Editor
Prof. Dr. Kazım Yalçın ARGA

gece kitaplığı

International Research and Evaluations in the Field of Bioengineering

October 2025

Genel Yayın Yönetmeni / Editor in Chief • Eda Altunel

Kapak & İç Tasarım / Cover & Interior Design • Gece Kitaplığı

Birinci Basım / First Edition • © Ekim 2025

ISBN • 978-625-388-782-7

Gece Kitaplığı / Gece Kitaplığı
Türkiye Adres / Turkey Address: Kızılay Mah. Fevzi Çakmak 1. Sokak
Ümit Apt No: 22/A Çankaya/ANKARA
Telefon / Phone: 0312 384 80 40

web: www.gecekitapligi.com
e-mail: gecekitapligi@gmail.com

Baskı & Cilt / Printing & Volume
Sertifika / Certificate No: 42488

INTERNATIONAL RESEARCH AND EVALUATIONS IN THE FIELD OF BIOENGINEERING

EDITOR

PROF. DR. KAZIM YALÇIN ARGA

gece kitaplığı

CONTENTS

Chapter 1

INFECTIONS IN THE TUMOR MICROENVIRONMENT AND THE ROLE OF METAGENOMICS IN CHARACTERIZATION AND THERAPEUTIC INSIGHTS

Ecem ÇETİNKAYA[1]

Bengisu ÇELİK[2]

Kazım Yalçın ARGA[3]

[1]Marmara University, Faculty of Engineering, Department of Bioengineering, İstanbul, Türkiye, ORCID: 0009-0004-7974-2708

[2]Marmara University, Faculty of Engineering, Department of Bioengineering, İstanbul, Türkiye, ORCID: 0009-0001-4423-3637

[3]Marmara University, Faculty of Engineering, Department of Bioengineering, İstanbul, Türkiye, ORCID: 0000-0002-6036-1348

1.Introduction

The tumor microenvironment (TME) is a complex and ever-changing ecosystem that includes malignant cells alongside immune cells, stromal components, vascular structures, elements of the extracellular matrix, and various signaling molecules. The complex interaction of these various components significantly impacts tumor initiation, progression, metastasis, and therapeutic response (Binnewies et al., 2018). In recent years, it has become increasingly evident that microbial communities—including bacteria, viruses, fungi, and protozoa—are essential components of the tumor microenvironment. Inflammation, immune evasion, metabolic pathways, and drug metabolism can all be modulated by their presence, which introduces additional layers of complexity.

Microorganisms inhabiting the tumor microenvironment often exhibit patterns that are specific to different types of cancer. For instance, Fusobacterium nucleatum has a significant correlation with colorectal cancer, playing a role in tumor progression via immune modulation and the activation of oncogenic signaling (Kostic et al., 2013). In the same way, the colonization of Helicobacter pylori is recognized as a significant factor in the development of gastric cancer, as it promotes chronic inflammation and induces genotoxic stress (Plummer et al., 2015). Viral pathogens are crucial in the development of certain cancers: human papillomavirus (HPV) is a significant cause of cervical and oropharyngeal cancers, whereas hepatitis B and C viruses are major factors in hepatocellular carcinoma due to their role in persistent inflammation and oncogenic transformation (de Martel et al., 2020). Furthermore, the Epstein–Barr virus has been associated with lymphomas and nasopharyngeal carcinoma through the encoding of viral proteins that interfere with cell cycle regulation and immune surveillance (Young et al., 2016). In addition to bacteria and viruses, certain fungal species like Malassezia have been associated with pancreatic cancer, possibly influencing tumor growth through the activation of the complement cascade (Aykut et al., 2019). The findings collectively emphasize that the tumor-associated microbiota plays an active role in cancer biology, rather than just a passive bystander.

The impact of the tumor microbiome significantly affects the response to therapy. Intratumoral bacteria have demonstrated the ability to metabolize chemotherapeutic agents, thereby directly influencing their efficacy. For example, Geller et al. (2017) showed that Gammaproteobacteria found in pancreatic tumors can metabolize gemcitabine, making it ineffective. Similarly, the gut microbiome has been shown to affect the systemic immune environment and impact the effectiveness of immune checkpoint inhibitors. A significant study conducted by Routy et al. (2018) demon-

strated that the presence of Akkermansia muciniphila in the gut was associated with enhanced responses to PD-1 blockade in epithelial cancers, underscoring a direct relationship between microbial composition and the effectiveness of immunotherapy. In the same way, Iida et al. (2013) found that dysbiosis caused by antibiotics diminished the effectiveness of cyclophosphamide in preclinical cancer models, highlighting the importance of interactions between microbiota and drugs.

Investigating the intricate interactions between microbes and tumors necessitates sophisticated methodologies. Conventional culture-based microbiology fails to adequately represent the complete diversity of microbes found within tumors, primarily due to low biomass, contamination risks, and the existence of unculturable species (Salter et al., 2014). Metagenomics, characterized by the high-throughput sequencing of microbial DNA directly from clinical or environmental samples, has emerged as a groundbreaking approach. This approach allows for the impartial identification of microbial taxa, the characterization of functional genes, and the analysis of host–microbe interactions without requiring cultivation (Riquelme et al., 2019; Poore et al., 2020). 16S rRNA gene sequencing and shotgun metagenomics are both extensively utilized in the study of the cancer microbiome, each offering different advantages and disadvantages. By integrating metagenomic insights with clinical and multi-omics datasets, a deeper understanding is emerging regarding how tumor-associated microbes influence oncogenesis, immune infiltration, therapeutic resistance, and clinical outcomes. This has created new opportunities for the advancement of diagnostic biomarkers, prognostic indicators, and therapies aimed at the microbiome (Nejman et al., 2020). Nonetheless, considerable obstacles persist, such as the need for standardized methodologies, effective contamination control, and the translation of research findings into clinical practice.

The topic of this section investigates into the infections present in the tumor microenvironment and examines how metagenomic technologies contribute to the characterization of these microbial communities. This chapter aims to provide a comprehensive overview of this rapidly evolving field at the intersection of oncology, microbiology, and computational biology by investigating the mechanisms that make tumors susceptible to infection, the methodological approaches used to investigate tumor-associated microbiota, and the therapeutic implications of these findings.

2.The Tumor Microenvironment

The tumor microenvironment is frequently colonized by diverse species including lots of different bacteria, viruses, fungi and other microorganisms, all of which can influence cancer initiation, progression and immune evasion. Among bacterial species, Fusobacterium nucleatum is usually associated with colorectal cancer, leading to promotion of tumor growth through altering host immunity and epithelial cell signaling. Moreover, Helicobacter pylori is highly seen in gastric cancer. It causes chronic inflammation and production of genotoxins in gastric cancer patients (Kostic et al., 2013; Plummer et al., 2015). Viral pathogens also play a significant role in the tumor microenvironment. For instance, while human papillomavirus is a key etiological agent in cervical and oropharyngeal cancers, hepatitis B and C viruses are implicated in hepatocellular carcinoma through mechanisms involving chronic inflammation and direct oncogenic transformation (de Martel et al., 2020). Epstein-Barr virus is another notable example, contributing to lymphomas and nasopharyngeal carcinoma by encoding viral proteins that disrupt host cell cycle control and immune surveillance (Young et al., 2016). In addition to bacterial and viral species, fungal communities are increasingly recognized for their impact in the tumor microenvironment. For example, Malassezia species have been linked to pancreatic cancer, potentially promoting tumorigenesis through activation of the complement cascade and immune modulation (Aykut et al., 2019).To sum up, interactions among bacteria, viruses, fungi and protozoa in the tumor microenvironment may lead to synergistic effects on inflammation, immune suppression and metabolic reprogramming, emphasizing the need for holistic microbiome profiling in cancer research.

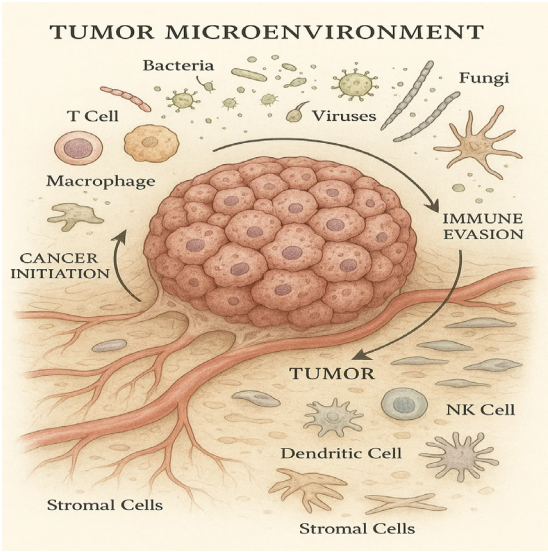


Figure 1: *The tumor microenvironment (TME)*

2.1. The Mechanisms of susceptibility to infections

The tumor microenvironment (TME) is prone to infections due to immunological, structural, and metabolic changes that distinguish it from normal tissues. These methods not only allow infections to survive within tumors, but they can also affect tumor biology.

2.1.1. The immunosuppressive environment

The TME is differentiated by the increase of regulatory T cells (Tregs), myeloid-derived suppressor cells (MDSCs), and tumor-associated macrophages (TAMs). These cells decrease antimicrobial T-cell and natural killer (NK) cell activities, allowing for microbial colonization (Joyce & Fearon, 2015; Binnewies et al., 2018). They also secrete immunosuppressive cytokines such IL-10 and TGF-β.

2.1.2. Hypoxia and metabolic stress

Tumors can cause hypoxia and metabolic stress due to rapid expansion and abnormal vasculature, resulting in nutrient-depleted zones. HIFs change the metabolism of immune cells, decrease the effectiveness of antimicrobial reactions, and increase microbial tolerance. Parallel to this, immunological surveillance is further compromised by an accumulation of lactate and acidosis (Noman et al., 2015).

2.1.3. Vascular abnormalities:

Tumor vasculature is disordered and leaky, affecting immune cells and antimicrobial peptide delivery to the tumor microenvironment. These anomalies in blood vessels also limit drug penetration, allowing microorganisms to escape detection and thrive (Jain, 2014).

2.1.4. Therapy-induced vulnerabilities

Cancer therapies including chemotherapy, radiation, and immunotherapy can impair mucosal barriers and reduce immune responses, leading to an increased risk of secondary infections. Furthermore, drugs and hospitalization in cancer patients might alter the natural microbiota, allowing opportunistic infections to thrive (Montassier et al., 2015).

Collectively, these pathways show how the TME's structural and immunological features make tumors highly susceptible to microbial invasion. This weakness not only promotes pathogen persistence but also allows these microbes to actively alter tumor growth, immune evasion, and treatment response.

3. Applications of Metagenomics

In cancer microbiome research, metagenomic techniques are crucial for analyzing microbial populations inside tumor tissues, surrounding non-malignant tissues, and biofluids. There are two primary sequencing strategies that are extensively used: 16S rRNA gene sequencing and shotgun metagenomics sequencing, all of which have their own processes, advantages, and limits.

3.1. 16S rRNA Gene Sequencing

The 16S rRNA sequencing targets the ribosomal RNA gene, which is found in nearly all bacterial species and is made up of both conserved and hypervariable regions. The conserved regions facilitate the application of universal primers, whereas the hypervariable regions (V1–V9) offer phylogenetic signals for species classification (Caporaso et al., 2011). In oncology research, particular hypervariable regions, such as V3–V4, are frequently amplified and read utilizing next-generation sequencing technologies. This approach is particularly valuable in cancer microbiome research, as it facilitates dependable taxonomic profiling of bacterial

communities, generally at the genus level, while its resolution at the species and strain levels is constrained. Moreover, it offers a cost-efficient and high-throughput approach, facilitating rapid screening of many samples. These properties make it particularly advantageous for comparative assessments of microbial composition, such as examining differences between tumor tissues and neighboring normal tissues or comparing variances among different patient cohorts. Nonetheless, 16S rRNA sequencing possesses limitations. It is not capable of detecting non-bacterial species, including viruses, fungi, or archaea, and is limited in its capacity to identify functional genes or metabolic pathways (Knight et al., 2018). Within the tumor microenvironment (TME), this indicates that significant interactions with fungi (e.g., Malassezia in pancreatic cancer) or viruses (e.g., HPV, HBV) may be disregarded when depending exclusively on 16S profiling.

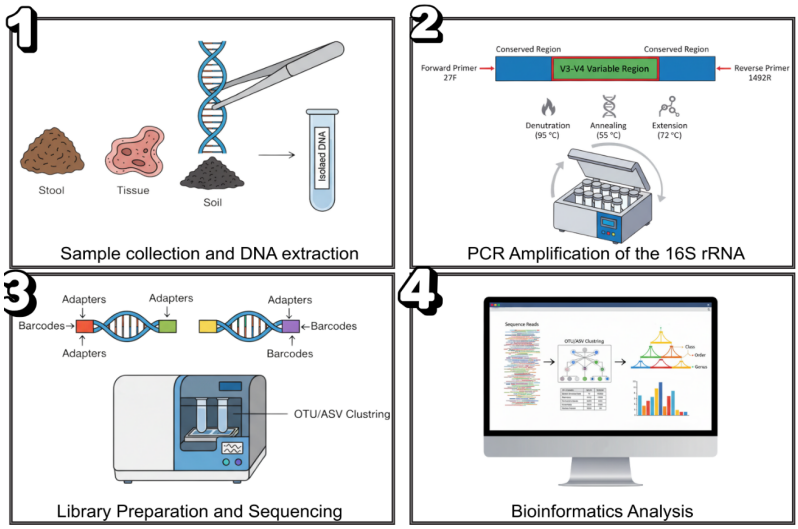


Figure 2: *The workflow of 16S rRNA gene sequencing*

3.2. Shotgun Metagenomics Sequencing

Conversely, shotgun metagenomics sequencing comprises sequencing all genetic material inside a sample without targeted amplification. The entire genetic content is sequenced using high-throughput technology after DNA extracted from tumor tissue or associated biofluids is randomly fragmented.

This strategy supplies many advantages. It facilitates high-resolution taxonomic profiling, down to the species and strain level, which is crucial for differentiating pathogenic subtypes (Ranjan et al., 2016).

This approach facilitates the characterization of functional genes, identifying microbial genes that play roles in metabolism, virulence, and resistance. It offers valuable insights into the ways microbes may affect tumor progression or contribute to therapy resistance (Beghini et al., 2021). Furthermore, shotgun sequencing facilitates the identification of a wide range of microbial kingdoms beyond just bacteria, encompassing viruses, fungi, protozoa, and archaea, thereby providing a comprehensive perspective on the tumor-associated microbiome (Wood et al., 2019). A significant benefit involves the integration with host data, enabling the investigation of host–microbe interactions, including the metabolism of chemotherapeutic agents by microbial enzymes and the disruption of host immunity by viral proteins.

While there are notable strengths, shotgun metagenomics encounter serious obstacles, such as low microbial biomass in tumor samples, increased host DNA contamination, and the requirement for advanced bioinformatics pipelines (Poore et al., 2020). To tackle these challenges, specific workflows—like host DNA depletion techniques and computational subtraction methods—are frequently utilized.

The analysis of tumor-associated microbial data in downstream bioinformatics necessitates the use of specialized computational tools that are specifically designed for each sequencing method. In the analysis of 16S rRNA datasets, it is common to utilize platforms including QIIME2 for tasks such as quality control, sequence clustering, and taxonomic classification. In contrast, shotgun metagenomics data are generally examined through more extensive pipelines, such as MetaPhlAn, HUMAnN, and Kraken2, which facilitate not only taxonomic profiling but also the reconstruction of functional pathways and gene-level annotation (Beghini et al., 2021; Wood et al., 2019).

