Anushka Pal MSc

by Anushka Pal

Submission date: 22-May-2026 09:01PM (UTC+0530)
Submission ID: 2967186700
File name: MSC_FINAL_THESIS.Anushka_Pal_1_1.docx (664.91K)
Word count: 18504
Character count: 111266

Identification Of Common Genes, miRNAs and Transcriptoin Factors through Comprehensive Network Analysis For Enhanced Management Of Gastritis and Stomach Cancer

Anushka Pal

Delhi Technological University , Delhi ,India

E-mail: Palanushka15@gmail.com

ABSTRACT

Gastritis is a condition marked by ongoing inflammation of the stomach lining. It is known to be a precursor to stomach cancer, or gastric carcinoma, which is a major cause of cancer-related deaths worldwide. This happens mainly because the disease is often detected late and progresses slowly. The shift from chronic issues like atrophic gastritis to cancer involves genetic mutations, changes in gene expression, and disrupted signaling pathways. Identifying common molecular mechanisms is critical for precise treatment and early intervention. This study uses a network-based bioinformatics approach to identify shared genes, microRNAs (miRNAs), and transcription factors that connect gastritis and stomach cancer. This could help improve clinical management of these gastric diseases. We gathered genes related to gastritis and stomach tumors from the Comparative Toxicogenomics Database (CTD). We selected the top 200 genes for each condition based on inference scores and analyzed them to find overlaps.Using NetworkAnalyst 3.0, we built networks to explore gene-miRNA interactions, co-expression specific to stomach tissue, and transcription factor-gene relationships. This revealed pathways involved in inflammation (like NF-κB signaling), cell cycle regulation, apoptosis, and oxidative stress. Key miRNAs affect groups of disease-related genes. Meanwhile, major transcription factors such as MYC, RELA, and STAT3 coordinate the shift from inflammation to cancer. These networks

show strong connections and specific patterns related to the stomach.The findings highlight how gastritis can progress to cancer, suggesting that these regulators could serve as biomarkers for early detection, prognosis, and monitoring in at-risk individuals. Targeting these pathways could lead to timely interventions that prevent cancer transformation and improve patient outcomes through tailored diagnostics and treatments. Future validation with clinical groups, functional tests, and long-term studies will help apply these findings in real-world settings. This analysis connects these diseases at a molecular level, enhancing diagnostics, therapies, and management of gastric diseases.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Stomach cancer, known as gastric cancer, is one of the most impactful types of cancer to affect people globally. Although efforts have been made to make incremental advances in the understanding of the disease and its treatment for decades, the disease continues to take a devastating toll, taking the lives of more than seven hundred thousand people and causing around one million new cases annually worldwide [16]. These are placed gastric cancer firmly in the top five causes of cancer-related death worldwide and give it undeniable public health priority. There are geographical inequities in the distribution of the disease with high levels of incidence in certain parts of the world and low levels in others, which have implications for both aetiology and health policy; for example, the incidence of the disease is much higher in areas with high burden in East Asia, Eastern Europe and Latin America than in North America and Western Europe, reflecting differences in the prevalence of infectious risk factors, dietary habits and in the depth of population-based surveillance infrastructure [16].
The prognosis of gastric cancer is closely linked to the stage at which the disease is first detected and this is one of the strongest arguments for improved and earlier detection of the disease. In specialized centres with good endoscopic surveillance programmes with, surgically resectable stage is close to 90% [16]. In most clinical contexts in the world, however – especially the vast majority of resource-poor countries that are most affected by the disease burden – the lack of regular monitoring results in most people arriving at the doctor's office with late-stage, metastatic disease where there is no point in attempting to cure them, and their chances of surviving are not good. This difference between what can be achieved in the biological world through early detection and what can be achieved in the clinical world through late presentation constitutes the central motivating tension of current studies on gastric cancer; the greater the difference, the more sustained the scientific interest in uncovering molecular biomarkers that can detect early change in the gastric mucosa before the onset of symptoms [15, 16].

The Cancer Genome Atlas Research Network (TCGA) has shed light on the molecular understanding of gastric carcinogenesis by using six parallel molecular profiling platforms to analyze a large collection of primary tumour samples from gastric adenocarcinoma patients, which uncovered four molecular types of gastric adenocarcinoma [16]. This classification proved to be so ground-breaking that it recognized that a disease that is clinically described as a single disease can actually be a series of molecularly distinct forms (neoplastic entities) which have a similar anatomical location, but are vastly different in mutational landscape, epigenetic state and therapeutic vulnerability. Such molecular heterogeneity makes any single-pathway and/or single-biomarker strategy for gastric cancer diagnosis and treatment fundamentally inadequate and requires systems-level analyses that are capable of capturing the full regulatory complexity that underlie gastric cancer phenotypic diversity [16, 25].

In most cases, a long period of chronic inflammation of the gastric mucosa termed gastritis precedes the development of gastric malignancy. Gastritis is not just a co-occurring condition with gastric cancer, but a continuous mechanism that creates and sustains the inflammatory mucosal environment in which the process of gastric cancer gradually accumulates molecular alterations during years to decades [15, 17]. It is therefore crucial to know the molecular characteristics of gastritis and gastric cancer, including the genes, microRNAs, transcription factors, and signalling pathways that are activated in both disease states, in order to understand how the process of gastric cancer develops, as well as to identify novel points of intervention where the development of the disease can be detected or interrupted. Even though this idea is clear, the published literature has largely focused on gastritis or gastric cancer as distinct research topics, and the common molecular basis of gastritis and gastric cancer is poorly understood [1, 2, 3, 12].

1.2 Definition of Key Constructs

1.2.1 Gastritis

Gastritis refers to inflammatory changes in the gastric mucosa with infiltration of immune cells, mostly neutrophils in the acute phase and lymphocytes, plasma cells and macrophages in the chronic phase, in the epithelial and the lamina propria layers of the stomach wall [15]. Gastritis, from a clinical and aetiological point of view, is a heterogeneous condition which can be caused by infection with Helicobacter pylori or auto-immune damage of the parietal cells, chemical and drug-induced damage, and idiopathic gastritis. In the present investigation, gastritis is defined as chronic inflammatory state of the gastric mucosa, which is most strongly associated with chronic H pylori colonisation, and therefore most closely and causally linked to the subsequent development of gastric adenocarcinoma along the Correa cascade pathway [14, 15]. From a molecular point of view, chronic gastritis has been linked to the chronic activation of pro-inflammatory transcription factors and signal transducer and activator of transcription 3, dysregulation of expression of microRNAs regulating proliferation and apoptosis, and progressive epigenetic remodelling of the gastric mucosal transcriptional programme [8, 9, 12].

1.2.2 Gastric Cancer

For the purposes of this investigation, stomachcancer is defined as primary malignant neoplasms originating from the epithelial lining of the stomach, and specifically as gastric adenocarcinoma, which makes up more than 90% of all gastric malignancies, and is the histological type most directly associated with the H. pylori-gastritis-carcinoma pathobiology. Gastric adenocarcinoma is also histologically subclassified by the Lauren classification, which classifies it into either an intestinal-type (architecturally cohesive, more closely related to chronic H. pylori infection and the Correa cascade) or a diffuse-type (less architecturally cohesive, more frequently related to CDH1 mutations and the genomically stable molecular subtype), and has a poor prognosis [16, 26]. Genomic instability is a hallmark of gastric cancer at the molecular level, along with the mutations which happen at particular time in life of individual of important tumor suppressor genes, such as ARID1A, as well as the loss of epigenetic regulation of growth-regulatory genes and the extensive deregulation of microRNA and transcription factor networks that normally maintain epithelial homeostasis [1, 2, 3, 16].

1.2.3 Gene co-Expression Network

A gene co expression network is a computational model of the directionality of the regulatory relationship between molecules (tfs, microRNAs, and/or target genes) in a biological system, represented as a directed graph in which nodes are molecular species and edges are regulatory relationships [25]. In biological systems, the topology of gene regulatory networks has been empirically found to be scale-free: that is, the number of connections of each node follows a power law distribution with a small minority of highly connected nodes called hubs that can regulate the behaviour of the network disproportionately, while the majority of nodes have relatively few connections [24, 25]. Hubs in GRNs are of particular interest in the context of disease biology as potential biomarkers and therapeutic targets because knocking them out or altering them causes large amplitude changes that ripple throughout the network and because their identity remains consistent across independent analyses of the same disease condition, despite the use of different datasets and/or analysis methods [24, 25].

1.2.4 MicroRNA

MicroRNAs form a class of endogenous, non-coding, single-stranded ribonucleic acids consisting of about eighteen to twenty-five nucleotide bases, with their main biological role being the regulation of the transcription of their target genes after transcription, which is accomplished by the specific targeting of their target sequences, located mainly within the 3-prime untranslated region of messenger RNAs, leading to the binding of the RNA-induced silencing complex for either translational repression or RNA degradation [23]. The regulatory reach of the microRNA system is broad; each microRNA can target hundreds of different kinds of messenger RNA and the expression of one target gene may be controlled by a number of convergent microRNAs, leading to a combinatorial, massive biological impact [23]. In the gastric disease setting, miR-21, miR-223, miR-218 and members of the miR-34 family have been suggested to be important post-transcriptional regulators of genes involved in cell growth, invasion and their dysregulated expression in the tissues of gastritis and gastric cancer has made them prime candidates for the development of non-invasive biomarkers [12, 13].

1.2.5 Transcription Factor

They are sequence-specific DNA-binding proteins that regulate the rate of transcriptional initiation from target gene promoters by either accelerating or decelerating this process in different cell types and in combination with different co-regulatory proteins [1, 2]. In this context, transcription factors are components of particular regulatory significance since they act as master coordinators of large gene expression programs: the action of a single transcription factor may directly affect the expression of dozens to hundreds of target genes, and the abnormal activation or repression of a transcription factor can therefore have a regulatory effect larger than the sum of the mutations of these genes [8, 9, 10]. Nuclear factor kappa B, signal transducer and activator of transcription 3, MYC proto-oncogene all serve as hub nodes within the gastric cancer regulatory network and are all candidates for therapeutic targeting, respectively, across independent analyses of gastric carcinogenesis [8, 9, 10, 11].

1.3 Importance and Significance of the Research Area

Several converging imperatives, across clinical, scientific and public health aspects of the problem, support the importance of research at the interface of gastritis and gastric cancer molecular biology. Clinically, the poor prognosis of advanced gastric cancer and the excellent prognosis of early stage disease makes the need for developing an early detection tool for detection of pre-malignant or early malignant transformation more urgent and unambiguous [16]. However, the current approach of endoscopic surveillance, which is effective in high compliance environments, is invasive, burdensome and has not been scalable to population level in most health systems. Molecularly informed, non-invasive biomarker panels, in particular, based on circulating microRNAs or other blood-based analytes, may represent an alternative of potential transformative power which could be applied at population level without the logistic and patient burden of endoscopy [13, 22].

The gastritis-to-gastric cancer transition is one of the most well-characterised models of inflammation-triggered cancer in human medicine, and hence, the molecular signatures of gastritis are not directly related to gastric cancer, but rather to carcinogenesis more generally. The same mechanisms of chronic inflammatory microenvironment-mediated oncogenesis that apply to the association of colitis and colorectal cancer, hepatitis and hepatocellular cancer, and pancreatitis and pancreatic cancer are also true of other associations between inflammation and cancer [14, 20, 21]. The conceptual and methodological implications of the systematic molecular characterization of gastritis and gastric cancer as co-disease states may then apply to these nearby research fields.

Systematic data-driven characterization of disease-associated gene regulatory networks, free from the constraints of time and expense of purely experimental approaches, has become more possible than ever before at the methodological level, thanks to the growing availability of comprehensive molecular databases like the Comparative Toxicogenomics Database [7] and integrative computational platforms like NetworkAnalyst [27]. This will be a methodological advance directly related to the shortcomings of the single-disease, single-layer analyses that have dominated published computational gastric cancer studies [1, 2, 3, 4, 5, 6] by being embedded in a unified network that incorporates the regulatory layers of microRNA, transcription factor, and co-expression.
The major theoretical frameworks used to facilitate the study are discussed.

1.4 Major Theoretical Frameworks Associated with the Research

1.4.1 Network Medicine Theory

The theoretical basis of the present investigation is network medicine, as formulated by Barabási, Gulbahce and Loscalzo, who suggest that diseases are not only disorders of individual genes or proteins, but disorders of molecular networks, and that the pathological phenotype of a disease is an emergent property of network-level perturbations, not a straightforward linear consequence of individual molecular alterations [24]. The human interactome, which represents the set of all protein-protein, protein-DNA and protein-RNA interactions in a cell, is envisioned as a map, on which the disease processes can be mapped as network perturbation patterns, called disease modules, that are located in specific topological regions of the interactome. The clinical and/or

aetiological similarity of diseases suggests that they share a significant molecular network substrate, which can be empirically tested across multiple disease pairs, and whose property of predicted shared disease modules is used in this study to predict a molecular network substrate shared by gastritis and gastric cancer [24, 25].

The network medicine approach also suggests that the most effective targets for biomarker and therapeutic development will not be randomly selected within the disease module, but will be concentrated around the hub nodes of the module, i.e. entities with a high degree of connectivity that have the most significant consequences when they are perturbed. The same small group of transcription factors and microRNAs has been found to be the most consistently identified regulators in different computational analyses of gastric carcinogenesis using different data sets, computational tools and patient populations [1, 8, 9, 10, 12]. Network medicine theory therefore offers both a conceptual argument for an analytical approach that focuses on hubs and a hypothesis to be tested: common hubs will be detectable in the gastritis and gastric cancer regulatory networks as a common attribute.Network medicine theory thus offers a conceptual justification for a "hub" focused approach to the analytical process, as well as a testable hypothesis for arriving at the common hubs: these hubs will be detectable as a common attribute in the gastritis and gastric cancer regulatory networks.

1.4.2 The Hallmarks of Cancer Framework

Cancer cells share certain characteristics.
The Hanahan and Weinberg hallmarks of cancer framework is a necessary conceptual superstructure to help understand the functional role of the molecular regulatory alterations found in gastric carcinogenesis. This approach categorises the biological capabilities that a malignant cell must gain into a list of canonical hallmarks, inducing angiogenesis and activating invasion and describes the cellular reprogramming that these capabilities entail [18]. The hallmarks are especially relevant in the current context because they provide a biological grammar for interpreting the pathway enrichment results obtained from network analysis: If the set of genes associated with a certain network module turns out to be enriched for genes involved in a particular pathway, such as cell cycle regulation or angiogenesis, the biological meaning of network topology can be directly translated to a particular cancer hallmark, and the pathway enrichment results can be directly functional rather than merely statistical [18]. Moreover, multifunctional transcription factors and microRNAs (such as HIF-1α and miR-21, respectively, which simultaneously control multiple programmes that are relevant to hallmarks) are the most promising targets for multi-target biomarker and therapeutic approaches [10, 11].

