurposing, Arrhythmogenic Cardiomyopathy (ACM), Random Walk with Restart (RWR), Disease Module, Transcriptomic Analysis.

Abstract in lingua italiana

La cardiomiopatia aritmogena (CMA) è una malattia ereditaria del muscolo cardiaco caratterizzata da progressiva sostituzione fibroadiposa e instabilità elettrica, per la quale gli attuali trattamenti farmacologici rimangono palliativi piuttosto che curativi. Questa tesi presenta un approccio di medicina di rete per identificare farmaci approvati dalla FDA o sperimentali in grado di modificare la progressione della patologia, trattando la CMA come una malattia da "fallimento di rete". Lo studio utilizza un modello di interattoma umano e applica l'algoritmo del Random Walk with Restart (RWR) per quantificare la rilevanza funzionale dei farmaci per il modulo di malattia della CMA. Per affrontare il bias topologico degli hub insito nelle reti scale-free, la metodologia integra un test di permutazione che preserva il grado con un punteggio di contesto basato sul fenotipo. Questo approccio a doppio filtro dà effettivamente priorità ai candidati che sono sia topologicamente significativi che fenotipicamente rilevanti per il contesto cardiaco. Il metodo ha dimostrato elevate prestazioni predittive, raggiungendo un'area sotto la curva (AUC) di 0.861 nel recupero di farmaci noti nel trattare la CMA, e la caratterizzazione topologica ha rivelato che i geni bersaglio efficaci non sono hub generici, ma piuttosto "ponti" di rete con elevate betweenness e eigenvector centrality. Inoltre, la validazione trascrittomica utilizzando la Connectivity Map (CMap) su cinque set di dati indipendenti ha identificato farmaci in grado di invertire le firme di espressione genica patogena. L'analisi ha dato priorità a *Isradipina* e *Fostamatinib* come nuovi candidati promettenti, insieme alla riscoperta di terapie consolidate come *Amiodarone* e *Flecainide*. Questo lavoro fornisce una pipeline computazionale open source e adattabile, nonché un elenco prioritario di farmaci che possono essere ulteriormente validati *in vitro*.

Parole chiave: Medicina di rete, Riposizionamento dei farmaci, Cardiomiopatia Aritmogena (CMA), Random Walk with Restart (RWR), Modulo malattia, Analisi trascrittomica.

Contents

Introduction

Arrhythmogenic Cardiomyopathy (ACM) is an inherited heart muscle disease characterized by a progressive loss of cardiomyocytes and their replacement with fibrofatty tissue, predisposing to life-threatening ventricular arrhythmias and sudden cardiac death, especially in young patients and athletes. The disease is fundamentally a "cell-to-cell junction cardiomyopathy", originating from genetically determined defects in the structural and signaling hub of the heart: the intercalated disc [1]. The clinical syndrome of ACM is highly heterogeneous, from asymptomatic carrier states to arrhythmic malignant phenotypes, reflecting the interaction between genetic predisposition, environmental factors and incompletely defined molecular pathways [2].

The genetic architecture of ACM is centered on mutations in genes encoding desmosomal proteins, which form the mechanical junctions between cardiomyocytes. A definitive curation by the international Clinical Genome Resource Expert Panel (ClinGen [3]) has clarified the genetic basis of ACM, distinguishing a core set of causative genes from a longer list of genes with limited or refuted evidence. The analysis confirmed only eight genes as having definitive or moderate evidence for causing ACM: PKP2 (Plakophilin-2), DSP (Desmoplakin), DSG2 (Desmoglein-2), DSC2 (Desmocollin-2), JUP (Plakoglobin), TMEM43 (Transmembrane protein 43), PLN (Phospholamban) and DES (Desmin) [4]. These eight genes are not just minor contributors, they represent the vast majority of the disease's genetic etiology, accounting for 97.4% of all pathogenic/likely pathogenic (P/LP) ACM variants cataloged in the ClinVar database [5]. In contrast, 18 other genes previously associated with ACM were found to have limited or no evidence, accounting for only 1.1% of P/LP variants [4].

Despite the advances in genetic characterization and clinical trials, there is no curative therapy for ACM yet. The current treatments are mostly symptomatic and preventive, including Implantable Cardioverter Defibrillators (ICDs), antiarrhythmic drugs and ablation procedures, which, though lowering the arrhythmic burden, do not address the basic underlying pathobiology [1]. This underlines the need for novel approaches that can modify the disease progression at a molecular level.
The genetic heterogeneity alone, coupled with observations of incomplete penetrance, vari-

able expressivity and mutations in multiple desmosome genes strongly suggest that ACM is not a simple monogenic disorder, rather, it is a disease of network failure. The various genetic insults (e.g., in PKP2, DSP or DSG2) converge on a common set of downstream cellular signaling pathways [1], that make ACM an ideal candidate for a network medicine approach.

Traditional drug discovery typically focuses on the "one target, one drug" paradigm: a single molecular target is identified, for which a drug candidate is designed to act upon it [6, 7]. While for simple, monogenic disorders, this model has been effective, for more complex ones, like ACM, it has proved to be highly inadequate [7]. The pathophysiology of ACM is not driven by a single defective protein; instead, it results from disruptions across an interconnected network of processes, including desmosomal dysfunction, aberrant signaling (such as Wnt/β-catenin), inflammation and fibrosis [1].
Network medicine offers a paradigm shift by conceptualizing diseases as the outcome of network perturbations rather than single-gene defects [6]. In this framework, genes, proteins and metabolites are represented as nodes in a network, connected by physical or functional interactions. Components that are associated with the same disease tend to aggregate within specific network neighborhoods or "disease modules". Thus, pathological processes emerge when disease-associated modules within the interactome are perturbed [6, 8].

An additional powerful strategy in network medicine is drug repurposing, also known as drug repositioning, which refers to the use of approved drugs for treating diseases beyond their original medical indication [9]. The rationale for this approach is strongly supported by the continuously rising costs and lengthy timelines associated with traditional drug development, where the process from discovery to market often exceeds a decade and requires investments of more than one billion USD [10].
Drug repurposing builds on the hypothesis that a drug effective in one disease may also be effective in another if the two conditions share similar pathobiological mechanisms. In other words, if diseases are linked by overlapping molecular pathways or targets, an existing drug can retain therapeutic efficacy across both [6]. Within the network medicine framework, this idea is applied by identifying drugs whose targets are located within, or in close proximity to, the disease module of interest. Such drugs are predicted to modulate disease-relevant pathways and thus represent promising candidates for repositioning [7].

The following chapters will explore in detail the concepts of network medicine and drug repurposing, alongside a comprehensive description of arrhythmogenic cardiomyopathy. Subsequently, the data collection and methodological framework will be illustrated, specifically focusing on the Random Walk with Restart algorithm, network separation analysis

and phenotype-driven context scoring. The study presents also a topological characterization of significant targets and applies a community detection strategy, the Louvain algorithm, to identify disease modules. Lastly, the computational predictions undergo transcriptomic validation leveraging the Connectivity Map and multiple disease signatures.

The main objective of this thesis is to create and validate an integrated computational tool for drug repurposing in arrhythmogenic cardiomyopathy; however, the developed pipeline is not specific to this condition, but rather a generalizable framework. Specifically, this work aims to integrate transcriptomic and protein-protein interaction data to apply a network-based algorithm and identify potential repurposable drug candidates that could be further validated experimentally.

1 | Network medicine and biological networks

1.1. From reductionism to network medicine

For centuries, Western scientific theory and biomedicine have been dominated by a reductionistic approach, which assumes that biological systems can be understood by disassembling them into their constituent parts, claiming that crucial biological factors operate in simple mechanistic associations to control disease pathobiology [7]. Although this paradigm has been historically successful in formalizing syndromic models and identifying monogenic disorders, it often oversimplifies disease phenotypes and fails to account for individual nuances in disease expression or the complex interdependencies of cellular components [6, 7].

Current biological knowledge suggests that a disease is rarely the result of a single gene anomaly [6]; rather, human disease phenotypes are almost always the result of multiple pathobiological pathways interacting through an intricate and interconnected network [7]. The discipline emerging from this idea is "network medicine", which provides a platform for systematically exploring this molecular complexity. It proposes that the molecular pathways governing human diseases consist of circuits that merge into complex, overlapping networks, and that understanding the topology and dynamics of these networks is essential for interpreting the pathogenic behavior of complex diseases that traditional methods fail to recognize [6, 7].

Network medicine can potentially describe and investigate the molecular complexity of human diseases as methods keep evolving, and it also provides computational methods to understand how this complexity controls prognosis and therapy [7]. The change from a single-gene-centered approach to a network-centered one is a major step in our understanding of human pathobiology [8].

1.2. The human interactome

A biological network can be generated by mapping a set of molecular entities, called "nodes", and their functional interconnections, i.e., their "links" or "edges" [7]. In the context of human biology, these nodes can represent genes, proteins, metabolites, or non-coding RNAs, while the links represent physical interactions, transcriptional induction, or enzymatic reactions [6, 7]. The totality of these interactions in a human tissue constitutes the "interactome" [6, 7].

This human interactome is a highly complex network, comprising approximately 25,000 protein-coding genes, around 1000 metabolites, and an indefinite number of distinct proteins resulting from splicing variants and post-translational modifications, easily exceeding 100,000 cellular components [6]. The number of functionally relevant interactions between these components is expected to be much higher, creating a subcellular interconnectivity that suggests that the effects of a particular genetic defect are not limited to the product of the mutated gene, but spreads along the network connections [6].

1.2.1. Topological characteristics

Biological networks are not random, but follow specific organizational principles, and a distinctive feature of the human interactome is that it is a scale-free network [6, 7]. In random networks, the majority of nodes have a similar number of connections, following a Poisson distribution. In contrast, the degree distribution of a scale-free network follows a power-law distribution, which means that while most nodes are sparsely connected, there are a few highly connected nodes, known as "hubs", that hold the network together [6, 7].

These hubs play a fundamental biological role within the network. In fact, evidence indicates that hub proteins are often encoded by essential and evolutionarily conserved genes that are crucial for cellular activities and embryonic survival [6, 7]. Consequently, the deletion of a hub gene is more lethal than the deletion of a non-hub gene [6]. This topological structure confers robustness to the network: scale-free systems are "overdetermined", i.e., able to adapt to perturbations (such as the loss of non-hub nodes) with minimal negative effects, while facilitating the rapid transfer of molecular information throughout the system [7].

1.3. Network topology and disease genes

A central question in network medicine is whether disease genes occupy specific positions within the interactome. Although one might assume that disease genes are hubs due to their extensive connectivity, empirical analysis suggests that essential genes, those whose dysfunction causes embryonic mortality, strongly tend to encode hubs and cluster at the functional center of the interactome [6]. In contrast, non-essential disease genes (which allow the host to survive to reproductive age) do not typically encode hubs, but instead tend to segregate to the functional periphery of the network [6, 7]. This distinction arises from evolutionary constraints: mutations in hubs are often incompatible with life, while mutations in peripheral nodes can persist in the population, manifesting as specific pathological phenotypes [6].

1.3.1. The local hypothesis and disease modules

If a gene is involved in a specific disease, its direct interactors may also play a role in the same pathological process. This concept is formalized as the "local hypothesis", according to which proteins associated with the same disease often interact with each other [6, 7]. Various studies confirm that the outputs of genes linked to the same disorder interact with each other 10 times more frequently than would be expected by chance [6]. This clustering suggests that for each disease there is a specific neighborhood within the interactome, a "disease module", in which the relevant components associated with the disease are located [6, 7]. A disease phenotype is therefore seen not as a single defect, but as the breakdown of a specific functional module [6], and thus the identification of these modules is one of the primary goals of network medicine, as the topological position of a disease gene can offer information about new associated partners and pathways that regulate the pathogenotype [7]. Defining these structures requires distinguishing between three modularity concepts [6]:

- **Topological modules:** locally dense neighborhoods where nodes are highly interconnected.
- **Functional modules:** clusters of nodes with related cellular functions (e.g., a signaling pathway).
- **Disease modules:** a group of network components whose perturbation results in a specific disease phenotype.

While these modules often overlap, a disease module is unique to a particular pathology [6].

1.4. Network dynamics and signaling

While topology provides a static map of the interactome, understanding diseases requires an assessment of network dynamics, i.e., the flow of information through these molecular circuits [7]. Cellular behavior, such as proliferation or apoptosis, is regulated by the integration of numerous input signals processed through signaling networks [11].

These networks are dynamic and regulated by mechanisms such as post-translational modifications (e.g., phosphorylation), which control protein-protein interactions and enzymatic activity [11]; moreover, the flow of information within a network can vary depending on the physiological context or disease state, effectively "reconfiguring" the functional connections between proteins [8]. For example, in cancer, aberrant signaling networks can become hyperactive or resistant to regulation, leading to disease states [11]. Therefore, network medicine seeks to understand not only the structure of the network, but also how the dynamic flow of signals is altered in disease [7].

1.5. Network pharmacology and drug repurposing

The principles of network medicine also have profound implications for drug discovery. Historically, drug design has followed a reductionist approach based on the search for a "magic bullet", a single target capable of reversing a phenotype [7]. However, the failure of many single-target drugs in the treatment of complex diseases suggests that therapeutic efficacy often depends on the modulation of multiple nodes (i.e., genes, proteins) within a network [7, 11].

1.5.1. Systems pharmacology

"Systems pharmacology" uses knowledge gained from networks to guide rational drug development. By mapping drug targets within the interactome, researchers have observed that many approved drugs do not target proteins actually associated with the disease, but rather those nearby within the network [7]. In addition, drugs often have "off-target" effects, unintended interactions with other proteins, that can lead to adverse events or, conversely, to new therapeutic benefits [7].

Network analysis allows these effects to be predicted. For example, creating a network map of drug interactions or projecting drug targets onto pathological modules can reveal how specific compounds might disrupt the network [7]. This approach moves away from the search for a single, highly specific compound and toward targeting the architecture of

the aberrant signaling network itself, potentially using combination therapies ("rational polypharmacy") to maximize efficacy and minimize toxicity [7, 11].

1.5.2. Drug repurposing

A key application of network medicine is drug repurposing (or repositioning). Since different diseases can share molecular links, a concept visualized in the "human diseasesome", drugs developed for one condition may be effective for another if they target overlapping modules [6, 7]. The "shared gene hypothesis" suggests that if two diseases share a gene or interact within the same network neighborhood, a drug that acts on that neighborhood in one disease could be repurposed for the other [6].

Computational approaches can predict these new indications by integrating gene expression signatures, protein-protein interaction networks and drug-target data [7]. For example, the identification of drugs that target proteins in the "vicinity" of the network of known disease genes has successfully predicted new therapeutic effects and side effects [7]. This strategy exploits the principle of "guilt by association", assuming that if two diseases are similar phenotypically or molecularly, they may respond to similar pharmacological interventions [7]. By targeting the disease module rather than a single symptom, network-based repositioning offers a cost-effective and efficient path to new therapies [7, 8].

1.5.3. Computational strategies in drug repurposing

Recent reviews show that approximately 30% of newly approved drugs in the United States are repurposed, confirming the growing importance of this strategy alongside traditional de novo development [9]. Computational drug repurposing, or *in silico* repurposing, can be divided into two main approaches: *drug-centric*, which aims to identify new indications for an existing drug, and *disease-centric*, which seeks to identify effective drugs for a specific condition [9].

To achieve this, three main computational strategies have been defined [9]:

- **Knowledge-based repurposing:** this strategy uses known properties of drugs, such as chemical structures, targets and adverse effects, to predict unknown mechanisms. It includes target-based methods (screening compounds against a specific protein), pathway-based methods (using metabolic or signaling pathways), and target mechanism-based methods (integrating omics data for precision medicine) [9].
- **Signature-based repurposing:** also known as connectivity mapping, this approach compares the gene expression "signatures" of a disease with those induced

by drugs. It looks for inverse relationships, for example, if a disease causes the upregulation of certain genes, a drug that downregulates those same genes could be therapeutic [9].

- **Phenotype-based repurposing:** this uses electronic health record (EHR) data and clinical observations to identify potential new indications or adverse effects that were not evident during initial trials [9].

1.6. Network medicine in cardiovascular diseases

The principles of network medicine have been applied with particular success to cardiovascular disease (CVD), revealing how network topology and dynamics can explain complex cardiac phenotypes and guide therapeutic interventions [7, 12]. Furthermore, the shift from a traditional reductionist approach to a global systemic perspective allows for the integration of multi-omic data, such as proteomics, into unified network models. This multi-layered approach is critical for discovering shared pathologies, identifying new biomarkers, and accelerating drug repositioning strategies [12, 13].

1.6.1. Cardiovascular disease modules

Network analysis has been instrumental in defining the molecular architecture of various cardiovascular conditions. By integrating Genome-Wide Association Studies (GWAS) with protein-protein interaction (PPI) networks, researchers have identified candidate genes for coronary artery disease susceptibility that were not evident from genetic data alone [7]. In addition, specific "disease modules" have been mapped for atherosclerosis, in-stent restenosis, cardiac hypertrophy and heart failure [7]. For example, a network-based analysis of pulmonary hypertension identified a critical disease module governed by a set of microRNAs (miRNAs) that regulate the intersection between hyperoxia, inflammation and signaling pathways, offering a new "network-based" therapeutic target [7].

1.6.2. Predicting cardiac side effects and off-target interactions

One of the most interesting applications of network medicine in cardiology is the prediction of adverse drug events due to "off-target" effects, such as arrhythmias, that conventional pharmacology often fails to predict. A crucial study on Long-QT Syndrome (LQTS), a disorder that predisposes individuals to fatal arrhythmias, showed that drugs that cause LQTS as a side effect do not necessarily act directly on the genes responsible for the disease; instead, these drugs act on proteins located within the specific network

"neighborhood" of LQTS genes [7]. By mapping this neighborhood, researchers were able to accurately predict which other approved drugs might carry a risk of arrhythmia, a prediction validated by FDA adverse event reports [7].

Similarly, network analysis clarified the failure of Torcetrapib, a cholesterol ester transfer protein (CETP) inhibitor. Despite improving cholesterol levels, the drug caused unexpected hypertension and increased mortality. A network-based investigation revealed that Torcetrapib had an off-target effect on the PPAR (peroxisome proliferator-activated receptor) pathway, a connection that could have been predicted through an analysis of the drug's network interactions [7].

These examples demonstrate that cardiovascular diseases are not isolated phenomena but are deeply interconnected with other biological processes through the interactome, and that network medicine provides the tools necessary to explore this complexity for safer drug development [7].

2 | Arrhythmogenic Cardiomyopathy

2.1. Definition and evolving nomenclature

Arrhythmogenic Cardiomyopathy (ACM) is currently defined as an inherited disease of the heart muscle characterized by the progressive loss of ventricular myocardium and its replacement by fibrofatty tissue. This structural alteration creates a substrate for electrical instability, predisposing patients to potentially lethal ventricular arrhythmias and sudden cardiac death (SCD), particularly in young people and athletes [1].

Historically, this condition was called "Arrhythmogenic Right Ventricular Dysplasia" (ARVD), reflecting the belief that it was a defect in the development of the right ventricular myocardium. However, subsequent understanding that the pathological process is genetically determined and progressive led to the adoption of the term "cardiomyopathy" (ARVC) [1]. More recently, the broader term "Arrhythmogenic Cardiomyopathy" (ACM) has been adopted to encompass the entire spectrum of phenotypic expressions, which include not only the classic right ventricular (RV) variant, but also biventricular and predominantly left ventricular (LV) forms [1].

2.2. Genetic basis and the intercalated disc

ACM is predominantly a genetic disease, typically transmitted in an autosomal dominant pattern, although there are autosomal recessive forms (such as Naxos disease and Carvajal syndrome) associated with skin abnormalities [1]. The disease is characterized by incomplete penetrance and variable expressivity, which means that clinical manifestations can vary significantly even among members of the same family who carry the same mutation [1].