The combination of these bioinformatics frameworks offers numerous significant benefits for the study of the cancer microbiome. In addition to identifying taxa, metagenomic analysis has the potential to reveal the functional characteristics of microbes associated with tumors, bring attention to rare or novel taxa that could have clinical significance, and demonstrate host–microbe interactions from a systems biology perspective. The comprehensive nature of this information enhances our comprehension of the roles that microbial communities play in carcinogenesis, tumor progression, immune modulation, and therapeutic responses.

3.3. Complementary Roles in Cancer Research

Although 16S rRNA sequencing serves as an economical and rapid method for assessing bacterial community structures, shotgun metagenomics provides a more thorough understanding, incorporating taxonomic, functional, and ecological perspectives. In practice, numerous studies on the cancer microbiome utilize a two-tiered approach, beginning with 16S rRNA sequencing for exploratory profiling, and subsequently employing shotgun metagenomics for comprehensive functional and taxonomic analyses of samples (Nejman et al., 2020; Riquelme et al., 2019).

Consequently, the use of metagenomic techniques in cancer research enhances our comprehension of the tumor microbiome and facilitates the identification of biomarkers, therapeutic targets, and treatment strategies informed by the microbiome.

4.Insights from Metagenomic Studies in Cancer

Recent studies have uncovered distinct microbial signatures associated with specific cancer types, revealing potential roles for intratumoral microbiota in cancer biology and therapy response. For instance, as mentioned before, fusobacterium nucleatum is consistently enriched in colorectal cancer tissues and it causes tumor progression and immune evasion in TME (Kostic et al., 2013). On the other hand, similarly, Helicobacter pylori is associated with gastric cancer development and lead to chronic inflammation and oncogenic signaling in the tumor microenvironment (Wroblewski et al., 2010). So, the tumor microenvironment possesses unique microbiota that modulates immune cell infiltration, cytokine profiles, and metabolic pathways, thereby influencing tumor behavior and treatment outcomes. For example, microbial metabolites within the tumor microenvironment can change the efficacy of immunotherapy and chemotherapy, as observed in pancreatic cancer where intratumoral bacteria metabolize gemcitabine, reducing its effectiveness (Geller et al., 2017). When come to the comparison between the tumor-resident microbiota with the gut microbiota, it seems that they are significantly different from each other in terms of composition and function. While the gut microbiota predominantly influences systemic immunity and inflammation, the tumor microbiota acts locally within the TME, contributing to heterogeneity in therapeutic responses (Routy et al., 2018; Nejman et al., 2020).

5.Challenges and Future Directions

The study of infections within the tumor microenvironment presents significant technical and methodological challenges, due to the low microbial biomass of tumor tissues, higher risk of contamination and the complex interplay among host, microbial and cancer signals. Standard culture-based methods often lack the sensitivity and resolution to capture the full diversity of microbial species within the tumor microenvironment (Nejman et al., 2020). Metagenomic sequencing, especially shotgun metagenomics, has emerged as a powerful tool for overcoming these challenges. Shotgun metagenomics allows the unbiased detection of bacterial, viral, fungal and archaeal genomes directly from tumor samples (Poore et al., 2020). Metagenomic sequencing, especially shotgun metagenomics, has thus become an effective way to address these challenges. Through the direct sequencing of microbial DNA extracted from tumor samples, shotgun metagenomics facilitates the impartial identification of genomes from bacteria, viruses, fungi, and archaea (Poore et al., 2020). By combining clinical and molecular datasets, these methods have revealed significant functions of tumor-resident microbes in shaping immune responses, affecting therapy resistance, and promoting tumor progression. These findings pave the way for the identification of biomarkers and the creation of innovative therapeutics aimed at the microbiome in various cancer types (Riquelme et al., 2019). However, the application of these insights in clinical settings is still limited due to differences among patients, the lack of standardized processes, and the pressing requirement for extensive validation studies. Integrative multi-omics approaches that integrate metagenomics with transcriptomics, proteomics, metabolomics, and spatial profiling offer potential for a comprehensive understanding of host–microbe–tumor interactions and could aid in the discovery of reliable biomarkers and therapeutic targets (Minot & Willis, 2019).

The field of study is expected to be shaped forward by several important directions. Initially, it is crucial to standardize sampling procedures and analytical workflows to improve reproducibility and facilitate significant cross-cohort comparisons. Creating standardized protocols for sample collection, processing, sequencing, and bioinformatic analysis will be crucial for the progression of clinical applications (Gilbert et al., 2025). Secondly, integrating spatial and longitudinal analyses can provide remarkable insights into the structure of microbial communities within tumors and their dynamic evolution throughout disease progression or treatment. Third, there is an increasing demand for functional validation using experimental model systems including patient-derived organoids and tumor-on-a-chip technologies, which can aid in establishing causal relationships between specific microbes and tumor phenotypes (Qasem

et al., 2024).

Another promising direction is the advancement of interventions that modulate the microbiome, such as probiotics, engineered microbial consortia, and fecal microbiota transplantation. These strategies may improve therapeutic responses; however, they necessitate thorough clinical evaluation to verify their efficacy and safety (Routy et al., 2024). The combination of artificial intelligence and machine learning with multi-omics datasets presents opportunities for identifying predictive microbial signatures, stratifying patients for personalized oncology, and discovering novel therapeutic targets (Gilbert et al., 2025). Ultimately, advancements in this area require meticulous consideration of regulatory, ethical, and biosafety factors. It is essential to tackle concerns surrounding patient consent, data privacy, and the potential long-term risks associated with microbiome manipulation to ensure the ethical clinical application of microbiome-based diagnostics and therapies.

6.Conclusion

In conclusion, infections within the tumor microenvironment play multifaceted roles in cancer progression, immune modulation and therapeutic response. Specific microbial populations have been shown to promote oncogenesis, influence immune infiltration and even modulate drug efficacy, highlighting the clinical significance of characterizing the tumor-associated microbiota (Riquelme et al., 2019; Kostic et al.,2013). In this context, metagenomics has emerged as a transformative tool in oncology, since it enables comprehensive, culture-independent profiling of microbial communities within tumors and their surroundings. Shotgun metagenomics offers high-resolution insights into taxonomic and functional diversity of microbial populations, providing a foundation for biomarker discovery, personalized therapy and novel treatment targets (Poore et al., 2020; Nejman et al., 2020). However, this field faces critical challenges including the absence of standardized sample collection protocols, contamination control, or data analysis pipelines. To fully harness the potential of microbiome research in cancer, there is an urgent need for standardized methodologies and more translational studies that validate microbiome-derived biomarkers and elucidate causal relationships (Salter et al., 2014; Minot & Willis, 2019). By integrating microbiome data with host multi-omics and clinical parameters, future research can more effectively translate these findings into actionable insights for precision oncology.

REFERENCES

Aykut, B., Pushalkar, S., Chen, R., Li, Q., Abengozar, R., Kim, J. I., Shadaloey, S. A., Wu, D., Preiss, P., Verma, N., Guo, Y., Saxena, A., Vardhan, M., Diskin, B., Wang, W., Leinwand, J., Kurz, E., Kochen Rossi, J. A., Hundeyin, M., Zambrinis, C., … Miller, G. (2019). The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. Nature, 574(7777), 264–267. https://doi.org/10.1038/s41586-019-1608-2

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife, 10, e65088. https://doi.org/10.7554/eLife.65088

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., Coussens, L. M., Gabrilovich, D. I., Ostrand-Rosenberg, S., Hedrick, C. C., Vonderheide, R. H., Pittet, M. J., Jain, R. K., Zou, W., Howcroft, T. K., Woodhouse, E. C., Weinberg, R. A., & Krummel, M. F. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nature Medicine, 24(5), 541–550. https://doi.org/10.1038/s41591-018-0014-x

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS, 108(Suppl 1), 4516–4522. https://doi.org/10.1073/pnas.1000080107

de Martel, C., Georges, D., Bray, F., Ferlay, J., & Clifford, G. M. (2020). Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. The Lancet Global Health, 8(2), e180–e190. https://doi.org/10.1016/S2214-109X(19)30488-7

Geller, L. T., Barzily-Rokni, M., Danino, T., Jonas, O. H., Shental, N., Nejman, D., Gavert, N., Zwang, Y., Cooper, Z. A., Shee, K., Thaiss, C. A., Reuben, A., Livny, J., Avraham, R., Frederick, D. T., Ligorio, M., Chatman, K., Johnston, S. E., Mosher, C. M., Brandis, A., … Straussman, R. (2017). Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science, 357(6356), 1156–1160. https://doi.org/10.1126/science.aah5043

Gilbert, J. A., Azad, M. B., Bäckhed, F., Blaser, M. J., Byndloss, M., Chiu, C. Y., Chu, H., Dugas, L. R., Elinav, E., Gibbons, S. M., Gilbert, K. E., Henn, M. R., Ishaq, S. L., Ley, R. E., Lynch, S. V., Segal, E., Spector, T. D., Strandwitz, P., Suez, J., Tropini,

C., … Knight, R. (2025). Clinical translation of microbiome research. Nature Medicine, 31(4), 1099–1113. https://doi.org/10.1038/s41591-025-03615-9

Iida, N., Dzutsev, A., Stewart, C. A., Smith, L., Bouladoux, N., Weingarten, R. A., Molina, D. A., Salcedo, R., Back, T., Cramer, S., Dai, R. M., Kiu, H., Cardone, M., Naik, S., Patri, A. K., Wang, E., Marincola, F. M., Frank, K. M., Belkaid, Y., Trinchieri, G., … Goldszmid, R. S. (2013). Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. Science, 342(6161), 967–970. https://doi.org/10.1126/science.1240527

Jain, R. K. (2014). Antiangiogenesis strategies revisited: from starving tumors to alleviating hypoxia. Cancer Cell, 26(5), 605–622. https://doi.org/10.1016/j.ccell.2014.10.006

Joyce, J. A., & Fearon, D. T. (2015). T cell exclusion, immune privilege, and the tumor microenvironment. Science, 348(6230), 74–80. https://doi.org/10.1126/science.aaa6204

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L. I., McDonald, D., Melnik, A. V., Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson, L. R., Tripathi, A., Xu, Z. Z., Zaneveld, J. R., … Dorrestein, P. C. (2018). Best practices for analysing microbiomes. Nature Reviews Microbiology, 16(7), 410–422. https://doi.org/10.1038/s41579-018-0029-9

Kostic, A. D., Chun, E., Robertson, L., Glickman, J. N., Gallini, C. A., Michaud, M., Clancy, T. E., Chung, D. C., Lochhead, P., Hold, G. L., El-Omar, E. M., Brenner, D., Fuchs, C. S., Meyerson, M., & Garrett, W. S. (2013). Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host & Microbe, 14(2), 207–215. https://doi.org/10.1016/j.chom.2013.07.007

Minot, S. S., & Willis, A. D. (2019). Subsampling to reduce false positives in differential abundance analysis of microbiome data. Frontiers in Microbiology, 10, 2241. https://doi.org/10.3389/fmicb.2019.02241

Montassier, E., Gastinne, T., Vangay, P., Al-Ghalith, G. A., Bruley

Des Varannes, S., Massart, S., Moreau, P., Potel, G., De, M. F., Cocheti Ere †, L., Batard, E., & Knights, D. (n.d.). Chemotherapy-driven dysbiosis in the intestinal microbiome. https://doi.org/10.1111/apt.13302

Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., Meltser, A., Douglas, G. M., Kamer, I., Gopalakrishnan, V., Dadosh, T., Levin-Zaidman, S., Avnet, S., Atlan, T., Cooper, Z. A., Arora, R., … Straussman, R. (2020). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science, 368(6494), 973–980. https://doi.org/10.1126/science.aay9189

Noman, M. Z., Hasmim, M., Messai, Y., Terry, S., Kieda, C., Janji, B., & Chouaib, S. (2015). Hypoxia: a key player in antitumor immune response. American Journal of Physiology - Cell Physiology, 309(9), C569–C579. https://doi.org/10.1152/ajpcell.00207.2015

Plummer, M., Franceschi, S., Vignat, J., Forman, D., & de Martel, C. (2015). Global burden of gastric cancer attributable to Helicobacter pylori. International Journal of Cancer, 136(2), 487–490. https://doi.org/10.1002/ijc.28999

Poore, G. D., Kopylova, E., Zhu, Q., Carpenter, C., Fraraccio, S., Wandro, S., Kosciolek, T., Janssen, S., Metcalf, J., Song, S. J., Kanbar, J., Miller-Montgomery, S., Heaton, R., Mckay, R., Patel, S. P., Swafford, A. D., & Knight, R. (2020). Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature, 579(7800), 567–574. https://doi.org/10.1038/s41586-020-2095-1

Qasem, H. H., & El-Sayed, W. M. (2024). The bacterial microbiome and cancer: development, diagnosis, treatment, and future directions. Clinical and Experimental Medicine, 25(1), 12. https://doi.org/10.1007/s10238-024-01523-9

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochemical and Biophysical Research Communications, 469(4), 967–977. https://doi.org/10.1016/j.bbrc.2015.12.083

Rehman, S. S. U., Nasar, M. I., Mesquita, C. S., Al Khodor, S.,

Notebaart, R. A., Ott, S., Mundra, S., Arasardanam, R. P., Muhammad, K., & Alam, M. T. (2025). Integrative systems biology approaches for analyzing microbiome dysbiosis and species interactions. Briefings in Bioinformatics, 26(4), bbaf323. https://doi.org/10.1093/bib/bbaf323

Riquelme, E., Zhang, Y., Zhang, L., Montiel, M., Zoltan, M., Dong, W., Quesada, P., Sahin, I., Chandra, V., San Lucas, A., Scheet, P., Xu, H., Hanash, S. M., Feng, L., Burks, J. K., Do, K. A., Peterson, C. B., Nejman, D., Tzeng, C. D., Kim, M. P., … McAllister, F. (2019). Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes. Cell, 178(4), 795–806.e12. https://doi.org/10.1016/j.cell.2019.07.008

Routy, B., Jackson, T., Mählmann, L., Baumgartner, C. K., Blaser, M., Byrd, A., Corvaia, N., Couts, K., Davar, D., Derosa, L., Hang, H. C., Hospers, G., Isaksen, M., Kroemer, G., Malard, F., McCoy, K. D., Meisel, M., Pal, S., Ronai, Z., Segal, E., … Zitvogel, L. (2024). Melanoma and microbiota: Current understanding and future directions. Cancer Cell, 42(1), 16–34. https://doi.org/10.1016/j.ccell.2023.12.003

Routy, B., Le Chatelier, E., Derosa, L., Duong, C. P. M., Alou, M. T., Daillère, R., Fluckiger, A., Messaoudene, M., Rauber, C., Roberti, M. P., Fidelle, M., Flament, C., Poirier-Colame, V., Opolon, P., Klein, C., Iribarren, K., Mondragón, L., Jacquelot, N., Qu, B., Ferrere, G., … Zitvogel, L. (2018). Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science, 359(6371), 91–97. https://doi.org/10.1126/science.aan3706

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology, 12, 87. https://doi.org/10.1186/s12915-014-0087-z

Schwabe, R. F., & Jobin, C. (2013). The microbiome and cancer. Nature Reviews Cancer, 13(11), 800–812. https://doi.org/10.1038/nrc3610

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biology, 20(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Wroblewski, L. E., Peek, R. M., Jr, & Wilson, K. T. (2010). Helicobacter pylori and gastric cancer: factors that modulate disease risk. Clinical Microbiology Reviews, 23(4), 713–739. https://doi.org/10.1128/CMR.00011-10

Young, L. S., Yap, L. F., & Murray, P. G. (2016). Epstein-Barr virus: more than 50 years old and still providing surprises. Nature Reviews Cancer, 16(12), 789–802. https://doi.org/10.1038/nrc.2016.92

99

Chapter 2

TRANSKRİPSİYON FAKTÖRLERİ VE GEN REGÜLATÖR AĞLARI: MEKANİZMADAN UYGULAMAYA

Erdem Burak KÜÇÜKKARA[1]

Elif Zeynep DEMİR[2]

Kazım Yalçın ARGA[3]

1 Marmara University, Faculty of Engineering, Department of Bioengineering, İstanbul, Türkiye, ORCID: 0009-0009-3263-1433
2 Marmara University, Faculty of Engineering, Department of Bioengineering, İstanbul, Türkiye, ORCID: 0009-0004-7805-7418
3 Marmara University, Faculty of Engineering, Department of Bioengineering, İstanbul, Türkiye, ORCID: 0000-0002-6036-1348

1. Giriş

Tanım ve Önem

Gen regülatör ağları (GRN), hücre içi transkripsiyonel düzenlemelerin network temelli gösterimidir. Bu ağlar, gen ekspresyonunu kontrol eden transkripsiyon faktörleri (TF) ile bu faktörlerin doğrudan veya dolaylı olarak etkilediği hedef genleri arasında kurulan kompleks etkileşimleri içerir (Davis ve ark., 2017). GRN'ler biyolojik süreçlerde kilit rol oynar: gelişim, homeostaz, metabolizma, davranış ve hastalık gibi fonksiyonların düzenlenmesinde önemli katkı sağlar. Sistem biyolojisi açısından da büyük öneme sahiptir çünkü genetik risklerin anlaşılması, ilaç hedeflerinin önceliklendirilmesi ve kişiselleştirilmiş tıp uygulamalarının geliştirilmesinde doğrudan kullanılırlar (Davis ve ark., 2017).