1.4.3 The Correa Multistep Carcinogenesis Model

The Correa cascade is a theoretical model of the disease and gives the biological context of the narrative in which all the molecular results of the present investigation have to be understood. The model conceptualises gastritis and gastric cancer as successive steps in a single continuum of cancer, providing the analytic space for comparing molecular analyses across the steps and thus for identifying the regulatory machinery active at each step, and thus characterising not only the endpoint of malignancy, but the molecular pathway along which it is traversed [15, 16]. It also has direct implications for the study of biomarkers for gastric cancer: if the molecular

changes that define gastric cancer occur early in gastritis, then the molecular changes should be observed in patients with gastritis, long before they develop malignant disease, which may be possible to treat [15, 17]. The current study is specifically to examine this implication, being designed to identify and characterize common genes, miRNAs and transcription factors shared by the gastritis and gastric cancer molecular landscapes.

1.5 Identification of the Research Gap

The research gap is identified through the following:
A survey of the available published literature shows that the field has a wealth of studies focusing on gastric cancer or gastritis alone, but only a few studies that jointly and systematically characterize the shared molecular regulatory architecture of both these conditions in a single computational framework. The prevalent analytical approach in literatures on gastric cancer has been to compare the gene expression patterns of the established gastric cancer specimens with normal adjacent gastric tissue to discover the expression changes of gastric cancer and find the different expression of hub proteins or miRNAs between gastric cancer and normal gastric tissue [1, 3, 4, 6]. Although such lesions have provided important mechanistic insights, this design has failed to reveal molecular events that precede the diagnosis of malignancy, which by definition are present during the gastritis stage of the Correa cascade, and may be the earliest detectable molecular signatures of malignant risk [15, 17].

Another gap in analysis is in the multi-layer integration of regulatory networks. Most published network analysis studies of gastric carcinogenesis focus on one layer of the network at a time either the microRNA-gene interaction network, protein with another protein network or the transcription factor gene association network, but not incorporate these layers into a single integrated multi-network model that reflects the hierarchical regulatory relationships among the genes of the gastric cancer transcriptome [1, 2, 5, 6]. Because microRNAs and transcription factors regulate the same target genes, with transcription factors controlling the production of the microRNAs and microRNAs suppressing the transcription factor production in both directions of regulation, the study of one type of regulator alone will inevitably miss the regulatory logic that arises from the interaction between these two dominant regulatory components [1, 2]. To fill this gap, the present study aims to simultaneously built and analyzed the microRNA-gene network, the transcription factor-gene network, and the co-expression network of the common gene set occurring in both gastritis and gastric cancer.

The third gap is methodological and is related to the source of disease-gene association data. Many bioinformatics studies published in the literature have used gene expression data from small patient cohorts from gene expression omnibus databases, data that may not capture the extent of molecular variation of gastric disease in ethnic and geographical populations. In contrast, the Comparative Toxicogenomics Database offers disease-gene associations from thousands of curated publications across many patient types, experimental systems and aetiological contexts and ranks them according to inference scores, which are derived from the breadth of supporting literature and its reproducibility [7]. This resource constitutes more comprehensive and population-representative data on disease-gene association for network

analysis than individual gene expression data sources, and is a methodological advance over the majority of the published literature.

Combined, the three gaps outlined above – no comparative investigation of gastritis and gastric cancer at the same time; single-layer network analysis is more common than multi-layer network analysis; and the limited representativeness of gene expression dataset-based approaches – constitute the space of research that the present investigation aims to fill. The investigation seeks to provide insights into the common molecular regulatory basis of gastritis and gastric cancer that are more mechanistically complete, more representative of the population, and more clinically translatable than knowledge offered by any single investigation of the three gaps considered [7, 24, 27].

1.6 Reserch targets of the Study

The specific targets of the present investigation were:

Objective 1: Find the top gene sets associated with gastritis and gastric cancer in the Comparative Toxicogenomics Database according to inference score, and the set of common genes that are shared between these two diseases is the molecular basis of the continuum between gastritis and gastric cancer [7].

Objective 2: To build and analyse a microRNA-gene interaction network for the selected common genes set using the NetworkAnalyst platform, to search for hub miRNAs that affect the highest number of common genes, with their functions assigned in gastric disease biology [27, 12, 13].

Objective 3: To conduct stomach tissue-specific gene co-expression network analysis of common gene set, identify key co-expression communities or modules in the network, and describe the functional biological processes enriched in each module using gene ontology and pathway enrichment analysis [3, 6, 25].

Objective 4: To build and analyse a TF-gene interaction network for the common set of genes, select the master TFs that control the most the common molecular programme and examine their known involvement in the NF-κB, STAT3, HIF-1α and MYC regulatory axes involved in gastric carcinogenesis [8, 9,]

Objective 5: To integrate the results from all three types of network analysis (microRNA-gene, co-expression and transcription factor-gene) to define the molecules and molecular modules most consistently located in multiple network representations as regulatory hubs in the gastritis-gastric cancer continuum; and to assess them as potential early-stage biomarkers or therapeutic targets in the gastritis-gastric cancer continuum [3, 15, 22].

1.7The scope and delimitations of the study

In the methodological perspective, the present investigation is computational: all results have been obtained from the analysis of biomedical database resources which were curated and publicly available gene expression and molecular interaction data, not from the collection or experimental analysis of original biological specimens. This scope is a conscious methodological decision to allow for the consolidation of the evidence from thousands of published studies in a single analytical framework, rather than any one experimental study, in order to achieve a breadth of evidential basis that no one study could have obtained within analogous resource restrictions. At the same time, it restricts the investigation in significant ways: the associations established for all molecular relationships are not causal, but are either curated from published evidence or predicted by a computer model; no experiments have been performed to conclusively demonstrate causality in the gastric tissue or in cell line models. This investigation is the first step in a research programme of which the validation of the identified hub microRNAs, TFs and co-expression modules in experimental samples – such as quantitative reverse transcription PCR of gastritis and gastric cancer tissue samples, functional knockdown and overexpression of these molecules in gastric cell lines, and clinical measurement of these molecules in samples from gastritis and stomach cancer patients – follows [13, 22].

The investigation is further restricted to examination of gastritis and gastric cancer, as broadly classified disease categories for analysis in the Comparative Toxicogenomics Database, without any molecular subtype, H. pylori infection status, geographic population or histological variant stratification. This delimitation is recognized as a potential source of biological imprecision as the molecular regulatory architecture of gastric disease is highly variable across these dimensions, as the Cancer Genome Atlas molecular classification has shown [16]. The goal of the investigation (characterising the shared molecular substrate that is most broadly relevant across the disease continuum rather than features specific to individual subtypes) and the pragmatic constraint that the Comparative Toxicogenomics Database disease-gene associations are a pool of evidence across these sources of variation [7] justify it. Subtype stratification should also be a key – and needed – analytical refinement for future investigations that follow up on the present findings.

1.8 Organisation of the Thesis

The present thesis is organised around an overall progression from the general contextual and theoretical underpinnings created in this introductory chapter, followed by a elaborately thematically organised and critically analysed review of the published literature, leading to a detailed methodological description of the computational analytical approach, and then to the presentation, discussion and synthesis of the results. Review of Literature is presented in Chapter Two and is organised thematically to the major research domains — epidemiology and pathology of gastric disease, network medicine methodologies, transcription factor regulatory networks, microRNA-mediated post-transcriptional regulation, hub gene identification studies and the contradictions and gaps identified in the existing evidence base. The Materials and Methods are presented in Chapter Three, and the database queries, parameters used for building

the networks, and the analytical tools used for this study are detailed to allow the study to be replicated. The Results are shown in Chapter Four for each of the analytical layers. The Discussion is given in Chapter Five where the results are placed in the context of the literature and the implications for mechanism, biomarkers and therapeutic applications are discussed. The conclusions and recommendations for further research are given in this.

CHAPTER 2

REVIEW OF LITERATURE

2. 1 Global Epidemiological and Pathological Context of the Gastric Disease

Gastric cancer remains a significant burden on global health systems, constituting one of the most deadly malignant neoplasms globally. Reported across varied and multiple oncology registries, there are more than a million cases diagnosed, with a consistently high number of deaths attributing the gastric cancer as the leading cause of cancer-assosiated deaths in the top five worldwide [16].

There is a large gap in the incidence and survival rates, which is even more pronounced in the low- and middle-income countries, where late diagnosis is not an exception, but rather a reflection of the lack of regular surveillance programme, thus less access to advanced diagnostic technologies [16]. Geographically, in high-burden regions in the east of Asia, Eastern Europe, parts of Latin America, the gastric cancer has been associated historically with a large proportion of the cancer-related mortalities, and this trend has been maintained even with slight and minor therapeutic methods' advances [15, 16].Epidemiological models of gastric malignancy often enshrine the linear transformation of the mucosal screen, conceptualized by Correa in 1992 and known now by the name Correa cascade [2]. According to this framework, gastric epithelium goes through specific and well-accepted histologically and molecularly characterised path: from chronic active inflammation to atrophic gastritis and intestinal metaplasia to the appearance of the dysplastic lesions, and finally reaching frank adenocarcinoma.

It is considered that each step of this cascade change doesn't merely only contain a morphological change but are actual accumulating of individual changes in the molecular level, which paralysing cellular homeostasis gradually. This initial model was postulated based of the population-level epidemiological observations, but further molecular analyses has proved significantly the mechanistic basis of this model, which shows the same pathways of inflammatory, oncogenic changing have got active on several levels of the cascade [3].The dominant role of H. pylori infection, a gram-negative bacterium, positioned by the International Agency for Research on Cancer as a Group 01 human carrot [3], is one of the leading reason behind the cascade's start in the vast majority of non-cardia gastric cancer patient's cases. The epidemiological data suggest that H. pylori infection was responsible for around 89% of intestinal-type gastric-carcinomas worldwide, but a few of the infected is going to malignancy, highlighting the impact of the host genetic in background, such as H. pylori virulence factors, co-exposure on the different foods, such as dietary salt, tobacco, and alcohol, for tumor oncogenic path [3, 2]. H. pylori promotion of carcinogenesis has complex mechanisms, it induces the direct mucosal inflammation induction via of virulence factors, such as Cagan and Vacalin; generation of oxidative stress to cause DNA non functional, and it induces the prolonged nuclear factor kappa B, and Janus kinase–signal transducer and activator of transcription signaling, which in turn make a mucosal environment for neoplastic change [3].

In the methodological context, early literature of gastric cancer was performing mainly based of population-based cohort studies, histopathological classification scheme, such as Lauren and World Health Organization schema and one-gene mutational analysis, but these studies have had limitation to obtain the complexity of a disease, which is characterized the vast molecular heterogeneity. The comprehensive molecular characterization of the gastric adenocarcinoma, published in 2014 by The Cancer Genome Atlas Research Network, is important methodological inflection point [2], has used six simultaneously molecular platforms, such as whole-exome sequencing, array-based copy number analysis, DNA methylation profile, messenger RNA, microRNA, reverse-phase protein arrays in a cohort of 295 primary gastric adenocarcinoma patients to identify four distinct molecular subtypes, tumours (Epstein–Barr virus), microsatellite unstable, the tumours genome stable with diffusely histologically enriched and chromosomally unstable, based on the integrative design of multi-platforms. This integrate framework for any other studies for the genetic cause of gastric adenocarcinomas. This study revealed important evidence that he heterogeneity at the phenotypic level is a direct consequence of the molecular heterogeneity and any classification to be used by a clinica point-of-care diagnosis must be the consideration of multiple level of the genomic information simultaneously [1]

2.2 Evolution of Bioinformatics and Network Medicine Frameworks in Gastric Disease Research

2.2.1 From Single-Gene to Systems-Level Approaches

In the 20 years that have elapsed, molecular gastric cancer research has progressed from genes-to-guts single gene hypothesis-driven studies to integrated, system-level studies that consider the biological phenomenon to be emergent properties of complex molecular networks instead of linear consequences of a single gene perturbation [24, 25]. This transition is driven by the emerging insight that the "oncogenic" phenotype of a cancer cell is not a direct consequence of the mutation or abnormal regulation of a single molecular entity but rather is the result of the global reorganisation of the topology of the regulatory network that is the context of its embedding. The concept of a systems perspective was fully developed in the network medicine framework, which stated that human diseases are perturbations of the human interactome, or the complete network of molecular interactions in a cell, and that diseases with similar aetiological or pathological characteristics lie on overlapping regions of the interactome, with common modules of molecular interactions that could be identified and characterised to gain insight into disease mechanisms and to uncover targets for intervention [24, 25].

The implementation of this framework in the field of stomach cancer studies has taken two main approaches:

In the field of stomach cancer research, the practical operationalisation of this framework has taken two methodologically main approaches. The first is building and analyzing protein-protein interaction networks based on curated molecular interaction databases, like STRING, BioGRID and the Human Protein Reference Database, which can systematically map the downstream effects of molecular interaction changes, for instance, the changes in protein activity caused by the dysregulation of a microRNA or transcription factor [25]. The second is the mining of curated biomedical literature databases to find disease-gene associations that are

statistically supported in published articles, for example, the Comparative Toxicogenomics Database, a manually curated database of chemical-gene and gene-disease, and chemical-disease relationships from the published biomedical literature that also provides inference scores reflecting the breadth and consistency of the supporting evidence for each association [7]. In fact these two tracks come together in integrated computational platforms like NetworkAnalyst offering the same functionality in one analytical environment for protein interaction network analysis, construction of tissue-specific co-expression networks and transcription factor-gene and microRNA-gene regulatory network analysis [27].

2.2.2 Identification of the protein–protein interaction networks in gastric cancer

In 2011, Tseng and colleagues used network-based reasoning to analyze microRNA-regulated protein interaction networks in gastric cancer molecular biology [25], one of the earliest uses of this technique in the field. Later network-based bioinformatics research has expanded upon this concept, utilizing increasingly complex analytical techniques. Wei's team extracted an integrated mRNA–microRNA–long non-coding RNA regulatory network in gastric tumor by mining various gene expression omnibus databases simultaneously and identified a common set of hub genes, CTHRC1, FNDC1 and INHBA, that were consistently up-regulated across gene expression omnibus datasets and independently correlated with poor patient survival [3]. This study's methodological design of cross-dataset integration to find differentially co-expressed genes, and multi-layer regulatory network construction, is a more mature expression of the network medicine paradigm than previous single-platform studies.
In 2023, a parallel systems biology analysis was conducted in BMC Gastroenterology, where the authors used gene and protein-with another protein interaction network reconstruction to select seven core genes in stomach adenocarcinoma and then investigated the expression pattern of these hub genes based on the GTEx portal, extending the network analysis approach to a developmental and ageing dimension [26].
These results highlight that the regulatory structure of gastric carcinogenesis is not just a two-dimensional one but a hierarchical multi-layer system where the circular RNAs and long non-coding RNAs function as the upstream regulator that affects the function of the downstream microRNA and transcription factor regulators.