The genetic architecture of ACM is fundamentally linked to the desmosome, a protein complex responsible for mechanical cell-to-cell adhesion. Mutations in genes encoding for

desmosomal proteins are present in approximately 50% of probands [1]. The key causal genes include:

- **Plakophilin-2 (PKP2):** the most frequently mutated gene in ACM cohorts.
- **Desmoplakin (DSP):** often associated with left ventricular involvement and cutaneous phenotypes in recessive forms.
- **Desmoglein-2 (DSG2)** and **Desmocollin-2 (DSC2):** other key desmosomal components.
- **Plakoglobin (JUP):** the first gene identified in the context of Naxos disease.

Although the "desmosomal hypothesis" remains central, recent discoveries have expanded the genetic horizon of ACM also to non-desmosomal genes. These include *CTNNA3* (encoding α-T-catenin), *PLN* (phospholamban), *TMEM43*, *LMNA* (lamin A/C), *DES* (desmin), and *FLNC* (filamin C) [1].

This genetic heterogeneity supports the reclassification of ACM as a "cell-to-cell junction cardiomyopathy". The disease is not limited to mechanical defects but involves the entire intercalated disc (ID), a complex structural and functional hub where mechanical junctions (desmosomes), adherens junctions and gap junctions interact. This protein network, sometimes referred to as the "connexome", integrates mechanical stability with electrical coupling and intracellular signaling pathways [1].

2.3. Pathophysiology: from mutation to phenotype

The pathological hallmark of ACM is the replacement of cardiomyocytes by fibrous and adipose tissue. This process typically begins in the epicardium and progresses toward the endocardium, often causing transmural lesions that can lead to aneurysms [1].
There are multiple mechanisms linking desmosomal mutations to this phenotype. Defective desmosomes compromise the mechanical integrity of the myocardium, leading to the death of myocytes under mechanical stress (e.g., during physical activity). Crucially, these structural defects also trigger aberrant intracellular signaling, in particular, disruption of the intercalated disc leads to the suppression of the canonical Wnt/β-catenin signaling pathway. The inhibition of this pathway promotes adipogenesis (fat formation) and fibrosis, resulting in the characteristic tissue replacement observed in ACM [1].

2.4. Clinical manifestations

The clinical course of ACM is traditionally divided into four phases, which highlight its progressive nature [1]:

1. **Concealed phase:** patients may be asymptomatic with subtle or no structural abnormalities, but are at risk of sudden cardiac death due to electrical instability. This highlights the critical "electrical-anatomical uncoupling" in which arrhythmias can occur before obvious structural damage.
2. **Overt electrical disorder:** it is characterized by palpitations, syncope and documented ventricular arrhythmias originating from the right ventricle (typically with a left bundle branch block morphology). Structural abnormalities become detectable by imaging.
3. **Right ventricular failure:** the diffuse replacement of myocardium leads to global RV contractile impairment and subsequent right ventricular failure, while the function of the left ventricle (LV) remains relatively preserved.
4. **Biventricular pump failure:** the final stage is characterized by significant parallel involvement of the left ventricle, resulting in biventricular systolic dysfunction and pump failure. At this advanced stage, ACM may mimic dilated cardiomyopathy.

Key diagnostic features include T wave inversion in precordial leads (V1-V4), Epsilon waves (low-amplitude signals between the end of the QRS and the T-wave), and ventricular ectopy [1].

2.5. Therapeutic challenges: a candidate for repurposing

Current therapeutic strategies for ACM, including lifestyle restrictions (avoiding competitive sports), beta-blockers, antiarrhythmic drugs and Implantable Cardioverter Defibrillators (ICDs), are primarily palliative. Their goal is to reduce symptoms and prevent sudden death, but they do not arrest the progression of the underlying disease [1].

This lack of curative therapies makes ACM an ideal candidate for drug repurposing through a network medicine approach. The disease results from a malfunction of a complex protein network (the intercalated disc) and involves known signaling perturbations (such as the Wnt/β-catenin pathway) [1]. By targeting these molecular networks rather than just the arrhythmic symptoms, repurposed drugs could potentially restore signaling

balance and halt fibrofatty replacement, addressing the root biological cause identified in the current literature.

3 | Data collection

3.1. The interactome

This analysis employed the human protein-protein interaction (PPI) network, obtained from the STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins, v12.0) [14] to construct a comprehensive model of the human interactome. Unlike databases that consider physical protein binding alone, the STRING database also considers functional associations, which are defined as two proteins participating in the same biological process.
The key feature of the downloaded STRING database is its scoring system, where each interaction is assigned a combined confidence score ranging between 0 and 1 (or 0 and 1000), which quantifies the likelihood that the interaction is biologically true given the available evidence. This score is not an arbitrary measure but it is a probability integration obtained from seven distinct evidence channels [14]:

- **High-throughput experiments:** physical interactions detected by lab techniques such as yeast two-hybrid or affinity purification.
- **Curated databases:** pathways and complexes from sources like KEGG, Reactome and BioCyc.
- **Co-expression:** genes that show similar expression patterns across varying conditions.
- **Text-mining:** associations extracted from the analysis of millions of scientific publications.
- **Neighborhood:** preservation of the genomic neighborhood, identifying genes that are located near each other in the genomes of multiple organisms (synteny).
- **Gene fusion:** identification of fusion events where two separate genes in one organism correspond to a single fused gene in another species.
- **Co-occurrence:** phylogenetic profiling that identifies genes that tend to be present

or absent together across different species, suggesting a functional link.

A very strict filtering criterion was applied to ensure the reliability of the subsequent network propagation algorithm and minimize the impact of false positives: only interactions with a combined score ≥ 0.7 were retained. In the STRING terminology, this threshold corresponds to "High Confidence" interactions [14] and, by selecting this cutoff, the resulting network prioritizes solid links supported experimentally or by computational methods, rather than low-confidence predictions.

3.2. Disease genes

A set of genes known to be causally associated with ACM was manually curated from clinical databases (Phenopedia [15], ClinVar [5]). This gene set (see Table 3.1), denoted as the disease module S, constitutes the "seed" nodes for the network propagation analysis, representing the core pathogenic perturbations.

Table 3.1: Seed genes associated with ACM.

Gene Symbol	NCBI Gene ID
PKP2	5318
DSG2	1829
TMEM43	79188
RYR2	6262
DSP	1832
DSC2	1824
DES	1674
TGFB3	7043
JUP	3728
PLN	5350
TTN	7273
CTNNA3	29119
LMNA	4000
CDH2	1000
FLNC	2318
SCN5A	6331
ACE	1636
ACTN2	88
LDB3	11155

Table 3.1: Seed genes associated with ACM.

Gene Symbol	NCBI Gene ID
OBSCN	84033
MYH7	4625

3.3. Drugs

Drug-target information was obtained from DrugBank [16], which provides experimentally verified relationships between pharmacological drugs and their target proteins. Target Uniprot IDs were converted into their corresponding Entrez gene IDs through the UniProt.map [17] function of the bioservices Python package [18] that queries the UniProt ID mapping service. The original database contained 14,687 small molecule drugs, 4,312 biotech (biologic) drugs and 135 nutraceutical drugs. Withdrawn and illicit drugs were excluded from the dataset, since they are not available and hence cannot be repurposed, and also veterinary drugs and compounds that are toxic to humans were discarded. The final dataset consisted of 5,089 drugs.

4 | Methods

4.1. Random Walk with Restart

The Random Walk with Restart (RWR) algorithm is a powerful network propagation method used to measure the functional proximity and influence between nodes in a network [19]. In the context of drug repurposing, it is used to quantify the relationship between a set of disease genes and all other proteins in the interactome, where the core intuition is that the influence of a disease is not confined to its core genes but diffuses through the network. A drug is considered a promising candidate if its targets lie within the "hot zones" of the network, i.e., regions that are highly influenced by the disease genes [20, 21].

The algorithm simulates a "random walker" that moves through the protein-protein interaction network. The walker starts on the disease genes and, at each step, either moves to a random neighbor or "restarts" by going back to one of the starting nodes. The final output is a probability score for every protein, representing how frequently it was visited by the walker, which provides a robust measure of its functional relevance to the disease.

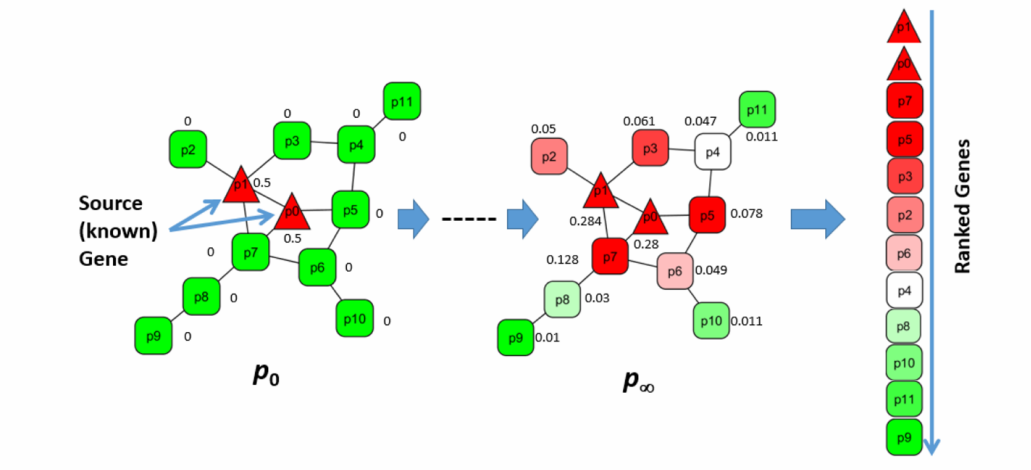

Figure 4.1: An illustration of disease candidate gene prioritization using the pure RWR algorithm (from [19]).

4.1.1. Network and input definition

The method requires three inputs:

- The human interactome: a graph $G(V, E)$, where V is the set of all proteins (nodes) and E is the set of interactions (edges). The graph can be weighted or unweighted and it is represented by an adjacency matrix A, which is column-normalized to obtain the transition matrix W. The element W_{ij} is the probability of going from node j to node i, which is defined as:

$$W_{ij} = \frac{A_{ij}}{\sum_j A_{ij}} \tag{4.1}$$

- The disease module S: a set of seed nodes corresponding to the known disease-associated genes.
- The restart probability γ: a scalar value between 0 and 1 that controls the probability of the walker going back to the seed nodes at each step.

An initial probability vector $\vec{p}_0$ is defined, which represents the starting position of the random walker and assigns a uniform probability to all nodes in the disease module S and zero to all other nodes:

$$\vec{p}_{0,i} = \begin{cases} 1/|S| & \text{if node } i \in S \\ 0 & \text{if node } i \notin S \end{cases} \tag{4.2}$$

where $|S|$ is the number of genes in the disease module.

4.1.2. The iterative propagation algorithm

The algorithm iteratively updates a probability vector $\vec{p}_t$, where each element $p_{t,i}$ represents the probability of the walker being at node i at step t. The update rule combines two actions: the random walk on the graph and the restart to the seed nodes. The iterative formula for Random Walk with Restart is:

$$\vec{p}_{t+1} = (1 - \gamma)W\vec{p}_t + \gamma\vec{p}_0 \tag{4.3}$$

where:

- $\vec{p}_t$ is the probability vector at step t.
- W is the transition matrix of the interactome.

- $(1-\gamma)W\vec{p}_t$ represents the probability distribution after one step of the random walk.
- $\gamma\vec{p}_0$ represents the probability of the walker "restarting" from the initial seed nodes.

This iteration continues until the probability vector converges to a steady state, i.e., when the difference between $\vec{p}_{t+1}$ and $\vec{p}_t$ (measured by the L1-norm) falls below a small threshold (e.g., 10^{-6}). The final steady-state vector, $\vec{p}_\infty$, contains the global, long-term visitation probability for every node in the network, and represents a comprehensive ranking of all proteins in the interactome based on their functional proximity to the ACM disease module.

Algorithm 4.1 Random Walk with Restart (RWR)

Require: Interactome graph $G(V, E)$, Seed genes $S \subset V$, Restart probability γ, Convergence threshold ϵ

Ensure: Steady-state probability vector $\vec{p}_\infty$

$W \leftarrow$ Column-normalized adjacency matrix of G {Transition matrix}
$\vec{p}_0 \leftarrow$ Initialize vector where $p_{0,i} = 1/|S|$ if $i \in S$, else 0
$\vec{p}_t \leftarrow \vec{p}_0$
repeat
  $\vec{p}_{t+1} \leftarrow (1-\gamma)W\vec{p}_t + \gamma\vec{p}_0$ {Iterative update rule}
  $\Delta \leftarrow ||\vec{p}_{t+1} - \vec{p}_t||_1$ {Measure convergence}
  $\vec{p}_t \leftarrow \vec{p}_{t+1}$
until $\Delta < \epsilon$
$\vec{p}_\infty \leftarrow \vec{p}_t$
return $\vec{p}_\infty$

4.2. Interactome pre-processing

Biological networks, unlike random graphs (e.g., Erdős-Rényi models), typically exhibit a scale-free architecture [6, 7]. The defining feature of a scale-free network is the distribution of node degrees (k), which follows a Power-Law distribution:

$$P(k) \sim k^{-\gamma} \tag{4.4}$$

where $P(k)$ is the probability that a node has exactly k connections and γ is the degree exponent (typically $2 < \gamma < 3$ for biological networks).

To verify this property in our specific network, the degree distribution was plotted on a log-log scale (Figure 4.2). The figure shows a clear linear decay in the tail of the

distribution, a hallmark of power-law behavior, which confirms that the interactome used in this study is indeed scale-free.

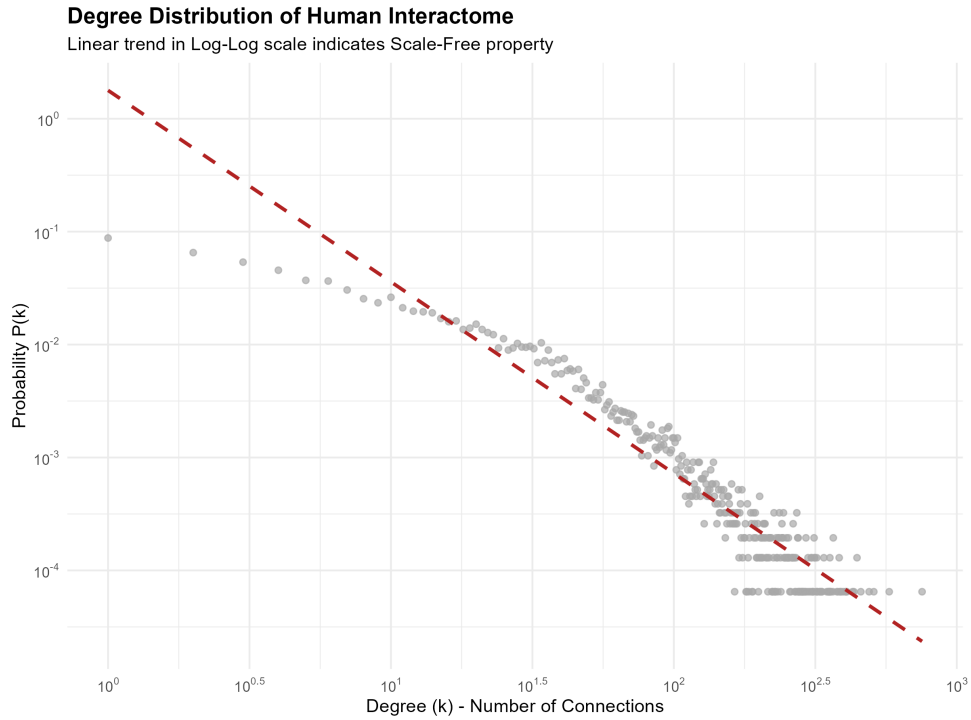

Figure 4.2: Degree distribution of the Human Interactome. The plot displays the probability $P(k)$ of a node having a degree k on a logarithmic scale. The data points (grey dots) follow a linear downward trend (red dashed line), confirming that the network topology follows a Power-Law distribution ($P(k) \sim k^{-\gamma}$), characteristic of scale-free networks.

The scale-free network is dominated by a small number of highly connected nodes, known as "hubs", while the vast majority of proteins have very few interactions. From a computational perspective, this topology creates a natural hub bias: propagation algorithms tend to go through these hubs and accumulate probability simply due to their high connectivity, regardless of their specific biological relevance to the disease module [22]. This observation justifies the need for the statistical correction implemented in this study, specifically the degree-preserving permutation test (Section 4.4) and the log-transformation of RWR scores (Section 4.3), which are designed to decouple specific disease proximity from generic network centrality.

Furthermore, the interactome loaded from STRING was pre-processed to ensure the topological integrity of the network and the viability of subsequent random walk algorithms. The interactome was modeled as a weighted undirected graph, where the weight w_{ij} of the edge connecting protein i and protein j corresponds to the STRING combined confidence

score. First, only interactions with a combined score ≥ 0.7 (High Confidence) were retained, and subsequently, the network was simplified by removing self-loops and duplicate edges. In addition, since network propagation algorithms are only effective on a connected graph, the Largest Connected Component (LCC) of the network was extracted. Disconnected components such as small islands of proteins isolated from the main network were removed to ensure that all proteins in the network were reachable from all other nodes. After processing, the LCC used for the analysis contained 15,437 proteins and 227,723 weighted interactions. All disease genes and drug targets not present in this LCC were excluded from the analysis. Figure 4.3 displays the ACM disease genes and their closest network neighborhood.

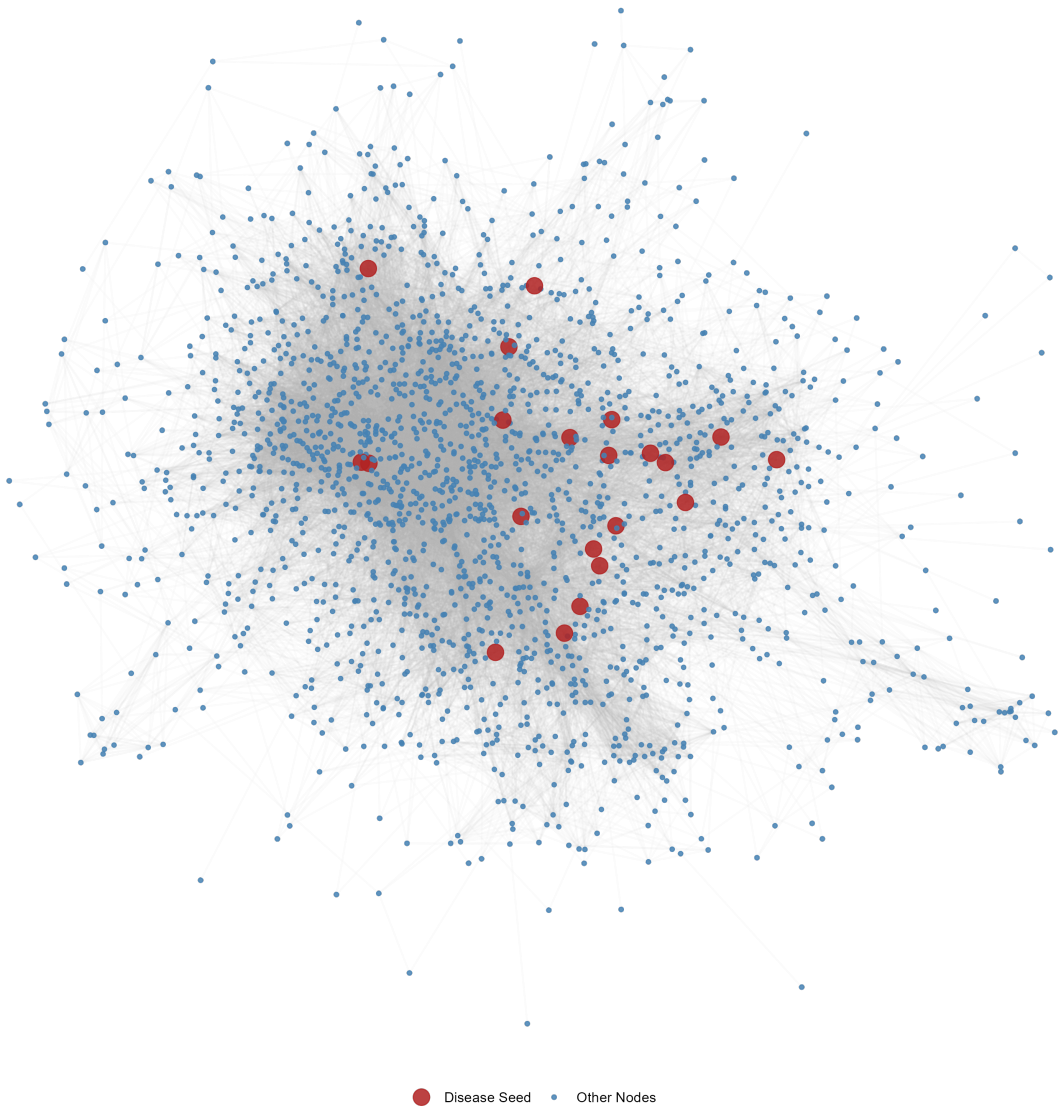

Figure 4.3: Network visualization of the ACM disease neighborhood.