Yeni nesil sekanslama teknolojileri ve hesaplamalı yaklaşımların gelişimiyle birlikte, GRN çıkarımı ve analizi giderek daha sistematik ve hassas hale gelmiştir. Özellikle genomik verilerin artmasıyla birlikte, TF-TG ilişkilerinin zamansal ve doku spesifik dinamiklerinin modellenmesi hem normal fizyolojik süreçlerin hem de patojenik durumların anlaşılmasında hayati önem kazanmıştır (Davis ve ark, 2017).

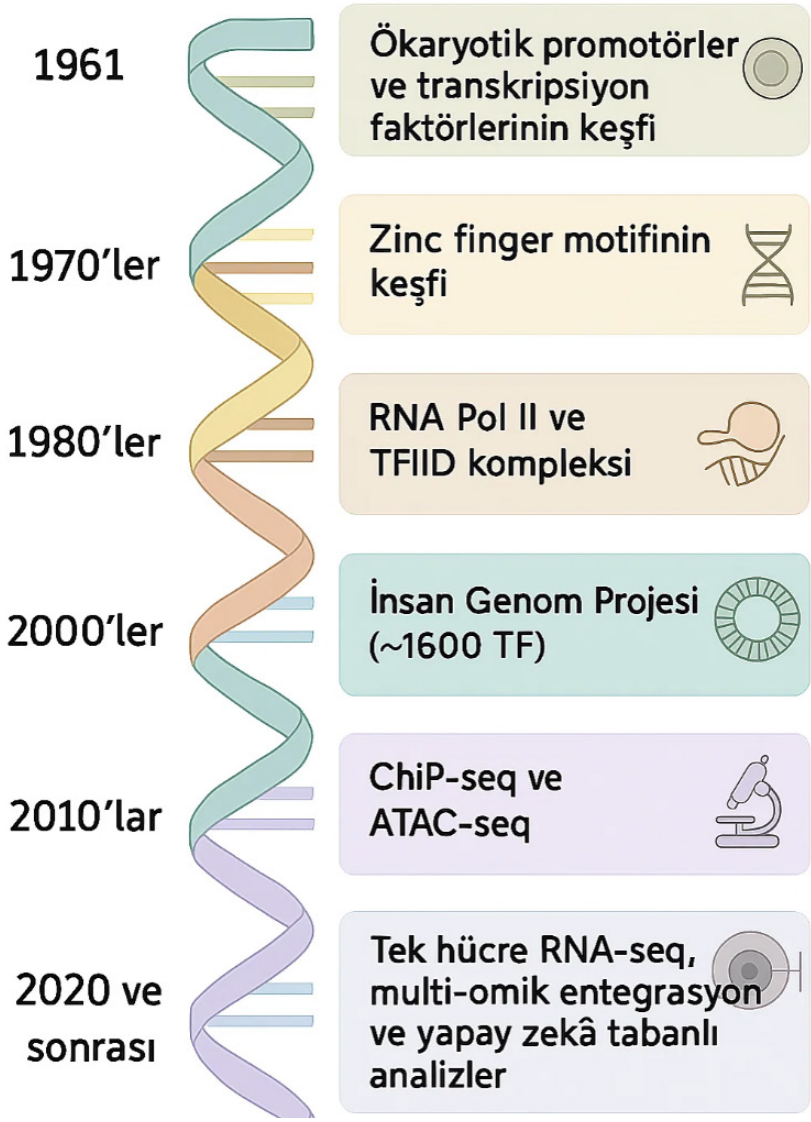
2. Transkripsiyon Faktörleri

2.1. Tarihçe

Transkripsiyon kontrolünü anlamak büyük bir çabadır. 1980'lerin başlarında, model ökaryotik genlerde doğru başlatma için çok sayıda değişkenin gerekli olduğu açıktı (Lis, 2019). Tek bir gen için RNA polimeraz II (Pol II) transkripsiyonuna dair kapsamlı bir mekanik bilgi, birçok bileşenin izole edilmesini ve uygun kromatin şablonları üzerinde doğru bir şekilde bir araya getirilmesini gerektirir. Bu çalışma alanı, moleküler düzeyde transkripsiyonun kapsamlı bir anlayışının önemli çaba ve yenilikçi teknoloji gerektireceğini kabul etti ve Pol II'nin tüm mRNA kodlayan genleri transkribe ettiğini, her birinin belirli düzenleyici proteinler gerektirdiğini kabul etti (Lis, 2019).

İnsanlarda yaklaşık 20,000 protein kodlayan genin hücreye özgü ifadesi, birçok biyolojik süreç için hayati öneme sahiptir ve öncelikle transkripsiyon düzeyinde düzenlenir. Transkripsiyonun mekanizmaları ve düzenlenmesi konusundaki ilk anlayış, çoğunlukla genetik araştırmalardan kaynaklanan bakteriler ve bakteriyofajlar üzerindeki çalışmalardan gelmiştir; özellikle Jacob ve Monod'un 1961'deki lac operonu üzerine yazdıkları dikkate değer makalede vurgulanmıştır (Roeder,

2019).



Figür 1. *TF Kullanımının Tarihi Gelişimi*

2.2. Yapısal Özellikleri

Transkripsiyon faktörleri, DNA'ya bağlanma alanları, aktivasyon veya represyon domainleri ve diğer proteinlerle etkileşim kurabilen bölgelerden oluşur. Yaygın DNA bağlama motifleri arasında zinc finger (C2H2 tipi), helix-turn-helix, leucine zipper ve bHLH (basic Helix-Loop-Helix) yapılar yer alır. Aktivasyon veya represyon domainleri, gen ekspresyonunu

artırıcı ya da baskılayıcı sinyaller gönderir. Ayrıca TF'ler epigenetik düzenleyicilere (HDAC, HAT, SWI/SNF kompleksleri) bağlanarak uzun vadeli genetik değişimleri tetikleyebilir. Bu yapısal farklılıklar, TF'lerin işlevsel çeşitliliğini ve regülasyon mekanizmalarındaki esnekliği sağlar (Spitz & Furlong, 2012).

TFIID kompleksi, transkripsiyon başlangıcında anahtar rol oynar. Bu kompleks, TBP (TATA-box bağlayan protein) ve çeşitli TAF (TBP Associated Factor) proteinlerinden oluşur. TAF'ler hem DNA hem da diğer TF'lerle etkileşim kurarak promotor bölgesindeki nükleozomal yapıyı açmak ve RNA polimeraz II'yi çekmek için görev alır. Bazı TAF'ler histon benzeri yapılar taşıyarak kromatin modellemesini destekler. Örneğin, TAFII250 histon asetiltransferaz (HAT) aktivitesine sahip olup promotör bölgesinin açık hale gelmesinde kritik rol üstlenir (Carlberg & Molnár, 2016).

2.3. Fonksiyonel Sınıflandırma

Transkripsiyon faktörleri, işlevlerine göre üç ana gruba ayrılabilir: aktivatörler, represörler ve gelişimsel anahtar faktörler. Aktivatör transkripsiyon faktörleri, örneğin CREB ve NF-κB gibi, gen ekspresyonunu teşvik ederken, represör TF'lerden REST ve GFI1 gibi faktörler gen ekspresyonunu bastırıcı etki gösterir. Gelişim sürecini doğrudan yönlendiren kritik TF'ler arasında OCT4, SOX2, FOXA1 ve PU.1 gibi faktörler yer alır. Bu gruplar, hücre farklılaşması ve doku gelişimi süreçlerinde hayati öneme sahip olup, gelişim biyolojisi ve patojenez bağlamında incelenmelidir (Spitz & Furlong, 2012).

Ayrıca dokuya özel transkripsiyon faktörleri de vardır ve bu faktörler, spesifik hücre tiplerinin işlevsel özelliklerini kazanmasında önemli rol oynar. Örneğin, hematopoetik hücrelerde PU.1, nöronlarda NRSF/REST, kas hücrelerinde MyoD ve karaciğer hücrelerinde HNF4 gibi faktörler belirli hücre tiplerinin genomik programlarını aktif hale getirir ya da susturur. Bu envanter, özellikle kanser araştırmaları ve gelişimsel bozuklukların anlaşılmasında temel bir referans kaynağıdır (Vaquerizas ve ark, 2009).

2.4. Bağlanma Dinamikleri

TF'lerin DNA'ya bağlanma özgüllüğü, motif tanıma yöntemleriyle analiz edilir. PWM (Position Weight Matrix) modelleri yaygın olarak kullanılır. Yüksek çözünürlüklü SELEX ve ChIP-seq verileri ile motif

çıkarımı yapılır. Enhancer bölgeleri ve cis-regülatuar modüller de TF bağlamasına önemli katkı sağlar. Ancak bazı motifler genomda işlevsel değildir veya nükleozomal DNA üzerinde erişilemez konumlarda olabilir; bu da motif tabanlı yaklaşımların sınırlılığını gösterir (Weirauch ve ark, 2014).

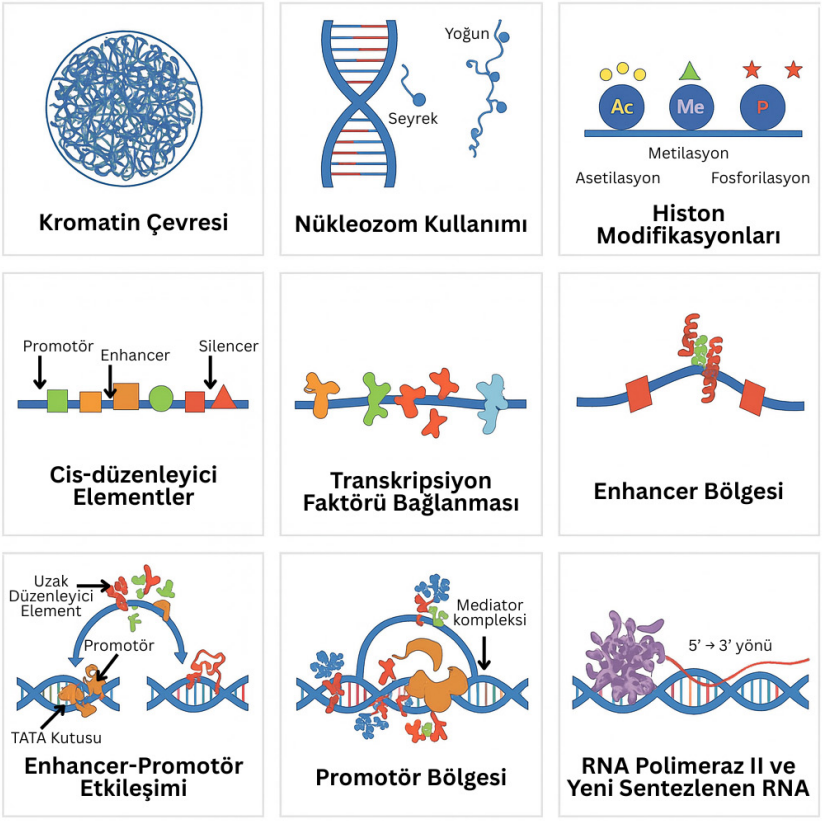
PWM modelleri, genomik sekanslarda TF bağlanma sitelerini tahmin etmede oldukça yaygındır. Ancak bu modellerin doğruluğu, motiflerin çevreleyen sekansların etkisiyle değişebilir. Özellikle enhancer bölgelerinde motiflerin çevresel dizilimi, TF bağlanma kapasitesini belirlemede önemlidir. Bazı TF'lerin bağlandığı motifler, genomda açıkça tanımlanmış olsa bile transkripsiyonel olarak pasif kalabilmektedir. Bu durum, motif varlığı ile işlevsellik arasında bir uyumsuzluk olduğunu gösterir (Weirauch ve ark., 2014).

Drosophila melanogaster embriyonik gelişiminde dorsal–ventral eksen patterning'i, GRN'in evrimsel olarak ne kadar korunmuş olduğuna dair güçlü bir örnektir. Maternal sinyallerin etkisiyle ortaya çıkan Dorsal gradyanı, Twi, Sna ve Dpp gibi TF'lerin işlevsel düzenlenmesini sağlar. Bu süreç, Notch, Wnt ve EGF sinyal yollarıyla da bağlantılıdır (Levine & Tjian, 2003).

2.5. Regülasyon Mekanizmaları

Transkripsiyon faktörleri post-translasyonel modifikasyonlarla (fosforilasyon, asetilasyon, ubiquitinasyon) düzenlenir. Bu modifikasyonlar nükleer lokalizasyonu, DNA bağlanma kapasitesini ve diğer proteinlerle etkileşimlerini etkiler. Protein-protein etkileşimleri, özellikle Mediator kompleksi ile birlikte, TF'lerin işlevsel kompleksler oluşturmasını sağlar. Bu düzenleyici mekanizmalar sayesinde TF'ler çevresel sinyallere hızlı şekilde yanıt verebilir ve hücrenin ihtiyaçlarına göre dinamik olarak aktif hâle gelebilir veya susturulabilir (Spitz & Furlong, 2012).