2.3 Transcription Factor Networks and Master Regulatory Architecture in Gastric Carcinogenesis

2.3.1The role of the NF-kB pathway as a central inflammatory-oncogenic switch.

Among the transcription factors that have been implicated in the transition from chronic gastric inflammation to is, one of the most important in maintaining the chronic inflammatory state of the mucosa, which predisposes to the process of carcinogenesis [8]. Liang et al. have summarized the various mechanisms by which NF-κB contributes to gastric carcinogenesis, which goes far beyond just cytokine induction [8].They include the ability to upregulate the cyclooxygenase-2, which induces the production of prostaglandins that antagonize apoptosis and stimulate angiogenesis; stimulate the secretion of anti-apoptotic proteins of the Bcl-2 and IAP families, which make cells resistant to cytotoxicity; and regulate matrix metalloproteinases

and adhesion molecules and stimulate epithelial-mesenchymal transition. Importantly, NF-κB activity is not an independent process, but is extensively cross-talked by concurrent oncogenic signalling pathways, such as the JAK/STAT3 pathway.

It has been shown that NF-κB and STAT3 can act synergistically as a cooperative transcriptional complex at promoters of common target genes to regulate the expression of pro-metastatic factors, such as IL-6 and VEGF, and that their synergistic function is highly potentiated and increases the carcinogenic effect of either pathway alone [9].

2.3.2 The triplet of oncogenic transcription factors HIF-1α, MYC, and the Oncogenic Transcription Factor Triad

In addition to NF-κB and STAT3, two transcription factors, HIF-1α and MYC, are major coordinators of the gastric oncogenic programme, each of which controls distinct but interconnected processes. Hypoxia-inducible factor alpha subunit (HIF-1alpha) is a key mediator of cellular response to hypoxia, and is constantly activated in various solid cancers including gastric cancer, through hypoxia-dependent and hypoxia-independent mechanisms, involving oncogenic signalling pathways through PI3K/AKT/mTOR and MAPK cascades. Ucaryilmaz Metin and Ozcan comprehensively reviewed the role of HIF-1α in gastric carcinogenesis and found that it participates in essentially all the hallmarks of cancer including sustained proliferative signalling, evasion of growth suppressors, induction of angiogenesis by the upregulation of VEGFA, activation of invasion and metastasis by the regulation of matrix metalloproteinases and epithelial-mesenchymal transition markers, and the reprogramming of cellular metabolism toward aerobic glycolysis [10,11].

2.4. MicroRNA-Mediated Post-Transcriptional Regulation in Gastritis and Gastric Cancer

2.4.1 Canonical miRNA Regulatory Framework.

MicroRNAs are endogenous non-coding RNA molecules of ~18-25n in length which regulate gene expression at the post-transcriptional level, by binding to the 3'UTR of their target mRNA's, through imperfect complementary base-pairing, either causing translational repression or the degradation of the mRNA depending on the degree of complementarity [23].
complexity which expands its regulatory effects beyond the capacity of single gene studies. Cancer biology: microRNAs are either oncomiRs, which are overexpressed to promote tumour development by targeting tumour suppressor genes, or tumour suppressor microRNAs, such that loss of expression allows oncogenic programmes to proceed unchecked. The dichotomy described above is not binary, and many miRNAs appear to have context-dependent functions that vary based on the cellular context, the set of available target transcripts and the stage of disease development. The hallmarks of cancer framework is a helpful conceptual tool for understanding the physiological consequences of microRNA dysregulation: microRNA dysregulation collectively disrupts the processes normally limiting uncontrolled proliferation, escape from apoptosis, evade immune surveillance and maintain genomic integrity – all of which are deregulated in cancer [23,24].

2.2.2 A key microRNA involved in the Gastritis-Gastric Cancer Continuum

The exploration of microRNA deregulation in gastric cancer has resulted in a huge and growing body of literature. The oncomiRs most consistently reported in independent studies are perhaps miR-21, which has been most extensively characterised in the gastrointestinal context. The authors measured the blood constituent concentrations of miR-21, miR-223 and miR-218 using quantitative reverse transcription (RT) PCR with synthetic spike-in controls for normalisation, and found that both miRNA were significantly elevated in cancer patients from all cancer stages, including stage I, whereas miR-218 was significantly decreased in cancer patients. When combining the ratio of (miR-the receiver operating characteristic curve area under the curve was 0.9531 and the sensitivity and specificity were 84.29% and 92.86% respectively, markedly better than the individual microRNAs and showed that panel-based methods were superior to single-marker approaches [12].
The adoption of microRNA biomarkers in clinical trials poses methodological hurdles and challenges.

The use of microRNA biomarkers in clinical trials presents methodological challenges and consistency.

This enthusiasm to turn microRNA expression signatures into clinically useful non-invasive diagnostics has uncovered a number of persistent methodological issues that help to account for the discrepancy between the level of mechanistic evidence supporting these efforts and the number of diagnosable assays that are established in the clinic. One of the key technical challenges with microRNA circulation is the normalisation issue, which occurs because of the variation in storage temperature, as well as the variation in endogenous microRNA concentrations among individuals and disease states, and the impact of haemolysis on the concentration of microRNAs during sample processing [13]. Different normalisation approaches (e.g. synthetic spike-in controls, global mean normalisation or ratios of endogenous reference microRNAs) lead to partially conflicting absolute expression values which makes direct comparisons between different studies and meta-analysis difficult [13, 22].Another potential complication in the stomach cancer context is the effect of infection itself on the expression of certain plasma miRNAs that may create a confounder that could give false positive signals for certain miRNAs in study design if H. pylori infection status is not accounted for [13]. However, from the available literature, it is clear that the multi-microRNA panel approaches can achieve much greater diagnostic accuracy (AUC values above 0.95) than single microRNA approaches, and that the diagnostic accuracy of multi-microRNA approaches is significantly greater than single microRNA approaches, with the former showing that the combinatorial regulatory logic of the underlying molecular networks is the reason for this [13, 22].
Computationally derived microRNA network hub candidates, when combined with plasma biomarker validation in the clinic, is the most promising avenue for early detection tests for gastric cancer that are clinically translatable [12, 13, 19, 22].

2.5.Hub Gene Identification and Functional Enrichment Analyses

2.5.1 Bioinformatics Approaches to Hub Gene Discovery

One of the very crucial goals of computational stomach cancer studies is the identification of hub genes: genes that are in a position of disproportionate connectivity and regulatory influence within molecular networks. The well-established methodological workflow for this approach consists of multiple consecutive steps: downloading the gene expression profile datasets from public data repositories like the Gene Expression Omnibus or Cancer Genome Atlas; performing differential expression analysis using statistical methods like limma, edgeR or DESeq2 to detect significant changes in the expression of transcripts; building protein-protein interaction networks on platforms like STRING or Cytoscape; calculating network centrality measures various important attributes to find hub genes in the network; and finally validation of the hub genes expression using survival analysis in large independent cohort using resources such as the Kaplan-Meier Plotter database. The workflow has been applied to various levels of rigor in the literature reviewed, and consistency of results across studies which used different datasets and analysis parameters gives some validation for the biological relevance of hub genes.A small number of key genes are consistently identified across independently performed hub gene identification studies using integrated bioinformatics analyses. The most frequently found down-regulated hubs in gastric cancer tissue are genes encoding tumour suppressor proteins such as TP53, CDKN2A (encoding p16) and PTEN, while oncogenic genes such as EGFR, AKT1, VEGFA, MYC, and BCL2 are most commonly identified as up-regulated hubs. The functional enrichment analysis of differentially expressed gene sets in gastric cancer, using Gene Ontology biological process categories and KEGG pathway databases, successfully implicates a set of core pathological processes: pro-inflammatory cytokine signalling intrinsic and extrinsic apoptotic pathway disruption;cell cycle checkpoint deregulation; reactive oxygen species-mediated oxidative stress response; and extracellular matrix remodelling processes, which are associated with invasion and metastasis [3, 8].

2.5.2 Co-expression Modules in tissues and their clinical relevance

One key methodological development with hub gene analysis has been to move from tissue-agnostic network analysis to stomach-specific co-expression analyses to reflect the transcriptional peculiarities of gastric mucosal tissue. Using curated stomach expression data from resources such as GTEx portal and the Human Protein Atlas, weighted gene co-expression network analysis (WGCNA) and tissue-specific combined expression modules have been used to identify sets of genes that are not only differentially expressed in cancer and normal tissue, but also genes with co-ordinated expression patterns that reflect functionally coherent transcriptional programmes that are uniquely operating within the gastric tissue microenvironment [26].

The first of these is enriched with genes that regulate innate immune activation and cytokine-mediated inflammatory signalling, pathways that are continuously activated in gastritis associated with H. pylori infection. The second module focuses on genes involved in the integrity of the epithelial barrier, tight junction assembly, and maintenance of the epithelial barrier, which are representative of the progressive loss of the epithelial architecture leading to an atrophic gastritis and intestinal metaplasia. The third module consists of genes more specifically linked to the processes of cell cycle progression, angiogenesis and extracellular

matrix remodelling, which relate to tumour establishment and invasive growth. This identification of this 'modular' organisation has direct implications for the design of a biomarker panel: markers should come from the three functional modules; information provided by markers from each module should be different, and not redundant, from markers of the other modules (1,3,8).

2.6 Contradictions, Inconsistencies, and Methodological Limitations in the Existing Literature

2.6.1 Algorithms for miRNA prediction and validation

The one of the most recurring conflicts in the microRNA research literature is the difference between the number of microRNA-target interactions predicted computationally and the number that have been validated by experiments. Studies that include only computationally predicted interactions without experimental evidence may therefore contain artefactual relationships, and miss biologically relevant interactions that are not detected by the algorithms. Experimental interaction data from curated databases, like miRTarBase containing only direct biochemical evidence of binding, including luciferase reporter assays, western blotting, and CLIP-seq, greatly enrich the biological context for network analyses, but unfortunately at the expense of network size and completeness [17].Also, the context-dependency of the interactions between miRNAs and their targets adds another methodological limitation as these interactions can differ significantly between cell types, disease stages, and experimental conditions. The binding of a specific target validated in a hepatocellular carcinoma cell line might result in completely different regulatory relationships in the gastric epithelial context, due to differences in the secondary structure of the RNA, competition for binding to the RNA by other proteins, and differences in the transcription availability of the target mRNA. The Cai et al. circHIPK3 study was explicit in stating that although there was strong correlative evidence that circHIPK3 bound with miR-124 and miR-29b, it was not directly validated in gastric cancer cells but upon the previously published data from other cancer types [12].

2.6.2 TCGA Subtype Stratification: One Unknown Dimension in the majority of Network Analyses

Perhaps the most important new development in the conceptualization of gastric oncology over the last decade has been the identification of gastric cancer as four distinct molecular subtypes by the Cancer Genome Atlas [16]. This is further turn on by the fact that different molecular properties of each subtype make them sensitive to the same molecular inputs differently. Tumours from patients with EBV are hypermethylated overall – hypermethylation is the highest reported in all cancer types in the Cancer Genome Atlas – with extreme methylation of genes such as CDKN2A, which epigenetically silences p16 and takes away a key brake on cell cycle progression [16]. Contrastingly, the mutational landscape of microsatellite unstable (MSI) tumours is one of high mutational load and hypermethylation of the mismatch repair (MMR) gene MLH1, which provides a completely different background for the function of microRNA and transcription factor regulatory networks. These molecular contexts—specific to each subtype—must be taken into account in network analyses in order to provide the most accurate translation of computational results and will be systematically addressed in future studies.

To explore the functional duality of key regulatory molecules :

A hub gene found as up-regulated in pooled gastric cancer tissue might actually be more relevant to the chromosomally unstable subtype, not so much to EBV-positive or the genomically stable subtypes, but then its biomarker or therapeutic value could get wrongly stretched to everyone if people don't do subtype stratified analysis. This issue shows up extra sharply in the microRNA biomarker literature, where most studies look at microRNA expression between gastric cancer patients overall and healthy controls, but they often don't check whether the proposed microRNA signature keeps working in the same way across all four molecular subtypes , or whether it is only informative for a certain subtype-defined subgroup [12, 13].
Because EBV-positive tumours tend to show extreme CpG island methylation that shuts down huge sets of microRNA gene loci, microsatellite unstable tumours carry higher frequencies of somatic mutations in microRNA processing genes, and the diffuse-type genomically stable tumours come from epithelial origins that are basically not the same as those of the intestinal-type chromosomally unstable tumours, it's pretty hard to believe that one single microRNA expression pattern would remain equally telling across every subtype [16]. Filling this gap by doing analyses that are stratified by subtype is, in other words, one of the main methodological steps that future studies should treat as a priority, especially when they build on the computational groundwork already laid by the current literature [16, 19, 21].

2.7 Identified Research Gaps & Emerging Frontiers

There is a need for a comparative analysis of Gastritis and Gastric Cancer .
The most significant missing element in the literature identified by the systematic review is the lack of published research that addresses the shared molecular regulatory network of gastritis and gastric cancer in a coherent computational model. Most published bioinformatic studies used a reference set of gastric cancer samples and compared them to normal adjacent or healthy gastric mucosa, but not using gastritis tissue or gastritis-related gene signatures as a separate reference [1, 3, 6]. In the design this is assumed that the transition from gastritis to cancer is a binary molecular transition, a progression rather than a continuum in which the inflammatory and oncogenic regulatory programmes gradually overlap and reinforce each other [15, 17]. The consequence is that molecular features initiating during the gastritis phase — which may represent the earliest detectable signatures of malignant transformation risk — are systematically excluded from analyses that are anchored exclusively to the cancer endpoint [12, 17].
Thus, the new approach of simultaneously querying the Comparative Toxicogenomics Database for disease-associated gene sets for gastritis and gastric cancer, followed by the identification of an overlap of gene sets as the shared molecular basis for these two diseases, and the development of multi-layered regulatory networks using the common gene set directly tackles this issue [7, 27]. This analysis, which is based on disease conditions at both inflammatory precursor and malignant endpoint, allows the identification of regulatory molecules and network modules that are active in both conditions, thus uncovering the molecular machinery by which chronic gastritis generates conditions for oncogenic transformation. The microRNAs and transcription factors that are hubs in this common regulatory network are especially promising as potential early-stage markers for gastric cancer

because they have been found to be dysregulated in both gastritis and gastric cancer and may be measurable in high-risk gastritis patients before the onset of malignancy [12, 13, 22].

2.7.1 Epigenetic Integration into network frameworks

The transcriptomic and protein interaction aspects of gastric carcinogenesis have been studied extensively in the context of network medicine; however, less systematic research has focused on the epigenetic regulatory component, defined as changes in DNA methylation -and changes in chromatin remodelling dynamics, during gastric carcinogenesis [16, 20]. Epigenetic dysregulation was shown to be a primary molecular characteristic because of extreme CpG island methylator phenotype in EBV-positive gastric cancer and DNA hypermethylation of the MLH1 mismatch repair gene as the mechanistic basis of microsatellite instability in the MSI molecular subtype [16]. The incorporation of genome-wide DNA methylation profiles, histone modification ChIP-sequencing data and chromatin accessibility information into models of gene regulatory networks is still technically and analytically challenging, and will be an important area for methodological innovation in this area [16, 21].