4.3. Network-based disease gene prioritization

The first objective was to quantify the functional relevance of all proteins in the interactome relative to the known ACM gene set. Thus, the Random Walk with Restart (RWR) algorithm, as described in Section 4.1, was applied.

The RWR algorithm was executed using the `igraph` package in the R statistical environment. Specifically, the `page_rank` function was employed with the `algo = "prpack"` option to handle the large-scale interactome efficiently. The customized restart vector was included via the `personalized` parameter, and edge weights from STRING were incorporated to guide the random walker's transition probabilities.

Algorithm 4.2 RWR Execution via igraph

1: $\vec{p}_\infty \leftarrow$ `igraph::page_rank`$(G, \text{algo}=\text{"prpack"}, \text{damping} = 1-\gamma, \text{personalized} = \vec{p}_0, \text{weights} = E(G)\$weight)$

The restart probability γ was not set a priori to a standard value (typically $\gamma \leq 0.30$). Instead, this hyperparameter was obtained through Leave-One-Out Cross-Validation (LOO-CV) on the disease module S. This procedure identified the value of γ (actually, its corresponding damping factor $1-\gamma$) that, on average, best recovered the known disease genes when they were left out of the seed set. More specifically, the goal was to identify the damping factor that maximizes the algorithm's ability to recover known disease genes at the top of the ranking. For each gene in the disease module, the gene was hidden (removed from the seed set), and the RWR algorithm was run using the remaining seeds and the rank of the hidden gene was then recorded. This was repeated for a grid of damping values ranging from 0.15 to 0.85. The parameter minimizing the mean rank of the held-out disease genes was selected for the final analysis.

Algorithm 4.3 LOO-CV for RWR Damping Factor Optimization

Require: Interactome $G(V, E)$, Seed genes S, Damping grid $D = \{0.15, \ldots, 0.85\}$
Ensure: Optimal damping factor d^*

for all $d \in D$ **do**
  for all $g_i \in S$ **do**
    $S' \leftarrow S \setminus \{g_i\}$ {Hide one seed gene}
    $\vec{p}_0 \leftarrow$ initialize restart vector with S'
    $\vec{p}_\infty \leftarrow$ run RWR$(G, \vec{p}_0, d)$
    $rank_i \leftarrow$ rank of g_i in $\vec{p}_\infty$
  end for
  $MeanRank(d) \leftarrow \frac{1}{|S|} \sum rank_i$
end for
$d^* \leftarrow \arg\min_{d \in D} MeanRank(d)$
return d^*

This optimization yielded an optimal damping factor of 0.60, which corresponds to a restart probability of $\gamma = 1 - 0.60 = 0.40$. This value for the restart probability is a significant finding because it indicates that the algorithm performs best when the walker has a 40% chance of returning to the seed nodes at each step, which is not exactly a standard value. This suggests that the known ACM disease genes form a compact and densely interconnected module. Consequently, the most relevant disease candidates are those found in the immediate vicinity of these seed nodes, and a standard, deeper exploration of the network (e.g., with $\gamma = 0.15$) is less effective for this specific disease context.

The initial probability vector $\vec{p}_0$ was constructed as defined in Equation 4.2: a uniform probability of $1/|S|$ was assigned to each of the ACM seed genes (S) present in the LCC, and a probability of 0 was assigned to all other nodes.

The algorithm's execution yielded a steady-state probability vector $\vec{p}_\infty$ which assigns a disease relevance score to every protein in the interactome, where a higher score indicates greater network proximity and functional association with the core ACM disease module.

Moreover, since biological interactomes are scale-free networks, the node degree follows a power-law distribution ($P(k) \sim k^{-\gamma}$) (see previous Section 4.2). Different studies have demonstrated that Random Walk probabilities are strongly correlated with node degrees, causing the steady-state probability vector $\vec{p}_\infty$ to inherit this heavy-tailed distribution [22, 23]. Consequently, the raw RWR scores span several orders of magnitude, with a few

nodes accumulating high probabilities while the vast majority of nodes retain negligible values. To mitigate this extreme skewness and enable robust statistical comparison, the raw scores were log-transformed:

$$logScore_t = \log_{10}(p_{\infty,t} + \epsilon) \tag{4.5}$$

where ϵ is a small constant to avoid undefined logarithms for zero-probability nodes.

4.4. Degree-preserving permutation test

In order to avoid false positives in the context of network analysis it is necessary to account for the network topology: high-degree nodes (hubs) naturally receive higher RWR scores regardless of the specific disease seed. This is why a degree-preserving permutation test was employed to rigorously quantify the significance of each drug and to identify candidates that are more relevant to the disease than expected by chance.

Algorithm 4.4 Degree-Preserving Permutation Test

Require: Drug targets T_D, RWR log-scores $\vec{L}$, Node degrees K, Permutations $N = 10000$
Ensure: Z-score, Empirical p-value

$S_{obs} \leftarrow \text{mean}(L_i)$ for all $i \in T_D$
$B \leftarrow$ bin nodes by their degree K
for $j = 1$ **to** N **do**
  $T_{rand} \leftarrow \emptyset$
  for all $t \in T_D$ **do**
    $b \leftarrow$ identify bin of node t
    $r \leftarrow$ sample 1 node from B_b
    $T_{rand} \leftarrow T_{rand} \cup \{r\}$
  end for
  $S_{rand,j} \leftarrow \text{mean}(L_r)$ for all $r \in T_{rand}$
end for
$Z \leftarrow (S_{obs} - \text{mean}(S_{rand}))/\text{std}(S_{rand})$
$p \leftarrow (\sum[S_{rand} \geq S_{obs}] + 1)/(N + 1)$
return Z, p

For each drug D with a set of targets T_D, the observed "proximity" score was computed

as the mean log-transformed RWR score of its targets:

$$S_{obs}(D) = \frac{1}{|T_D|} \sum_{t \in T_D} logScore_t \tag{4.6}$$

To construct a null distribution, 10,000 sets of random targets for each drug were generated. It is important to note that these random sets were not extracted uniformly from the network; instead, each real target was replaced by a random node with equal degree. The nodes were grouped according to their degree and random samples were drawn from the corresponding groups. This ensures that statistical significance reflects specific proximity to the disease rather than generic network centrality.
For each drug, a Z-score was computed:

$$Z_D = \frac{S_{obs} - \mu_{rand}}{\sigma_{rand}} \tag{4.7}$$

where S_{obs} is the observed proximity score of a drug, μ_{rand} and σ_{rand} are the mean and standard deviation of the scores from the 10,000 permutations. An empirical p-value was calculated based on the fraction of permutations achieving a score higher than the observed one. The p-values were subsequently adjusted using the Benjamini-Hochberg (FDR) procedure.

4.5. Network separation analysis

To better understand the mechanism of action of the different drugs, the network separation metric s_{AB} [24] was used. It measures the topological relationship between the drug target module A and the disease module B, specifically it quantifies how distinct two sets of nodes are within a network. The metric is given by the following equation:

$$s_{AB} = \langle d_{AB} \rangle - \frac{\langle d_{AA} \rangle + \langle d_{BB} \rangle}{2} \tag{4.8}$$

where $\langle d_{AB} \rangle$ is the average shortest distance between the drug targets and the disease genes, $\langle d_{AA} \rangle$ and $\langle d_{BB} \rangle$ are the average internal distances within the drug target set and the disease set respectively.
The sign and magnitude of the separation metric s_{AB} offer valuable information regarding the potential mechanism of action of a drug. Based on the comparison of the distance between the two modules to their internal diameters, there are three distinct topological classes for the drug candidates:

- $s_{AB} < 0$ **(Overlap):** the drug targets and disease genes form overlapping neighborhoods. These drugs probably interact with the disease module directly.
- $s_{AB} \approx 0$ **(Proximity):** positive but small values show that these drugs are in the immediate proximity and can modulate the disease via signaling pathways.
- $s_{AB} > 0$ **(Separation):** the drug targets are topologically separated from the disease module.

4.6. Phenotype-driven context scoring

While the RWR scores identify drugs that target the nodes most affected by the disease module, they do not account for the broader phenotypic context. In order to refine the prioritization strategy, a "Context Score" was defined that favors drugs targeting genes known to be involved in diseases genetically similar to ACM. This strategy exploits the "guilt by association" principle, assuming that if two diseases are similar at a phenotypic level, they may respond to similar pharmacological therapies [7]. This is further supported by the "shared gene hypothesis", which suggests that if two diseases share a gene or interact within the same network neighborhood, they likely share common pathobiological mechanisms, as textually stated by Barabási et al. (2011) [6]: "if the same gene is linked to two different disease pathophenotypes, this linkage is often an indication that the two diseases have a common genetic origin".

The information on different diseases was obtained from the Phenopedia database [15], which aggregates known gene-disease associations and more than 1000 diseases were analyzed. For each disease in the database, the Jaccard Similarity Index (J) was computed with the ACM disease module (see Equation 4.9). The use of the Jaccard Similarity Index to quantify phenotypic similarity based on shared genetic architecture is well-established in network medicine [24, 25]. Specifically, diseases sharing a higher proportion of causal genes are likely to share similar molecular mechanisms, justifying the use of this metric to weight the therapeutic context.

$$J(ACM, D_x) = \frac{|G_{ACM} \cap G_{D_x}|}{|G_{ACM} \cup G_{D_x}|} \tag{4.9}$$

G_{ACM} is the set of ACM seed genes and G_{D_x} is the gene set for disease x. Diseases with $J > 0$ (sharing at least one gene) were selected as the phenotypic context, yielding a total of 844 diseases.

This metric enabled to rank the entire Phenopedia database based on genetic similarity

to ACM, and the analysis successfully identified a cluster of pathologies characterized by overlapping arrhythmic and structural phenotypes. The top 10 diseases identified are listed in Table 4.1. Notably, the ranking placed *Bundle-Branch Block* and *Brugada Syndrome* at the top, reflecting the strong electrical component of ACM. Furthermore, the inclusion of *Dilated Cardiomyopathy* and *Sudden Cardiac Death* confirms that the Jaccard-based selection correctly captures the dual nature, structural degeneration and arrhythmic risk of the disease.

Disease	**Jaccard Score**	**Overlap Genes (N)**
Bundle-Branch Block	0.240	6
Cardiomyopathy, Hypertrophic, Familial	0.140	6
Myopathies, Structural, Congenital	0.129	4
Brugada Syndrome	0.125	9
Cardiomyopathy, Dilated	0.097	15
Heart Arrest	0.091	14
Death, Sudden, Cardiac	0.086	14
Atrioventricular Block	0.086	3
Cardiomyopathy, Restrictive	0.080	2
Atrial Flutter	0.077	2

Table 4.1: Top 10 diseases genetically similar to ACM, ranked by Jaccard Similarity Index.

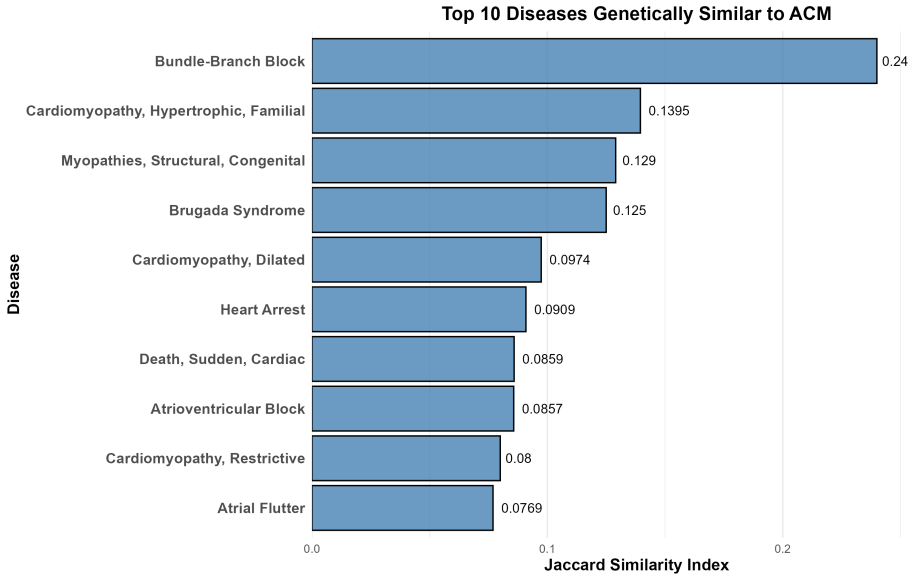

Figure 4.4: Barplot of the top 10 diseases genetically similar to ACM.

A weight w_g was then assigned to each gene g in the interactome, proportional to the similarity of the diseases in which it is involved. This weight was computed as the sum of the Jaccard Similarity Indices of all the context diseases D_x associated with that gene:

$$w_g = \frac{\sum_{D_x \in \text{Context}} (\mathbb{1}(g \in D_x) \cdot J(ACM, D_x))}{w_{max}} \tag{4.10}$$

where $\mathbb{1}(g \in D_x)$ indicates the presence of the gene in disease x, and w_{max} is a normalization factor corresponding to the maximum observed score. This method ensures that genes associated with highly similar conditions (e.g., *Brugada Syndrome*, *Bundle Branch Block*) contribute more significantly to the final context score than genes associated with loosely related conditions.

Finally, for each drug D, a Context Score $S_{context}(D)$ was computed as the average weight of its targets:

$$S_{context}(D) = \frac{1}{|T_D|} \sum_{t \in T_D} w_t \tag{4.11}$$

where T_D is the set of targets for drug D, and $|T_D|$ is the total number of targets. This ensures that the score rewards drugs with high specificity for the cardiac context.

5 | Results

The results are reported in the following chapters, moving from the global identification of drug candidates to their specific topological and functional characterization. In this chapter, the final ranking of drug candidates obtained through the integration of network propagation and phenotypic scoring is first presented, and the selection is validated against known clinical benchmarks. Next, Chapter 6 investigates why these drugs are effective by analyzing the topological properties of their targets compared to non-significant ones. Moving from individual nodes to modules, Chapter 7 applies a community detection algorithm to analyze the internal structure of the disease module, identifying specific functional pathways targeted by the drugs. Lastly, Chapter 8 provides an orthogonal validation by cross-referencing our topological predictions with transcriptomic data to identify compounds capable of reversing pathogenic gene expression signatures.

5.1. Selection of repurposing candidates

The prioritization strategy integrated the statistical robustness of the degree-preserving permutation test with the biological specificity of the phenotype-driven context score. For this reason, a dual filtering criterion was applied to ensure the selection of high-confidence candidates:

1. **Statistical significance:** an adjusted p-value < 0.05 from the permutation test, ensuring that the drug's targets did not randomly obtain high relevance scores (RWR scores).

2. **Phenotypic consistency:** a Context Score > 0. While this threshold technically acts as a strict consistency filter to exclude drugs with no relevance to the cardiac phenotypic context, a score of exactly zero is highly uncommon for candidates that have already demonstrated topological significance. Therefore, the primary biological utility of this metric is not mere exclusion, but rather establishing the final hierarchical ranking of the candidates based on their specific phenotypic context.

This procedure identified 118 promising candidates. The top 20 are presented in Table

5.1, ranked by Context Score. For the complete ranking, see Table B.1.

Drug Name	Z-Score	Adj. P-val	Context Score	s_{AB}
Lisinopril	4.80	1.82×10^{-2}	0.58	1.10
Enalaprilat	4.81	2.24×10^{-2}	0.52	1.29
Ramipril	4.77	2.88×10^{-2}	0.52	1.29
Candoxatril	4.54	4.39×10^{-2}	0.52	1.12
Ilepatril	4.51	4.02×10^{-2}	0.52	1.12
Omapatrilat	4.51	1.82×10^{-2}	0.52	1.12
Gallopamil	5.67	7.83×10^{-3}	0.38	0.60
Hydrochlorothiazide	4.01	2.88×10^{-2}	0.35	0.74
Captopril	4.33	1.32×10^{-2}	0.31	0.82
Propafenone	4.49	2.24×10^{-2}	0.26	0.47
Dyclonine	4.73	2.88×10^{-2}	0.20	0.74
Hexylcaine	4.81	3.21×10^{-2}	0.20	0.74
Azimilide	4.29	3.21×10^{-2}	0.18	0.87
Procainamide	4.35	2.88×10^{-2}	0.18	0.11
Aprindine	5.30	7.83×10^{-3}	0.18	0.31
Flecainide	7.04	7.83×10^{-3}	0.17	0.47
Vernakalant	6.02	7.83×10^{-3}	0.14	0.77
Neratinib	3.42	4.74×10^{-2}	0.13	0.59
Canertinib	3.80	2.65×10^{-2}	0.12	0.62
Carvedilol	4.78	7.83×10^{-3}	0.12	0.58

Table 5.1: Top 20 significant drug repurposing candidates for ACM, ranked by Context Score.

Furthermore, the relationship between statistical significance (Z-score), Context Score and topological location (Network Separation s_{AB}, see section 4.5) is illustrated in Figure 5.1. Interestingly, the analysis revealed that no approved drugs directly overlap with the ACM disease module ($s_{AB} < 0$); instead, the majority of the significant candidates cluster in the "Proximal" region ($s_{AB} \approx 0.2 - 1$), and so they target proteins that are neighbors of the disease genes.