Post-translasyonel modifikasyonlar, TF'lerin işlevsel aktivitelerini doğrudan etkiler. Örneğin fosforilasyon, TF'lerin nükleusa girişini ya da çıkışını düzenleyebildiği gibi onların bağlanma yeteneğini de artırabilir ya da azaltabilir. Benzer şekilde asetilasyon genellikle TF'lerin transaktivasyon kapasitesini yükseltirken, ubiquitinasyon genellikle protein yıkımını tetikler. Bu mekanizmalar, hücrenin kısa ve uzun vadede nasıl tepki vereceğini belirler (Spitz & Furlong, 2012).



Figür 2. *Ökaryotik Hücrelerde Transkripsiyon Regülasyonu*

2.6. İnsan TF Envanteri

İnsan genomunda yaklaşık 1600 transkripsiyon faktörü olduğu tahmin edilmektedir. Bunların çoğu doku spesifik ifade profillerine sahiptir. Evrimsel olarak korunmuş TF'ler dışında yeni ortaya çıkan faktörler de vardır. Bu envanter, kanser araştırmaları ve gelişimsel bozuklukların anlaşılmasında temel kaynaktır. Özellikle genomik veri tabanları ve hesaplamalı yaklaşımlarla TF'lerin sınıflandırılması ve işlevlerinin çıkarımı giderek daha sistematik hâle gelmektedir (Vaquerizas ve ark., 2009).

Bazı transkripsiyon başlangıç faktörleri, örneğin TFIID kompleksinde yer alan TAFII150, TAFII68 ve TAFII250 gibi, genom boyunca yüksek oranda bulunurlar. Bu faktörlerin bağlandığı promotör bölgelerinin işlevsel analizi, gen ekspresyonunun düzenlenmesinde çok önemli katkılar sunmaktadır. Ayrıca bu faktörlerin mutasyonları bazı gelişimsel bozukluklara ve kanser türlerine bağlı olabilir (Carlberg & Molnár, 2016).

2.7. Hastalık Mekanizmalarında Rolü

Mutant transkripsiyon faktörleri birçok hastalığın altında yatar. Örneğin kanserde EWS-FLI1 füzyonu, mutant p53 ve MYC aşırı ekspresyonu gibi olaylar onkogenezde rol oynar. Nörolojik hastalıklarda FoxP2 geni dil gelişim bozukluklarında anahtar rol oynarken, REST disfonksiyonu nörodejeneratif hastalıklara yol açar. Kalıtsal hastalıklarda ise ZFHX3 mutasyonları aritmi gibi kalp ritim bozukluklarında görülmektedir. Bu bulgular, TF'lerin patojenezdeki merkezi rolünü vurgulamakta ve tedavi stratejilerinin geliştirilmesi için potansiyel hedefler sunmaktadır (Spitz & Furlong, 2012).

Boeva, V. (2016) tarafından yapılan çalışmada, ChIP-seq verileri üzerinden TF bağlama bölgelerinin işlevsel doğrulanmasının nasıl yapıldığı detaylı olarak açıklanmıştır. Bu yöntemler, özellikle enhancer hijacking gibi onkogenik süreçlerin aydınlatılmasında büyük rol oynamıştır (Boeva, 2016).

3. Gen Regülatör Ağları

3.1. GRN'lerin Temel Unsurları

GRN'ler dört ana bileşenlerden oluşur: cis-regülatuar elementler, transkripsiyon faktörleri, hedef genler ve miRNA ve lncRNA gibi post-transkripsiyonel düzenleyiciler. Bu unsurlar, gen ekspresyonunun zamansal ve mekânsal düzenlenebilmesini sağlar. Promotörler ve enhancer bölgeleri, TF'lerin bağlandığı anahtar sitelerdir. Aynı zamanda non-kodlayan RNA'lar da GRN yapısına dahildir ve gen ekspresyonunu negatif ya da pozitif yönde düzenlerler (Davis ve ark., 2017).

Gen ifadesinin ifadesinin başlatılmasında promotörler zorunlu elementlerdir. Promotörlerden uzak mesafelerde ise Enhancer bölgeler bulunabilir ve bunlar üç boyutlu kromatin yapısının yeniden düzenlenmesiyle gen ekspresyonunu artırabilirler (Helmsauer ve ark., 2020). Enhancer bölgeleri, özellikle gelişimsel süreçlerde kritik öneme sahiptir. Bu bölgeler, TF'lerin bağlanmasıyla birlikte promotörlere uzak mesafelerden sinyal ileten cis-regülatuar elementlerdir. Bir enhancer, birden fazla TF ile etkileşime girerek gen ekspresyonunu hassas şekilde ayarlayabilir. Bu nedenle enhancer organizasyonu, GRN'lerin işlevsel bütünlüğü açısından hayati öneme sahiptir (Spitz & Furlong, 2012).

3.2. GRN Oluşturma Yöntemleri

3.2.1. Deneysel Yaklaşımlar

ChIP-seq ve motif analizi ile TF bağlanma bölgeleri tanımlanır. Co-immunoprecipitation verileriyle protein-protein etkileşim haritalama yapılır. Bu teknikler, TF-TG ilişkilerini doğrudan gözlemlemek için kullanılır (Davis ve ark., 2017). Ayrıca motif tanıma ve bağlama özgüllüğü üzerine yapılan çalışmalar, yüksek çözünürlüklü verilerle motif doğrulamasının önemini göstermiştir (Weirauch ve ark., 2014).

Yüksek çözünürlüklü SELEX ve in vitro seçici bağlanma deneyleri, TF'lerin bağlama özgüllüğünü tanımlamak ve motiflerin işlevsel doğrulanmasını sağlamak amacıyla kullanılmıştır. ChIP-seq verileri ile motiflerin genomdaki gerçek konumları ve işlevsel etkinlikleri de belirlenebilmektedir. Bu nedenle motif eşleştirme ve skorlama işlemleri, bağlama ilişkilerinin doğrulanmasında tek başına yeterli olmayabilir (Weirauch ve ark., 2014).

3.2.2. Hesaplamalı Yaklaşımlar

Korelasyon ve mutual information temelli yöntemler (minet, ARACNe), Bayesian ağları, Granger nedensellik yaklaşımı, dinamik diferansiyel denklemlerle modelleme ve probabilistik Boolean ağlar yüksek boyutlu omik verilerden TF-TG ilişkilerini çıkarabilmektedir (Marbach ve ark., 2012). Bu yaklaşımlar, özellikle deneysel verilere ulaşım sınırlı olduğunda, TF-TG ilişkilerinin çıkarımında değerli bir alternatif sunar. Bir Boolean ağında, bir varlık iki farklı duruma ulaşabilir: aktif (1) veya inaktif (0). Bir gen, herhangi bir anda ifade edilmiş veya ifade edilmemiş olarak tanımlanabilir. Her varlığın seviyesi, belirlenmiş bir Boolean fonksiyonu aracılığıyla çeşitli varlıkların seviyelerine göre değiştirilir (Karlebach & Shamir, 2008).

Diferansiyel denklem tabanlı modeller ve bir ana denklemden türetilen stokastik modeller (Thattai ve van Oudenaarden 2001), biyokimyasal olayları ve doğrusal olmayan etkileşimleri etkili bir şekilde tasvir eder; ancak mevcut verilerle uyumlu olmayabilirler. Bayes ağları, Markov ağları ve faktör grafikleri de dahil olmak üzere olasılıksal grafik modeller, düzenleyici ağların modellenmesi için sağlam çerçeveler görevi görür (Friedman 2004, Markowetz ve Spang 2007, Segal ve ark. 2005) ve ölçümlerdeki gürültü ve belirsizliği yönetmede ustadır (Thompson ve ark., 2015).

ARACNe-AP ve minet gibi yöntemler, gen ekspresyon verilerinden yola çıkarak TF-TG ilişkilerini tahmin eder. Özellikle yüksek korelasyona sahip gen çiftlerinin motif analiziyle birleştirilmesi, bağlama ilişkilerinin doğrulanmasını kolaylaştırır. Ayrıca tek hücre düzeyinde verilerin kullanılmasıyla, heterojen hücre popülasyonlarındaki TF-TG ilişkileri daha hassas şekilde çıkarılabilir hale gelmiştir (Marbach ve ark., 2012). Margolin ve arkadaşlarının (2006) geliştirdiği ARACNe yöntemi, genom çapında TF-TG ilişkilerinin çıkarılmasında kullanılan güçlü bir araçtır. Mutual information temelli yaklaşımlar, dolaylı TF-TG etkileşimlerini de modelleyebilir ve böylece ağ çıkarımında daha gerçekçi sonuçlar elde edilebilir.

Thompson, W.A. & Levine, M. tarafından yapılan çalışma, kantitatif cis-regülatuar modellerin , özellikle Drosophila'daki gen ekspresyonu dinamiklerinin modellenmesinde nasıl kullanıldığını açıklamaktadır. Bu modeller, enhancer bölgelerinin işlevsel katkısını ve zaman içindeki değişimi anlamada önemli bir adım olmuştur (Karlebach & Shamir, R. (2008).. Zamansal transkripsiyon faktörü aktiviteleri, dinamik Bayesian ağları aracılığıyla mikrodizi zaman serisi verilerinden çıkarılır. Uzman bir sezgisel yöntem, yeni düzenleyicileri ve düzenleyici bağlantıları belirlemeyi amaçlar (Karlebach & Shamir, 2008).

3.2.3. Çok-Omik Veri Entegrasyonu

Eşleştirilmiş tek hücreli çoklu omik verileri kullanan son teknoloji gen düzenleyici ağ çıkarım yaklaşımlarının oluşturulması, dizileme teknolojisindeki son gelişmelerden kaynaklanmıştır. Bu yöntemler, daha kapsamlı ve doğru gen düzenleme ağları oluşturmak için çeşitli matematiksel ve istatistiksel teknikler kullanır (Kim ve ark., 2023). Transkriptomik, epigenomik ve proteomik verileri birleştirilerek GRN'ler daha hassas şekilde çıkarılabilir. Metilasyon profilleri ve single-cell verileri hücre içi heterojenliğin modellenmesine olanak tanır (Davis ve ark., 2017). Bu entegrasyon, TF'lerin bağlanma bölgelerinin işlevsel olarak aktif olup olmadığını da belirlemeye yardımcı olur. Örneğin, bir enhancer bölgesinin açık (aktivasyon markırları taşıyan) ya da kapalı (represyon işaretleri taşıyan) olması, TF'lerin oradaki bağlanma kapasitesini doğrudan etkiler (Spitz & Furlong, 2012).

Metilasyon profilleri genellikle CpG zengin bölgelerde yoğunlaşmıştır. Bu bölgelerin hipometile olması, gen ekspresyonunu başlatan TF'lerin bağlanma bölgelerini de etkileyebilir. Bu nedenle epigenomik verilerin GRN çıkarımına entegre edilmesi, regülasyonun dinamik yönünü daha net yansıtmaktadır (Spitz & Furlong, 2012). He & Hannon (2004) tarafın-

dan önerilen Bayes tabanlı yaklaşımlar, mikrodizi verilerinden yola çıkarak GRN çıkarımında oldukça başarılı olmuştur. Bu yöntemler, özellikle düşük boyutlu veri setlerinde bile güçlü tahminler sunmuştur.

3.3. Araçlar ve Kaynaklar

GRN analizinde kullanılan araçlar ve kaynaklar büyük ölçüde gelişmiştir. TRANSFAC, JASPAR, ENCODE, PlantPAN3.0, AGRIS ve iRegulon gibi veritabanları geniş çaplı TF bağlanma motifleri ve hedef gen verilerini barındırır. Gene Ontology (GO), KEGG, MSigDB ve BioCarta gibi kaynaklar ise fonksiyonel zenginleştirme analizlerinde kullanılır. Peak anotasyon araçları arasında ChIPpeakAnno, LOLA, GREAT ve PAVIS yer alırken, ağ çıkarımı için ARACNe-AP, minet, Inferelator, VULCAN ve TF2Network yaygın olarak tercih edilir. Cytoscape ve Gephi gibi yazılımlar ise oluşturulan ağların görselleştirilmesi açısından önemlidir (Davis ve ark., 2017).

TRANSFAC ve JASPAR, TF bağlama motiflerinin doğrulanmasında yaygın olarak kullanılan veri tabanlarıdır. Bu kaynaklara entegre edilen motif verileri, motif tanıma ve motif eşleştirme işlemlerinde doğrudan kullanılmaktadır. Cytoscape, ağların görselleştirilmesi yanında alt-ağların analizi ve modülsel yapıların çıkarımı için de güçlü bir platform sağlar (Davis ve ark, 2017). Chan ve ark., 2018 çalışmasında, circHIPK3 ve CDR1as gibi circular RNA'ların ceRNA (competing endogenous RNA) olarak davranarak TF-TG ilişkilerini modüle ettiğini göstermişlerdir. Bu süreç, özellikle kanser gelişiminde gen ekspresyonunu yeniden şekillendiren önemli bir mekanizmadır.

Tek hücreli RNA dizilimi için optimize edilmiş istatistiksel teknikler ve biyoenformatik araçlar, yeni biyolojik bilgiler sağlamıştır; ancak kararlı hücre durumlarının altında yatan farklı ve güçlü gen düzenleyici ağların tanımlanıp tanımlanamayacağı belirsizliğini korumaktadır. Bu zor olabilir çünkü tek hücre düzeyinde, gen ifadesi, transkripsiyonel patlamalardan ve diğer faktörlerden kaynaklanan stokastik varyasyon nedeniyle transkripsiyon faktörü girdilerinin dinamiklerinden kısmen ayrışmış olabilir (Aibar ve ark., 2017)

3.4. Zorluklar ve Sınırlılıklar

GRN çıkarımı ve analizi bazı önemli zorluklar içerir. Hücre/doku özgüllüğü ve dinamizm, farklı koşullarda değişen regülasyon mekanizmaları açısından büyük bir sınırlılıktır. TF-TG ilişkilerinde bağla-

ma özgüllüğünün eksikliği, yanlış pozitif veya negatif tahminlerin yapılmasına neden olabilir. Tekrarlayan DNA elemanlarının bazen TF süngesi gibi davranması da analizlerde gürültüye yol açabilir. Ayrıca çevresel faktörlerin (örneğin oksijen seviyesi) TF aktivitesine olan etkisi, sabit şartlarda yapılan çalışmalarda ihmal edilebilir bir durum değildir (Davis ve ark., 2017).

TF-TG ilişkilerinde bağlama özgüllüğünün düşük olması, özellikle hesaplamalı yaklaşımların sınırlılıklarından biridir. Bu nedenle motif eşleştirme ve motif skorlaması gibi işlemler, bağlama ilişkisini doğrulamak için yalnız başına yeterli olmayabilir. Ek olarak, bir TF'nin bağlandığı tüm bölgelerin aynı anda aktif olması beklenemez; bazı bölgeler işlevsel olmayabilir ya da geçici olarak inaktif hâlde olabilir (Weirauch ve ark., 2014).

3.5. Evrimsel Perspektif

Ortoloji tabanlı GRN çıkarımı, TF bağlama motiflerinin evrimini ve regülatör ağların korunmuşluğunu incelemek için kullanılır. Özellikle insan ve model organizmalar arasındaki karşılaştırmalar, evrimsel yeniliklerin anlaşılmasına yardımcı olur. Bazı TF-TG ilişkileri çok uzak türler arasında bile yüksek derecede korunurken, bazıları tamamen yeni ortaya çıkan düzenleyici sistemlerdir (Davis ve ark., 2017).