Given the emerging evidence of the role of specific microRNA expression changes in epigenetic reprogramming events in gastric cancer, this new layer of network complexity is particularly relevant, one that is currently not included in single-layer analyses [20, 21]. Examples of transcription factor dysregulation leading to altered epigenetic remodelling that can persist without the original upstream stimulus include aberrant activation of STAT3 by the epigenetic kinase, MSK1, leading to epigenetic changes in the phosphorylation of H3 at target gene promoters [20]. The use of these transcription factor-epigenetic regulatory feedback loops in multi-layer network models of gastric carcinogenesis opens a fascinating opportunity for future investigations, to take advantage of the basic computational studies presented above [20, 21].

2.7.2 The validation of clinical translation and liquid biopsy.

The largest disconnect between the current body of computational literature and its clinical use is the lack of prospective clinical validation of computationally derived molecular signatures in actual patients. While the pipeline of discoveries that include the identification of genes or microRNAs, based on databases, move to network analysis and then to a clinical deployment as a diagnostic or prognostic biomarker assay is fraught with many validation steps that are of increasing stringency, from retrospective cross sectional case-control studies to prospective cohort studies to multi-centre population-based screening trials, few computational discoveries have made it past this pipeline to reach clinical practice [13, 22]. The ability of biomarker panels defined in single-centre retrospective studies to discriminate between different groups may not be the same in multi-centre cohorts of prospective H. pylori patients with different demographic and H. pylori epidemiological features; and the analytical platforms that are needed to quantify microRNA molecules are not fully standardized across clinical laboratory settings [13, 22].

Despite these challenges, the combination of plasma circulating microRNA quantification, methylation profile of plasma circulating tumour DNA and analysis of plasma-derived exosome nucleic acids with computational identified biomarker signatures appears to be the most promising way to move forward towards clinical translation [19, 22]. Overall, the stability of the plasma and serum microRNAs during standardised collection and storage, the availability of sensitive quantitative PCR and next-generation sequencing platforms for detection, and

preliminary clinical evidence of high AUCs for composite panels of microRNAs identified through computational analyses suggest that commercially available, clinically deployable liquid biopsy assays of hub microRNAs are feasible [13, 22]. The future, multi-centre, population-stratified clinical validation of such panels to determine their sensitivity, specificity and positive predictive value to validate in real-world gastrocancer and gastritis surveillance would be highly desirable [19, 21, 22].

SUMMARY AND SYNTHESIS

The body of literature sampled in this chapter collectively demonstrates that gastric cancer is not a pathology of discrete molecular events but a systems-level disease resulting from the stepwise reorganisation of interlinked gene regulatory networks functioning simultaneously at multiple biological layers. The epistemological basis for this conceptualisation lies in the Correa cascade conceptualising the progression of chronic gastritis through intestinal metaplasia to invasive adenocarcinoma as a continuous molecular slope, primarily fuelled by a persistent Helicobacter pylori infection rather than a black-and-white onogenic switch [15,16]. The systematic molecular subtyping of gastric adenocarcinoma into four biologically distinct entities by The Cancer Genomics Atlas validated that the phenotypic heterogeneity observed at the clinical stage are but the direct manifestation of dissimilar regulatory architectures manifested at the molecular level and no single biomarker or therapeutic target could comprehensively encompass the full spectrum of the disease [16]. The methodological transformation of the field has left an appreciable impact and served as a paradigm shift from single gene based hypothesis testing to network based models, wherein oncogenesis is perceived as an emergent property of network topology, not linear outcome of single gene perturbations [24,25]. Studies developing protein-protein interaction networks microRNA-gene regulatory networks, transcription factor-gene interaction networks, circRNAs, miRNAs and single cell RNA-seq, and last but not theleast the triple-layer circRNA-miRNA-mRNA regulatory architecture have shown that a small and reproducible set of hub molecules- most consistently miR-21, STAT3, NF-kappa-B, HIF1A, MYC, TP53 and VEGFA hold positions linked to disproportionate regulatory influence in the case when these network are generated independently using different databases, cohorts and analytical methods [1,3,8,10]. The cross-methodological convergence on the same nodes represent best evidence available regarding hub molecules in gastric carcinogenesis, making them as the most prominent candidates for biomarker and therapeutic target developments. The study identifies on the level of transcription factors cooperative NF-kappa_B and STAT3 axis being the predominant regulatory switch linking H. Pylori induced mucosal inflammatory response to oncogenic transcriptional program of established gastric cancer [8,9]. The HIF1A expanded these transcriptional controls to angiogenic and metabolic reprogramming dimensions of the cancer hallmarks while the MYC up-regulated proliferative and ribosomal biogenesis programs in an reciprocally self reinforcing feedback loop with STAT3 [11,10]. At the post-transcriptional

level, miR 21 and miR-223/ miR-218 axis controlled the overlapping set of tumour suppressor targets and its dysregulation was observable in blood of gastritis patients as well, identifying it as markers of early stage of malignant transformation from established stages rather than entities for established disease [12,13]. The mechanistic fertility has been accompanied with gaps in the literature and which were specifically addressed in the present investigation. Firstly, most studies focused on gastric cancer exclusively separate of its inflammatory precursor, consequently losing the mechanistic continuity of the Correa cascade and fail to explore the molecular features of tumourigenicity active during the phase of gastritis which are the most clinically important early biomarkers [15,17]. Secondly, the major studies focused on one layer of regulatory network e.g microRNA-gene or transcription factor-gene or protein interaction networks and do not combine these layer together within a unified multi network framework to depict the hierarchical regulatory logic which arise due to their interplay [1,2,5]. Finally, single gene expression dataset obtained from minimal patient cohort compromises the population representativeness of the computational findings while the disease-gene association databases like CTD construct the best evidential base for construction of network [7,28] (an online learning module to study the evolution of gene expression in different species available at, https://online-learning.ctdbase.org/). The translational vista of which all the aforementioned computational findings is for the development of the non-invasive liquid biopsy biomarker panels based on the hub miRNAs and transcription factors signatures identified by network analysis that are able to detect the pre-malignant transformation in gastritis patients prior the development of clinically overt cancer [13,22]. The present investigation is a contribution for the above translational vista by presenting the computationally validated catalogue of shared molecular regulatory machinery of gastritis and gastric cancer which are developed employing an integrated multi-layer network analytical framework through anchoring to most comprehensive available resource for disease-gene associations as a stringent background for further experimental validation and clinical translation research [7,27,29,30].

MATERIAL AND METHODS

CHAPTER 3

3.1 Data Source: Comparative Toxicogenomics Database (CTD)

Gene-disease association data for chronic gastritis and gastric cancer (stomach neoplasms) were retrieved from the Comparative Toxicogenomics Database ,a publicly accessible access, manually curated bioinformatic repository that systematically catalogs chemical-gene-disease interactions compiled from the peer-reviewed biomedical literature. The decision in favor of CTD as the primary data source was made based upon three defining characteristics that distinguish CTD from other publicly available gene-disease databases.

First, CTD employs an expert manual curation model, where trained biomedical scientists directly assemble molecular interaction data from primary published sources, and manually annotate genes, diseases and chemicals with standardized ontological terms . This ensures much higher annotation accuracy and biological specificity than could be feasible if using automated text-mining pipelines.

Second, CTD makes use of controlled vocabularies at all levels. For instance, disease entities are mapped to Medical Subject Headings (MeSH) terminology; gene identifiers are cross-referenced to NCBI Gene; and chemical entities are harmonized with MeSH and ChemIDplus nomenclature. The ontological consistency allows unambiguous comparative analyses between disease categories.

Third, CTD calculates an inference score for each gene–disease pair, a proprietary statistical measure based on co-occurrence frequencies in the curated literature, allowing evidence-weighted gene prioritization rather than simple frequency-based ranking.

3.1.1 Gene Retrieved and Association Filtering

For each disease of interest (chronic gastritis and gastric cancer, MeSH: D005756 and D013274, respectively), we searched the CTD gene-disease association module. The extraction was limited to curated direct associations of two types:

• Markers/Mechanisms: Genes that have been experimentally demonstrated as disease biomarkers or functionally implicated in the molecular mechanisms of the disease, including differentially expressed genes, genes harboring somatic mutations or copy number variations, and genes reported to undergo post-translational modification in the context of the disease.

• Therapeutic Interactions: Genes identified as molecular targets of therapeutic agents used in disease management, or genes whose modulation has shown a disease modifying effect in experimental models, thereby enriching the candidate gene set with clinically actionable and functionally validated molecular targets

The first extraction depended on biological specificity, and indirect or computationally inferred

associations, based on intermediate chemical or molecular relationships rather than direct experimental evidence, were omitted.

3.2. Ranking by Two Criteria and Top 200 Genes Selected

The targeted CTD extraction protocol as described in the previous section resulted in rather large raw gene lists for each disease. The query on chronic gastritis produced several thousand candidate gene entries, while the query on gastric cancer produced an even larger initial set, together reflecting the extensive and heterogeneous body of published research accumulated on both conditions over decades of molecular investigation. While this breadth is indicative of the depth of current scientific engagement with these diseases, it also presents a major analytical challenge: not all reported gene-disease associations carry the same weight of evidence, and the lower-ranked entries in such comprehensive lists often contain weakly supported, context-limited or methodologically narrow associations that may not robustly reflect core disease biology.

To meet overcoming this obstacle and to reduce the raw gene universe to an analytically manageable, high-confidence subset, we systematically applied a dual-criterion evidence-weighted ranking strategy to each disease-specific gene list prior to any downstream analysis. This approach was aimed to simultaneously reward statistical breadth of association (measured through an inference-based metric) and experimental depth of documentation (measured through a direct curation count), so identifying genes that are broadly and specifically implicated in disease biology.

Criterion (a)
The CTD Inference Score is a proprietary, computationally derived statistical metric that quantifies the strength of a gene–disease association based on systematic patterns of co-occurrence in the biomedical literature. The score is calculated by analyzing the frequency with which a given gene and a given disease term are mentioned in proximity across indexed PubMed records, and normalizing this observed co-occurrence frequency against the expected background frequency of citation for each entity individually.
This normalization step is important: it corrects for the confounding effect of publication bias, in which highly studied genes (such as TP53 or VEGFA) and highly studied diseases tend to co-occur more frequently in the literature by virtue of their overall citation volume, irrespective of any specific biological relationship

A high CTD Inference Score thus means that a gene is co-mentioned with the disease in a significantly larger number of independent publications than would be expected by chance, giving strong statistical evidence that the gene–disease association is not simply an artifact of general research popularity but is indicative of a true biological association repeatedly reported in the literature.
The inference score, in the context of this study, was a measure of the breadth and statistical robustness of the gene-disease relationship across the aggregate literature and reflects associations that have been documented independently across diverse study designs, laboratories, and experimental systems.

Criterion (b) – Curation of Literature Volume
The Literature Curation Volume was defined as the total number of unique peer-reviewed publications in which a carefully selected causal relationship across a particular gene and the disease subject of study has been manually annotated by CTD biocurators. The curation volume is an empirical count of publications in which a trained human expert read the primary research article, made a judgment that the gene-disease association was experimentally supported, and recorded the association with

controlled vocabulary in the CTD database. In contrast to the inference score, which is a statistical construct based on automated co-occurrence analysis.
This measure therefore represents an entirely unique aspect of evidence for association: replication and research consensus. If a gene has been recorded in numerous independent publications, then there have been multiple findings demonstrating a correlation in multiple experiments, on multiple patient populations, on multiple types of cells or animals or using multiple analytical techniques, suggesting that it is not a one-off discovery or method-specific finding, but a biologically replicable result. The measure of the number of literature curations is therefore a measure of the confidence of the scientific community in a specific gene-disease association, and can be used as a complement to the inference score.

3.4 Intersection of Diseases: Identification of Shared Genes

One of the main target of this study was the elucidation of common molecular pathways that exist between chronic gastritis and gastric cancer, two pathologies whose clinical/histopathological sequence (Correa cascade) presupposes the presence of common molecular drivers behind neoplastic transformation of inflamed tissues. With this objective in mind, a cross-disease analysis of the two top-200 genes identified in the previous section was performed.

Intersection of two gene sets in terms of the presence of identical genes was done via a set-based comparison, i.e., genes in the two lists were considered elements of mathematical sets which are compared for overlap. Gene identity was determined only on the basis of NCBI Gene identifier (Entrez Gene ID)—the unique numerical identifier assigned to each locus of every human gene by the National Center for Biotechnology Information.

It was decided to use NCBI Gene IDs as the criteria for gene matching to minimize errors of identification, since the literature on biomedical topics has long been notorious for variability in naming conventions for any particular gene—a gene can have more than one symbol across various articles and literature streams, e.g., HER2 can also be found under the names ERBB2 or NEU, while CDKN2A can also be called p16INK4a or MTS1.

Using only gene names could result not only in underestimating the number of shared genes in two lists (due to absence of identical names) but also in overestimation (confusing two separate but similar genes based on names). Thus, using NCBI Gene IDs as a common denominator for gene matching allows one to avoid such inconsistencies and ensure precision of results.

Table 3.1 Gene Symbol And Full Gene Names Of Common Genes Between Gastritis And Stomach Cancer

S.No.	Gene Symbol	Full Gene Name
1	IL6	Interleukin-6
2	MAPK1	Mitogen-Activated Protein Kinase 1 (ERK2)
3	TP53	Tumour Protein p53
4	RGS2	Regulator of G-Protein Signalling 2
5	CDKN1A	Cyclin-Dependent Kinase Inhibitor 1A (p21)
6	JUN	Jun Proto-Oncogene
7	IL1B	Interleukin-1 Beta
8	ALB	Albumin
9	ENO1	Enolase 1 (Alpha-Enolase)
10	SOD2	Superoxide Dismutase 2
11	CCND1	Cyclin D1
12	GAST	Gastrin
13	IL1RN	Interleukin-1 Receptor Antagonist
14	PTGS2	Prostaglandin-Endoperoxide Synthase 2 (COX-2)
15	GADD45A	Growth Arrest and DNA-Damage-Inducible Alpha
16	CEBPB	CCAAT/Enhancer-Binding Protein Beta
17	HMOX1	Haem Oxygenase 1
18	CASP8	Caspase-8
19	RELA	RELA Proto-Oncogene (NF-κB p65 Subunit)
20	MAPK3	Mitogen-Activated Protein Kinase 3 (ERK1)
21	NOS3	Nitric Oxide Synthase 3 (Endothelial)
22	SERPINE1	Serpin Family E Member 1 (PAI-1)
23	MYC	MYC Proto-Oncogene (c-Myc)

3.5 Network Construction and Analysis Using NetworkAnalyst

This will enable to build and analyze a network using NetworkAnalyst 3.0.
All network construction, topology characterisation and visualization was performed in the web-based, integrated, computational environment specifically designed for multi-omics network analysis and visualization of biological networks: NetworkAnalyst 3.0 .
NetworkAnalyst 3.0 offers a unified environment for analysis that can generate a variety of classes of molecular interaction networks from a database of curated, experimentally validated and computationally predicted interactions.NetworkAnalyst 3.0 is intended to create a range of classes of molecular interaction networks using a curated set of experimentally validated and computationally derived interaction databases. This had the advantage that it allowed the multi-network approach of systems biology to be applied in this study, as it facilitated amalgamation of multiple data sets in a single, repeatable analysis.