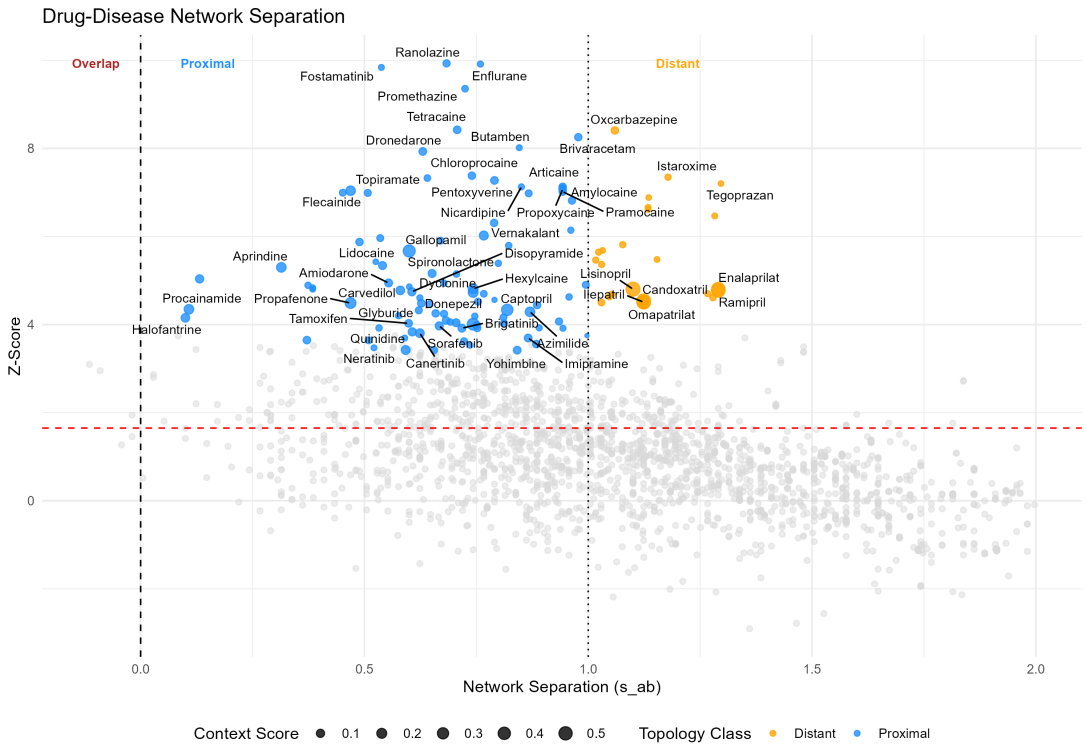

Figure 5.1: Network-based drug repurposing map. The x-axis shows the Network Separation (s_{AB}), where negative values indicate topological overlap with the disease module. The y-axis shows the proximity Z-score derived from the permutation test. The bubble size reflects the Context Score relevance.

The ranking successfully recovered widely used antiarrhythmic drugs such as *Flecainide*, *Amiodarone*, *Dronedarone* and *Propafenone* in the proximal zone (blue). The distant zone (orange) revealed a cluster of Renin-Angiotensin-Aldosterone System (RAAS), specifically Angiotensin-Converting Enzyme (ACE) inhibitors, such as *Lisinopril*, *Ramipril* and *Omapatrilat*, which are being studied for their potential antifibrotic effect [26]. Moreover, it is important to note that, although topologically separated from the area immediately surrounding the disease ($s_{AB} > 1$), these candidates reached the top positions in the final ranking because the RWR algorithm is able to identify genes relevant to the disease even over greater distances, unlike algorithms that rely solely on the minimum distance from disease genes.

As illustrated in Figure 5.2, the network visualization confirms that the targets of significant drug candidates densely cluster within the immediate topological vicinity of the ACM seed genes.

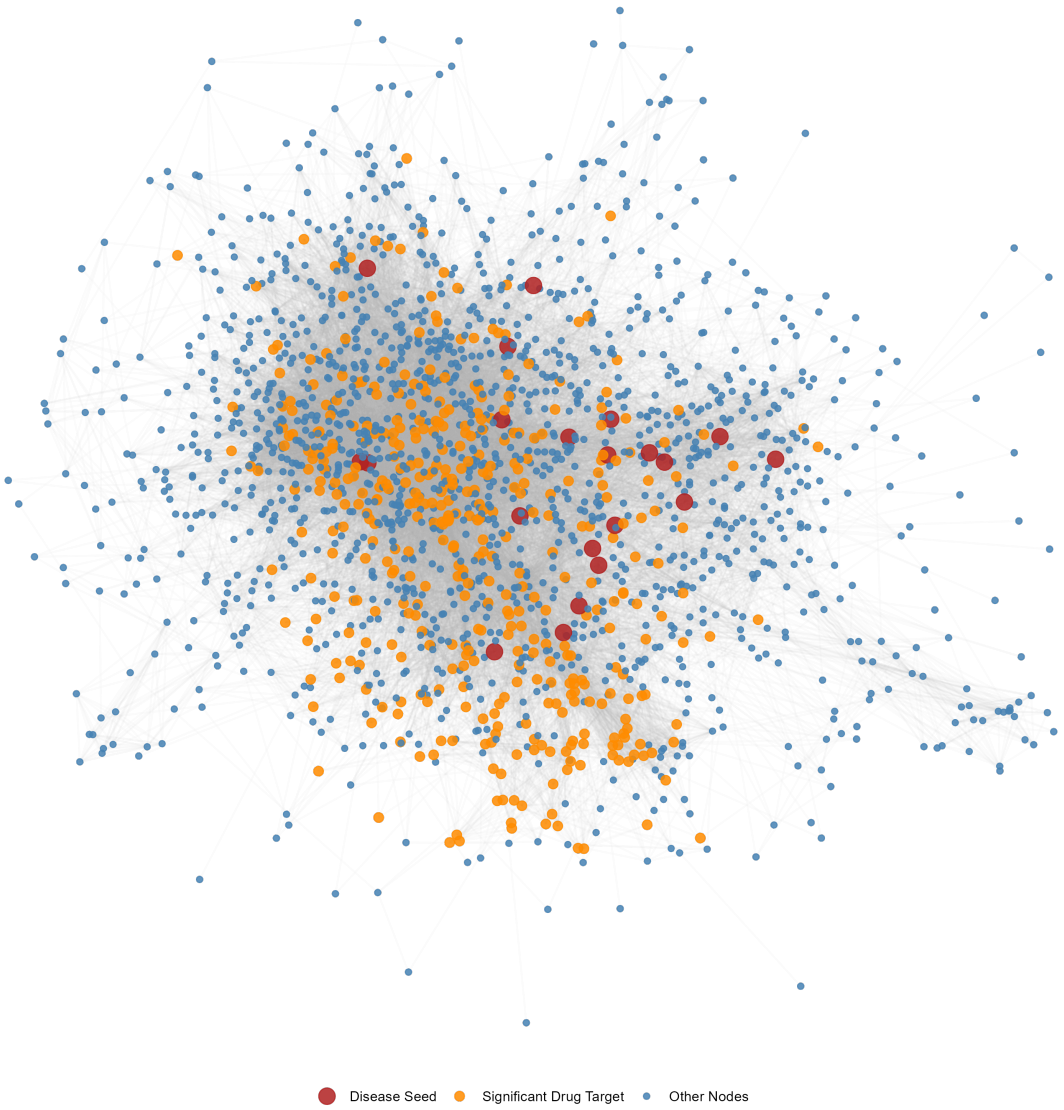

Figure 5.2: Network visualization of the ACM disease neighborhood with target genes of the 118 significant candidates.

5.2. Validation of ranking performance

To assess whether the ranking methodology is meaningful and capable of identifying relevant compounds, a validation analysis was conducted using a "gold standard" set of drugs. A list of drugs with known clinical relevance or use in treating arrhythmias and heart failure associated with arrhythmogenic cardiomyopathy was manually curated. This set included β-blockers (e.g., *Sotalol*), ACE inhibitors (e.g., *Captopril*) and other antiarrhythmics (e.g., *Flecainide*) [27–29]. This list served as the "ground truth" of positive examples against which the ranking was benchmarked. The performance of the method

was quantified using Receiver Operating Characteristic (ROC) curve analysis, which evaluates the ability of a classifier to distinguish between two classes: "known ACM drugs" (positives) and all other drugs (negatives). The computed Z-Scores were used as predictor variables, and the binary labels were set to 1 for drugs in the gold standard set and 0 for all others. The resulting ROC curve is shown in Figure 5.3.

The model achieved an Area Under the Curve (AUC) of 0.861. This result suggests that this network-based ranking performs significantly better than random chance and successfully enriches for known relevant drugs, validating the model's predictive power.

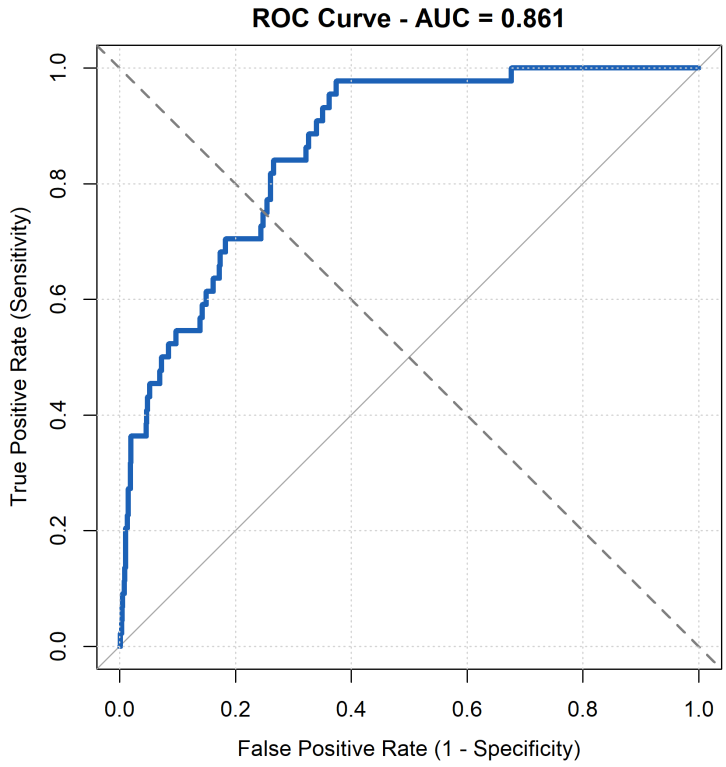

Figure 5.3: ROC curve validating the performance of the drug ranking methodology. The y-axis represents the True Positive Rate (Sensitivity), and the x-axis represents the False Positive Rate (1 - Specificity). A diagonal line (AUC = 0.5) would represent a random ranking, while a curve pushed to the top-left corner (AUC = 1.0) would represent a perfect ranking.

6 | Network-based characterization of significant targets

To better understand the properties of the identified repurposing candidates, a topological analysis was performed on the target genes of the 118 statistically significant drugs. Two main questions were addressed:

1. Are the targets of significant drugs topologically distinct from the targets of non-significant drugs? (Section 6.1)
2. Which topological features best predict the efficacy of a drug target? (Section 6.2)

6.1. Univariate topological analysis

It was first hypothesized that effective drug targets may possess distinct topological properties, such as being central "hubs" or "bridges" in the network. The targets of the significant drugs list ($pvalue_{adj} < 0.05$) were compared against a background set of targets from non-significant drugs ($pvalue_{adj} > 0.8$).

Five key metrics were analyzed:

- **Degree:** the number of direct neighbors of a protein.

$$k_i = \sum_{j \in V} A_{ij} \tag{6.1}$$

  where A is the adjacency matrix and $A_{ij} = 1$ if nodes i and j are connected, 0 otherwise.

- **Betweenness Centrality:** a measure of how often a protein lies on the shortest path between two other proteins, indicating its importance as a "bridge".

$$BC_i = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{6.2}$$

where σ_{st} is the total number of shortest paths from s to t, and $\sigma_{st}(i)$ is the number of those paths passing through i.

- **Eigenvector Centrality:** a measure of influence, where connections to high-scoring nodes contribute more to the score.

$$x_i = \frac{1}{\lambda} \sum_{j \in V} A_{ij} x_j \tag{6.3}$$

where λ is the largest eigenvalue of the adjacency matrix A.

- **Clustering Coefficient:** a measure of how interconnected a protein's immediate neighbors are, indicating its role in a densely connected local cluster.

$$C_i = \frac{2e_i}{k_i(k_i - 1)} \tag{6.4}$$

where e_i is the number of edges existing between the k_i neighbors of node i.

- **Distance to Disease:** the shortest path length from a drug-target to the nearest ACM seed gene.

$$d_{min}(i, \mathcal{S}) = \min_{s \in \mathcal{S}} (d(i, s)) \tag{6.5}$$

where $d(i, s)$ is the shortest path distance between node i and a disease gene s.

A Wilcoxon rank-sum test for univariate analysis showed the presence of different trends (Table 6.1). As expected, the most striking difference lies in the distance to disease (Figure 6.1), where significant targets are much closer to the disease module ($p < 10^{-16}$). In terms of centrality measures, significant targets exhibit higher betweenness and eigenvector centrality, and interestingly, they show a lower clustering coefficient, suggesting that they do not tend to cluster in dense, local structures; instead, the degree showed no significant difference ($p = 0.315$).

Metric	Median (Sig)	Median (Non-Sig)	P-value	Signif
Degree	30.0	31.0	3.15×10^{-1}	ns
Betweenness	1.51×10^{-4}	1.12×10^{-4}	4.46×10^{-4}	***
Eigenvector	5.60×10^{-4}	4.38×10^{-4}	2.03×10^{-3}	**
Clustering Coeff.	0.27	0.32	4.66×10^{-8}	***
Dist. to Disease	3.09	3.60	1.98×10^{-41}	***

Table 6.1: Univariate comparison of topological properties between targets of significant drugs ($pvalue_{adj} < 0.05$) and non-significant drugs ($pvalue_{adj} > 0.08$). Significance levels: ** $p < 0.01$, *** $p < 0.001$, ns = not significant.

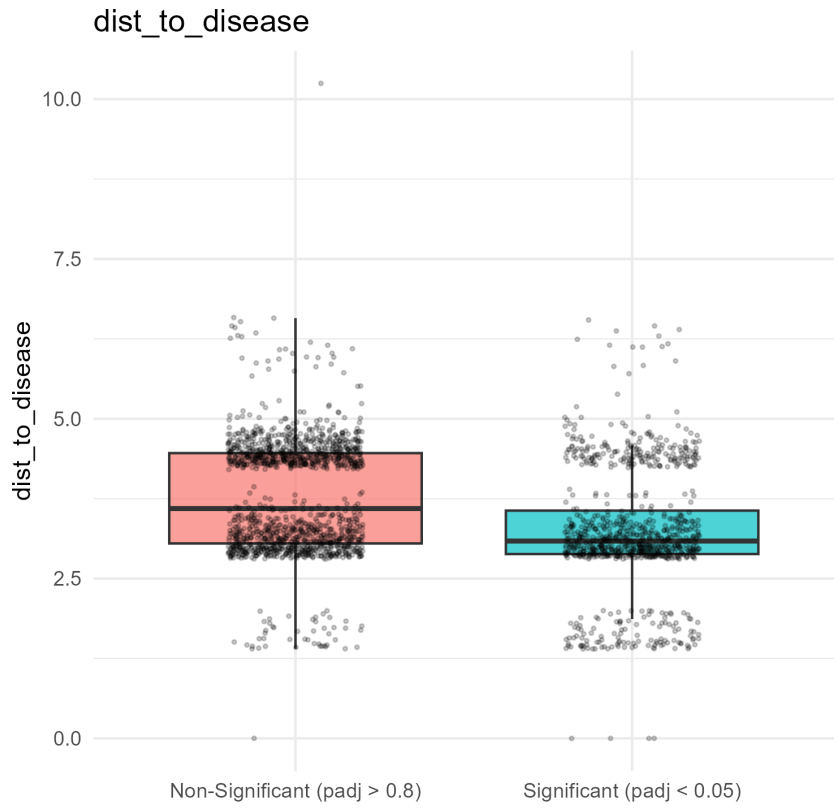

Figure 6.1: Boxplot comparison of the distance to disease genes. See appendix A to see all the boxplots.

Prior to multivariate analysis, a univariate logistic regression was performed for each metric independently (Table 6.2) to quantify the strength of the association between each topological feature and the probability of a target being significant. Formally, for each standardized topological feature X, the model is:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot X \tag{6.6}$$

where p is the probability of the target being significant, β_0 is the intercept and β_1 represents the change in the log-odds for a one-standard-deviation increase in the metric X. Before standardization, the degree was log-transformed ($x' = \ln(x+1)$) to mitigate the skewness of its power-law distribution, while all other metrics were standardized directly.

The Odds Ratios (OR) align with the Wilcoxon tests: the distance to disease is the strongest negative predictor ($OR = 0.535$, $p < 10^{-37}$), confirming that proximity to the disease module increases the likelihood of a target being significant. Notably, betweenness centrality shows a strong positive association ($OR = 1.436$), and the clustering coefficient appears to be a significant negative predictor ($OR = 0.826$), suggesting that lower clustering is associated with better targets. However, degree remains non-significant ($p = 0.23$) when analyzed in this univariate setting, implying that the raw number of connections alone does not distinguish good targets from the background.

Predictor	Odds Ratio	2.5 %	97.5 %	P-value	Signif
Dist. to Disease	0.535	0.486	0.588	4.17×10^{-38}	***
Betweenness	1.436	1.241	1.661	1.16×10^{-6}	***
Clustering	0.826	0.758	0.900	1.29×10^{-5}	***
Eigenvector	1.086	1.001	1.179	4.70×10^{-2}	*
Degree	1.051	0.968	1.142	2.36×10^{-1}	ns

Table 6.2: Univariate logistic regression results (unadjusted Odds Ratios). An $OR > 1$ indicates a positive association with target significance, while $OR < 1$ indicates a negative association. Significance levels: * $p < 0.05$, *** $p < 0.001$, ns = not significant.

6.2. Multivariate analysis

While univariate analysis identifies single metric performance, topological features are often highly correlated, for instance, nodes with a large number of connections (high degree) naturally tend to have high betweenness, potentially introducing multicollinearity into statistical models. Indeed, the Spearman correlation matrix in Figure 6.2 clearly shows these strong correlations between the metrics. Consequently, the unadjusted Odds Ratios observed in Table 6.2 might be biased by these dependencies. In order to separate the independent contribution of each metric, a multivariate logistic regression model was implemented to understand which features best predict the significance of a drug target.

The Variance Inflation Factor (VIF) was computed for all predictors to ensure the stability of the regression coefficients. All the features yielded VIF values lower than 5.0, falling below the critical threshold, confirming that, while correlations exist, they do not reach the level of severe multicollinearity that would compromise the model; consequently, the resulting Odds Ratios (see Table 6.3) can be interpreted as the unique predictive power of each topological feature, independent of the others.

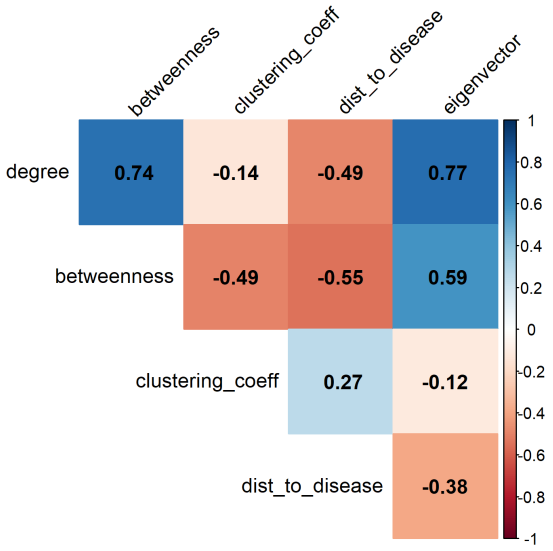

Figure 6.2: Spearman correlation matrix of network metrics. Strong correlations can be observed between Degree vs Betweenness, and Degree vs Eigenvector centrality

The multivariate logistic regression predicts the probability of a target being "significant" based on the scaled topological features. The model is defined as:

$$\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{deg} + \beta_2 \cdot \text{betw} + \beta_3 \cdot \text{eigen} + \beta_4 \cdot \text{clust} + \beta_5 \cdot \text{dist} \tag{6.7}$$

The results of the multivariate analysis (Table 6.3) clarify these relationships, isolating the independent effects masked in the univariate tests:

- **Distance to disease ($OR = 0.448$):** this remains the strongest predictor. The Odds Ratio < 1 confirms that increasing the distance drastically reduces the probability of a target being relevant, so proximity to the disease module is the primary driver of repurposing potential.