Model organizmalar üzerinden yapılan evrimsel karşılaştırmalar, bir TF'nin bağlama bölgesindeki değişimin ne kadar işlevsel etkiler yarattığını gösterir. Örneğin, bazı motiflerin bağlanma kapasitesindeki küçük değişiklikler, genin ifade edilme zamanını ya da bağlanma gücünü önemli ölçüde etkileyebilir. Bu nedenle motiflerin evrimsel olarak korunmuşluğu, regülasyonun evrimsel sürekliliği hakkında bilgi sağlar (Spitz & Furlong, 2012).

Alvarez-Buylla ve arkadaşlarının (2007) çalışmasında, bitki sistemlerinde GRN'lerin işlevsel organizasyonu ve evrimsel dinamikleri detaylı olarak incelenmiştir. Bitki meristem gelişimi ve WUS/CLV ağı gibi örnekler, cis-regülatuar kontrollerin nasıl işlediğini ve TF-TG ilişkilerinin evrimsel olarak nasıl değişebileceğini göstermiştir.

Thompson & Levine'nin kantitatif modelleri, Drosophila'da enhancer organizasyonunun işlevsel analizini ve cis-regülatuar kontrolün dinamiklerini anlamada kritik bir rol oynamıştır. Bu modeller, GRN'lerin işlevsel ve evrimsel dinamiklerini analiz etmede yeni yaklaşımlar sunmuştur (Karlebach & Shamir, 2008). Drosophila melanogaster em-

briyonik gelişiminde dorsal–ventral eksen patterning'i, GRN'in evrimsel olarak ne kadar korunmuş olduğuna dair güçlü bir örnektir. Maternal sinyallerin etkisiyle ortaya çıkan Dorsal gradyanı, Twi, Sna ve Dpp gibi TF'lerin işlevsel düzenlenmesini sağlar. Bu süreç, Notch, Wnt ve EGF sinyal yollarıyla da bağlantılıdır (Levine & Tijan, 2003).

4. Uygulamalar

4.1. İnsan Hastalıklarında

Gen regülatör ağları ve transkripsiyon faktörleri, birçok hastalığın altında yatan moleküler mekanizmaları anlamada önemli roller üstlenir. Kanser araştırmalarında onkogenik TF füzyonları ve genetik risklerin analizinde GRN'ler sıklıkla kullanılır. Nörolojik hastalıklarda FoxP2 geni gibi TF'ler dil bozukluklarında anahtar rol oynarken, REST disfonksiyonu nörodejeneratif hastalıklara yol açar. İlaç geliştirme sürecinde de TF-TG ağlarından yola çıkarak hedef genlerin önceliklendirilmesi mümkün olmaktadır (Spitz & Furlong, 2012).

Kanser terapisi bağlamında transkripsiyon faktörleri doğrudan hedef olarak değerlendirilebilir. STAT3 ve NF-κB gibi sinyal iletimine duyarlı TF'ler, özellikle hematolojik malignitelerde ve solid tümörlerde anormal şekilde aktif bulunmuştur. Bu faktörlerin inhibitörleri preklinik çalışmalarda umut verici sonuçlar göstermiştir. Ancak TF'lerin doğrudan ilaçlanmasının zorluğu, bunların düzenlenmesi için dolaylı stratejilerin geliştirilmesini gerekli kılmaktadır (Darnell, 2002).

4.2. Bitki Biliminde

Bitkilerde stres yanıtları, gelişimsel süreçler ve tarımsal iyileştirme çalışmalarında GRN'ler ve TF'ler önemli uygulama alanları sunar. Kuraklık, tuzluluk ve diğer çevresel streslere yanıt veren TF'ler, bitkilerin direnç mekanizmalarının anlaşılması açısından değerlidir. Kök gelişimi, çiçeklenme ve diğer gelişimsel süreçlerde GRN dinamiklerinin incelenmesi, bitki biyolojisinde yeni keşiflere yol açmıştır. Tarım sektöründe TF mühendisliği ile verim artışı ve çevresel direncin artırılması hedeflenmektedir (Davis ve ark., 2017).

TF'lerin stres yanıtlarındaki rolleri, özellikle kuraklık ve tuz toleransı gibi özelliklerin artırılmasında değerlendirilmiştir. WRKY ve bZIP ailesi TF'ler, stres koşullarında aktive olan genleri düzenlerler. Bu TF'lerin işlevsel karakterizasyonu, transgenik bitki üretiminde doğrudan

kullanılmaktadır (Davis ve ark., 2017). Alvarez-Buylla ve arkadaşları (2007) tarafından yapılan çalışmada, bitki meristem gelişiminde WUS/CLV düzenleyici ağı ve GRN'lerin işlevsel organizasyonu ayrıntılı olarak açıklanmıştır. Bu çalışma, TF'lerin işlevsel katkısının evrimsel olarak nasıl değiştiğini ve bitki sistemlerinde GRN çıkarımında hangi yaklaşımların kullanıldığını göstermektedir.

4.3. İlaç Geliştirmede

Temel araştırma hedefleri ve rejeneratif tıp için hücresel kimliği yeniden yapılandırmak amacıyla, gen ifadesini seçici olarak artırma kapasitesi etkili bir yöntem sunmaktadır. Son bulgular, yapay indüksiyon için birden fazla faktöre ihtiyaç duyulmasına rağmen, bireysel transkripsiyon faktörleri olan nörogenin 2 (NGN2 veya NEUROG2) ve nörojenik farklılaşma faktörü 1'in (NEUROD1) ekzojen aktivitesinin, insan iPSC'lerinin indüklenmiş nöronlara (iNöronlar) farklılaşmasını kolaylaştırmak için yeterli olduğunu göstermektedir (Chavez ve ark., 2019).

STAT3 geleneksel olarak "ilaçla tedavi edilemez" olarak kabul edilmiştir. STAT3 inhibitörlerinde ileri teknolojilerin ve yenilikçi tedavi yaklaşımlarının devam eden ilerlemesi, "ilaçla tedavi edilebilir" bir statüye doğru ilerlemektedir. Şu anda etkili bir tedavi yöntemi bulunmamaktadır (Hu & Liu, 2024). IL-6, otokrin/parakrin bir şekilde endokrin etkiler ortaya çıkarmak için proinflamatuar veya bağışıklık hücreleriyle etkileşime girebilir ve hücre çoğalması, invazyon, hafifletme, hayatta kalma ve metastaz dahil olmak üzere doğal pro-tümörojenik etkiler gösterebilir. Ayrıca, STAT3'ün aktivasyonu yoluyla mikro çevreyi ve kanserle ilişkili inflamasyonu etkilemek için stromal hücreleri modüle ederek dışsal pro-tümörojenik etkiler gösterir (Hu & Liu, 2024).

NF-κB sinyallemesinin daha derin ve kapsamlı bir şekilde anlaşılması, bu yolu gen manipülasyonu, hücresel davranış modülasyonu ve terapötik müdahale için kullanma konusundaki güvencemizi artırır (Guo et al., 2024). Bu nedenle, çeşitli B hücresi malignitelerinde NF-κB›nin terapötik olarak hedeflenmesi için önemli gerekçeler vardır. Bruton tirozin kinazı (BTK), BCR sinyallemesinin ardından NF-κB aktivasyonu ve B hücresi sağkalımı için kritik öneme sahiptir ve bu da onu çeşitli hematolojik malignitelerde terapötik stratejiler için bir hedef haline getirir (Verzella ve ark., 2022).

5. Gelecek Nesil Yaklaşımlar

Yeni nesil teknolojiler ve analiz yöntemleri, GRN ve TF araştırmalarında büyük ilerlemeler sağlamaktadır. Single-cell düzeyinde GRN çıkarımı, hücre içi heterojenliğe dayalı daha hassas modellerin oluşturulmasını sağlar. Dinamik zaman serisi modelleri sayesinde gen ekspresyonundaki değişimler takip edilebilir hâle gelmiştir. Yapay zekâ destekli motif tanıma ve TF-TG tahmini, hesaplamalı biyolojinin ön saflarında yer almaktadır. Ayrıca açık erişilebilir ve kullanıcı dostu yazılımların geliştirilmesi, bu alanın daha geniş kitlelerce kullanılabilir olmasını sağlamaktadır (Marbach ve ark., 2012). Mesela He, F. ve arkadaşları tarafından geliştirilen Bayesian tabanlı yöntemler, mikrodizi verilerinden TF-TG ilişkilerinin çıkarılmasında oldukça başarılı olmuştur. Bu yöntemler, özellikle düşük boyutlu veri setlerinde bile güçlü tahminler sunmuştur (He & Hannon, 2004).

Single-cell RNA-seq verileri, her bir hücrenin benzersiz transkripsiyonel profilini tanımlamakta ve TF-TG ilişkilerinin hücre tipine özel analizlerini olanaklı kılmaktadır. Bu verilerle yapılan hesaplamalı yaklaşımlar, daha önce karışık hücre popülasyonlarıyla yapılan çalışmalardan daha isabetli sonuçlar vermekte ve hücresel heterojenliğin etkisini minimize etmektedir (Marbach ve ark., 2012).

Dinamik sistem modelleri, özellikle sistem biyolojisi alanında yeni yaklaşımlar sunmakla kalmaz, özellikle tek molekül düzeyindeki verilerle birlikte, gen ekspresyonunun zamansal düzenlenmesi açısından hayati öneme sahiptirler. Deterministik ve stokastik diferansiyel denklemlerle modellenen GRN'ler, gen ekspresyonundaki değişimleri takip edebilir hâle gelmiştir. rFBA ve STOCKS gibi yazılımlar, bu modellerin çıkarımında önemli rol oynamaktadır (Thompson ve ark., 2015).

6. Sonuç ve Öneriler

Özetle, transkripsiyon başlangıç faktörleri ve Mediator kompleksi, GRN'de doğrudan düzenleyici fonksiyonlara sahiptir. Bu faktörlerin genomdaki bağlanma bölgeleri ve işlevsel katkısı, özellikle gelişimsel süreçler ve kanser araştırmalarında değerlendirilmelidir. Dolayısıyla bu tür TF'lerin işlevsel karakterizasyonu, hastalık mekanizmalarının ve ilaç hedeflerinin çıkarımında hayati önem taşımaktadır (Carlberg & Molnár, 2016).

Gen regülatör ağları ve transkripsiyon faktörleri, sistemsal biyoloji ve kişiselleştirilmiş tıp alanında büyük potansiyele sahiptir. Ancak bu al-

anın daha da gelişebilmesi için büyük veri kaynaklarının standartlaştırılması, hesaplamalı stratejilerin iyileştirilmesi ve açık erişilebilir veri platformlarının geliştirilmesi gerekmektedir. Gelecekte yapılacak çalışmalar özellikle tek hücre düzeyinde dinamik ağların modellenmesi, yapay zekâ destekli motif analizleri ve çevresel faktörlerin etkilerinin dahil edilmesi yönünde yoğunlaşmalıdır (Davis ve ark., 2017).

7. Kaynakça

Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. Nature Methods, 14(11), 1083–1086. https://doi.org/10.1038/nmeth.4463

Alvarez-Buylla, E. R., Benitéz, M., Dávila, E. B., Chaos, A., Espinosa-Soto, C., & Padilla-Longoria, P. (2007). Gene regulatory network models for plant development. Current Opinion in Plant Biology, 10(1), 83-91. https://doi.org/10.1016/j.pbi.2006.11.008

Boeva, V. (2016). Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. Frontiers in Genetics, 7. https://doi.org/10.3389/fgene.2016.00024

Carlberg, C., & Molnár, F. (2016). Mechanisms of Gene Regulation. Springer Netherlands. https://doi.org/10.1007/978-94-017-7741-4

Chan, J. J., & Tay, Y. (2018). Noncoding RNA: RNA Regulatory Networks in Cancer. International Journal of Molecular Sciences, 19(5), E1310. https://doi.org/10.3390/ijms19051310

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B. E., Perrimon, N., Weiss, R., Aach, J., Collins, J. J., & Church, G. M. (2015). Highly efficient Cas9-mediated transcriptional programming. Nature Methods, 12(4), 326–328. https://doi.org/10.1038/nmeth.3312

Darnell, J. E. (2002). Transcription factors as targets for cancer therapy. Nature Reviews Cancer, 2(10), 740–749. https://doi.org/10.1038/nrc906

Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., Hilton, J. A., Jain, K., Baymuradov, U. K., Narayanan, A. K., Onate, K. C., Graham, K., Miyasato, S. R., Dreszer, T. R., Strattan, J., Jolanki, O., Tanaka, F. Y., & Cherry, J. (2017). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Research, 46(D1), D794–D801. https://doi.org/10.1093/nar/gkx1081

He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nature Reviews Genetics, 5(7), 522–531. https://doi.org/10.1038/nrg1379

Helmsauer, K., Valieva, M. E., Ali, S., Chamorro González, R., Schöpflin, R., Röefzaad, C., Bei, Y., Dorado Garcia, H., Rodriguez-Fos, E., Puiggròs, M., Kasack, K., Haase, K., Keskeny, C., Chen, C. Y., Kuschel, L. P., Euskirchen, P., Heinrich, V., Robson, M. I., Rosswog, C., & Toedling, J. (2020). Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-19452-y

Hu, Y., Dong, Z., & Liu, K. (2024). Unraveling the complexity of STAT3 in cancer: molecular understanding and drug discovery. Journal of Experimental

& Clinical Cancer Research, 43(1). https://doi.org/10.1186/s13046-024-02949-5

Guo, Q., Jin, Y., Chen, X., Ye, X., Shen, X., Lin, M., Zeng, C., Zhou, T., & Zhang, J. (2024). NF-κB in biology and targeted therapy: new insights and translational implications. Signal Transduction and Targeted Therapy, 9(1), 53. https://doi.org/10.1038/s41392-024-01757-9

Karlebach, G., & Shamir, R. (2008). Modelling and analysis of gene regulatory networks. Nature Reviews Molecular Cell Biology, 9(10), 770–780. https://doi.org/10.1038/nrm2503

Kim, D., Tran, A., Hani Jieun Kim, Lin, Y., Yee, J., & Yang, P. (2023). Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. Npj Systems Biology and Applications, 9(1). https://doi.org/10.1038/s41540-023-00312-6

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. Nature, 424(6945), 147–151. https://doi.org/10.1038/nature01763

Lis, J. T. (2019). A 50 year history of technologies that drove discovery in eukaryotic transcription regulation. Nature Structural & Molecular Biology, 26(9), 777–782. https://doi.org/10.1038/s41594-019-0288-9

Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Kellis, M., Collins, J. J., & Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. Nature Methods, 9(8), 796–804. https://doi.org/10.1038/nmeth.2016

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., & Califano, A. (2006). ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. BMC Bioinformatics, 7(Suppl 1), S7. https://doi.org/10.1186/1471-2105-7-s1-s7

Roeder, R. G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. Nature Structural & Molecular Biology, 26(9), 783–791. https://doi.org/10.1038/s41594-019-0287-x

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics, 13(9), 613–626. https://doi.org/10.1038/nrg3207

Thompson, D., Regev, A., & Roy, S. (2015). Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution. Annual Review of Cell and Developmental Biology, 31(1), 399–428. https://doi.org/10.1146/annurev-cellbio-100913-012908

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. Nature Reviews Genetics, 10(4), 252–263. https://doi.org/10.1038/nrg2538

Verzella, D., Cornice, J., Arboretto, P., Vecchiotti, D., Di Vito Nolfi, M., Capece, D., Zazzeroni, F., & Franzoso, G. (2022). The NF-κB Pharmacopeia: Novel Strategies to Subdue an Intractable Target. Biomedicines, 10(9),

2233. https://doi.org/10.3390/biomedicines10092233

Weirauch, Matthew T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., Najafabadi, Hamed S., Lambert, Samuel A., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J.-C., Galli, M., Lewsey, M. G., Huang, E., Mukherjee, T., Chen, X., & Reece-Hoyes, John S. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell, 158(6), 1431–1443. https://doi.org/10.1016/j.cell.2014.08.009

99

Chapter 3

DRUG-TARGET INTERACTION NETWORKS AS TOOLS FOR PERSONALIZED MEDICINE

Bengisu Çelik[1]

Kazım Yalçın Arga[2]

[1]Marmara University, Faculty of Engineering, Department of Bioengineering, İstanbul, Türkiye, ORCID: 0009-0001-4423-3637

[2]Marmara University, Faculty of Engineering, Department of Bioengineering, İstanbul, Türkiye, ORCID: 0000-0002-6036-1348

1. Introduction

0.1 Emergence of Personalized Medicine in Modern Biomedical Science

Personalized medicine (PM) is a method which focuses on comprehending the differences between patients with the same condition and represents a departure from the "one size fits all" model. According to this concept, relevant therapies should be chosen for specific patient groups. PM enables the prediction of whether a certain therapy will be helpful for a specific patient (Stefanicka-Wojtas & Kurpas, 2023). The recent development of high-throughput sequencing technology and computational biology has triggered a paradigm change from comprehensive treatment methods to personalized medicine. Personalized therapy is designed to personalize medical options, practices, and therapies to each patient's particular characteristics, such as their genetic profile, transcriptome data, proteomic patterns, and environmental exposures. This transition is particularly important in diseases with various etiologies, such as cancer, autoimmune disorders, and neurodegenerative diseases.