We used the common gene set as the common query input for the three network construction modules described below. This method provided us with a consistency of analysis and enabled us to compare the same topologically related, biologically prioritized molecular candidates across networks.

3.5.1Gene–miRNA Interaction Network
These are some categories you can use in order to build your network.These are the types of nodes you can create a network from.
The common gene set was used as an input query to create a gene–miRNA interaction network in the miRNA interaction module of NetworkAnalyst 3.0. Experimentally verified miRNA–target relationships were obtained from two main complementary databases for constructing a network:

Manually curated miRNA-target interaction database miRTarBase includes experimentally validated miRNA-target interactions, identified by a series of rigorous functional assays, , transcription PCR (qRT-PCR) and crosslinking immunoprecipitation followed by high throughput sequencing (CLIP-seq). The interactions of miRNAs with targets in miRTarBase are the most confident ones, and each one should be supported by direct experimental evidence of a functional regulatory relationship between the miRNA and the target, so that each interaction is based on two independent experiments.

A complementary and curated database of experimentally-supported miRNA–gene regulatory interactions, and also represents those miRNA pairs that are known to regulate the same gene and are not necessarily included in miRTarBase, to ensure that as large as possible portion of the common set of genes in the interaction landscape is covered.

One methodological decision that was taken was to focus on purpose the experimentally reported interaction data from these two databases, instead of the computationally predicted miRNA–target pairs generated by the seed-sequence matching algorithms, aiming to maximize the biological reliability and specificity of the output network. While useful for making predictions, which can be used to test hypotheses, these predictions have high false positive rates and can cause non-biological regulatory connections to be shown in the network topology.

The identification of the hub microRNA was performed by topological analysis.
After the network construction, the topology of the gene–miRNA interaction network was analyzed based on a system of well-established graph-theoretic centrality measures representing different facets of node-level importance in the network topology:

Degree Centrality: Number of direct interaction partners (edges) connected to each node, which represents a simple measure of local connectivity. High degree centrality nodes are connected to a high number of other members of the network and therefore can exert a regulatory influence over a number of targets at once.

Betweenness Centrality: The proportion of shortest communication paths (geodesic paths) of a node that connect other nodes in the network. These nodes with high betweenness centrality are important information bottlenecks, regulatory connectors, or bridges in networks and their perturbations have a disproportionate effect on the network-wide flow of information and signal propagation.

Closeness Centrality: The average inverse shortest path length from a node to every other node in the network, measuring the efficiency with which a node in a network can communicate with, and be reached by, the other nodes in the network. A node having high closeness centrality has a structurally central position in the network that allows for a fast propagation of regulation throughout the network. The Hub miRNAs were operationally defined as nodes that had above average values for both degree centrality and betweenness centrality in the network, similar to the definition used for Hubs in biological network analysis.

The dual criterion definition assures that nodes classified as hubs are not only densely connected within their local neighbourhoods but also are nodes of structural significance in the global topology of the network. Furthermore, the resulting network had a scale-free topology and was evaluated by a fit to a power-law distribution for degrees, in which a good fit to a power-law degree distribution was a prerequisite for the constructed network to have the topological properties of a biologically realistic regulatory network (which are, that most nodes have small numbers of links, while a few are hub nodes, with many more links than the rest).

3.5.2Tissue-Specific Co-expression Network

The biological rationale and database sources are discussed.The biological rationale and database sources are discussed.
A tissue-specific co-expression network was created with stomach data from two large-scale, publicly available transcriptomic repositories, NetworkAnalyst 3.0, to interrogate the coordinated transcriptional behaviour of the common genes in the specific physiological context of the gastric tissue; and thus to ground the analysis in tissue-relevant biology, rather than pan-tissue or cell-line derived transcriptomics.

GTEx (Genotype-Tissue Expression Project): A comprehensive reference resource for bulk RNA sequencing data in 54 tissue types and cell types from more than 900 donors, which allows tissue-resolved co-expression relationships by bulk RNA sequencing data to be derived that reflect in vivo transcriptional regulation in histologically normal human tissue.

ARCHS4 (All RNA-seq and ChIP-seq Sample and Signature Search): A uniformly processed compendium of publicly deposited RNA-seq datasets available from the Gene Expression Omnibus (GEO) and the Sequence Read Archive (SRA) that offers a complementary and quantitatively rich

transcriptomic reference that complements GTEx with additional expression profiles derived from the stomach, including those in disease context.

Experience the exclusive benefits of Network Construction and Edge Criteria.
The resulting co-expression network consists of nodes which correspond to a common gene, and edges that correspond to statistically significant co-expression relationships between pairs of genes, identified by the Pearson correlation coefficient above a certain threshold and corrected for multiple testing .The correlation threshold was set to be strict, with the nodes that are not co-expressed at that level removed from the network after multiple testing correction, thus excluding spurious correlations that could mask module detection and interpretation of function.

Find community structure and characterize functions on modules.Detect community structure and functional characterization of modules.
A co-expression network was then created and analyzed using community detection methods based on the Louvain modularity optimization algorithm, a graph partitioning algorithm that detects communities (or modules) by maximizing the score Q, which quantifies the difference between the proportion of edges within a detected community and the proportion of edges expected by a null model with the same degree sequence. The Louvain algorithm was chosen due to its computational speed and its ability to identify meaningful gene co-expression modules in large-scale transcriptomic networks that are biologically interpretable.

The modules were defined based on co-expression modules; that is gene clusters that have a high edge density within the module and a low density of edges between different modules, which have been reported as related modules of genes with similar functions that are co-regulated by common upstream regulatory mechanisms. Functional enrichment analysis of each identified module was performed by using Gene Ontology (GO) biological processes terms and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway annotations, and significance determined by hypergeometric test with Benjamini–Hochberg correction. This enrichment analysis allowed the biological interpretation of each module of co-expression, as a collective group of genes representing a group of cellular functions, metabolic pathways, or signalling cascades.

3.5.3 Transcription Factor–Gene Interaction Network
Having experience with Network Construction and Database Sources.Experience in Network Construction and Database Sources.
The common gene set was used to build a transcription factor–gene regulatory interaction network using curated regulatory interaction data that was downloaded and incorporated into the NetworkAnalyst 3.0 transcription factor network module. The main data sources were two complementary, authoritative databases of transcription factor–target gene relationships:

ENCODE (Encyclopedia of DNA Elements): A large-scale functional genomics consortium making available experimentally derived data on transcription factor-binding sites from the human genome in a variety of human cells, primary cell types, and tissue types by way of chromatin immunoprecipitation followed by high-throughput sequencing .The ENCODE ChIP-seq data offer direct experimental evidence of transcription factor binding at specific genomic regions, and identify regulatory relationships between transcription factors and their target genes at high confidence in a cell-type-specific fashion.

JASPAR: curated, open access non-redundant transcription factor DNA-binding profiles as position

weight matrices (PWMs) and position frequency matrices (PFMs) from in vitro and in vivo binding experiments. JASPAR PWMs allow the computational prediction of transcription factor binding to target genes at the regulatory regions (promoters, enhancers) based on models derived from the sequence-level affinity of transcription factors, thus complementing the regulatory relationship coverage provided by the ENCODE ChIP-seq data, especially for transcription factors that are not yet profiled; they allow to identify putative binding events at the regulatory loci of the common genes. The incorporation of both ENCODE experimental binding data and JASPAR computational prediction data resulted in a comprehensive view of the regulatory interaction landscape for the common set of genes, no matter which source of data was used.

Understanding the Topological Analysis and Master Regulator Identification.
Similar graph-theoretic, centrality analyses were performed on the resulting transcription factor–gene regulatory network as in the gene–miRNA interaction network (Section 2.3.1), consisting of computing degree centrality, betweenness centrality, and closeness centrality for each node in the network. Transcription factors with high regulatory control were identified as transcription factors with high degree and betweenness centrality scores compared to the distribution of the rest of the transcription factors in the network (high-centrality transcription factors), that is, transcription factors that regulate multiple common genes at the same time, covering functionally distinct biological pathways and processes.

The identification of master regulators is of particular analytical significance within the context of this study, as transcription factors occupying structurally central positions within the regulatory network are capable of orchestrating coordinated transcriptional programmes across multiple downstream target genes.
Somatic mutations, epigenetic silencing and pharmacological modulation of their perturbation can have pleiotropic effects on the molecular landscape of gastric pathophysiology. Given their topological centrality in the network and their independent biological functions in gastric inflammation, mucosal homeostasis, and neoplastic transformation, as described in the main literature, master transcription factors were thus taken as key targets to investigate their mechanism of action.

RESULTS

CHAPTER 4

4.1 Common Gene Identification

4.1.1 Shared Molecular Targets Across Chronic Gastritis and Gastric Cancer

By comparing the top 200 CTD curated disease related genes for chronic gastritis and gastric cancer , a specific subset showed up that seems to be shared by both conditions. This overlapping gene collection, sort of the core starting point, became the analytical base for every later network mapping step and the enrichment work done in this study.

Those shared genes were not just scattered in a random way across biological space, instead they appear to cluster at functional hubs that relate to inflammation, cellular proliferation , apoptotic control and epithelial remodeling . These are the kinds of processes that are often tied to the stepwise changes in gastric mucosa that happen as the disease worsens. The overall distribution, and also the exact names of these overlapping genes, are laid out in (Table 3.1)where the intersection is treated as candidates with dual pathological meaning , not simply as markers unique to one disease alone.

Also, a few of the higher confidence genes in the shared group have already been linked to cytokine driven signaling and immune microenvironment management. So , that lines up with what you would expect when chronic mucosal inflammation gradually turns into more oncogenic progression, kind of at that overlap where both sides interact.

4.1. 2 Functional assignment of common gene set

Gene Ontology , (GO) enrichment and KEGG pathway annotation of the shared gene set showed statistically significant overrepresentation of several biological processes and routes, each one can be considered a well established sign of oncogenic progression, especially when chronic gastrointestinal inflammation is involved

• Pro-inflammatory cytokine signaling : there was a clear enrichment for genes encoding or being regulated by interleukins (IL-6, IL-1β) and tumour necrosis factor (TNF). this matched the idea of a chronic NF-κB- and JAK-STAT driven inflammatory environment in H. pylori associated gastritis, and that environment helps build a carcinogenic microhabitat, via ongoing oxidative DNA damage , plus epigenetic reconfiguration.

• Apoptotic pathway regulation : genes involved in both intrinsic apoptotic pathways (mitochondrial) and extrinsic apoptotic cascades (death receptor mediated) were enriched. this included pro-apoptotic effectors, together with anti-apoptotic regulators whose mismanagement, can prevent the proper removal of mucosal cells that have genomic damage.

• Cell cycle checkpoint regulation : a strong overrepresentation was observed for genes connected to G1/S and G2/M checkpoint control, and especially those tied to cyclin dependent kinase regulators.

this suggests a steady skipping of cell cycle surveillance controls, which is typical for the gastritis to dysplasia shift.

• Oxidative stress response : there was enrichment for genes that code for reactive oxygen species (ROS) scavenging components and cytoprotective enzymes. overall this points to oxidative mucosal injury being central in H. pylori mediated gastric inflammation and also in the gradual emergence of m• DNA damage repair: Overrepresentation of nucleotide excision, and mismatch repair pathway components, that kind of thing matches with the genomic instability that keeps building up due to chronic inflammatory oxidative stress, and then this later leads to malignant transformation

A few hub genes in the overlapping set, like TP53 CDKN1A, MYC, BCL2, and VEGFA, were already shown in earlier transcriptomic and proteomic work on gastric adenocarcinoma. so that helps confirm the biological credibility and also the analytical soundness of what we see in the current computational results. Table 1 below gives a broad annotation for those common hub genes. including their degree, and betweenness centrality scores, inside the downstream gene–miRNA network, and their specific mechanistic contribution to the gastritis-cancer transition

4.2 Gene–miRNA Interaction Network Analysis

4.2.1 Network Topology and Hub miRNA Identification

The gene–miRNA interaction network assembled in NetworkAnalyst 3.0, combining together experimentally supported regulatory links from miRTarBase, and TarBase, had a total of nodes and edges, with hub gene targets shown as circular nodes, and their matching regulatory miRNAs shown as square nodes. Looking at the network (Figure 4.1) right away suggested a heterogeneous, non-random layout, where only a small number of very connected nodes really dominated the whole interaction space, this is a kind of structural signature that fits well with scale free network topology, alignant traits .

Node Classification and Visual Representation

In the network, nodes were sort of visually teased apart by both their shape and size. The circular ones stood for protein coding hub genes, with the node diameter being roughly proportional to degree centrality, while square nodes meant regulatory miRNAs. Also the color intensity had this extra sorting role for connectivity, where the largest, deep red circles marked the most highly connected gene hubs, orange circles showed genes with intermediate connectivity, and the blue squares corresponded to individual miRNA regulators. This whole visual pecking order basically mirrored how the topology was dominated by selected genes and miRNAs inside the post transcriptional regulatory circuitry, so it looked like the network itself was telling you the story.

Identification of Hub Genes

Then for hub genes, the topological analysis picked out five genes as main high centrality hubs, based on their very large node size and strong edge density inside the network: CCND1, CDKN1A, MYC, TP53, and IL6. Out of these, CCND1 appeared as the most strongly connected node, and it sat in the lower central part of the network, showing the largest node diameter and the highest count of converging miRNA edges, which points to CCND1 as the main post transcriptional regulatory target within this system. CDKN1A and MYC were in the central cluster, with comparably large node sizes

and very dense edge patterns, suggesting they are key regulatory targets too. TP53 and IL6 were a bit smaller in diameter but still showed notable connectivity, especially toward the miRNA dense core of the network, and that made them linger in importance.
Peripheral orange nodes, including MAPK1, SOD2, ENO1, IL1B, JUN, PTGS2, HMOX1, NOS3, SERPINE1, CASP8, CEBPB, and RGS2, acted as secondary hub genes with moderate degree centrality. These nodes were linked through shared miRNA regulators, and together they helped build the broader functional architecture of the network, more like the supporting cast than the lead role, but still clearly there.