- **The centrality paradox (Degree vs. Betweenness/Eigenvector):** the multivariate model reveals that the degree is a negative predictor ($OR = 0.589$, $p < 10^{-16}$) when controlling for other centralities, and it demonstrates that the effect observed in the univariate case was driven by confounding factors from other metrics with which the degree is highly correlated (Figure 6.2). This finding is fully consistent with the degree-preserving permutation strategy used for candidate prioritization (Section 4.4): by strictly controlling for node degree during the statistical validation, the method systematically penalized generic hub-genes with many connections but low specific relevance to the disease module (hub bias). Conversely, betweenness ($OR = 1.454$) and eigenvector centralities ($OR = 1.202$) are significant positive predictors. This suggests that "bridging" capabilities and connections to other influential nodes are the true drivers of therapeutic effect, rather than the raw number of connections.
- **Clustering coefficient:** once controlled for centrality and distance, local clustering becomes non-significant ($p = 0.769$), therefore the difference observed in the univariate test was likely a confounding effect of the other metrics.

Predictor	Odds Ratio	2.5 %	97.5 %	Signif
(Intercept)	0.622	0.569	0.680	***
Degree (Scaled)	0.589	0.520	0.667	***
Betweenness (Scaled)	1.454	1.209	1.750	***
Dist. to Disease (Scaled)	0.448	0.399	0.503	***
Eigenvector (Scaled)	1.202	1.098	1.314	***
Clustering (Scaled)	0.986	0.895	1.086	ns

Table 6.3: Multivariate logistic regression results (Odds Ratios). An $OR > 1$ indicates a positive association with target significance, while $OR < 1$ indicates a negative association. Significance levels: ** $p < 0.01$, *** $p < 0.001$, ns = not significant.

These results suggest a specific topological signature for effective ACM drug targets: the ideal target is not a "super-hub" with the highest number of raw connections (degree), as high-degree nodes might be too generic or "promiscuous" within the cell. Moreover, as noted by Yu et al. [30], such high-degree nodes often correspond to "party hubs", i.e., hubs with low betweenness, that dominate static complexes, making them less specific as regulatory control points. In contrast, the authors identified bottlenecks as the key determinants in signal transduction pathways. Our analysis indeed favors nodes with "strategic centrality": proteins that act as bridges (betweenness) or are connected to

other influential nodes (eigenvector), provided they are located in the vicinity of the disease module. This aligns with the definition of network bottlenecks as nodes that control the information flow between functional clusters [30], supporting a therapeutic strategy that targets these dynamic connectors rather than disrupting massive protein complexes.

7 | Community detection

Diseases are often caused by the dysfunction of specific biological modules or communities within the interactome, rather than by scattered genes. Therefore, a drug is likely to be effective if its targets are not just "near" the disease genes, but if they are strategically positioned to modulate the same communities that are perturbed by the disease [31]. In particular, in this chapter the analysis focuses on the expanded disease module to investigate its internal heterogeneity, and it will be shown how the RWR algorithm successfully identified as "relevant" genes that are part of many biological pathways related to ACM. The following approach formalizes these concepts by mapping the functional landscape of the disease module using the Louvain algorithm.

7.1. Louvain algorithm

The Louvain method is an efficient, two-step, greedy optimization process for community detection in large networks [32, 33]. Its goal is to partition a network into a set of communities by optimizing a metric called modularity, which measures how dense the connections are within communities compared to between communities. A high modularity score for a partition indicates dense connections within communities and sparse connections between them. The modularity Q is defined as:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{7.1}$$

where:

- A_{ij} is the element of the adjacency matrix, equal to 1 if nodes i and j are connected, and 0 otherwise;
- k_i and k_j represent the degrees of nodes i and j, respectively;
- $m = \frac{1}{2} \sum_{i,j} A_{ij}$ is the total number of edges in the network;
- $\delta(c_i, c_j)$ is the Kronecker delta, equal to 1 if nodes i and j belong to the same

community, and 0 otherwise.

The Louvain algorithm operates through the following steps (see Figure 7.1):

1. **Step 1 - Local optimization**:
   At the beginning, every single node in the network forms one community. The algorithm goes through each node, one by one, and makes a "greedy" decision: it considers moving the current node out of its current community and into one of its neighboring communities. For each possible move, it calculates the change in modularity [33]:

$$\Delta Q = \left[\frac{\Sigma_{in} + k_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m}\right)^2\right] - \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right] \tag{7.2}$$

   where Σ_{in} is the sum of the weights of the links inside C, Σ_{tot} is the sum of the weights of the links incident to nodes in C, k_i is the sum of the weights of the links connected to node i (degree), and $k_{i,in}$ is the sum of the weights of the links from i to nodes in C.
   The algorithm moves the node into the neighboring community that results in the greatest positive increase in modularity, i.e. only if $\Delta Q > 0$. If no move increases modularity, the node stays where it is.
   This process is repeated for all nodes until no single node movement can further improve the overall modularity of the network.

2. **Step 2 - Network aggregation**:
   The communities found in Step 1 are now treated as single "super-nodes." A new, smaller network is built where the nodes are these super-nodes. The weight of the edge between two super-nodes is the sum of the weights of all the edges that connected the original nodes between those two communities.
   The algorithm now repeats Step 1 on this new, aggregated network.

This two-step process is repeated until the modularity can no longer be increased, resulting in a final, stable community structure.

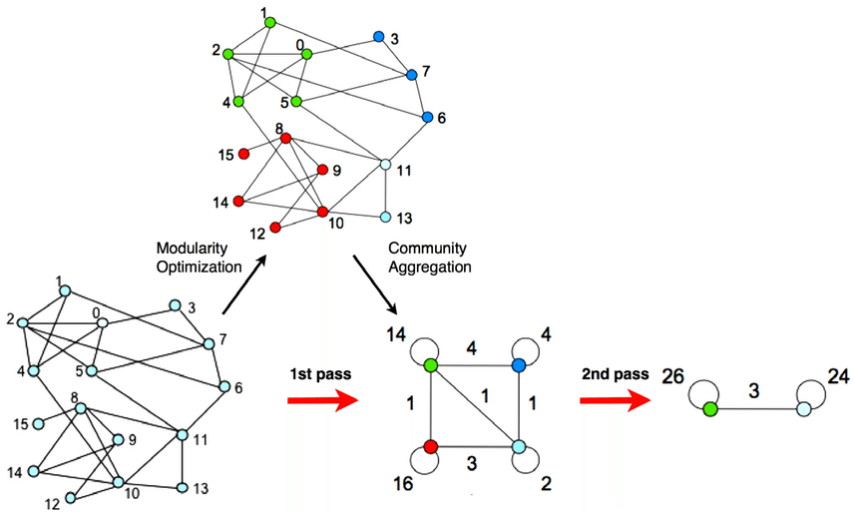

Figure 7.1: Louvain Algorithm (from [32])

Louvain algorithm is computationally very fast, making it suitable for large networks like the human interactome, and it produces high-quality and biologically meaningful communities.

7.2. Disease module and characterization

It is important to note that, in the previous chapters, the terms "disease gene" or "disease module" S referred exclusively to the seed nodes, i.e., the causative genes of ACM shown in Table 3.1. From now on, the concept of disease module will be expanded to include all genes that have obtained a high score from the Random Walk with Restart.

7.2.1. Expanded disease module

The boundaries of the new disease module were established using the "elbow method" by plotting the ranked RWR scores. A cutoff was chosen at the point of maximum curvature in the distribution, which separates the few highly ranked nodes, that retain most of the random walker's probability, from the long tail of low-probability nodes. This procedure identified a final ACM Disease Module consisting of 65 nodes.

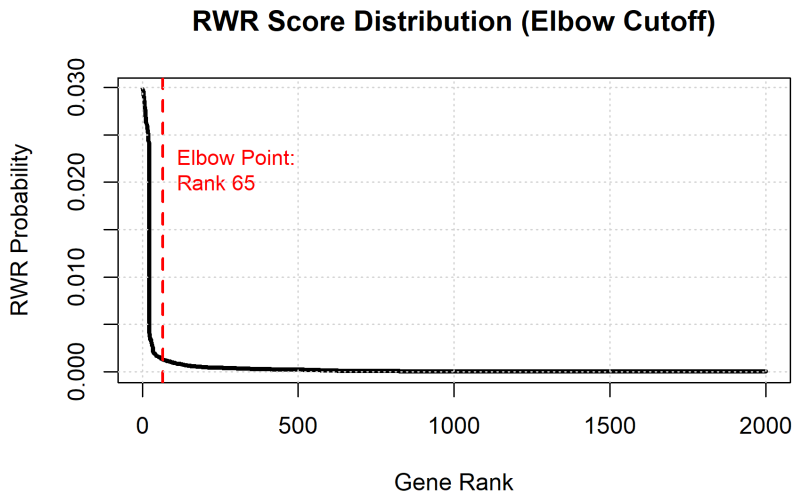

Figure 7.2: Ranked RWR scores for the definition of the ACM Disease Module. Cutoff at rank 65 (red line).

Crucially, a cross-reference analysis reveals that a subset of the 118 significant drug candidates identified in Section 5.1 directly targets nodes within this expanded module of 65 genes. Specifically, 9 of the ranked drugs act directly upon these high-score genes: *Carbamazepine*, *Dronedarone*, *Topiramate*, *Tetracaine*, *Carvedilol*, *Gallopamil*, *Ramipril*, *Nicardipine* and *Nifedipine*.

7.2.2. Functional partitioning via Louvain

Although the RWR algorithm provides a global ranking of all proteins in the interactome, a community analysis on the entire interactome could weaken the disease-specific biological signal. This is why the following analysis focuses on the genes most affected by the disease in the immediate vicinity of the seed nodes: the new expanded disease module.

The community structure of the disease module was investigated to better understand its biological heterogeneity using Metascape [34], in order to know whether it is composed of distinct "functional units" (e.g., a cluster for cell adhesion, another for muscle contraction, etc.).
First, the Louvain algorithm, executed using the `cluster_louvain` function from the `igraph` R package, was applied to the induced subgraph of 65 nodes. It optimized modularity by iteratively grouping nodes to maximize the density of edges within communities relative to edges between communities.
This method partitioned the ACM disease module into 4 distinct communities, which

were subsequently subjected to functional enrichment analysis using Metascape [34] to determine their specific biological roles in the pathogenesis of ACM.

Below the detailed characterization of one community, labeled as Module 1, is reported as a representative case study (the results for the other three modules are reported in appendix A). Figure 7.3 illustrates the structural and functional composition of this community made of 15 genes. It displays the protein-protein interaction subgraph, which highlights the dense local connectivity among these nodes, and also the specific pathways and processes performed by these genes.

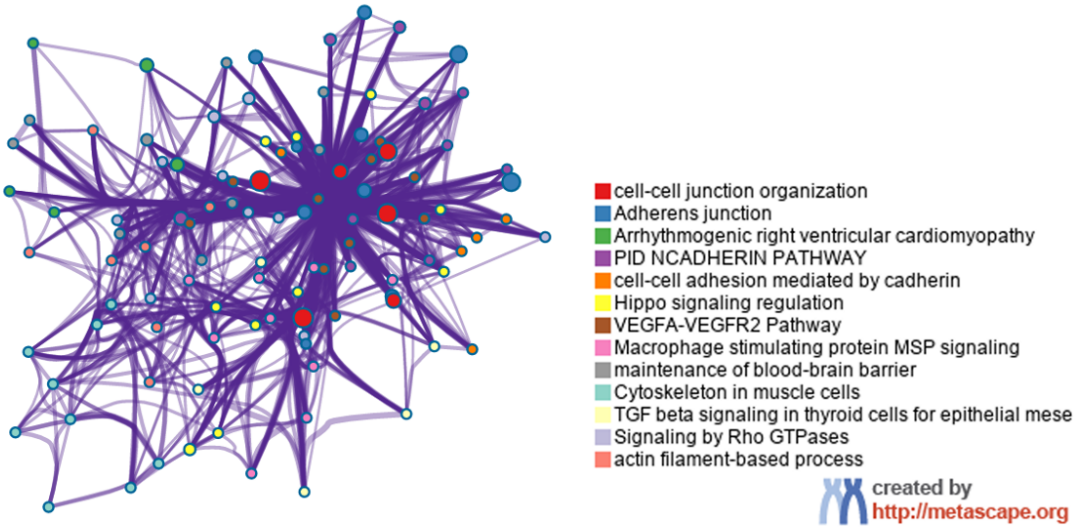

Figure 7.3: Network of enriched terms: colored by cluster ID, where nodes that share the same cluster ID are typically close to each other. [34]

This visualization also provides an overview of the mechanisms of the disease. Each node in the graph represents an enriched biological term (e.g., a pathway) and the edges connect terms that share a significant number of genes. The clustering of these terms (indicated by color in Figure 7.3) shows that the disease module does not affect isolated processes, but rather alters coherent biological pathways, in particular, the dense clustering of terms related to "cell-cell junction organization" and "adherent junctions" (red/blue cluster) confirms that the primary defect lies in structural integrity. However, the connection of these structural nodes to signaling pathways such as "Hippo" and "TGF-β" (yellow clusters) provides the functional link that explains how mechanical detachment translates into the fibro-adipose remodeling observed in patients.

Figure 7.4 and Figure 7.5 provide a hierarchy of these biological processes based on their statistical significance ($-\log_{10}(Pvalue)$).

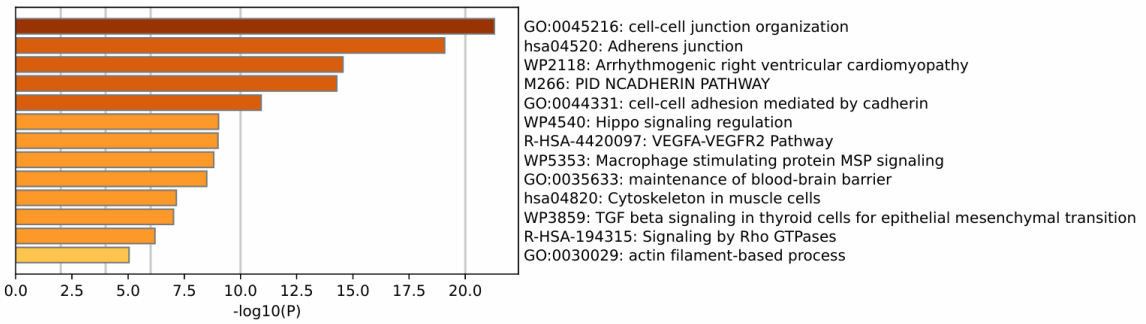

Figure 7.4: Bar graph of top enriched terms across input gene lists. [34]

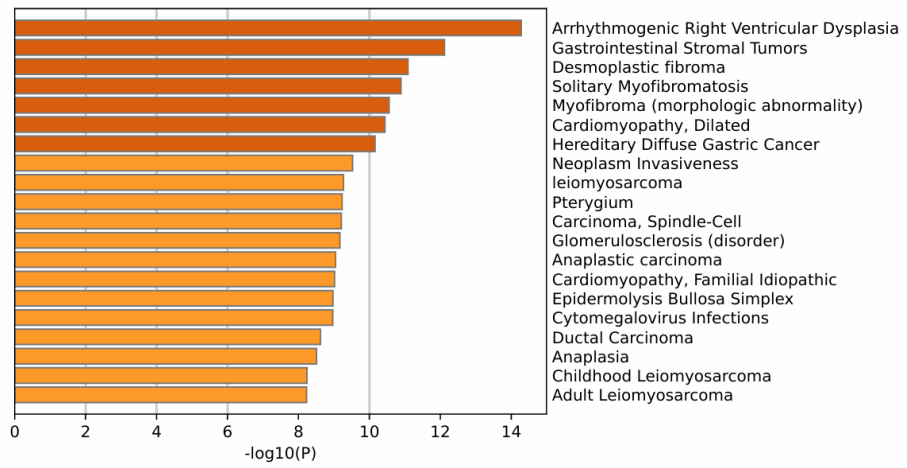

Figure 7.5: Summary of enrichment analysis in DisGeNET. [34]

In Figure 7.4, the terms "cell-cell junction organization" and "adherens junctions" appear as the most statistically significant and confirm that the structural impairment is the primary driver of the disease module.

Furthermore, the DisGeNET enrichment in Figure 7.5 serves as an unbiased validation of the method. The fact that "Arrhythmogenic Right Ventricular Dysplasia" emerges as the single most significant disease association indicates that the RWR algorithm effectively isolated the specific pathogenic module from the global interactome. Interestingly, the retrieval of "Dilated Cardiomyopathy" and "Myofibromatosis" reflects the biventricular and fibro-fatty nature of the ACM phenotype.

8 | Transcriptomic validation analysis

In order to complete the network-based prioritization strategy, a transcriptomic validation was performed. While the network approach relies on static protein-protein interactions, transcriptomic analysis captures the dynamic changes in gene expression driven by the disease state. This orthogonal approach aims to identify drugs capable of reversing the global gene expression signature associated with ACM, a concept known as "signature reversion".

8.1. The Connectivity Map (CMap) and L1000 platform

The Connectivity Map (CMap) is a large-scale catalog of cellular responses to chemical, genetic and disease perturbations [35]. The core principle of CMap is "pattern matching": by comparing the gene expression signature of a disease, i.e., a list of up- and down-regulated genes, against the signatures induced by thousands of small molecules, it is possible to identify compounds that induce an opposing transcriptional profile.
A drug is considered a potential therapeutic candidate if its induced gene expression profile is negatively correlated with the disease signature, i.e., it down-regulates genes that are up-regulated in the disease and vice versa. This is quantified by a "Connectivity Score" (or τ-score), ranging from -100 to +100. A score approaching -100 indicates a strong potential to reverse the disease phenotype [35]. In this work, drugs with a score lower than -80 were selected as plausible candidates for reversing the disease signature.
To generate this massive library of perturbations, the Broad Institute developed the L1000 high-throughput gene expression assay [35]. Unlike traditional microarrays or RNA-Seq that measure the entire transcriptome, L1000 measures the expression of 978 "landmark genes" that capture the majority of cellular transcriptional variance. The expression of the remaining non-landmark genes is computationally inferred based on established correlations. This cost-effective approach has allowed the expansion of CMap (via the

LINCS project) to over 1.3 million profiles, covering several cell lines and perturbations [35].

8.2. ACM disease signatures

Five distinct transcriptomic datasets were selected from the literature to ensure a robust validation that captures the heterogeneity of ACM. These datasets represent different stages of the disease (early vs. end-stage), different biological materials (heart tissue vs. stromal cells) and different detection technologies (RNA-Seq vs. Microarray).

8.2.1. Dataset 1 (Rouhi et al. 2022)

The first signature was obtained from the study of Rouhi et al. (2022) [36]. This dataset is particularly valuable as it represents the early, arrhythmic phase of the disease, prior to the onset of clear heart failure.

- **Source:** Right Ventricular (RV) endomyocardial biopsies from patients with truncating mutations in Desmoplakin (DSP).
- **Methodology:** High-throughput RNA-Sequencing (RNA-Seq).
- **Key pathological features:** The study identified suppression of the canonical Wnt and Hippo signaling pathways and activation of the EP300/TP53 pathway. These changes were linked to altered intercalated disc structures and were validated in autopsy samples.

Top up-regulated and down-regulated genes from this study were extracted from GEO repository GSE156869 [37].

8.2.2. Dataset 2 (Rainer et al. 2018)

The second dataset focused on the non-myocyte fraction of the heart, recognizing that ACM is not solely a cardiomyocyte disease. Rainer et al. (2018) characterized the transcriptome of Cardiac Stromal Cells (CStCs), which are known to differentiate into adipocytes in ACM [38].