0.2 The Limitations of Conventional Drug Development Processes

Conventional drug development is recognized as time-consuming, expensive, and inefficient. A single new treatment takes an average of 10 to 15 years and more than $2.5 billion to create, and the success rate from clinical trial to market approval is believed to be less than 12% (DiMasi, Grabowski, & Hansen, 2016).

Target selection, lead compound screening, preclinical testing, and staged clinical trials make up this conventional pipeline, which is extremely linear and unsuitable to the complexity of human diseases. Furthermore, it typically applies to a "one-drug-fits-all" strategy that ignores genetic background, disease heterogeneity, and individual differences in drug metabolism (Collins & Varmus, 2015).

As a result, several drugs demonstrate less than ideal effectiveness or cause negative side effects in particular patient subgroups. These drawbacks show how urgently additional integrative approaches like networks of drug-target interactions and systems pharmacology are needed to maximize therapeutic development and promote customized therapy.

0.3 Overview of Drug-target Interaction Networks as tools for Personalized Therapy

Drug-target interaction (DTI) networks are a system-level framework for modeling the complex interactions between chemical substances and biological targets such proteins, enzymes, and nucleic acids. DTI networks may expose known and potential connections through the combination of various types of biological information, such as proteomics, cheminformatics, and genomes. All of this can assist identify drug resistance mechanisms, off-target effects, and prospects for drug repurposing (Yamanishi et al., 2008). These networks are especially useful for individualized therapy, where more precise drugs selection is required due to individual differences in genetic and molecular profiles. Individualized treatments may be developed by categorizing patients according to their predicted drugs response using network-based approaches (Barabási, Gulbahce, & Loscalzo, 2011). Furthermore, machine learning and deep learning algorithms have improved the prediction capacity of DTI models, allowing for more rapid identifying of novel interactions that conventional methods may ignore. As a result, DTI networks are a useful computational technique for advancing precision medicine while reducing the trial-and-error aspect of conventional medication.

2. Drug-Target Interaction (DTI) Networks: Conceptual Background

2.1. Definitions and Key Concepts

1.0.1. Target-based Therapy Strategies

In target-based therapy, drugs are designed or selected to target specific molecular targets related to a disease. This method gained popularity as molecular biology and genomics advanced, making it possible to identify genes and proteins associated with disease. Researchers intend to enhance treatment accuracy and minimize off-target effects by concentrating on well-defined targets (Imming, Sinning, & Meyer, 2006). In immunology and oncology, monoclonal antibodies and tyrosine kinase inhibitors are well-known instances of effective target-based treatments. However, this approach has drawbacks as well, such as the simplifying of intricate disease networks, particularly in multiple disorders, and the establishment of resistance mutations.

1.0.2 Polypharmacology and Multi-target Drugs

Polypharmacology is a concept that a molecule is designed to attach to two or more targets at the same time, resulting in a greater therapeutic impact than binding to a single target. It has many advantages compared to the conventional single-targeting molecules. A multi-targeting medicine is substantially more efficacious because of its cumulative efficacy at all its individual targets, making it much more effective in complex and multifactorial diseases such as cancer, where several proteins and pathways have been associated in disease origin and progression. To be polypharmacologic, a molecule must have promiscuity, which is the capacity to interact with numerous targets while avoiding binding to antitargets, which would result in off-target undesirable effects (Kabir & Muth, 2022).

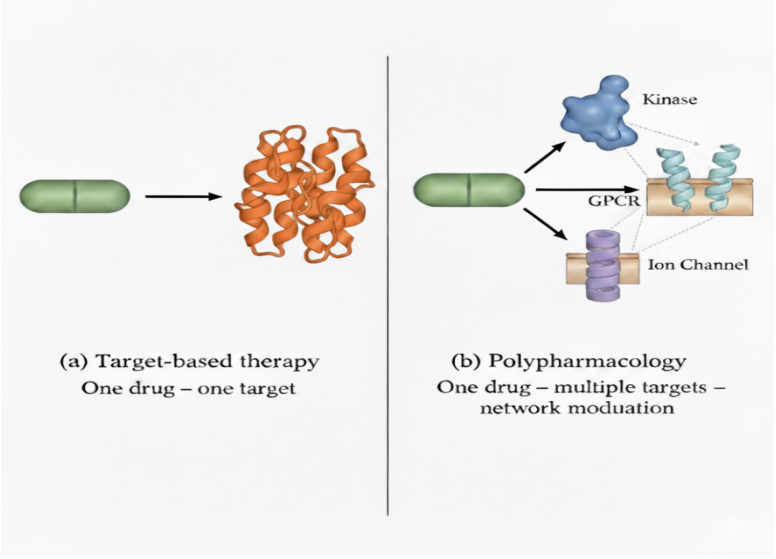


Figure 1: *Target-based Therapy vs Polypharmacology*

2.2. Network-Based Approaches

1.1.1 Biological networks

Biological networks are computational models depicting the interactions among many biomolecules, including genes, proteins, and metabolites, that guide cellular activities. By offering a systems-level perspective on biology, these networks help scientists comprehend the intricate relationships that underline processes that are both physiological and pathological. By encapsulating the dynamic and related character of molecular biology, biological networks contribute to recognizing of important

regulatory nodes, disease modules, and new therapeutic targets in the context of drug development as well as personalized therapy (Barabási & Oltvai, 2004). Signaling pathways and protein–protein interaction (PPI) networks are two important parts of biological networks associated with drug-target interactions.

1.1.1.1 **Protein-protein interactions (PPI)**

Protein–protein interactions (PPIs) are the functional and physical connections between proteins that are necessary for nearly every biological activity, such as metabolism, gene expression, and transmission of signals. PPIs are particularly important in drug development because disrupting or stabilizing specific interactions can affect disease phenotypes. The MDM2–p53 interaction in cancer biology is among the best-characterized cases demonstrating the therapeutic potential of modulating protein–protein interactions. By directly attaching to the tumor suppressor p53's N-terminal transactivation domain, MDM2 (Murine Double Minute 2) serves as a crucial negative regulator of the tumor suppressor under normal physiological conditions. Under unstressed situation, this association maintains p53 at low levels by promoting its ubiquitination and proteasomal degradation in addition to inhibiting its transcriptional activity. However, MDM2 is overexpressed in many cancers, especially those that maintain wild-type TP53. This causes pathological inactivation of p53, which permits uncontrolled cell survival and proliferation. Targeting the MDM2-p53 connection provides a key point of intervention for restoring the p53 function. Nutlins are a type of small-molecule inhibitors that competitively bind to the p53-binding pocket of MDM2 and mimic important p53 residues, as discovered in an important study by Vassilev et al. (2004). By disrupting the PPI interface, these compounds stabilize and activate p53 in vivo, which causes its target genes involved in cell cycle arrest and apoptosis (e.g. BAX, PUMA) to be transcriptionally upregulated. Targeting intracellular PPIs that had been believed to be "undruggable" is now clinically feasible thanks to the pharmacological reactivation of p53, which caused tumor regression in xenograft mice without causing significant toxicity to normal tissues (Vassilev et al., 2004).

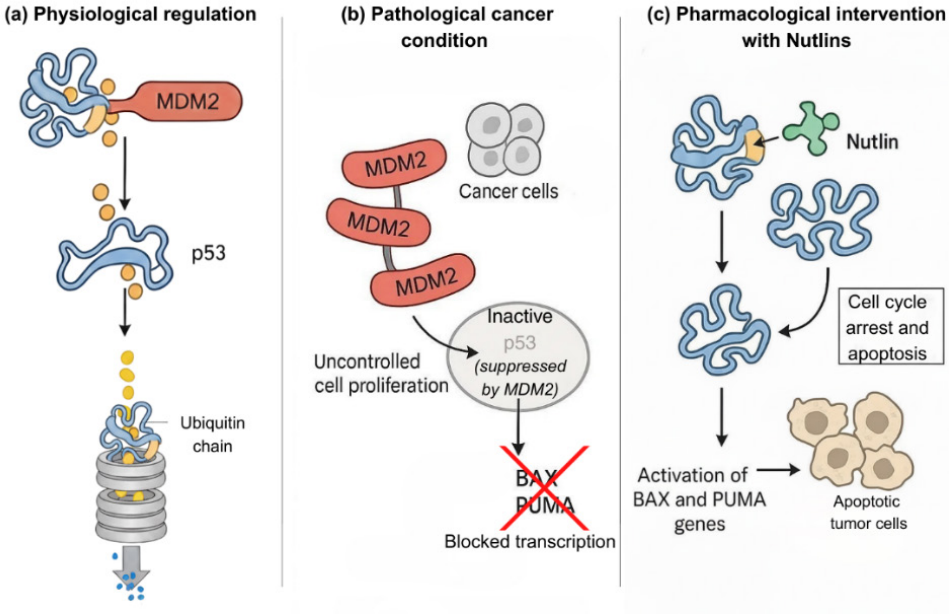


Figure 2: *Regulation of the MDM2-p53 interaction in physiological, pathological, and pharmacological conditions.* **(a) Physiological regulation.** *Under normal conditions, MDM2 binds to the N-terminal transactivation domain of p53 and promotes its poly-ubiquitination (yellow). Poly-ubiquitinated p53 has been identified and translocated into the proteasome (gray barrel-shaped complex) for proteolytic breakdown into tiny peptides. This process prevents the unnecessary activation of stress-response genes by maintaining low nuclear p53 levels.* **(b) Pathological cancer condition.** *In numerous cancers, MDM2 is overexpressed and significantly inhibits p53 function. Consequently, p53 remains inactive and does not initiate the transcription of pro-apoptotic genes like BAX and PUMA, resulting in inhibited apoptosis and unregulated cell growth.* **(c) Pharmacological intervention with Nutlins.** *Nutlins competitively bind the p53-binding pocket of MDM2 (yellow domain), displacing p53 and inhibiting its ubiquitination. Stabilized p53 accumulates and induces the transcription of BAX and PUMA, facilitating cell cycle arrest and apoptosis in cancerous cells.*

In systems pharmacology, PPI networks are formed by representing proteins as nodes and their interactions as edges, resulting in complex interaction maps. These networks enable the discovery of "hub" proteins—nodes with high levels of connectivity—which are frequently used as master regulators in biological systems. Hubs are usually required for cell survival and are connected to disease characteristics, making them appealing, if difficult, therapeutic targets (Jeong, Mason, & Barabási, 2001). Furthermore, PPI networks can identify "bottlenecks" and modular structures, such as protein complexes and pathways, to aid in the rational design of targeted therapeutics.

Integrating PPI data into drug-target interaction (DTI) networks improves the prediction capacity of computational models. Researchers can identify polypharmacological profiles, prioritize drug repositioning prospects, and detect indirect or secondary medication effects using algorithms such as random walk with restart (RWR), neighbourhood scoring, or network diffusion (Chen et al., 2016). Public databases like STRING, BioGRID, IntAct, and DIP offer high-confidence, curated PPI datasets allowing for the creation of physiologically relevant interaction landscapes (Szklarczyk et al., 2021). Such network-informed techniques have proven critical for understanding the systems-level impact of therapeutic interventions and customizing therapies to patients' particular biological circumstances.

PPIs have a considerable structural and functional variety, and their classification provides important details about their biological activities and druggability. PPIs can occur between identical or non-identical protein subunits, resulting in homo- or hetero-oligomeric complexes. Homo-oligomers can associate isologously—using identical contact surfaces with rotational symmetry—or heterologously, utilizing distinct interfaces that can promote unlimited polymerization unless restricted (Nooren & Thornton, 2003).

Additionally, PPIs are classified as obligatory or non-obligatory. Individual subunits of obligatory complexes are not stable in isolation and must fold or operate properly within the complex, such as the Arc repressor dimer, which is required for DNA binding. On the other hand, non-obligate interactions involve independently stable proteins that bind together quickly as seen in many signaling complexes, receptor-ligand interactions, and enzyme-inhibitor pairs. The duration of the interaction is another important aspect of classification.

While transient PPIs actively link and disassociate in response to environmental signals or chemical triggers, permanent PPIs form persistent, frequently physically obligatory assemblies. Weak, reversible linkages and more durable relationships that rely on nucleotide or post-translational modifications are examples of transient interactions. Importantly, these categories are not absolute; many PPIs exist on a continuum, and their characteristics may change depending on cellular circumstances or context. Accordingly, structural information, folding behavior, subcellular localization, and physiological significance are frequently considered for appropriate categorization (Nooren & Thornton, 2003). In drug development, it is essential to comprehend the variety of interaction types, especially when creating compounds that aim to stabilize or disrupt certain protein assemblies in disease-related pathways.