Hub miRNA Identification

In the miRNA layer of the network, a few square nodes had this kind of disproportionately high connectivity, like they were linked to multiple hub genes at once, which made them look like hub miRNAs with really wide regulatory influence.
The miRNAs that stood out most topologically, in what you can actually see in the network, included hsa-miR-106a-5p, hsa-miR-34a-5p, hsa-miR-21-5p, hsa-miR-195a-5p, hsa-let-7b-5p, hsa-let-7a-5p, -hsa-let-7c, hsa-miR-26b-5p, hsa-miR-129a-5p, hsa-miR-6873-5p,and hsa-miR-10b-5p. They basically sat in densely connected areas, bridging several hub genes, and they were especially converging on CCND1, CDKN1A, MYC, and TP53 which suggests, in a broad sense, that these miRNAs together make up the core post transcriptional regulatory layer controlling the expression of those key oncogenic plus tumor-suppressive targets.

Then there's betweenness centrality, which kind of reinforced the same idea, because their intermediate placement on shortest routes between gene nodes implies a gatekeeper role, in how "information" flows around this network,, formed a sort of noticeable sub cluster in the mid network area, targeting several overlapping gene nodes, and that points toward cooperative regulation or maybe redundant control over shared downstream targets.

Network Structural Interpretation

Overall, the network architecture kind of shows a tightly interconnected core , anchored by CCND1, CDKN1A, MYC, and TP53 — with a more peripherally distributed collection of secondary gene and miRNA nodes. It looks like several miRNAs converge on the same hub genes, and then each hub miRNA can reach multiple targets at once. That combination suggests a rather robust , buffered regulatory system, where if one hub miRNA gets dysregulated it might mess with the expression of several genes that are central to gastric disease progression. So these hub miRNAs, then, become top candidates for mechanistic follow-up , and also possible therapeutic targeting in the setting of gastritis related gastric carcinogenesis.

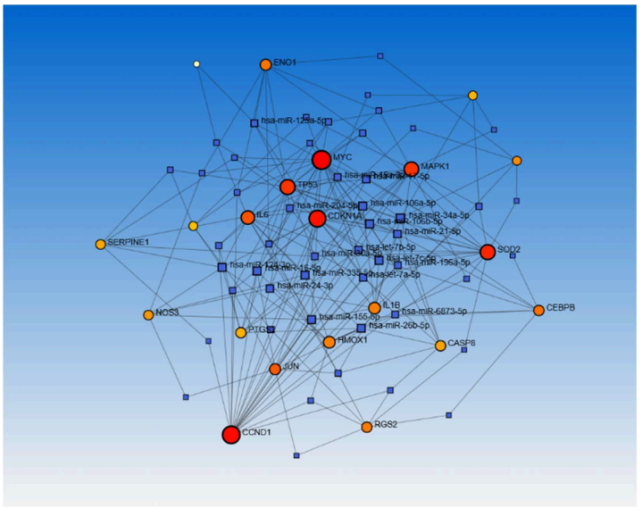

Figure 4.1: Gene–miRNA interaction network illustrating high-degree hub miRNAs and their target common genes. Node size reflects degree centrality; edge thickness denotes regulatory confidence score.

4.2.2 Functional Characterization of Hub miRNAs

The hub miRNAs that turned up from the topological check of the gene–miRNA interaction network—mostly hsa-miR-21-5p, hsa-miR-34a-5p, , hsa-miR-155-5p, hsa-miR-195a-5p, hsa-miR-204-5p, hsa-miR-24-3p, hsa-miR-26b-5p, hsa-miR-335-5p, hsa-miR-129a-5p, hsa-miR-10b-5p, and members of the let-7 set (let-7a-5p, let-7b-5p, let-7c)— sort of jointly focused on the five highest-centrality hub genes sitting in the network: CCND1, CDKN1A, MYC, TP53, and IL6

hsa-miR-21-5p

This showed connections to several hub genes inside the central cluster, like CDKN1A and IL6. Since CDKN1A gives rise to the cyclin-dependent kinase inhibitor p21, basically a strict enforcer for G1-phase cell cycle arrest , the miR-21-5p knockdown of this target looks like a pretty straightforward route where post-transcriptional tuning could bypass the usual cell cycle checkpoint during gastric mucosal transformation.

hsa-miR-34a-5p

Which appears as a big square node with edges converging on TP53-related targets, is often described as a transcriptional player lying downstream of p53 signaling. Its spot in the network— acting like a bridge between TP53 and CCND1 and CDKN1A— hints at a regulatory loop where p53 activation boosts miR-34a-5p expression, and then miR-34a-5p suppresses CCND1-driven proliferative signaling.

hsa-miR-106a-5p

This is sitting in a central part of the network, with edges pointing toward CDKN1A and MYC, which are two regulators that basically pull in opposite directions for cellular proliferation. The fact that one miRNA can simultaneously target a cell cycle brake (CDKN1A) and a transcriptional ignition switch for growth programs (MYC) really highlights how hub exert coordinated, bidirectional control over proliferative balance within gastric epithelial cells .

The let-7 family members

They were located in a spatially coherent sub-cluster in the middle part of the network (Figure 1), and their edges tended to concentrate on MYC and MAPK1. MYC is one of the largest-diameter nodes in the network and is a master transcriptional regulator of cellular growth with multiple let-7 isoforms controlling its activity, indicating cooperative suppression of oncogenic transcriptional output, which may become increasingly lost as the result of chronic H. pylori-induced inflammation.

hsa-miR-155-5p

was associated in the network to IL6 and IL1B, which encode pro-inflammatory cytokines which are peripheral nodes in Figure 1. This edge configuration is congruent with a feed-forward inflammatory signaling loop that amplifies cytokine-mediated signaling via induction of miR-155-5p and that regulates immune regulatory targets, thus perpetuating pre-neoplastic inflammatory mucosal disease. The hsa-miR-195a-5p and hsa-miR-24-3p, which are located in the middle to lower part of the network cluster, had regulatory connections to the most densely connected gene hub, CCND1. The convergent targeting of CCND1 by several miRNAs such as miR-195a-5p and miR-24-3p indicates that the network was structurally prioritizing mitosis,meosis as very important regulatory outputs of this miRNA-gene system.

Together the functional relationships coded in the network topology of Figure 1 define a coherent post-transcriptional regulatory program in which hub miRNAs function together to regulate cell cycle progression (via CCND1, CDKN1A), oncogenic transcription (via MYC), tumor suppression (via TP53), inflammatory cytokine output (via IL6, IL1B), and stress-response signaling (via MAPK1, HMOX1, SOD2) — all functional categories that together span the molecular hallmarks of chronic gastric inflammation as well as early neoplastic transformation.

Table 4.1: Node Table Gene–miRNA Interaction

Id	Label	Degree	Betweenness	Expression
2023	ENO1	8	78.55	0
MIMAT0000443	hsa-miR-125a-5p	5	71.19	0
4609	MYC	24	440.51	0
5594	MAPK1	14	195.54	0
7157	TP53	15	196.18	0
MIMAT0000265	hsa-miR-204-5p	5	40.58	0
MIMAT0000103	hsa-miR-106a-5p	7	56.55	0
1026	CDKN1A	19	311.37	0
MIMAT0000255	hsa-miR-34a-5p	8	126.06	0
3569	IL6	12	119.85	0
MIMAT0000076	hsa-miR-21-5p	6	42.73	0
5054	SERPINE1	6	36.36	0
MIMAT0000069	hsa-miR-16-5p	7	64.42	0
MIMAT0000765	hsa-miR-335-5p	7	61.41	0
MIMAT0000080	hsa-miR-24-3p	6	43.45	0
6648	SOD2	15	236.52	0
MIMAT0000226	hsa-miR-196a-5p	5	26.07	0
MIMAT0000646	hsa-miR-155-5p	7	76.74	0
1051	CEBPB	7	81.02	0
4846	NOS3	6	46.60	0
5743	PTGS2	7	31.85	0
MIMAT0000083	hsa-miR-26b-5p	7	73.40	0
3162	HMOX1	9	62.95	0
3725	JUN	8	111.17	0
595	CCND1	21	343.39	0
5997	RGS2	7	70.00	0
841	CASP8	7	38.48	0

4.3 Tissue-Specific Co-expression Network

Stomach tissue-specific co-expression network of gastric gene expression data in NetworkAnalyst 3.0 showed a strong modular structure with very large and noticeable clusters of highly co-expressed genes clustered into functionally coherent modules with high intra-module connectivity and low inter-module connectivity. This topological setting is typical of co-expression networks found in biology that are meaningful and functional, where sets of clusters of genes function as semi-autonomous gene regulatory regions in a specific tissue microenvironment and not as a undifferentiated gene mass. Figure 2 shows that the network is structured in several modules

.

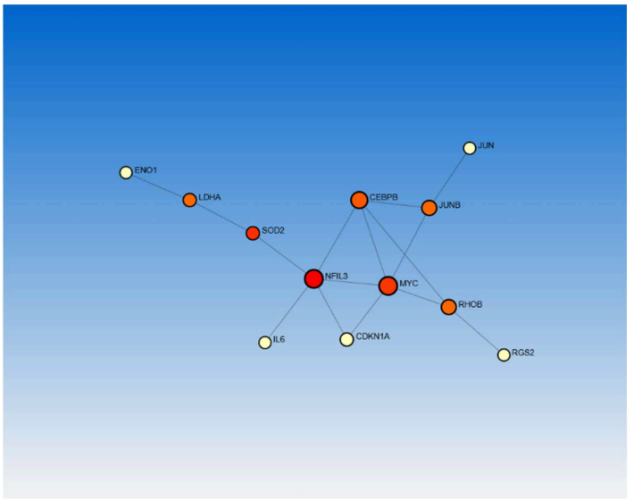

Figure 4.2: Tissue Specific co-expressionin map in stomach tissues

Characterization of Co-expression Modules

Within the co-expression network, the common set of genes were found to be divided into three structurally and functionally distinct modules, that were each connected to a specific hub gene and shared by common co-expression edges (Figure 2).

Module -1Transcriptional Regulation and Immune-Metabolic Coordination

The first and the most topologically dense module was anchored by two highly central modules in the module network: NFIL3 and MYC, with the largest module diameters in the network, and the highest number of direct edges. NFIL3, located at the geometric center of the network, had direct links to the most connected genes of the whole network, such as SOD2, CEBPB, MYC, CDKN1A and IL6. This structural dominance makes NFIL3 the central component of the module, having co-expression connections with oxidative stress response (SOD2), transcription factor activity (CEBPB), oncogenic regulation (MYC), cell cycle control (CDKN1A), and cytokine signaling (IL6). The diversity of these co-expression relationships suggests that NFIL3 is positioned to regulate multiple transcriptional programs that are both immune and metabolic as well as proliferative and all active in gastric mucosal disease.

Module 2 — bZIP Transcription Factor Cluster and AP-1 Axis

The second module was a spatially coherent sub-cluster of three basic leucine zipper (bZIP) genes, namely CEBPB, JUNB and JUN, which are known to regulate inflammatory gene expression. CEBPB was the largest node in this sub-cluster, and had direct connections to both MYC and NFIL3, and to JUNB. JUNB (see Fig. 2) was linked to JUN, which was a peripheral lower-connectivity node at the upper right of the network. The co-expression of CEBP, JUNB and JUN, as well as the known functional interaction between the C/EBP and AP-1 family members for the regulation of cytokine gene promoters, suggest that this type of transcriptional interaction may be of particular importance in H. pylori stimulated gastric epithelial and stromal cells. This sub-cluster is connected to MYC via CEBPB, which further associates AP-1-mediated inflammatory transcription with oncogenic proliferative signaling in the same co-expression neighborhood.

Module 3 — Metabolic Reprogramming and Peripheral Regulatory Genes, students delve into the mechanisms driving metabolic shifts and how these changes influence the function of peripheral regulatory genes.

The third module was a group of nodes with fewer connections, such as " ENO1", SOD2 ,RHOB, RGS2, and IL6, located in the periphery of the network with fewer inter-nodal connections. LDHA and ENO1 seemed to form a loosely coupled pair in the upper left part of the network, connected to the central hub NFIL3 via the role of SOD2.
The co-expression of the glycolytic enzymes encoded by LDHA and ENO1 in the context of this network suggests that metabolism may be a coordinated transcriptional output of the core module, in addition to its immune and proliferative functions. Single-edge connectivity with MYC and the rest of the cluster, and their location, periphery, right-side of the network, were reflected in the lower centrality members of the network, namely RHOB and RGS2, which participate in cytoskeletal

remodeling and G-protein signal termination, respectively, processes associated with the invasive phenotype of stomach pre-malignant cells. Although biologically important as a pro-inflammatory cytokine, IL6 emerged as a low connectivity PN with only one edge to NFIL3, indicating that its transcriptional coupling is relatively selective when compared to the rest of the co-expressed genes in this subnetwork context.

Together, the three co-expression modules outlined in Figure 2 suggest a multi-layered transcriptional organization with a core module of NFIL3 and MYC that regulates inflammatory, proliferative and metabolic gene expression programs, as well as a sub-network of bZIP transcription factors, and a collection of peripheral metabolic and signaling genes whose co-expression patterns collectively capture the transcriptional diversity associated with progression of gastric disease.

Table 4.2 Node Table Tissue Specific Coexpression

Id	Label	Degree	Betweenness	Expression
3725	JUN	8	116.41	0
4609	CEBPB	26	521.52	0
6648	MYC	15	232.65	0
1051	IL6	8	105.43	0
5743	CDKN1A	7	31.34	0
3569	RGS2	12	113.54	0
3553	SOD2	11	150.77	0

4.4 Transcription Factor–Gene Interaction Network

4.4.1 Network Architecture and Identification of Master Regulators

.

The transcription factor–gene interaction network generated by NetworkAnalyst 3.0 with curated regulatory information from ENCODE and JASPAR consisted of 38 nodes and the regulatory edges linking them (Figure 3). In the network, a few different node types were differentiated visually: red circles stand for target genes, and cyan diamonds for transcription factor regulators. Degree centrality was used to scale the size of the nodes, creating a visually stratified network with the high degree hubs (MYC, RELA, JUN, CDKN1A, PTGS2) easily distinguished from the peripheral lower degree nodes.

The overall network topology had a hub-and-spoke pattern, with a tightly knit, high degree central group of nodes, and a progressively lower-connectivity group of gene nodes and transcription factor nodes at the periphery. The central cluster was well connected with many common edges and the most central hubs of the network were comprised of MYC, RELA, JUN, CDKN1A, PTGS2, and TP53, showing the largest node diameters in the network.

The distribution of the transcription factor diamond nodes across the network was done with the edge shape suggesting the direction of the regulation, away from or toward the gene target nodes; these nodes included STAT3, STAT1, SP1, ESR1, JUND, AR, CREB1, RUNX3, CEBPA, CEBPB, CEBPD, HIF1A, TWIST1, SPI1, DNMT1, and MEN1. Peripheral genes (such as MAPK3, ALB, IL1RN, NOS3, SOD2, GADD45A, SERPINE1, and GAST) had a single or low-degree connectivity, which is consistent with (and located at the outer margins of) the low betweenness centrality values for these genes in the node table

The characterization of Master Transcription Factors.