- **Source:** Patient-specific CStCs derived from endomyocardial biopsies (mostly PKP2 mutation carriers).
- **Methodology:** RNA-Seq for coding genes and TaqMan Arrays for miRNAs.
- **Key pathological features:** The study identified deregulation in genes involved

in extracellular matrix organization, cell adhesion and lipid transport. Notably, the authors found significant upregulation of genes like LUM and DCN (decorin), highlighting the fibrotic drive of stromal cells.

This dataset provides a stromal/fibro-adipogenic signature, which allows the identification of drugs targeting the structural remodeling aspect of ACM. Gene expression results were directly downloaded from the additional files of the article [38].

8.2.3. Dataset 3 (Gaertner et al. 2012)

The third dataset, from Gaertner et al. (2012) [39], reflects the terminal stage of the disease.

- **Source:** Explanted hearts from patients undergoing transplantation (NYHA IV). Samples were taken from both Right (RV) and Left (LV) ventricles.
- **Methodology:** Affymetrix Microarrays.
- **Key Pathological Features:** The study revealed that end-stage ACM is a biventricular disease with a transcriptome clearly distinct from Dilated Cardiomyopathy (DCM). Interestingly, classical fibrosis genes were not specifically regulated compared to controls in this terminal stage, suggesting an end phase of the disease.

This dataset yields a late-stage signature, useful for identifying drugs that might be effective even in advanced disease stages. The datasets (for both right and left ventricles) were downloaded from the additional material of the article [39].

8.2.4. Dataset 4 (De Bortoli et al., 2023)

This dataset provides a unique perspective on the genetic determinants of the disease phenotype, distinguishing between pathogenic manifestation and simple mutation carriage [40].

- **Source:** Human induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CMs) generated from a single familial cohort carrying the PKP2 exon 4 deletion.
- **Methodology:** Transcriptional profiling was performed using high-throughput mRNA-Sequencing.
- **Key Pathological Features:** The study revealed a specific "penetrance signature" characterized by the upregulation of adipogenic modulators (e.g., SFRP1) and the downregulation of lipid transport genes (e.g., CPT1C), suggesting that an intrinsic

metabolic inability to handle lipids distinguishes patients from healthy carriers.

This dataset can be found in the GEO repository GSE222793 [37].

8.2.5. Dataset 5 (Lippi et al., 2023)

The dataset from Lippi et al. (2023) [41] captures the contribution of the non-myocyte fraction and the metabolic alterations in ACM.

- **Source:** Cardiac Mesenchymal Stromal Cells (CMSCs) isolated from endomyocardial biopsies of ACM patients and healthy controls.
- **Methodology:** High-throughput RNA-Sequencing combined with methylome analysis.
- **Key Pathological Features:** This study highlighted profound alterations in mitochondrial homeostasis and chromatin organization, pathways previously underappreciated in ACM. Specifically, ACM-CMSCs exhibited an upregulation of genes related to mitochondrial respiration (e.g., COX7C) and Epithelial-to-Mesenchymal Transition (EMT) (e.g., S100A11), accompanied by a downregulation of cell cycle progression genes (e.g., CCND2), delineating a pro-fibrotic and metabolically hyperactive stromal phenotype.

The dataset was downloaded from GEO repository GSE233780 [37].

8.3. Cross-Validation and consensus scoring

For each of the five disease signatures described above, a query was submitted to the CMap/L1000 database (via the CLUE.io platform [35]) to retrieve a list of "reverter" drugs (Connectivity Score < -80). First, for each dataset, significant differentially expressed genes (DEGs) were selected using a threshold of adjusted p-value < 0.05 and $|\log_2 \text{Fold Change}| > 1$. Fold Change is the ratio that measures how much a gene's expression level has changed between two conditions (e.g., a "treated" sample versus a "control" sample). It is measured as:

$$FC = \frac{\text{Expression}_{\text{Treated}}}{\text{Expression}_{\text{Control}}} \tag{8.1}$$

If $FC > 1$, the gene is upregulated (more expressed in the treated sample), if $FC < 1$, the gene is downregulated, and if $FC = 1$, there is no change.
The $\log_2$ FC is the most common metric in bioinformatics, and it is used because the

raw Fold Changes are asymmetric. For instance, an increase from 10 to 20 is a FC of 2, while a decrease from 20 to 10 is a FC of 0.5. This makes decreases harder to visualize and analyze statistically, therefore, the "$\log_2$" transformation fixes this by making the scale symmetric around zero. Specifically, doubling expression ($FC = 2$) becomes $+1$, halving expression ($FC = 0.5$) becomes -1, and now $\log_2 \text{FC} > 0$ indicates upregulation, $\log_2 \text{FC} < 0$ indicates downregulation and if $\log_2 \text{FC} = 0$ there's no difference.

Furthermore, a "consensus scoring system" was developed to integrate the network-based findings with this transcriptomic evidence. The 118 significant drug candidates identified by the RWR algorithm (Section 5.1) were cross-referenced against the CMap output lists, as shown below in Table 8.1.

Dataset 1	**Dataset 2**	**Dataset 3**	**Dataset 4**	**Dataset 5**
Carbamazepine	Gallopamil	Dronedarone	Carvedilol	Spironolactone
Isradipine	Disopyramide	Tetracaine	Disopyramide	Ponatinib
Nimodipine	Lacidipine	Carbamazepine	Isradipine	Amiloride
Topiramate	Dalfampridine	Ritodrine	Verapamil	Tivozanib
Fostamatinib	Digoxin	Isradipine	Topiramate	
Sorafenib	Canertinib	Nifedipine	Lacidipine	
Canertinib	Flunarizine	Zonisamide	Istaroxime	
Imipramine		Trimebutine	Nicardipine	
		Nitrendipine	Fostamatinib	
		Digoxin	Amiodarone	
		Regorafenib	Amitriptyline	
		Phenytoin	Tivozanib	
		Ramipril		
		Nintedanib		
		Lamotrigine		
		Imipramine		
		Flunarizine		

Table 8.1: Intersection of RWR prioritized drugs with CMap revertant signatures across the five datasets.

A discrete score ($S_{consensus} \in \{0, 1, 2, 3, 4, 5\}$) was then assigned to each candidate based on the number of transcriptomic signatures it successfully reversed:

$$S_{consensus}(D) = \sum_{i=1}^{5} \mathbb{I}(D \in \text{CMap}_i) \tag{8.2}$$

where CMap_i is the set of drugs with a significant negative connectivity score for Dataset i and $\mathbb{I}$ is the indicator function (1 if true, 0 if false). A higher score implies that the candidate drug is more robust: it is not only topologically significant but it is also able to reverse the disease signature. A score equal to 0 means that the drug is topologically relevant but does not show transcriptomic reversion in the analyzed datasets. Therefore, this multi-layer filtering ensures that the final proposed candidates are not only located in the network neighborhood but are also functionally capable of modulating the global gene expression perturbations of ACM.

Applying this consensus scoring system to the few candidates in Table 8.1 resulted in the stratification of drugs based on their ability to reverse pathogenic gene expression signatures across the five datasets. Although no single compound achieved a score of 4 or 5, a predictable result given the profound biological heterogeneity of the datasets, a subset of highly reliable candidates emerged with a consensus score of 3 or 2, indicating robust efficacy in several distinct biological contexts.

The top-ranking drugs validated by this transcriptomic screen include:

Drug Candidate	Score	D1	D2	D3	D4	D5
High Confidence (Score ≥ 2)						
Isradipine	**3**	✓		✓	✓	
Canertinib	2	✓	✓			
Carbamazepine	2	✓		✓		
Digoxin	2		✓	✓		
Disopyramide	2		✓		✓	
Flunarizine	2		✓	✓		
Fostamatinib	2	✓			✓	
Imipramine	2	✓		✓		
Lacidipine	2		✓		✓	
Tivozanib	2				✓	✓
Topiramate	2	✓			✓	
Validation Hits (Score 1)						
Amiodarone	1				✓	
Carvedilol	1				✓	
Ramipril	1			✓		

Drug Candidate	Score	D1	D2	D3	D4	D5
Spironolactone	1					✓
Verapamil	1				✓	

Table 8.2: Consensus scoring of prioritized candidates. The table lists all candidates with Score ≥ 2 and selected clinically relevant candidates with Score 1. Checkmarks indicate successful signature reversion. For the complete table see Table B.3 in appendix.

8.4. Transcriptomic results

The consensus scoring system highlighted *Isradipine* ($S = 3$) as the most robust novel candidate by integrating the results from the five datasets. Additionally, ten compounds achieved a score of 2, showing recurrence across independent datasets. The screening also identified several drugs currently used to treat ACM, primarily in the subset of candidates with $S = 1$. The retrieval of *Amiodarone*, *Carvedilol*, *Ramipril* and *Spironolactone* serves as a strong biological validation of the proposed network-based algorithm. Indeed, *Carvedilol* (beta-blocker) and *Amiodarone* are fundamental in ACM therapy to manage ventricular arrhythmias and reduce adrenergic stress [27, 42]. Their presence in the output confirms that the RWR algorithm correctly prioritized the gene modules related to excitability and adrenergic signaling. *Spironolactone* (Mineralocorticoid Receptor Antagonist) and *Ramipril* (ACE inhibitor) are widely used to manage heart failure and limit myocardial fibrosis [43, 44], reinforcing the algorithm's ability to detect anti-fibrotic mechanisms. In addition to validation with known drugs, the analysis revealed *Isradipine* and *Fostamatinib* as promising new candidates. *Isradipine* ($S = 3$) is a calcium antagonist and acts on the intracellular overload of Ca^{2+} typical of ACM cardiomyocytes, potentially reducing the trigger factor of DAD (Delayed Afterdepolarizations)-mediated arrhythmias [45, 46]. *Fostamatinib* ($S = 2$), on the other hand, is a Syk inhibitor and offers a targeted approach to the structural remodeling aspect of the disease, particularly inflammation and fibro-adipose replacement, which standard antiarrhythmic drugs fail to reverse [47, 48].

9 | Conclusions and further developments

The work presented in this thesis addresses the urgent clinical need for new therapeutic strategies for arrhythmogenic cardiomyopathy (ACM), a complex disease for which the conventional "reductionist" drug research has proven ineffective. This study developed and validated a computational approach for drug repurposing based on the principles of network medicine, shifting the focus from the traditional single-target intervention to a broader systems-level perspective.

The central hypothesis of this work, namely that the effects of genetic alterations in ACM are not confined but rather propagate through specific modules of the human interactome, was confirmed by the analysis. The application of the Random Walk with Restart (RWR) algorithm enabled to prioritize proteins that, although not always direct neighbors of the disease genes, are functionally relevant for the disease module. The proposed prioritization pipeline based on network propagation, degree-preserving permutation test and phenotype-based context scoring identified 118 statistically significant drug candidates. The accuracy of this ranking was validated with a ROC-AUC analysis, which achieved an AUC of 0.861 and successfully recovered known therapeutic drugs such as antiarrhythmics (*Amiodarone*), beta-blockers (*Carvedilol*) and ACE inhibitors (*Ramipril*). Moreover, the integration of CMap/L1000 information provided a set of candidates capable of reversing the pathological gene expression signatures from five independent patient datasets. Notably, the topological characterization of target genes revealed that the efficacy of a drug is not driven by targeting high degree nodes (hubs), but rather by "network bridges" with high betweenness and eigenvector centrality.

The flexibility of the proposed approach, combined with the innovative nature of the topic, opens the way for future research and methodological improvements. Since the current framework is based on the integration of static protein-protein interactions (the interactome) and dynamic gene expression data (transcriptomics), a significant refinement would be the integration of proteomic data, that would enable to confirm that the drug

targets are actually expressed and active also at the protein level in the target tissue. In this way, the edges of the interactome could be weighted dynamically according to transcriptomic expression levels of genes, or proteomic abundance and phosphorylation states observed in patient samples. Furthermore, the computational model itself could be also refined through the use of neural networks. Specifically, deep learning approaches are able to learn nonlinear representations of nodes and capture latent topological features, potentially revealing complex associations between drugs and diseases that linear methods might overlook.

The ultimate goal is to translate the *in silico* predictions into clinical reality, indeed, the most crucial step is the biological validation of the 118 identified candidates. This thesis, in fact, provides a ranked list of drugs that actually forms the basis for new experimental studies at Centro Cardiologico Monzino (IRCCS). Moving from *in silico* to *in vitro* models, the therapeutic efficacy of the top candidates will be tested, and these experiments will assess whether the drugs can restore the cellular phenotype.
Beyond the necessary biological validation of the main candidates, the pipeline has also the potential to evolve into a patient-specific diagnostic tool. By mapping the specific mutations and transcriptomic profile of an individual patient onto the network, this framework could directly predict the most effective targeted therapy for that individual, thus moving from a generalized therapeutic approach to precision medicine.

In conclusion, a key strength of the methodology is its adaptability. The developed pipeline is not specific to arrhythmogenic cardiomyopathy, but rather a generalizable framework for network-based drug repurposing. This means that by simply replacing the input files, such as the list of seed genes, the drug list and the interactome, the entire algorithm can be used to identify drug candidates for other complex diseases. In order to facilitate this generalization and ensure reproducibility, the code implemented in this thesis for the drug candidates search has been documented and uploaded to GitHub `https://github.com/AuroraVido/NetMedWalkeR`. This open-source repository allows the scientific community to use the method for different input data or even improve the algorithm to better tackle the challenges of modern drug discovery.

Bibliography

[1] Judge DP Corrado D, Basso C. Arrhythmogenic cardiomyopathy. *Circulation research*, pages 784–802, 2017.

[2] Corrado D Thiene G Basso C, Bauce B. Pathophysiology of arrhythmogenic cardiomyopathy. *Nat Rev Cardiol*, 2011.

[3] Riggs E. R. Buchanan A. H. Ceyhan-Birsoy O. DiStefano M. Dwight S. S. Goldstein J. Ghosh R. Seifert B. A. Sneddon T. P. Wright M. W. Milko L. V. Cherry J. M. Giovanni M. A. Murray M. F. O'Daniel J. M. Ramos E. M. Santani A. B. Scott A. F. Plon S. E. ... Berg J. S. Strande, N. T. Evaluating the clinical validity of gene-disease associations: An evidence-based framework developed by the clinical genome resource. *American journal of human genetics*, 100(6):895–906, 2017.

[4] Cynthia A. James, Jan D.H. Jongbloed, Ray E. Hershberger, Ana Morales, Daniel P. Judge, Petros Syrris, Kalliopi Pilichou, Argelia Medeiros Domingo, Brittney Murray, Julia Cadrin-Tourigny, Ronald Lekanne Deprez, Rudy Celeghin, Alexandros Protonotarios, Babken Asatryan, Emily Brown, Elizabeth Jordan, Jennifer McGlaughon, Courtney Thaxton, C. Lisa Kurtz, and J. Peter van Tintelen. International evidence based reappraisal of genes associated with arrhythmogenic right ventricular cardiomyopathy using the clinical genome resource framework. *Circulation: Genomic and Precision Medicine*, 14(3):e003273, 2021.

[5] Chitipiralla S. Brown G. R. Chen C.-Gu B. Hart J. Hoffman D. Jang W. Kaur K. Liu C. Lyoshin V. Maddipatla Z. Maiti R. Mitchell J. O'Leary N. Riley G. R. Shi W. Zhou G. Schneider V. Maglott D. ... Kattman B. L. Landrum, M. J. Clinvar: improvements to accessing data. *Nucleic acids research*, 2020.

[6] Loscalzo J Barabási A-L, Gulbahce N. Network medicine: a network-based approach to human disease. *Nat Rev Genet*, 2011.

[7] Loscalzo J. Chan, S. Y. The emerging paradigm of network medicine in the study of human disease. *Circulation research*, 2012.

[8] Soler-López M. Aloy P. Zanzoni, A. A network medicine approach to human disease. *FEBS letters*, 2009.

[9] Park K. A review of computational drug repurposing. *Translational and clinical pharmacology*, 2019.

[10] Hansen RW DiMasi JA, Grabowski HG. Innovation in the pharmaceutical industry: New estimates of rd costs. *Health Econ.*, 2016.

[11] Linding R. Pawson, T. Network medicine. *FEBS letters*, 2008.

[12] Abhijeet Rajendra Sonawane, Elena Aikawa, and Masanori Aikawa. Connections for matters of the heart: Network medicine in cardiovascular diseases. *Frontiers in Cardiovascular Medicine*, Volume 9 - 2022, 2022.

[13] Cristina Banfi, Dominic M. Desiderio, Joseph A. Loo, and Igor A. Kaltashov, editors. *Cardiovascular Proteomics Techniques: Applications for Clinical and Laboratory Research*. John Wiley & Sons, 2026.

[14] Kirsch R. Koutrouli M. Nastou K. Mehryary F.-Hachilif R. Gable A. L. Fang T. Doncheva N. T. Pyysalo S. Bork P. Jensen L. J. von Mering C. Szklarczyk, D. The string database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, 2023.

[15] Muin J. Khoury Wei Yu, Melinda Clyne and Marta Gwinn. Phenopedia and genopedia: Disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics*, 2010.

[16] Klinger CM et al. Knox C, Wilson M. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic Acids Res*, 2024.

[17] The UniProt Consortium. Uniprot: the universal protein knowledgebase in 2023. *Nucleic Acids Research*, 51(D1):D523–D531, 11 2022.

[18] Thomas Cokelaer, Dennis Pultz, Lea M. Harder, Jordi Serra-Musach, and Julio Saez-Rodriguez. Bioservices: a common python package to access biological web services programmatically. *Bioinformatics*, 29(24):3241–3242, 09 2013.

[19] Duc-Hau Le. Random walk with restart: A powerful network propagation algorithm in bioinformatics field. pages 242–247, 2017.

[20] Horn D Robinson PN Köhler S, Bauer S. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet*, 2008.

[21] Alberto Valdeolivas, Laurent Tichit, Claire Navarro, Sophie Perrin, Gaëlle Odelin, Nicolas Levy, Pierre Cau, Elisabeth Remy, and Anaïs Baudot. Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics*, 35(3):497–505, 07 2018.

[22] N. Litvak, W. R. W. Scheinhardt, and Y. Volkovich. In-degree and pagerank: Why do they follow similar power laws? *Internet Mathematics*, 4(2-3):175–198, 2007.

[23] Gopal Pandurangan, Gopal, Raghavan, Prabhakara, Upfal, and Eli Upfal. Using pagerank to characterize web structure. volume 3, 08 2002.

[24] Sharma A. Kitsak M. Ghiassian S. D. Vidal-M. Loscalzo J. Barabási A. L. Menche, J. Uncovering disease-disease relationships through the incomplete interactome. *Science*, 2015.

[25] Valle D et al. Goh K-I, Cusick ME. The human disease network. *Proc Natl Acad Sci*, 2007.

[26] Papa A. A. Williams E. A. Rago A.-Palladino A. Politano L. Nigro G. Russo, V. Ace inhibition to slow progression of myocardial fibrosis in muscular dystrophies. *Trends in cardiovascular medicine*, 2018.

[27] Calkins H. Gaine, S. P. Antiarrhythmic drug therapy in arrhythmogenic right ventricular cardiomyopathy. *Biomedicines*, 11(4), 2023.

[28] Gerstenfeld E. P. Svetlichnaya Y. Scheinman M. M. Ermakov, S. Use of flecainide in combination antiarrhythmic therapy in patients with arrhythmogenic right ventricular cardiomyopathy. *Heart rhythm*, 14(4):564–569, 2017.