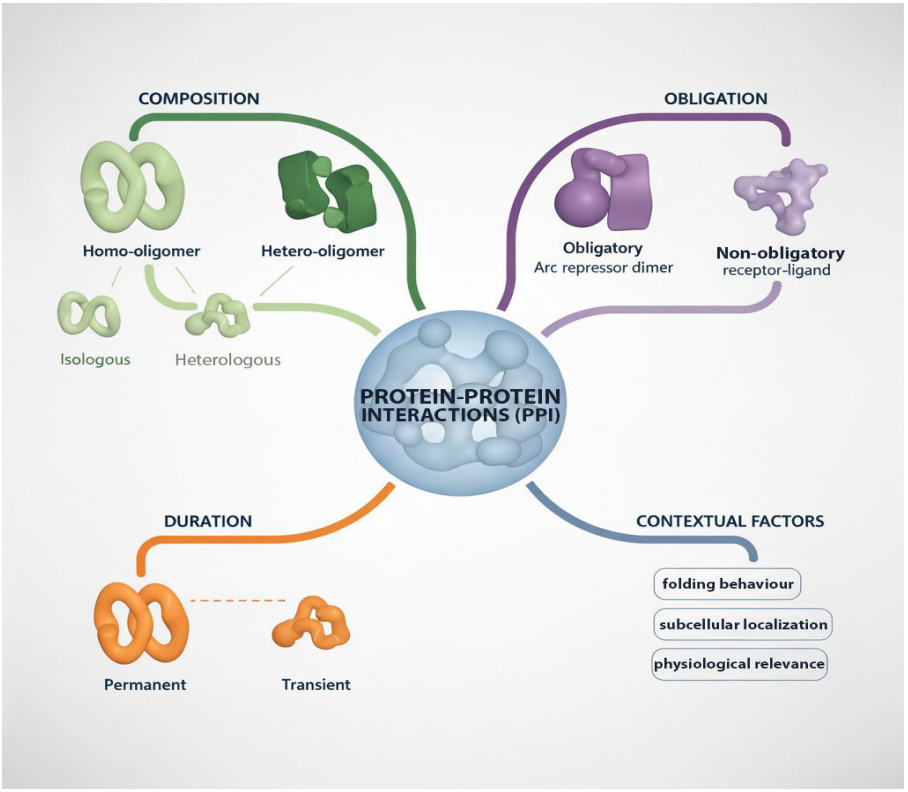


Figure 3: *Classification of Protein–Protein Interactions (PPIs).*

2.2.1.2. Signaling pathways

Signaling pathways are hierarchical networks of molecular interactions that coordinate cellular responses to both external and internal stimuli. These cascades are critical for maintaining cellular homeostasis and coordinating complicated biological activities such as differentiation, proliferation, apoptosis, and immune response. Numerous diseases, including cancer, autoimmune problems, and metabolic disorders, are characterized by dysregulation of signaling pathways, which can occur through mutations, overexpression, or post-translational changes. As a result, pathway components such as kinases, phosphatases, and transcription factors are among the most targeted compounds in current drug research (Hanahan & Weinberg, 2011).

Personalized medicine can identify druggable vulnerabilities and altered circuit architecture by incorporating patient-specific information, such as gene expression profiles, mutational landscapes and phosphoproteomics, into signaling network models. Computational techniques such

as Pathifier and SPIA (Signaling Pathway Impact Analysis) measure the extent of pathway perturbation in individual samples (Calzone et al., 2010). Drug-target interaction networks that include signaling pathway data provide a mechanistic basis for understanding drug activity beyond the key target. By mapping pharmacological effects onto signaling networks, researchers can assess downstream functional results, forecast side effects, and discover new indications for current medications. This integrated approach bridges the gap between molecular pharmacology and clinical practice, allowing for more accurate and effective medicines.

1.1.2. Key Network Terms: nodes, edges, centrality

Understanding the underlying components of biological networks is critical for comprehending drug-target interaction (DTI) networks and associated pharmacological models. These networks are represented mathematically as graphs G (V, E), where V is the collection of nodes (vertices) and E is the set of edges (links) that define their connections.

Nodes in biological contexts frequently represent molecular entities including proteins, genes, drugs, and metabolites. Edges represent the interactions or relationships between these components, which might include binding events, transcriptional regulation, enzymatic inhibition, or co-expression. Edges can be undirected physical protein-protein interactions) or directed (signaling pathways, where information flows from receptor to effector).

Several graph-based centrality metrics are commonly employed to quantify the functional relevance of individual nodes within a biological network:

Degree centrality measures the number of direct connections a node has. High-degree nodes, also known as hubs, are frequently associated with important proteins engaged in biological processes.

Betweenness centrality indicates how frequently a node is on the shortest path between other node pairs. High betweenness suggests that a node serves as an information bottleneck, making it an important control point in signaling or metabolic pathways.

Closeness centrality represents a node's average distance from all other nodes in the network, providing information on how quickly a node may disseminate signals or respond to perturbations.

1.1.3. Systems Biology and Network Pharmacology

Systems biology is an interdisciplinary study that studies the emergent features of biological systems by combining molecular components into a coherent framework of interactions. Unlike reductionist methods, which focus on individual molecules in isolation, systems biology stresses the dynamic and nonlinear interactions that underlie complex biological functions. This is accomplished by combining high-throughput omics data (transcriptomics, proteomics, and metabolomics) with computer modeling and mathematical simulation (Kitano, 2002).

Network pharmacology expands on this framework by bringing a systems-level approach to drug discovery. Traditional pharmacological models frequently assume a "one drug-one target-one disease" connection. However, accumulating evidence suggests that most drugs interact with multiple targets, and that diseases, particularly chronic or complex ones including cancer, neurodegeneration, and autoimmune disorders, result from perturbations in interconnected molecular networks rather than single molecular defects (Hopkins, 2007).

Network pharmacology is interested in organizing medicines, targets, and disease-associated molecules into a cohesive network to better understand how therapeutic interventions influence biological systems. This integrative technique offers several significant applications:

· Off-target effects and unfavorable medication responses can be predicted using the network's topology.

· Designing combination medicines involves discovering synergistic targets in convergent or parallel pathways.

· Drug repurposing is based on common molecular substructures or pathway involvement in several illnesses.

· Individual-specific genetic profiles are mapped into established biological networks to divide patients and generate personalized therapies.

Systems biology and network pharmacology demonstrate an advance toward mechanism-based, data-driven therapeutic design, moving away from symptomatic therapy and toward therapies that target the underlying cause of disease. These fields provide the conceptual and methodological foundation for personalized medicine by combining biological complexity with computational understanding.

2. Construction of Drug-Target Interaction Networks

The construction of drug-target interaction (DTI) networks is a critical step in systems pharmacology, allowing researchers to map, evaluate, and forecast the intricate interactions between medicinal drugs and their biological targets. These networks are constructed on curated biological data, omics integration, and computational prediction models, and serve as the foundation for target discovery, medication repurposing, and customized therapeutic techniques.

3.1. Data Sources

2.0.1 Databases: DrugBank, ChEMBL, STITCH, BindingDB

High-quality, well-curated datasets are critical for establishing strong DTI networks. DrugBank is one of the most comprehensive sites, including complete chemical, pharmacological, and molecular data on allowed and investigational medications, as well as their known protein targets (Wishart et al., 2018). ChEMBL provides additional bioactivity data, such as IC_{50} and EC_{50} values, derived from literature and high-throughput screening experiments (Gaulton et al., 2017). STITCH combines a wide range of chemical-protein connections by integrating experimental evidence with text mining and route data to provide both direct and indirect interaction information (Szklarczyk et al., 2016). Meanwhile, BindingDB focuses on empirically observed binding affinities, making it ideal for quantitative modeling of interaction strength (Gilson et al., 2016). These databases serve as the foundation for algorithms with supervised learning and facilitate validation in computational prediction pipelines.

2.0.2 Integration of Omics Data: Genomics, Transcriptomics, Proteomics

While static interaction data is necessary for scaffolding DTI networks, integrating omics data provides biological context and patient distinctiveness. Genomics can assist in finding mutations and polymorphisms that affect target structure or treatment responsiveness, especially in cancer and not common disorders. Transcriptomic data acquired from RNA-seq or microarrays show gene expression patterns across tissues or disease states, providing information about condition-specific treatment responsiveness. Proteomic profiles provide further resolution by identifying protein abundance, changes, and interaction dynamics. These multi-omics layers are critical for changing typical networks into context-aware and customized DTI models, which will help with precision medicine efforts (Hasin, Seldin, & Lusis, 2017).

3.2. Computational Prediction Methods

2.1.1 Machine Learning models for DTI Prediction

Traditional machine learning (ML) approaches are commonly used to estimate unknown DTIs from known interactions. Support vector machines (SVMs), random forests, and gradient boosting use features from chemical fingerprints, molecular descriptors, and sequence similarity to identify or predict drug-target combinations (Chen et al., 2016). Using matrix factorization and neighborhood-based models have also proven useful in capturing interaction patterns in large-scale datasets. While these models are often interpretable and computationally efficient, their reliance on feature engineering may restrict generalizability, particularly when dealing with new chemicals or protein targets.

2.1.2 Deep learning frameworks (DeepDTI, DeepPurpose)

Deep learning (DL) improves performance by learning to recognize features automatically from raw input data. This phenomenon has been demonstrated in several notable models.

· DeepDTI uses convolutional neural networks (CNNs) to learn from drug molecular graphs and protein sequences, capturing complicated interaction patterns (Wen et al., 2017).

· DeepPurpose is a versatile DL toolkit that can handle various encoding schemes, such as SMILES and FASTA, and perform regression and classification tasks (Hepp et al., 2021).

These models outperform classical machine learning (ML) methods, particularly in cold-start scenarios, and are being rapidly incorporated into virtual screening procedures.

2.1.3 Molecular Docking and Virtual Screening

For modeling drug–target interactions, structure-based techniques like molecular docking maintain to be essential tools beside data-driven approaches. Docking predicts binding poses and estimates affinity scores by simulating the physical binding process between a small chemical and a protein. Virtual screening pipelines frequently employ programs like AutoDock, Glide, and MOE to rank possible ligands against a target protein structure (Pagadala, Syed, & Tuszynski, 2017). The quality and availability of protein structures, as well as the limits of scoring algorithms,

frequently restrict docking's capacity to yield useful mechanistic insights. As a result, it is most effective when combined with ML/DL models, acting as a validation tool or a supplementary input source.

3. The Role of Drug-Target Interaction (DTI) Networks in Personalized Therapy

Drug-target interaction (DTI) networks provide a framework at the systems level for understanding and predicting the molecular causes of disease and the effectiveness of treatment. DTI networks have become effective tools in personalized therapy as the medical paradigm changes from population-wide treatment approaches to individualized management. In a variety of clinical situations, they facilitate patient-specific target discovery, drugs repurposing, and optimized treatment decision-making by integrating molecular data with network-based algorithms.

4.1. Patient-Specific Targeting

3.0.1 The Impact of Genetic Variation on Drug Response

Genetic variation can have a significant impact on interindividual heterogeneity in drugs response. Somatic mutations, structural variations, and single nucleotide polymorphisms (SNPs) can change the binding affinity of a drugs to its target, change the function of proteins, or impact drug metabolism and clearance (Rodriguez-Antona & Ingelman-Sundberg, 2006).

Polymorphisms in the cytochrome P450 enzyme family have major effects on drugs metabolism and patient response. For example, CYP2D6 genetic variations affect the metabolic rate of opioids and antidepressants, with poor metabolizers having a higher risk of toxicity and ultrarapid metabolizers having subtherapeutic effects (Haufroid & Hantson, 2015). Similarly, CYP2C19 phenotypes have been linked to antidepressant effectiveness and tolerability, with poor metabolizers having greater blood drug levels and less consistent response rates (Li et al., 2024). Additionally, polymorphisms like CYP3A4*1B have been associated with the altered metabolism of anticancer prodrugs such ifosfamide, which might contribute to poor treatment results (Torres et al., 2021). When combined with DTI networks, these pharmacogenomic data enable the discovery of genetically variable targets, improving the capacity to anticipate and customize medication effectiveness and toxicity depending on an individual's molecular makeup.

3.0.2 Drug Selection Based on Individual Molecular Profiles

In addition to genetic variations, patient-specific transcriptome and proteomic data helps in drugs selection using DTI network-based classification. High-throughput molecular profiling tools enable clinicians to identify the active biological pathways in an individual's disease state. Aligning these molecular features with drug-target information provided by DTI networks allows us to select drugs whose targets are not just expressed but also physiologically meaningful in that setting (Hodos et al., 2016). This method has had a particularly major influence oncology, where the discovery of tumor-specific signaling requirements has allowed the creation of targeted therapies such as EGFR inhibitors in lung cancer and PARP inhibitors in BRCA-mutated breast cancer.

3.1. Drug Repurposing and Network-Based Discovery

3.1.1. Discovery of New Therapeutic Indications via Network Analysis

Network pharmacology has contributed major contributions to drugs repurposing by focusing on biological systems' interconnectivity. Unlike traditional repurposing efforts, which frequently rely on spontaneous clinical findings or retrospective analysis, network-based solutions provide a methodical, hypothesis-driven framework. Drug-target interaction (DTI) networks allow researchers to find similar biological targets, signaling pathways, or disease modules that connect different diseases (Cheng, Kovács, & Barabási, 2019). For example, network proximity analysis has been used to propose that some anti-inflammatory medicines may alter biochemical pathways involved in neurodegenerative illnesses, indicating possible novel indications (Wang et al., 2024). Similarly, drugs like metformin, which were originally approved for type 2 diabetes, have shown anticancer activity due to interactions with energy metabolism regulators and the mTOR signaling pathway link discovered using integrative network modelling (Liu et al., 2024). The following examples demonstrate the potential of DTI networks to expand the therapeutic range of existing medications, whilst also reducing the time, cost, and failure rate associated with de novo drug discovery by leveraging established safety and pharmaceutical characteristics.

4.3. Clinical Applications and Case Studies

3.2.1 Use of DTI Networks in Oncology and Precision Medicine

The complexity of cancer biology, characterized by genetic variability and adaptive resistance, necessitates the utilization of precise treatment design techniques. The utilization of DTI networks in oncological contexts entails the integration of genetic alterations with medication response profiles, thereby facilitating the identification of novel and alternative targets (Wu, Wang, Chen, & Li, 2015). For instance, in cases where cancers exhibit resistance mutations in primary targets, network analysis has the potential to identify downstream or parallel targets for combination treatment. Furthermore, multi-target DTI networks have been utilized to create personalized medication prescriptions that simultaneously block multiple oncogenic drivers, thereby enhancing efficacy while minimizing the likelihood of relapse. These techniques are consistent with clinical initiatives such as The Cancer Genome Atlas (TCGA) and the NCI-MATCH study, the objective of which is to match patients to therapies based on molecular profiles.

3.2.2 Application of DTI Networks During the COVID-19 Pandemic

The COVID-19 pandemic underlined the critical importance of effective treatment discovery, as well as the relevance of drug-target interaction (DTI) networks in infectious disease response. Traditional medication development pipelines, which generally last more than a decade, were inadequate for the global health emergency posed by SARS-CoV-2. In this situation, DTI networks provided a quick, cost-effective, and system-level approach to identifying repurposable medicines with antiviral activity.

By integrating the known drug-target relationships with the SARS-CoV-2 host-virus protein-protein interaction (PPI) network, researchers were able to identify current medications capable of altering host pathways crucial for viral replication. In a significant development, Gordon et al. conducted an innovative study that identified the physical interactions between 26 SARS-CoV-2 proteins and 332 human proteins (Gordon et al., 2020). This finding has significant implications for the identification of potential treatment targets. Utilizing this network, medications that had been authorized for alternative applications were expeditiously evaluated for their proximity to viral host characteristics.

One of the most remarkable achievements of this method was the discovery of baricitinib, a Janus kinase (JAK) inhibitor initially used in rheumatoid arthritis, as a potential COVID-19 therapy. Baricitinib was predicted by network-based models to reduce viral entry by modifying JAK-STAT-mediated cytokine signaling and decreasing AAK1-mediated

endocytosis, which was later verified in experimental research and clinical application (Richardson et al., 2020). The FDA eventually authorized the medicine Emergency Use Authorization (EUA).

Furthermore, several multi-omics-based DTI models were used to integrate transcriptomic and proteomic data from COVID-19 patients with drug-target networks, allowing therapeutic candidates to be categorized by disease stage, tissue specificity, and immune response profiles (Barrera et al., 2020). These findings provided support to move away from direct antiviral mechanisms and toward host-directed treatment.