The node table was used to compute degree and betweenness centrality values, which revealed a hierarchy of regulatory influence; the top positions were occupied by MYC, RELA, JUN, CDKN1A, PTGS2, and TP53 (Table 4.3).
The node with the upraised degree (14) and the highest betweenness centrality (147.93) was MYC, which is clearly at the top of the regulatory hierarchy. It has a betweenness centrality value of almost twice that of the second ranked node RELA, which shows that not only does it link to many targets, but it is also a structurally irreplaceable node on the shortest regulatory pathways between the majority of other nodes on the network. In the network figure, MYC was positioned in the middle with the rest of the nodes extending outwards, which visually verified its position as the main transcriptional integrator in this network, while also being surrounded by CCND1, CDKN1A, PTGS2, JUN, TP53, CEBPB, IL6, and other transcription factor diamond nodes (SP1, STAT3, ESR1).

The network position of RELA, with a highest degree (14), but a lower betweenness centrality (78.64), was in the lower center of the network, linked to IL1B, IL6, PTGS2, CCND1, MYC, JUN, SERPINE1, SOD2, NOS3, and IL1RN. This edge distribution encompasses genes involved in inflammatory cytokines, oxidative stress regulators, and fibrinolytic targets, a gene signature similar to the NF-κB transcriptional program. RELA edges occur on some genes in common with SP1 and STAT3 (e.g. IL6, PTGS2), indicating that within this network, there is co-regulation of transcription at overlapping gene promoters.

JUN was the third most central within the network based on its above criteria, and was connected to other heightened connected nodes in the network, such as MYC, RELA, CDKN1A, PTGS2, IL1B, CEBPB, MAPK1 and GAST. What is particularly of structural significance is its connectivity to

MAPK1, which has a betweenness centrality of 76.72 for a degree of 3, which is much higher than the average betweenness centrality for nodes of that degree.Its connectivity to MAPK1, which is a node with a disproportionately high betweenness centrality of 76.72 for its low degree of 3, is of particular structural significance, as MAPK1 is a node in the network with relatively few direct links to nodes that are connected to it, yet at the same time it is a node with a high betweenness centrality.

Among all the transcription factor diamonds, SP1 is the high degree diamond node (with 10 regulatory edges) and the highest betweenness centrality (40.42) in the network, with regulatory edges linked to MYC, CDKN1A, CCND1, TP53, CASP8, PTGS2, HMOX1, ENO1, and SERPINE1. SP1 is the most broadly active transcriptional regulator in the diamond node layer of the network, since it regulates a wide variety of target genes, including genes involved in proliferation, apoptosis, metabolism and stress response.

The central-upper part of the network was dominated by two nodes, STAT3 (degree 9, betweenness 35.76) and STAT1 (degree 7, betweenness 20.97), which are connected by overlapping edges pointing to common gene targets: CDKN1A, CCND1, MYC, and CASP8. The spatial closeness and the connectivity of target genes seen in the figure also hint at some degree of functional redundancy or cooperation at common sites of target gene regulation, which is consistent with the ability of STAT3 and STAT1 to compete with each other for binding to shared promoter elements in the context of the upstream cytokine signal.

But the degree (2) was low and the betweenness centrality was high at 35.00, suggesting it as a structural bridge node in the lower left periphery of the network, in an area where there were few links. It had very few connections in the figure connected to it, with two being linked to MAPK1 and one to CEBPB, and was found to be part of a bottleneck position in the network, with a high betweenness as it lies in the center of the network between the peripheral MAPK3–MAPK1 axis and the central hub cluster, it acts as a relay function in the propagation of the transcription signal despite the relatively small number of target genes it connects to.

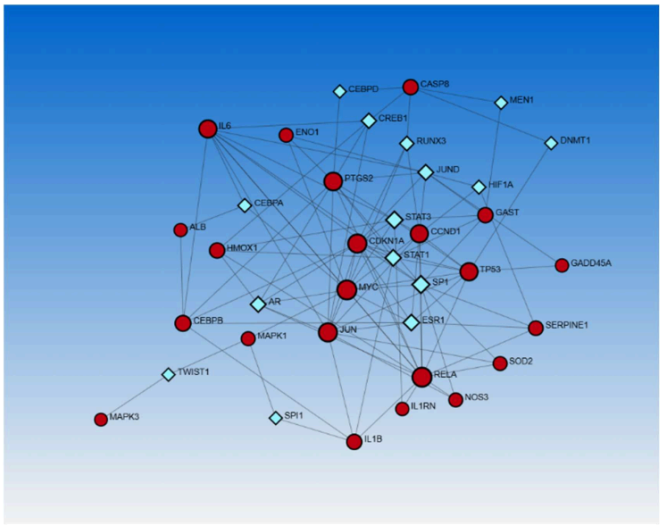

Figure 4.3: Transcription factor–gene interaction map

4.5 A protein–drug interaction network and its clinical implications.

4.5.1 Network Architecture

The protein–drug interaction network built from the hub gene set included nodes for proteins (circles) and nodes for drugs (squares) and the size of the nodes was proportional to degree centrality (Figure 4). Most highly connected protein targets were shown as red circular nodes, while orange and yellow circular nodes were shown for proteins of moderate and lower connectivity, respectively, such as SERPINE1, ENO1, CCND1, HMOX1, NOS3, JUN, MYC, CDKN1A, CASP8, and ALB.

The nodes of the network were labeled with the names of the drugs and the edges were labeled with the protein targets/sites of each drug.The nodes were labelled with the name of the drug and the edges were labelled with the name of the drug's protein target or targets. One with uprised connectivity (highest number of links) in the network was the visually identified arsenic trioxide.

Figure 4.4: A protein–drug interaction network map

The recent network shown in Figure 4 .4 provides a useful framework for the analysis of computed results for this study and links each hub protein target to existing drugs and other drugs under investigation, that have been demonstrated to have binding activity to the corresponding protein. The network structure defines three levels of clinical relevance in terms of drug–target connectivity, target protein type, and the presence of clinical stage drugs. The purpose of this work is to determine the chemopreventive capability of acetylsalicylic acid and sulindac.

The top part of the network was dominated by the node of the protein encoded by PTGS2, which is responsible for the production of the protein cyclooxygenase-2 (COX-2), and the edges linked to it were Acetylsalicylic acid, Bryostatin 1, and Sulindac. Of these, Acetylsalicylic acid and Sulindac are both proven (NSAIDs) whose some inhibitory activity are well characterized. The network position of PTGS2 – which shares network edges with CASP8, CCND1, HMOX1 and MAPK3 – suggests that targeting PTGS2 pharmacologically would not have an isolated effect, but rather would affect a wider interconnected network of hub gene clusters.

The multi-target adjacency here gives a network level justification for the observed association between regular NSAID use and lower incidence of gastric cancer in the epidemiology literature, beyond a simple COX-2 enzyme blocker; it supports the notion that it might be the coordinated blockage of a co-regulated gene neighborhood in inflammation that is the epidemiologically observed chemopreventive effect of regular NSAID use.

Arsenic Trioxide, Colforsin and Minocycline are also under investigation to treat these cancers.

MAPK3
was the most central node in the network in terms of red node diameter, and had the greatest number of drug edges, including those of Arsenic trioxide, Acetylsalicylic acid, Colforsin, Minocycline, Bryostatin 1, and Sulindac. MAPK1, which was connected to both Colforsin and the rest of the central cluster, was a functionally coupled pair with MAPK3, which was adjacent to it. The MAPK3 protein node is the most druggable target found in this analysis, since it is targeted by six different pharmacological agents.

The most "visible" drug node in the figure (arsenic trioxide) was connected to MAPK3, JUN, and SERPINE1 at the same time, indicating that arsenic trioxide was the only drug in the network that had edges on both the MAPK signaling axis and the AP-1/fibrinolytic regulatory nodes. The multi-target network footprint makes arsenic trioxide a candidate agent with the ability to disrupt convergent oncogenic signaling by targeting multiple signaling nodes during a single drug exposure; this property is relevant for its use in haematological malignancies and the growing evidence indicating its potential for use in solid tumours. Minocycline, a tetracycline class antibiotic with secondary activities such as anti-inflammatory and anti-proliferative, linked to both MAPK3 and IL1B within the network, indicating it may have two modes of action: anti-inflammatory and kinase-modulatory – both of which could be relevant to the inflammatory component of gastric carcinogenesis. Colforsin was observed to be a direct adenylate cyclase activator, which was related to MAPK3 and SERPINE1, suggesting that it may play a regulatory role in the modulation of the activity of the associated pathway and the expression of the plasminogen activator inhibitor through cAMP.

Find the details about IL1B and IL1RN — Rilonacept and the Inflammatory Cytokine Axis.

IL1B was a moderately connected red node that connected to Rilonacept and Minocycline, and IL1RN (encoding the endogenous interleukin-1 receptor antagonist) was a low-connectivity peripheral node that connected only to Rilonacept. Structurally informative: Rilonacept, an IL-1 cytokine trap, targets both IL-1α and IL-1β and its ability to interact with both the cytokine (IL1B) and its endogenous inhibitor (IL1RN) within this drug network reflects a pharmacodynamic mechanism that reduces the amount of active cytokines while also modulates the regulatory balance of the IL-1 signaling axis. Considering the IL1B promoter polymorphisms are one of the best replicated genetic susceptibility factors for gastric cancer with respect to H. pylori infection, the presence of Rilonacept as a network-mapped pharmacological agent acting to target this node has direct preventive and therapeutic relevance.

In the bottom right part of the network, HMOX1 and NOS3 were linked as they shared an edge with Ferroheme and Isopropyl alcohol. One of these is HMOX1, which encodes heme oxygenase-1, an anti-inflammatory cytoprotective enzyme that is up-regulated in the presence of stress and chronic inflammation (stress and chronic inflammation are hallmark features of gastric mucosa in infection). It has been suggested that there is a linked regulatory relationship between heme catabolism, nitric oxide bioavailability, and oxidative stress management in the gastric mucosal microenvironment, given that NOS3 co-localizes with its network and both are linked to Ferroheme. This indicates that iron–heme metabolic balance could be a targetable axis in the oxidative part of gastric inflammatory disease, as a direct substrate-related ligand in this context.

These are all metabolic and proteomic targets that are relevant to the ENO1, ALB and Artenimol products.

ENO1 and ALB were placed in the top of the network, and were linked by Artenimol and Copper respectively. Glycolytic enzyme ENO1, which is known to moonlight as a plasminogen receptor on cancer cell surfaces, was coupled with an artemisinin derivative, Artenimol, which has anti-glycolytic and anti-proliferative properties. Artemisinin-related compounds are known to be safe and have been shown to have anti-cancer activity, and the inclusion of Artenimol in the drug network as a candidate agent targeting ENO1 indicates that metabolic reprogramming of glycolytic flux – in particular, at the enolase step – may be pharmacologically accessible.

In Figure 4.4, three proteins with high centrality in the network (CCND1, MYC, CDKN1A) were identified that did not have direct connections to drugs, but instead were connected with other proteins in the network with edges pointing outward and connected with a square node with edges pointing inward. This lack of direct pharmacological annotation for 3 out of the 10 most important gene targets in its network (cyclin CCND1 involved in G1/S transition, the master transcription amplifier MYC and the major cellular cycle brake CDKN1A) shows that there is a disconnect between the importance of these gene targets in the network and the current availability of approved or well characterized pharmaceuticals annotated within the database.
This finding itself is a clinically relevant result of the network analysis as it identifies MYC, CCND1, and CDKN1A as high priority targets for new drug discovery programs targeting gastric pre-malignancy, where there is a clear lack of drugs covering the nodes' biological importance.

4.5.1 Integrated Therapeutic Interpretation

The protein–drug interaction network of Figure 4 represents a pharmacological landscape in which several known drugs that bind to the hub gene targets identified in the upstream network and enrichment analysis of this study are represented, including multiple anti-inflammatory drugs (e.g., NSAIDs such as Acetylsalicylic acid and Sulindac), cytokine inhibitors (e.g., Rilonacept), kinase modulators (e.g., Arsenic trioxide and Colforsin), and drugs with anti-inflammatory activity such as antibiotics (e.g., Minocycline), artemisinin derivatives (e.g., Artenimol) and metabolic substrate analogs (e.g., Ferroheme). Several agents target MAPK3 and PTGS2, implying the capacity for multi-drug perturbation of the MAPK–COX-2 axis, which is also structurally adjacent to other hub genes, suggesting the potential for greater network perturbation by multi-drug intervention targeting the MAPK–COX-2 axis than by single agents. On the other hand, the lack of coverage of the current network for drug targets such as MYC, CCND1, and CDKN1A underlines the most urgent unmet therapeutic needs in this molecular network defined by this study.

CONCLUSION

CHAPTER -5

In the present study, the shared molecular landscape of gastritis and gastric cancer is analysed in a comprehensive, integrative network-based approach based on a systems biology approach to elucidating the regulatory architecture that drives the progression from chronic mucosal inflammation to malignant neoplasia. This systematic analysis revealed common genes, miRNAs, and master transcription factors across these two states of diseases, and defined the general molecular circuitry involved in the progressive acquisition of cancerous cellular properties through persistent inflammatory stimuli. A set of common genes with the same functional properties, including an enrichment of genes involved in inflammatory signaling, apoptotic regulation, ROS and reinforces the biochemical continuity of the Correa cascade at the molecular level and provides a computational model validated for gastric carcinogenesis. Based on the gene–miRNA interaction network analysis, a group of high-centrality pro-oncogenic and tumor-suppressive miRNAs was identified, whose altered expression is mechanistically associated with the repression of tumor suppressor gene expression and the promotion of pro-oncogenic signaling. The tissue-specific co-expression network also provided information on the modular organization of the common genes in the gastric tissue context, which identified distinct functional groups of genes related to immune activation, disruption of the epithelial barrier and neoplastic growth. Using TF network analysis, we identified NF-κB, STAT3, HIF-1α and MYC as master regulators of the shared gene regulatory program, suggesting these regulators as key orchestrators of the inflammatory-to-oncogenic transcriptional transition. Together, these results make a significant contribution to the molecular understanding of the progression of H. pylori-induced chronic inflammation to stomach cancer, and offer a multi-dimensional catalog of candidate biomarkers and therapeutic targets derived from a computational analysis. These computational predictions will need to be validated in future prospective clinical trials, experimental studies of functional genomics, and preclinical therapeutic models to bring these tools into the clinic for early detection, risk stratification, and precision treatment of gastric malignancies. The study is therefore a valuable contribution to the body of evidence that supports the use of principles of network medicine to elucidate complex oncogenic processes driven by inflammation.