[29] Wu L. Zheng L. Liu S. Sheng L. Liu L. Zhu Z. Yao Y. Tu, B. Angiotensin-converting enzyme inhibitors/angiotensin receptor blockers: Anti-arrhythmic drug for arrhythmogenic right ventricular cardiomyopathy. *Frontiers in cardiovascular medicine*, 2021.

[30] Kim P. M. Sprecher E. Trifonov V. Gerstein M. Yu, H. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS computational biology*, 2007.

[31] Vidal M Barábasi AL Guney E, Menche J. Network-based in silico drug efficacy screening. *Nat Commun.*, 2016.

[32] Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre.

Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008.

[33] G. Rajendran D. Dhanalakshmi. Comparative study of louvain, leiden, and infomap algorithms for community detection in complex networks. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 2025.

[34] Zhou B. Pache L. Chang M. Khodabakhshi A. H. Tanaseichuk O. Benner C. Chanda S. K. Zhou, Y. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications*, 2019.

[35] Narayan R. Corsello S. M. Peck D. D. Natoli T. E. Lu X. Gould J. Davis J. F. Tubelli A. A. Asiedu J. K. Lahr D. L. Hirschman J. E. Liu Z. Donahue M. Julian B. Khan M. Wadden D. Smith I. C. Lam D. Liberzon A. ... Golub T. R. Subramanian, A. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 2017.

[36] Fan S. Cheedipudi S. M. Braza-Boïls A. Molina M. S. Yao Y. Robertson M. J. Coarfa C. Gimeno J. R. Molina P. Gurha P. Zorio E. Marian A. J. Rouhi, L. The ep300/tp53 pathway, a suppressor of the hippo and canonical wnt pathways, is activated in human hearts with arrhythmogenic cardiomyopathy in the absence of overt heart failure. *Cardiovascular research*, 2022.

[37] Lash A. E. Edgar R., Domrachev M. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res*, 2002.

[38] Meraviglia V. Blankenburg H. Piubelli C.-Pramstaller P. P. Paolin A. Cogliati E. Pompilio G. Sommariva E. Domingues F. S. Rossini A. Rainer, J. The arrhythmogenic cardiomyopathy-specific coding and non-coding transcriptome in human cardiac stromal cells. *BMC genomics*, 2018.

[39] Schwientek P. Ellinghaus P. Summer H.-Golz S. Kassner A. Schulz U. Gummert J. Milting H. Gaertner, A. Myocardial transcriptome analysis of human arrhythmogenic right ventricular cardiomyopathy. *Physiological genomics*, 2012.

[40] Meraviglia V. Mackova K. Frommelt-L. S. König E. Rainer J. Volani C. Benzoni P. Schlittler M. Cattelan G. Motta B. M. Volpato C. Rauhe W. Barbuti A. Zacchigna S. Pramstaller P. P. Rossini A. De Bortoli, M. Modeling incomplete penetrance in arrhythmogenic cardiomyopathy by human induced pluripotent stem cell derived cardiomyocytes. *Computational and structural biotechnology journal*, 2023.

[41] Maione A. S. Chiesa M. Perrucci-G. L. Iengo L. Sattin T. Cencioni C. Savoia M.

Zeiher A. M. Tundo F. Tondo C. Pompilio G. Sommariva E. Lippi, M. Omics analyses of stromal cells from acm patients reveal alterations in chromatin organization and mitochondrial homeostasis. *International journal of molecular sciences*, 2023.

[42] Lukas M. A. Naccarelli, G. V. Carvedilol's antiarrhythmic properties: therapeutic implications in patients with left ventricular dysfunction. *Clinical cardiology*, 2005.

[43] Moreau A. Sleiman Y. Charrabi-A. Delinière A. Bessière F. Gardey K. Richard S. Chevalier P. Reisqs, J. B. Spironolactone as a potential new treatment to prevent arrhythmias in arrhythmogenic cardiomyopathy cell model. *Journal of personalized medicine*, 2023.

[44] Mitchell L. B. Pogue J.-Bosch J. Dagenais G. Yusuf S. HOPE Investigators Teo, K. K. Effect of ramipril in reducing sudden deaths and nonfatal cardiac arrests in high-risk individuals without heart failure or left ventricular dysfunction. *Circulation*, 2004.

[45] R.D Zühlke, A Bouron, N.M Soldatov, and H Reuter. Ca2+ channel sensitivity towards the blocker isradipine is affected by alternative splicing of the human 1c subunit gene. *FEBS Letters*, 427(2):220–224, 1998.

[46] Faris P. Iengo L. et al. Maione, A.S. Ca2+ dysregulation in cardiac stromal cells sustains fibro-adipose remodeling in arrhythmogenic cardiomyopathy and can be modulated by flecainide. *J Transl Med*, 2022.

[47] Wu Y. Zhou C. Xie J.-Zhang Y. Yang X. Xiao J. Wang D. W. Shan C. Zhou X. Xiang Y. Yang B. Zhang, B. Hyperactivation of atf4/tgf-1 signaling contributes to the progressive cardiac fibrosis in arrhythmogenic cardiomyopathy caused by dsg2 variant. *BMC medicine*, 2024.

[48] Can G. Ayvaz S. Karaca-T. Pamuk G. E. Demirtas S. Tsokos G. C. Pamuk, O. N. Spleen tyrosine kinase (syk) inhibitor fostamatinib limits tissue damage and fibrosis in a bleomycin-induced scleroderma mouse model. *Clinical and experimental rheumatology*, 2015.

A | Additional figures

Additional figures for the univariate analysis (Section 6.1):

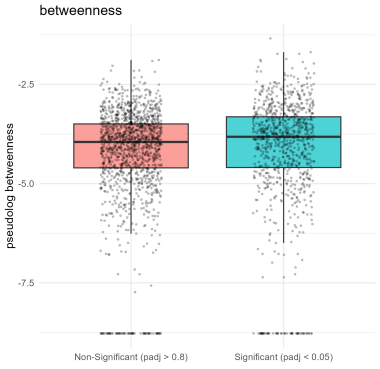

Figure A.1: Boxplot comparison of the Betweenness Centrality.

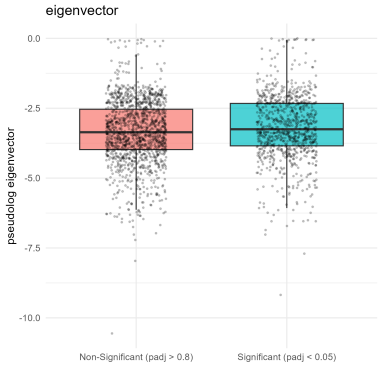

Figure A.2: Boxplot comparison of the Eigenvector Centrality.

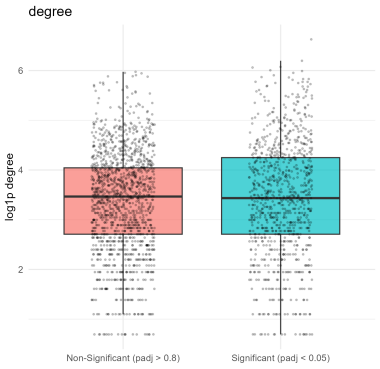

Figure A.3: Boxplot comparison of the Degree.

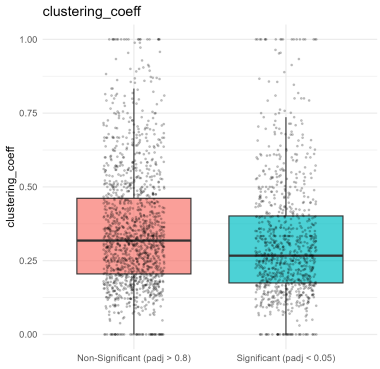

Figure A.4: Boxplot comparison of the Clustering Coefficient.

Below, the additional figures from the enrichment analysis on Metascape of the disease modules from Section 7.2.2 are reported in the following order: Module 2, 3, 4.

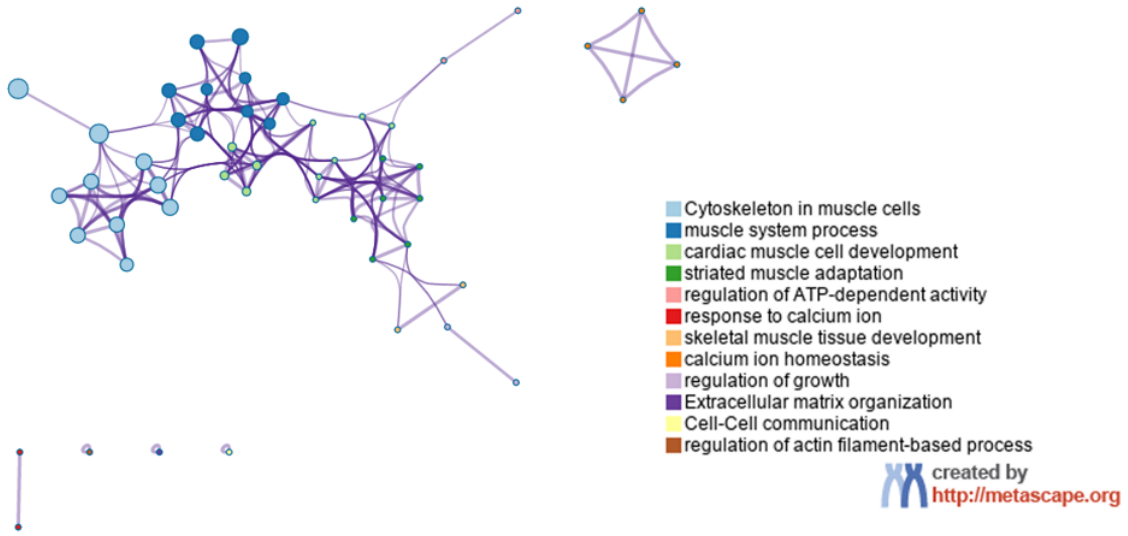

Figure A.5: Network of enriched terms (Module 2): (a) colored by cluster ID, where nodes that share the same cluster ID are typically close to each other. [34]

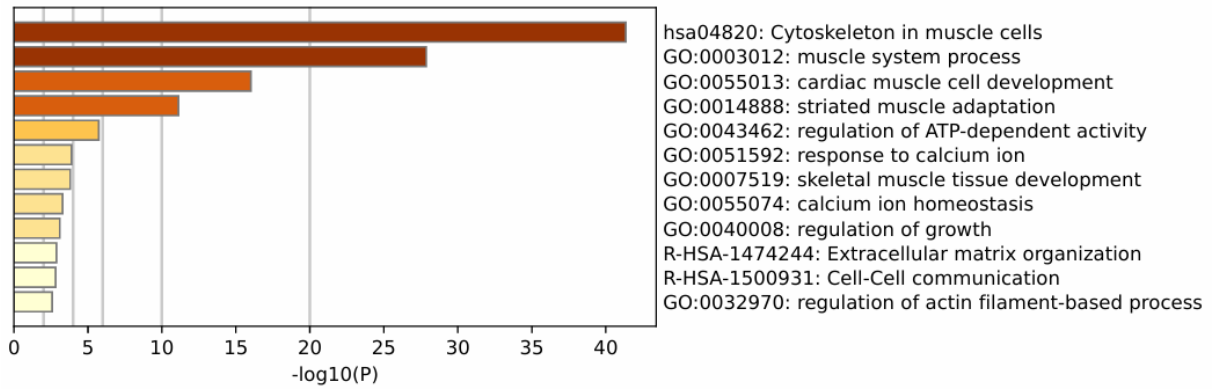

Figure A.6: Bar graph of top enriched terms across input gene lists (Module 2). [34]

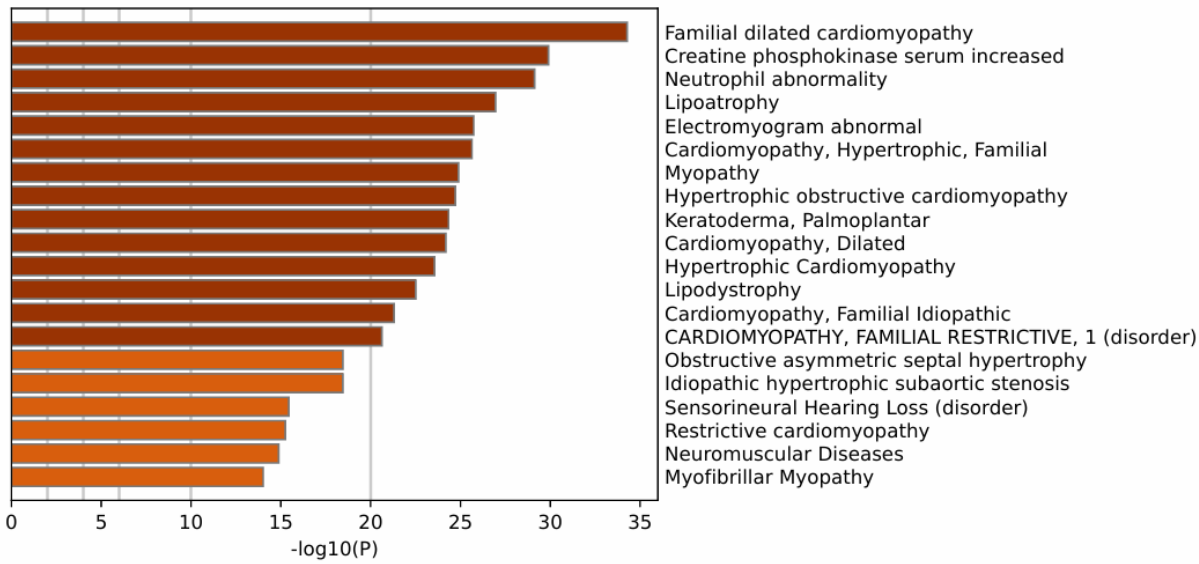

Figure A.7: Summary of enrichment analysis in DisGeNET (Module 2). [34]

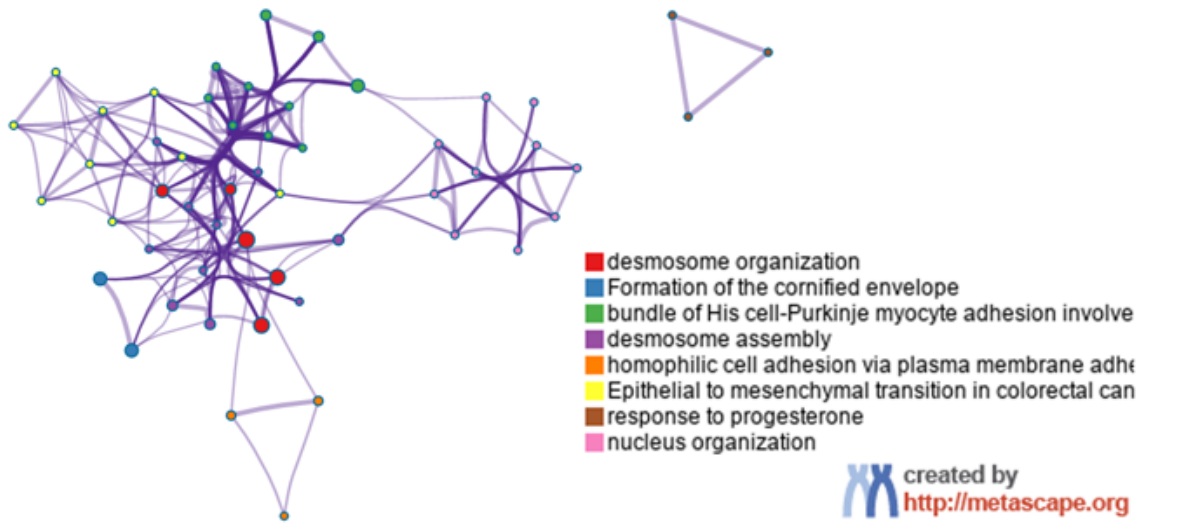

Figure A.8: Network of enriched terms (Module 3): (a) colored by cluster ID, where nodes that share the same cluster ID are typically close to each other. [34]

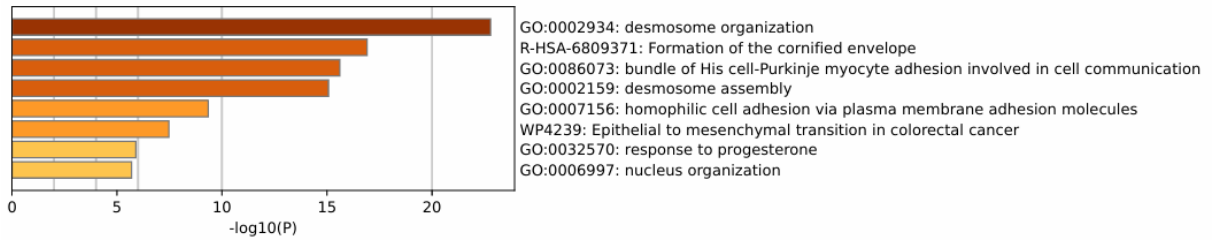

Figure A.9: Bar graph of top enriched terms across input gene lists (Module 3). [34]

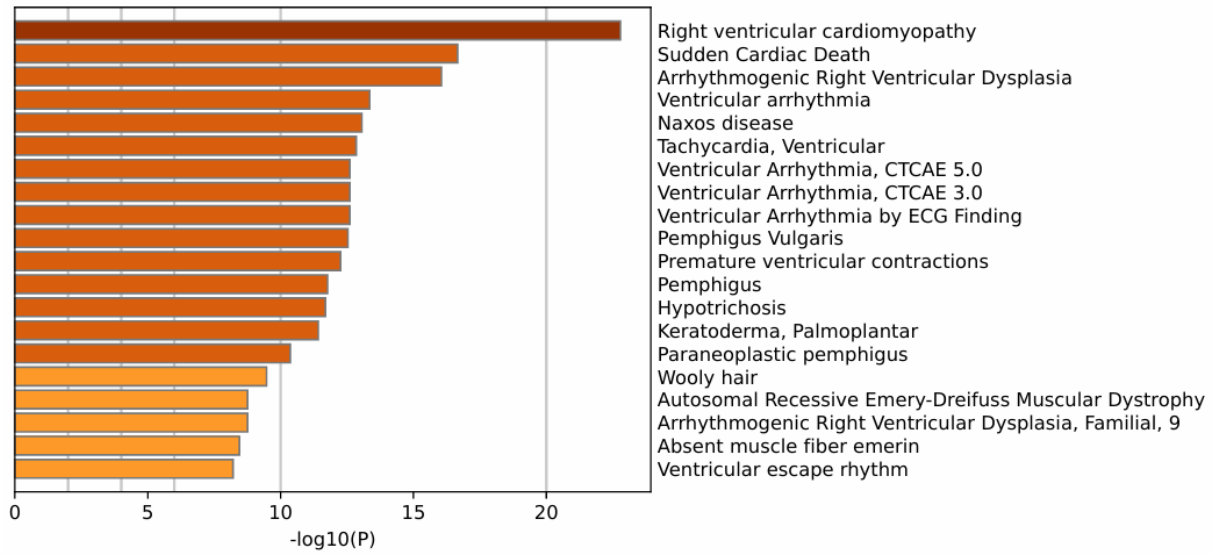

Figure A.10: Summary of enrichment analysis in DisGeNET (Module 3). [34]

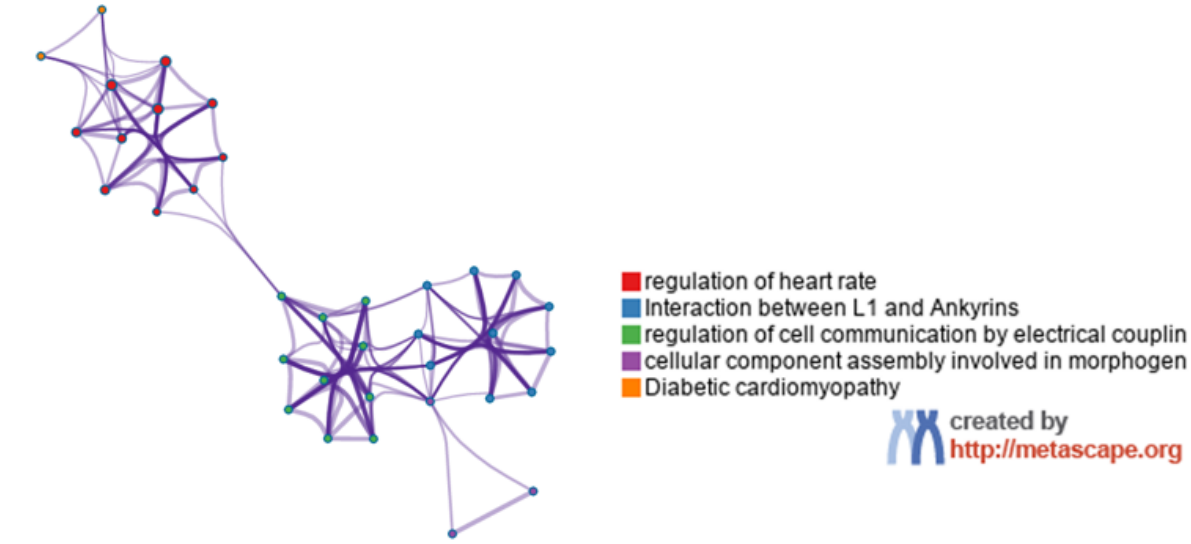

Figure A.11: Network of enriched terms (Module 4): (a) colored by cluster ID, where nodes that share the same cluster ID are typically close to each other. [34]

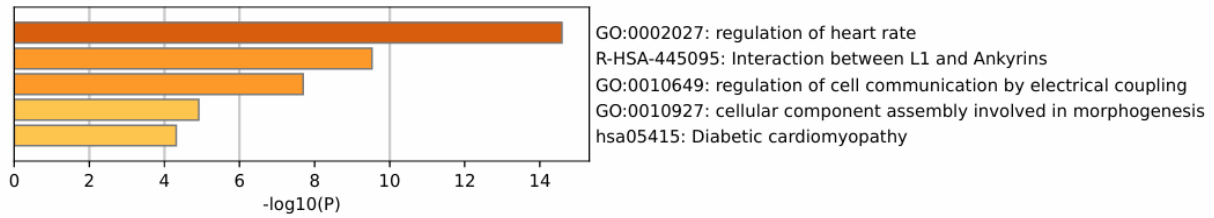

Figure A.12: Bar graph of top enriched terms across input gene lists (Module 4). [34]

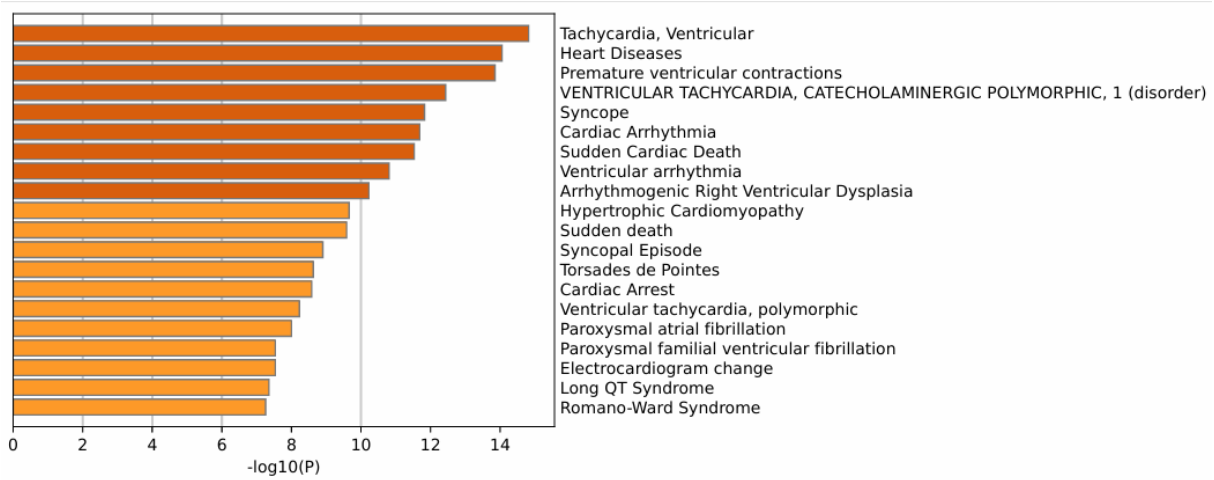

Figure A.13: Summary of enrichment analysis in DisGeNET (Module 4). [34]

B | Additional tables

All drug candidates from the network-based method (Section 5).

Table B.1: Significant drug repurposing candidates for ACM, ranked by Context Score.

Drug Name	Z-Score	Adj. P-value	Context Score	s_{AB}
Lisinopril	4.80	1.82×10^{-2}	0.58	1.10 (Distant)
Enalaprilat	4.81	2.24×10^{-2}	0.52	1.29 (Distant)
Ramipril	4.77	2.88×10^{-2}	0.52	1.29 (Distant)
Candoxatril	4.54	4.39×10^{-2}	0.52	1.12 (Distant)
Ilepatril	4.51	4.02×10^{-2}	0.52	1.12 (Distant)
Omapatrilat	4.51	1.82×10^{-2}	0.52	1.12 (Distant)
Gallopamil	5.67	7.83×10^{-3}	0.38	0.60 (Proximal)
Hydrochlorothiazide	4.01	2.88×10^{-2}	0.35	0.74 (Proximal)
Captopril	4.33	1.32×10^{-2}	0.31	0.82 (Proximal)
Propafenone	4.49	2.24×10^{-2}	0.26	0.47 (Proximal)
Dyclonine	4.73	2.88×10^{-2}	0.20	0.74 (Proximal)
Hexylcaine	4.81	3.21×10^{-2}	0.20	0.74 (Proximal)
Azimilide	4.29	3.21×10^{-2}	0.18	0.87 (Proximal)
Procainamide	4.35	2.88×10^{-2}	0.18	0.11 (Proximal)
Aprindine	5.30	7.83×10^{-3}	0.18	0.31 (Proximal)
Flecainide	7.04	7.83×10^{-3}	0.17	0.47 (Proximal)
Vernakalant	6.02	7.83×10^{-3}	0.14	0.77 (Proximal)
Neratinib	3.42	4.74×10^{-2}	0.13	0.59 (Proximal)
Canertinib	3.80	2.65×10^{-2}	0.12	0.62 (Proximal)
Carvedilol	4.78	7.83×10^{-3}	0.12	0.58 (Proximal)
Halofantrine	4.15	2.88×10^{-2}	0.12	0.10 (Proximal)
N-(6-Aminohexyl)-5-Chloro-1-Naphthalenesulfonamide	5.04	1.32×10^{-2}	0.10	0.13 (Proximal)
Quinidine	3.83	2.24×10^{-2}	0.10	0.61 (Proximal)
Disopyramide	4.74	7.83×10^{-3}	0.10	0.61 (Proximal)
Sorafenib	3.97	1.32×10^{-2}	0.09	0.67 (Proximal)
Spironolactone	5.16	7.83×10^{-3}	0.09	0.65 (Proximal)

Drug Name	Z-Score	Adj. P-value	Context Score	s_{AB}
Lidocaine	5.34	7.83×10^{-3}	0.09	0.54 (Proximal)
Donepezil	4.04	2.65×10^{-2}	0.08	0.70 (Proximal)
Amiodarone	4.94	7.83×10^{-3}	0.08	0.55 (Proximal)
Imipramine	3.70	2.65×10^{-2}	0.08	0.87 (Proximal)
Brigatinib	3.92	2.24×10^{-2}	0.08	0.72 (Proximal)
Yohimbine	3.42	4.39×10^{-2}	0.08	0.84 (Proximal)
Dronedarone	7.93	7.83×10^{-3}	0.07	0.63 (Proximal)
Pentoxyverine	7.27	7.83×10^{-3}	0.07	0.79 (Proximal)
Glyburide	4.48	7.83×10^{-3}	0.07	0.63 (Proximal)
Tamoxifen	4.03	1.32×10^{-2}	0.07	0.60 (Proximal)
Tetracaine	8.42	7.83×10^{-3}	0.07	0.71 (Proximal)
Trifluoperazine	3.65	3.21×10^{-2}	0.07	0.37 (Proximal)
Aranidipine	3.56	2.88×10^{-2}	0.07	0.88 (Proximal)
Minocycline	4.09	1.32×10^{-2}	0.07	0.68 (Proximal)
Oxcarbazepine	8.41	7.83×10^{-3}	0.07	1.06 (Distant)
Chloroprocaine	7.38	7.83×10^{-3}	0.06	0.74 (Proximal)
Bepridil	5.87	7.83×10^{-3}	0.06	0.49 (Proximal)
Carbamazepine	6.31	7.83×10^{-3}	0.06	0.79 (Proximal)
Pramocaine	7.02	7.83×10^{-3}	0.06	0.94 (Proximal)
Propoxycaine	7.07	7.83×10^{-3}	0.06	0.94 (Proximal)
Amylocaine	7.09	7.83×10^{-3}	0.06	0.94 (Proximal)
Articaine	7.13	7.83×10^{-3}	0.06	0.94 (Proximal)
Regorafenib	4.25	1.32×10^{-2}	0.06	0.66 (Proximal)
Amitriptyline	4.02	1.82×10^{-2}	0.06	0.81 (Proximal)
Brivaracetam	8.25	7.83×10^{-3}	0.06	0.98 (Proximal)
Nisoldipine	4.44	1.82×10^{-2}	0.06	0.89 (Proximal)
Ponatinib	3.99	2.65×10^{-2}	0.06	0.75 (Proximal)
Cenobamate	6.81	7.83×10^{-3}	0.06	0.96 (Proximal)
Ranolazine	9.93	7.83×10^{-3}	0.06	0.68 (Proximal)
Nintedanib	3.92	2.88×10^{-2}	0.06	0.75 (Proximal)
Tivozanib	3.62	3.21×10^{-2}	0.06	0.72 (Proximal)
Lamotrigine	4.07	1.32×10^{-2}	0.06	0.93 (Proximal)
Ibutilide	4.24	1.32×10^{-2}	0.05	0.68 (Proximal)
Arsenic trioxide	3.42	4.02×10^{-2}	0.05	0.66 (Proximal)
Ritodrine	6.99	7.83×10^{-3}	0.05	0.51 (Proximal)
Acetylsalicylic acid	4.32	7.83×10^{-3}	0.05	0.62 (Proximal)
Zinc sulfate, unspecified form	4.52	7.83×10^{-3}	0.05	0.75 (Proximal)
Verapamil	6.98	7.83×10^{-3}	0.05	0.87 (Proximal)

Drug Name	Z-Score	Adj. P-value	Context Score	s_{AB}
Isradipine	4.90	7.83×10^{-3}	0.05	0.99 (Proximal)
Valproic acid	5.96	7.83×10^{-3}	0.05	0.54 (Proximal)
Nifedipine	5.90	7.83×10^{-3}	0.05	0.67 (Proximal)
Myristic acid	3.64	3.63×10^{-2}	0.05	0.51 (Proximal)
Cinnarizine	4.16	7.83×10^{-3}	0.04	0.81 (Proximal)
Felodipine	6.99	7.83×10^{-3}	0.04	0.45 (Proximal)
Phenytoin	4.50	1.82×10^{-2}	0.04	1.03 (Distant)
Nimodipine	4.70	7.83×10^{-3}	0.04	0.77 (Proximal)
Tetrabromo-2-Benzotriazole	4.46	7.83×10^{-3}	0.04	0.65 (Proximal)
Promethazine	9.36	7.83×10^{-3}	0.04	0.72 (Proximal)
Colforsin	4.94	7.83×10^{-3}	0.04	0.68 (Proximal)
Topiramate	7.33	7.83×10^{-3}	0.04	0.64 (Proximal)
Triamterene	4.61	1.32×10^{-2}	0.04	1.28 (Distant)
Zonisamide	5.15	7.83×10^{-3}	0.03	0.71 (Proximal)
Lacidipine	4.64	7.83×10^{-3}	0.03	1.05 (Distant)
Calcium	4.89	7.83×10^{-3}	0.03	0.37 (Proximal)
Amiloride	3.92	2.88×10^{-2}	0.03	0.53 (Proximal)
Trimebutine	5.79	7.83×10^{-3}	0.03	0.82 (Proximal)
Cilnidipine	4.69	7.83×10^{-3}	0.03	1.05 (Distant)
Drotaverine	4.63	1.82×10^{-2}	0.03	0.96 (Proximal)
Bisindolylmaleimide I	4.06	1.32×10^{-2}	0.03	0.69 (Proximal)
Prenylamine	4.20	2.88×10^{-2}	0.03	0.58 (Proximal)
Levomenthol	5.39	7.83×10^{-3}	0.03	0.80 (Proximal)
Lanoteplase	4.19	2.24×10^{-2}	0.03	0.75 (Proximal)
Istaroxime	7.35	7.83×10^{-3}	0.03	1.18 (Distant)
Nilvadipine	5.81	7.83×10^{-3}	0.03	1.08 (Distant)
Manidipine	5.46	7.83×10^{-3}	0.03	1.02 (Distant)
Nitrendipine	6.14	7.83×10^{-3}	0.03	0.96 (Proximal)
Mibefradil	5.37	7.83×10^{-3}	0.03	1.03 (Distant)
Benidipine	5.64	7.83×10^{-3}	0.03	1.02 (Distant)
Flunarizine	3.93	2.65×10^{-2}	0.03	0.89 (Proximal)
Enflurane	9.91	7.83×10^{-3}	0.03	0.76 (Proximal)
Dequalinium	4.61	7.83×10^{-3}	0.02	0.62 (Proximal)
Nicardipine	7.13	7.83×10^{-3}	0.02	0.85 (Proximal)
Fluciclovine (18F)	3.92	1.32×10^{-2}	0.02	0.94 (Proximal)
Staurosporine	3.53	2.24×10^{-2}	0.02	0.74 (Proximal)
Ouabain	4.70	2.24×10^{-2}	0.02	1.27 (Distant)
Butamben	8.02	7.83×10^{-3}	0.02	0.85 (Proximal)

Drug Name	Z-Score	Adj. P-value	Context Score	s_{AB}
Tetraethylammonium	3.69	4.74×10^{-2}	0.02	0.59 (Proximal)
Phosphonothreonine	3.47	2.88×10^{-2}	0.02	0.52 (Proximal)
Fostamatinib	9.84	7.83×10^{-3}	0.02	0.54 (Proximal)
Rostafuroxin	6.47	7.83×10^{-3}	0.02	1.28 (Distant)
Efonidipine	6.61	7.83×10^{-3}	0.02	1.13 (Distant)
Ergocalciferol	6.66	7.83×10^{-3}	0.02	1.13 (Distant)
Dalfampridine	5.48	7.83×10^{-3}	0.02	1.15 (Distant)
Digoxin	5.68	1.32×10^{-2}	0.02	1.03 (Distant)
Calcium citrate	4.84	7.83×10^{-3}	0.02	0.38 (Proximal)
Calcium phosphate dihydrate	4.80	7.83×10^{-3}	0.02	0.38 (Proximal)
Tegoprazan	7.20	7.83×10^{-3}	0.02	1.30 (Distant)
Artenimol	4.86	7.83×10^{-3}	0.02	0.60 (Proximal)
Phenethyl Isothiocyanate	5.43	7.83×10^{-3}	0.01	0.53 (Proximal)
Bisacodyl	6.88	7.83×10^{-3}	0.01	1.14 (Distant)
Rimacalib	4.56	1.82×10^{-2}	0.01	0.79 (Proximal)
CYT997	3.75	3.21×10^{-2}	0.00	1.00 (Proximal)

The raw summary of the multivariate logistic regression model is reported below (Section 6.2).

Table B.2: Multivariate Logistic Regression Coefficients. The table reports the raw beta coefficients (β), standard errors, z-statistics and P-values. Significance codes: *** $p < 0.001$, ns = not significant.

Predictor	Estimate (β)	Std. Error	z-value	P-value
(Intercept)	-0.475	0.045	-10.50	$< 2 \times 10^{-16}$ ***
Degree (Scaled)	-0.529	0.063	-8.37	$< 2 \times 10^{-16}$ ***
Betweenness (Scaled)	0.375	0.094	3.97	7.13×10^{-5} ***
Dist. to Disease (Scaled)	-0.802	0.059	-13.63	$< 2 \times 10^{-16}$ ***
Eigenvector (Scaled)	0.184	0.046	4.01	6.00×10^{-5} ***
Clustering (Scaled)	-0.014	0.049	-0.29	0.769 (ns)

Complete table for the transcriptomic results (Section 8.3):

Table B.3: Complete table for the transcriptomic consensus scoring (Section 8.3). The table lists all drugs showing successful signature reversion in at least one dataset ($N = 36$). Checkmarks indicate the specific datasets where reversion was observed.

Drug Candidate	Score	D1	D2	D3	D4	D5
High Confidence (Score ≥ 2)						
Isradipine	**3**	✓		✓	✓	
Canertinib	2	✓	✓			
Carbamazepine	2	✓		✓		
Digoxin	2		✓	✓		
Disopyramide	2		✓		✓	
Flunarizine	2		✓	✓		
Fostamatinib	2	✓			✓	
Imipramine	2	✓		✓		
Lacidipine	2		✓		✓	
Tivozanib	2				✓	✓
Topiramate	2	✓			✓	
Validation Hits (Score 1)						
Amiloride	1					✓
Amiodarone	1				✓	
Amitriptyline	1				✓	
Carvedilol	1				✓	
Dalfampridine	1		✓			
Dronedarone	1			✓		
Gallopamil	1		✓			
Istaroxime	1				✓	
Lamotrigine	1			✓		
Nicardipine	1				✓	
Nifedipine	1			✓		
Nimodipine	1	✓				
Nintedanib	1			✓		
Nitrendipine	1			✓		
Phenytoin	1			✓		
Ponatinib	1					✓
Ramipril	1			✓		
Regorafenib	1			✓		

Drug Candidate	Score	D1	D2	D3	D4	D5
Ritodrine	1			✓		
Sorafenib	1	✓				
Spironolactone	1					✓
Tetracaine	1			✓		
Trimebutine	1			✓		
Verapamil	1				✓	
Zonisamide	1			✓		

List of Figures

List of Tables

Acknowledgements

Desidero ringraziare il mio relatore, il Prof. Zunino, per il costante supporto e per la fiducia riposta in me nello sviluppo di questa tesi. Un ringraziamento sincero va anche alla Prof.ssa Paganoni per avermi dato l'opportunità concreta di lavorare a questo progetto. Desidero inoltre ringraziare le Dottoresse Cristina Banfi e Elena Sommariva, il cui costante entusiasmo per il lavoro svolto è stato per me un grande stimolo.

Ringrazio la mia famiglia e i miei amici per il supporto incondizionato. Siete stati fondamentali per alleggerire la tensione nei momenti più complessi e per avermi sempre spronato ad andare avanti.

Grazie ad Ale, Ali, Fede, Ludo e Mati per essere state al mio fianco negli ultimi mesi e per tutto ciò che abbiamo vissuto insieme in questi dieci anni.

Grazie a Cami, per accompagnarmi da sempre in ogni singolo passo, per non farmi mai mancare il sorriso e per il sostegno incondizionato.

Infine, grazie ad Andre, per essere sempre una presenza costante e il mio punto fermo, grazie per la pazienza, per l'ironia e per supportarmi (e soprattutto sopportarmi) ogni giorno.