REFERENCES

1.Zhang X, Ni Z, Duan Z, et al. Overexpression of E2F mRNAs associated with gastric cancer progression identified by the transcription factor and miRNA co-regulatory network analysis. *PLoS ONE*. 2015;10(2):e0116979.

2. Xu J, et al. Transcription factor and miRNA co-regulated genes in gastric cancer invasion. *PMC*. 2015.

3. Wei G, Dong Y, He Z, et al. Identification of hub genes and construction of an mRNA-miRNA-lncRNA network of gastric carcinoma using integrated bioinformatics analysis. *PLoS ONE*. 2021;16(12)

4. Tian Y, Xing Y, Zhang Z, et al. Bioinformatics analysis of key genes and circRNA-miRNA-mRNA regulatory network in gastric cancer. *BioMed Research International*. 2020;2020:2862701.

5.Bioinformatics prediction and machine learning on gene expression data for lncRNA-miRNA-mRNA signaling axes in gastric cancer. *PMC*. 2022.

6. Hub gene identification in gastric adenocarcinoma using STRING and Cytoscape. *PMC*. 2023. https://pmc.ncbi.nlm.nih.gov/articles/PMC10454997/

7. Davis AP, Wiegers TC, Johnson RJ, et al. Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Research*. 2023;51(D1):D1257–D1262.

8. Liang WH, Li N, et al. NF-κB signaling in gastric cancer. *American Journal of Cancer Research*. 2017;7(5):1055–1067.

9. Luo N, et al. Synergistic interaction between NF-κB and STAT3 promotes gastric cancer migration and invasion. *Oncology Letters*. 2013;5(6):1897–1903.

10. Ucaryilmaz Metin C, Ozcan G. HIF-1α as a potent inducer of the hallmarks in gastric cancer. *Cancers*. 2022;14(11):2711.

11.Kim WH, Kim MJ, Jin JO, Lee PCW. IDF-11774 induces cell cycle arrest and apoptosis by inhibiting HIF-1α in gastric cancer. *Pharmaceutics*. 2023;15(12):2772.

12. Prediction of blood miRNA-mRNA regulatory network in gastric cancer. *PMC*. 2021. /[13.Li BS, Zhao YL, Guo G, et al. Plasma microRNAs, miR-223, miR-21 and miR-218, as novel potential biomarkers for gastric cancer detection. *PLoS ONE*. 2012;7(7):e41629.

14.Pyroptosis caused by H. pylori in gastric cancer. *Annals of Medicine and Surgery*. 2024.

15. Correa P, et al. Gastric precancerous process in a high risk population: cohort follow-up. *Cancer Research*. 2010.

16. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513(7517):202–209.

17.Watari J, et al. H. pylori associated gastric mucosal alterations. *PubMed*. 2015.

18. Ooi CH, et al. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. *ScienceDirect*. 2022.

19. Molecular markers study in gastric cancer. *Scientific Reports*. 2026.

20. Gastric cancer transcription factor regulatory network study. *Discover Oncology*. 2026.

21. MicroRNA study in gastric cancer progression. *Cancer Research and Treatment*. 2026.

22. Biomarker identification study in gastric cancer. *BMC Cancer*. 2025.

23. Bandyopadhyay S, Mitra R, Maulik U, Zhang MQ. Development of the human cancer microRNA network. *Silence*. 2010

24. Cai H, et al. MicroRNA regulatory network analysis in gastric cancer. *Scientific Reports*. 2015;5:8270.

25. Tseng CW, Lin CC, Chen CN, Huang HC, Juan HF. Integrative network analysis reveals active microRNAs and their functions in gastric cancer. *BMC Systems Biology*. 2011;5:99.

26. Cheng J, Zhuo H, Xu M, et al. Regulatory network of circRNA-miRNA-mRNA contributes to histological classification and disease progression in gastric cancer. *Journal of Translational Medicine*.

27. NetworkAnalyst and bioinformatics computational analysis framework. *arXiv preprint* 2015.

Anushka Pal MSc

ORIGINALITY REPORT

6%	4%	5%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	pmc.ncbi.nlm.nih.gov Internet Source	<1%
2	www.mdpi.com Internet Source	<1%
3	genome.cshlp.org Internet Source	<1%
4	www.frontiersin.org Internet Source	<1%
5	academic.oup.com Internet Source	<1%
6	www.dovepress.com Internet Source	<1%
7	Evelina Charidemou, Christos Papaneophytou. "Non-Coding RNAs in Cancer: Decoding Regulatory Networks for Liquid Biopsy Applications", Genes, 2026 Publication	<1%
8	m.amedeo.com Internet Source	<1%

9	Advances in Experimental Medicine and Biology, 2016. Publication	<1 %
10	respiratory-research.biomedcentral.com Internet Source	<1 %
11	www.ijsr.net Internet Source	<1 %
12	www.cjnmcpu.com Internet Source	<1 %
13	Gang Wei, Youhong Dong, Zhongshi He, Hu Qiu, Yong Wu, Yongshun Chen. "Identification of hub genes and construction of an mRNA-miRNA-lncRNA network of gastric carcinoma using integrated bioinformatics analysis", PLOS ONE, 2021 Publication	<1 %
14	Muhammad Harith Zulkifli, Zafirah Liyana Abdullah, Nur Intan Saidaah Mohamed Yusof, Fazlin Mohd Fauzi. "In silico toxicity studies of traditional Chinese herbal medicine: A mini review", Current Opinion in Structural Biology, 2023 Publication	<1 %
15	www.hindawi.com Internet Source	<1 %

16	Hamid D. Ismail. "Bioinformatics of Autoimmune Diseases", CRC Press, 2026 Publication	<1%
17	shivdasanilab.dana-farber.org Internet Source	<1%
18	pdfs.semanticscholar.org Internet Source	<1%
19	www.eurasianbiochem.org Internet Source	<1%
20	dadun.unav.edu Internet Source	<1%
21	Gabriel D. Dakubo. "Cancer Biomarkers in Body Fluids", Springer Science and Business Media LLC, 2017 Publication	<1%
22	Qin Tang, Qiong Zhang, Yao Lv, Ya-Ru Miao, An-Yuan Guo. "SEGreg: a database for human specifically expressed genes and their regulations in cancer and normal tissue", Briefings in Bioinformatics, 2018 Publication	<1%
23	Richard Jones. "The search for genenotype/phenotype associations and the phenome scan", Paediatric and Perinatal Epidemiology, 7/2005 Publication	<1%

24	elifesciences.org Internet Source	<1%
25	lib.smu.edu.cn Internet Source	<1%
26	insight.jci.org Internet Source	<1%
27	Rossi, Ana Flávia Teixeira, Aline Cristina Targa Cadamuro, Joice Matos Biselli-Périco, Kátia Ramos Moreira leite, Fábio Eduardo Severino, Patricia P. Reis, José Antonio Cordeiro, and Ana Elizabete Silva. "Interaction between inflammatory mediators and miRNAs in Helicobacter pylori infection. : Inflammatory mediators and miRNAs in H. pylori infection", Cellular Microbiology, 2016. Publication	<1%
28	mdozmorov.github.io Internet Source	<1%
29	mymedr.afpm.org.my Internet Source	<1%
30	pesquisa1.bvsalud.org Internet Source	<1%
31	Chunhong Li, Yuhua Mao, Yi Liu, Jiahua Hu, Chunchun Su, Haiyin Tan, Xianliang Hou, Minglin Ou. "Machine learning-based	<1%

integration develops a multiple programmed cell death signature for predicting the clinical outcome and drug sensitivity in colorectal cancer", Anti-Cancer Drugs, 2024
Publication

32	Jaya Thomas, Dongmin Seo, Lee Sael. "Review on Graph Clustering and Subgraph Similarity Based Analysis of Neurological Disorders", International Journal of Molecular Sciences, 2016 Publication	<1 %
33	Xing Yu, Lin Luo, Yibo Wu, Xiuchong Yu et al. "Gastric juice miR-129 as a potential biomarker for screening gastric cancer", Medical Oncology, 2013 Publication	<1 %
34	cancermatters.cancer.org.my Internet Source	<1 %
35	rbmb.net Internet Source	<1 %
36	"Algorithms in Computational Molecular Biology", Wiley, 2011 Publication	<1 %
37	Luigina Micolucci, Most Mauluda Akhtar, Fabiola Olivieri, Maria Rita Rippo, Antonio Domenico Procopio. "Diagnostic value of microRNAs in asbestos exposure and	<1 %

malignant mesothelioma: systematic review and qualitative meta-analysis", Oncotarget, 2016
Publication

38	bura.brunel.ac.uk Internet Source	<1%
39	maayanlab.cloud Internet Source	<1%
40	results2021.ref.ac.uk Internet Source	<1%
41	www.bellequipment.com Internet Source	<1%
42	www.coursehero.com Internet Source	<1%
43	C. Imrie, M. Rowland, B. Bourke, B. Drumm. "Is Helicobacter pylori Infection in Childhood a Risk Factor for Gastric Cancer?", PEDIATRICS, 2001 Publication	<1%
44	Elif AYDIN, Ayse KOCAK SEZGIN, Meliha KOLDEMIR GUNDUZ, Birkan AÇIKGOZ, Güllü KAYMAK, Sercan SIMSEK. "Natural supplements effectively suppress Helicobacter pylori colonization and improve gastric antioxidant and antiinflammatory	<1%

responses in an indomethacin-induced ulcer model", Microbial Pathogenesis, 2026
Publication

45	Karin Van Amsterdam, Arnoud H. M. Van Vliet, Johannes G. Kusters, Arie Van Der Ende. " Of microbe and man: determinants of -related diseases ", FEMS Microbiology Reviews, 2006 Publication	<1 %
46	Suvi M. Kuosmanen, Juha Hartikainen, Mikko Hippeläinen, Hannu Kokki, Anna-Liisa Levonen, Pasi Tavi. "MicroRNA Profiling of Pericardial Fluid Samples from Patients with Heart Failure", PLOS ONE, 2015 Publication	<1 %
47	file.scirp.org Internet Source	<1 %
48	www.econferences.ru Internet Source	<1 %
49	www.omicsdi.org Internet Source	<1 %
50	Diogo Libânio. " and microRNAs: Relation with innate immunity and progression of preneoplastic conditions ", World Journal of Clinical Oncology, 2015 Publication	<1 %

51	J. V. Ponomarenko. "rSNP_Guide, a database system for analysis of transcription factor binding to target sequences: application to SNPs and site-directed mutations", Nucleic Acids Research, 01/01/2001 Publication	<1%
52	Md. Kawsar, Tasnimul Alam Taz, Bikash Kumar Paul, Kawsar Ahmed, Md. Ahsan Habib, Touhid Bhuyian. "Identification of vital regulatory genes with network pathways among Huntington's, Parkinson's, and Alzheimer's diseases", Network Modeling Analysis in Health Informatics and Bioinformatics, 2020 Publication	<1%
53	Meng Shao, Fang Wu, Jie Zhang, Jiangtao Dong et al. "Screening of potential biomarkers for distinguishing between latent and active tuberculosis in children using bioinformatics analysis", Medicine, 2021 Publication	<1%
54	Priyanka Mehta, Partha Chattopadhyay, Varsha Ravi, Bansidhar Tarai, Sandeep Budhiraja, Rajesh Pandey. "SARS-CoV-2 infection severity and mortality is modulated by repeat-mediated regulation of alternative splicing", Microbiology Spectrum, 2023 Publication	<1%

55	Tian-jiao Zhou, Jun-feng Liu, Ping Wang, An-na Hu, Lin-lin Chen, Jun-feng Zan. "Identification of Targets and Active Components of Yiqi SanJie Formula Against Lung Neoplasms Based on Network Pharmacology Analysis and Molecular Docking", Natural Product Communications, 2021 Publication	<1%
56	Zhongyu Liu, Yanzhi Guo, Xuemei Pu, Menglong Li. "Dissecting the regulation rules of cancer-related miRNAs based on network analysis", Scientific Reports, 2016 Publication	<1%
57	cgp.iiarjournals.org Internet Source	<1%
58	cuestionesdefisioterapia.com Internet Source	<1%
59	dipot.ulb.ac.be Internet Source	<1%
60	link.springer.com Internet Source	<1%
61	mdpi-res.com Internet Source	<1%
62	oup.silverchair-cdn.com Internet Source	<1%

63	www.ncbi.nlm.nih.gov Internet Source	<1%
64	www.readbag.com Internet Source	<1%
65	www.science.gov Internet Source	<1%
66	www.tandfonline.com Internet Source	<1%
67	Chachami, G.. "Bacterially produced human HIF-1@a is competent for heterodimerization and specific DNA-binding", Biochemical and Biophysical Research Communications, 20050603 Publication	<1%
68	Everton Cruz dos Santos, Paulo Rohan, Renata Binato, Eliana Abdelhay. "Integrated Network Analysis of microRNAs, mRNAs, and Proteins Reveals the Regulatory Interaction between hsa-mir-200b and CFL2 Associated with Advanced Stage and Poor Prognosis in Patients with Intestinal Gastric Cancer", Cancers, 2023 Publication	<1%
69	Tianyi Wei, Yulai Liang, Claire Anderson, Ming Zhang, Naishuo Zhu, Jun Xie. "Identification of candidate hub genes correlated with the	<1%

pathogenesis, diagnosis, and prognosis of prostate cancer by integrated bioinformatics analysis", Translational Cancer Research, 2021
Publication

---

70 "UEG Week 2013 Poster Presentations", United European Gastroenterology Journal, 2013. <1%
Publication

---

71 Igor Bado, Fotis Nikolos, Gayani Rajapaksa, Jan-Åke Gustafsson, Christoforos Thomas. "ERβ decreases the invasiveness of triple-negative breast cancer cells by regulating mutant p53 oncogenic function", Oncotarget, 2016 <1%
Publication

---

72 Jingchao Wei, Yinghao Yin, Qiancheng Deng, Jun Zhou, Yong Wang, Guangming Yin, Jianfu Yang, Yuxin Tang. "Integrative Analysis of MicroRNA and Gene Interactions for Revealing Candidate Signatures in Prostate Cancer", Frontiers in Genetics, 2020 <1%
Publication

---

73 Marcie R. Williams, Noriyuki Kataoka, Yumiko Sakurai, Christina M. Powers, Suzanne G. Eskin, Larry V. McIntire. "Gene Expression of Endothelial Cells due to Interleukin-1 Beta <1%

Stimulation and Neutrophil Transmigration", Endothelium, 2009
Publication

74 Ran Xu, Jacopo Martelossi, Morgan Smits, Mariangela Iannello et al. "Multi-tissue RNA-Seq Analysis and Long-read-based Genome Assembly Reveal Complex Sex-specific Gene Regulation and Molecular Evolution in the Manila Clam", Genome Biology and Evolution, 2022 <1%
Publication

75 Sheetanshu Gupta, Sarfraj Ahmad Siddiqui, Praneeth Juvvi, Ng. Joykumar Singh, Lokesh Kumar Mishra. "Food and Cancer", CRC Press, 2026 <1%
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		