Collectively, these case studies demonstrate how DTI networks may dramatically speed up the discovery and validation of repurposed medications during public health crises. These networks represent invaluable instruments in the realm of pandemic preparedness and response strategies, as they facilitate the rapid identification of chemicals with relevance to mechanistic processes.

5. Challenges, Ethical Considerations, and Future Perspectives

4.1 Current Limitations and Challenges in the Field

Despite their promising potential, drug-target interaction (DTI) networks have various limitations that restrict their translational value. The incompleteness and bias of existing data pose considerable challenges. Many interaction databases (e.g., DrugBank, STITCH) rely on curated or expected interactions that may lack experimental validation and are frequently biased toward well-studied proteins and pathways, restricting the discovery of uncommon illnesses and new targets. Furthermore, static network models cannot always account for changing biological circumstances such tissue-specific expression, post-translational modifications, and disease states. Another problem is the interpretability of machine learning and deep learning models utilized in DTI prediction. While algorithms like DeepPurpose and GraphDTA are highly predictive, they frequently behave as "black boxes" making it impossible to explain or validate the biological plausibility of their predictions (Zitnik, Agrawal, & Leskovec, 2018). Furthermore, the absence of set standards and assessment criteria makes it difficult to replicate findings across investigations.

4.2 Ethical and Societal Considerations

Integration of patient-specific data, such as genetic variations and

transcriptome profiles, into DTI networks brings ethical and privacy problems. Data privacy, informed permission, and safe data storage are critical, particularly in clinical contexts where genetic data is utilized to provide individualized therapeutic recommendations. There is also concern about algorithmic bias, especially if training datasets are not diverse in terms of race, gender, or simultaneous diseases, which might lead to unequal treatment recommendations (Obermeyer, Powers, Vogeli, & Mullainathan, 2019).

Access to network-driven therapeutics may increase healthcare inequities if such technologies are predominantly offered in high-resource areas. As DTI networks become more integrated into clinical decision-making, open regulatory frameworks and public involvement will be critical in maintaining trust and equality.

4.3 Future Directions and Emerging Opportunities

The future of DTI network research relies on multimodal integration and real-time flexibility. Combining proteomics, metabolomics, and single-cell data with standard genomic information will result in more complex, context-aware networks. Advances in graph neural networks (GNNs) and explainable AI (XAI) present opportunities for interpretable and dynamic modelling of drug-target systems in specific diseases situations.

Furthermore, digital twins—computational duplicates of patients—could use DTI networks to simulate drug reactions in silico, significantly lowering the requirement for trial and error in clinical therapy (Topol, 2019). Regulatory acceptance of network-based biomarkers and therapies, as well as multinational partnerships and data-sharing initiatives, will be critical in advancing the area to clinical maturity.

6. Conclusion

Drug-target interaction (DTI) networks represent an important change in medical studies by providing a systems-level framework for understanding complex molecular interactions and optimizing treatment techniques. Their integration with multi-omics datasets and computational tools, including machine learning and deep learning, has greatly enhanced the prediction of novel drug-target relationships, improved drug repurposing efforts, and enabled patient-specific therapy design. These networks have proved particularly useful in areas requiring rapid

responses, such as cancer and infectious disease epidemics like COVID-19.

However, this field is still facing significant obstacles, such as data incompleteness, algorithmic bias, poor interpretability of deep learning models, and ethical considerations over patient privacy. Addressing these limits through improved data curation, explainable AI models, and accessible regulatory norms will be critical to widespread clinical usage. Looking ahead, the combination of DTI networks, digital twin technologies, graph neural networks, and real-time patient data offers enormous potential for taking customized medicine from idea to practical reality. Thus, DTI networks are not only improving our understanding of drug action, but also transforming how we create, assess, and administer medicines in the age of precision medicine.

References

Barrera, A., Ramos, H., Vera-Otarola, J., Fernández-García, L., Angulo, J., Olguín, V., Pino, K., Mouland, A. J., & López-Lastra, M. (2020). Post-translational modifications of hnRNP A1 differentially modulate retroviral IRES-mediated translation initiation. Nucleic acids research, 48(18), 10479–10499. https://doi.org/10.1093/nar/gkaa765

Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nature reviews. Genetics, 12(1), 56–68. https://doi.org/10.1038/nrg2918

Barabási, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. Nature reviews. Genetics, 5(2), 101–113. https://doi.org/10.1038/nrg1272

Calzone, L., Tournier, L., Fourquet, S., Thieffry, D., Zhivotovsky, B., Barillot, E., & Zinovyev, A. (2010). Mathematical modelling of cell-fate decision in response to death receptor engagement. PLoS computational biology, 6(3), e1000702. https://doi.org/10.1371/journal.pcbi.1000702

Chen, X., Yan, C. C., Zhang, X., Zhang, X., Dai, F., Yin, J., & Zhang, Y. (2016). Drug-target interaction prediction: databases, web servers and computational models. Briefings in bioinformatics, 17(4), 696–712. https://doi.org/10.1093/bib/bbv066

Cheng, F., Kovács, I. A., & Barabási, A. L. (2019). Network-based prediction of drug combinations. Nature communications, 10(1), 1197. https://doi.org/10.1038/s41467-019-09186-x

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. The New England journal of medicine, 372(9), 793–795. https://doi.org/10.1056/NEJMp1500523

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of health economics, 47, 20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

Espindola, L. M. T., López, M. J. C., Flores, A. U., Espinosa, L. R., Granados, J., Pacheco, J. L. C., García, M. P., Cervantes, M. T. R., Méndez, V. C. B., Doño, S. H., & Ruíz Gómez, D. (2019). Genetic polymorphism of CYP3A4 is associated with poor response to ifosfamide treatment in children with solid embryonic tumors. Archives of medical science: AMS, 17(6), 1766–1771. https://doi.org/10.5114/aoms.2019.86648

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magariños, M. P., Overington, J. P., Papadatos, G., Smit, I., & Leach, A. R. (2017). The ChEMBL database in 2017. Nucleic acids research, 45(D1), D945–D954. https://doi.org/10.1093/nar/gkw1074

Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., & Chong, J. (2016). BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic acids research, 44(D1), D1045–D1053. https://doi.org/10.1093/nar/gkv1072

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Hüttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., Kim, M., … Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature, 583(7816), 459–468. https://doi.org/10.1038/s41586-020-2286-9

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. Cell, 144(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. Genome biology, 18(1), 83. https://doi.org/10.1186/s13059-017-1215-1

Haufroid, V., & Hantson, P. (2015). CYP2D6 genetic polymorphisms and their relevance for poisoning due to amfetamines, opioid analgesics and antidepressants. Clinical toxicology (Philadelphia, Pa.), 53(6), 501–510. https://doi.org/10.3109/15563650.2015.1049355

Hepp, B., Da Cunha, V., Lorieux, F., & Oberto, J. (2021). BAGET 2.0: an updated web tool for the effortless retrieval of prokaryotic gene context and sequence. Bioinformatics (Oxford, England), 37(17), 2750–2752. https://doi.org/10.1093/bioinformatics/btab082

Hodos, R. A., Kidd, B. A., Shameer, K., Readhead, B. P., & Dudley, J. T. (2016). In silico methods for drug repurposing and pharmacology. Wiley interdisciplinary reviews. Systems biology and medicine, 8(3), 186–210. https://doi.org/10.1002/wsbm.1337

Hopkins, A. L. (2007). Network pharmacology. Nature biotechnology, 25(10), 1110–1111. https://doi.org/10.1038/nbt1007-1110

Imming, P., Sinning, C., & Meyer, A. (2006). Drugs, their targets and the nature and number of drug targets. Nature reviews. Drug discovery, 5(10), 821–834. https://doi.org/10.1038/nrd2132

Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. Nature, 411(6833), 41–42. https://doi.org/10.1038/35075138

Kabir, A., & Muth, A. (2022). Polypharmacology: The science of multi-targeting molecules. Pharmacological research, 176, 106055. https://doi.org/10.1016/j.phrs.2021.106055

Kitano, H. (2002). Systems biology: a brief overview. Science (New York, N.Y.), 295(5560), 1662–1664. https://doi.org/10.1126/science.1069492

Li, D., Pain, O., Fabbri, C., Wong, W. L. E., Lo, C. W. H., Ripke, S., Cattaneo, A., Souery, D., Dernovsek, M. Z., Henigsberg, N., Hauser, J., Lewis, G., Mors, O., Perroud, N., Rietschel, M., Uher, R., Maier, W., Baune, B. T., Biernacka, J. M., Bondolfi, G., … Lewis, C. M. (2024). Metabolic activity of CYP2C19 and CYP2D6 on antidepressant response from 13 clinical studies using genotype imputation: a meta-analysis. Translational psychiatry, 14(1), 296. https://doi.org/10.1038/s41398-024-02981-1

Liu, S., Xu, M., Yang, Z., Li, Y., Wu, D., & Tang, X. (2024). Network pharmacology-based investigation and experimental validation of the

mechanism of metformin in the treatment of acute myeloid leukemia. European journal of medical research, 29(1), 475. https://doi.org/10.1186/s40001-024-02022-z

Nooren, I. M., & Thornton, J. M. (2003). Diversity of protein-protein interactions. The EMBO journal, 22(14), 3486–3492. https://doi.org/10.1093/emboj/cdg359

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science (New York, N.Y.), 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. Biophysical reviews, 9(2), 91–102. https://doi.org/10.1007/s12551-016-0247-1

Richardson, P., Griffin, I., Tucker, C., Smith, D., Oechsle, O., Phelan, A., Rawling, M., Savory, E., & Stebbing, J. (2020). Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet (London, England), 395(10223), e30–e31. https://doi.org/10.1016/S0140-6736(20)30304-4

Rodriguez-Antona, C., & Ingelman-Sundberg, M. (2006). Cytochrome P450 pharmacogenetics and cancer. Oncogene, 25(11), 1679–1691. https://doi.org/10.1038/sj.onc.1209377

Stefanicka-Wojtas, D., & Kurpas, D. (2023). Personalised Medicine-Implementation to the Healthcare System in Europe (Focus Group Discussions). Journal of personalized medicine, 13(3), 380. https://doi.org/10.3390/jpm13030380

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research, 49(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074

Szklarczyk, D., Santos, A., von Mering, C., Jensen, L. J., Bork, P., & Kuhn, M. (2016). STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic acids research, 44(D1), D380–D384. https://doi.org/10.1093/nar/gkv1277

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., & Liu, E. A. (2004). In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science (New York, N.Y.), 303(5659), 844–848. https://doi.org/10.1126/science.1092472

Wang, X., Fu, X., Luo, X., Lai, Y., Cai, C., Liao, Y., Dai, Z., Fang, S., & Fang, J. (2023). Network Proximity Analysis Deciphers the Pharmacological Mechanism of Osthole against D-Galactose Induced Cognitive Disorder

in Rats. Molecules (Basel, Switzerland), 29(1), 21. https://doi.org/10.3390/molecules29010021

Wen, M., Zhang, Z., Niu, S., Sha, H., Yang, R., Yun, Y., & Lu, H. (2017). Deep-Learning-Based Drug-Target Interaction Prediction. Journal of proteome research, 16(4), 1401–1409. https://doi.org/10.1021/acs.jproteome.6b00618

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., … Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Research, 46. https://doi.org/10.1093/nar/gkx1037

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics (Oxford, England), 24(13), i232–i240. https://doi.org/10.1093/bioinformatics/btn162

Yella, J. K., Yaddanapudi, S., Wang, Y., & Jegga, A. G. (2018). Changing Trends in Computational Drug Repositioning. Pharmaceuticals (Basel, Switzerland), 11(2), 57. https://doi.org/10.3390/ph11020057

Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics (Oxford, England), 34(13), i457–i466. https://doi.org/10.1093/bioinformatics/bty294

99

Chapter 4

FROM INTERACTIONS TO INDICATIONS: A NETWORK-BASED PERSPECTIVE ON DRUG REPOSITIONING

Sena ÖZYİĞİT[1], Prof. Kazım Yalçın ARGA[2], Assoc. Prof. Beste TURANLI[3]

1 Biruni University, Faculty of Engineering and Natural Sciences, Department of Biomedical Engineering, Istanbul, Türkiye

https://orcid.org/0000-0003-0973-0758
2 Marmara University, Faculty of Engineering, Department of Bioengineering, Istanbul, Türkiye

https://orcid.org/0000-0002-6036-1348
3 Marmara University, Faculty of Engineering, Department of Bioengineering, Istanbul, Türkiye

https://orcid.org/0000-0003-1330-9712

INTRODUCTION

The conventional drug discovery process requires a lot of time and resources. It typically requires 10–15 years of development and trial periods and costs exceeding $1 billion USD to bring a new therapeutic agent to market (DiMasi et al., 2003). Moreover, failures in late-stage clinical trials caused by insufficient efficacy or toxicity problems worsen the inefficiencies of this traditional approach. As a result, alternative strategies like drug repositioning, also known as drug repurposing, therapeutic repurposing, drug re-profiling, drug re-tasking, or drug recycling, have emerged as viable solutions.

Drug repositioning seeks to delineate novel therapeutic applications for previously approved or discontinued pharmacologic agents, capitalizing on their well-documented safety and pharmacokinetic properties. This strategy simultaneously shortens drug development timelines while increasing both clinical success rates and financial returns (Ashburn & Thor, 2004). Numerous successful cases demonstrate the translational impact of this approach. However, traditional drug repositioning efforts often relied on serendipitous discoveries or isolated clinical observations. The complexity of human diseases, driven by multiple factors and interactions at the molecular and cellular levels, demands a more systematic approach. This is where network biology enables a fundamental shift in perspective. Network biology frameworks represent biological systems as interconnected graphs comprising genes, proteins, drugs, and diseases, offering a unified analytical perspective to: investigate pharmacological mechanisms, identify novel therapeutic applications, and predict off-target effects (Barabasi et al., 2011).

This review outlines the foundational concepts of network biology and their implementation in drug repositioning strategies. We critically evaluate principal computational techniques, characterize available analytical resources and repositories, and discuss persistent challenges alongside emerging research trajectories in this field.

1 FUNDAMENTALS OF NETWORK BIOLOGY

Biological networks represent complex biological systems as graphs, with nodes corresponding to proteins, genes, and drugs, while edges capture their physical interactions, regulatory relationships, or functional associations. Network topologies exist as directed/undirected graphs with weighted/binary edges, featuring either homogeneous (single-type) or heterogeneous (multi-type) node configurations.

Key network properties include degree, betweenness, modularity and community structure, closeness, and shortest path length metrics. In drug repositioning studies, critical network metrics include betweenness centrality (identifying bottleneck proteins crucial for disease propagation (Barabasi et al., 2011), closeness centrality (highlighting efficient information spreaders as multi-target drug candidates (Csermely et al., 2013), node degree (revealing highly connected targets with potential side-effect risks (Jeong et al., 2001), shortest path lengths (quantifying drug-disease proximity for therapeutic potential (Yildirim et al., 2007), modular structure (exposing functional target clusters for combination therapies (Menche et al., 2015), and edge-betweenness (pinpointing key pathway interactions to disrupt (Morselli Gysi et al., 2021) - collectively enabling data-driven drug prioritization (Hodos et al., 2016).

1.1 Types of Biological Networks

Protein-protein interaction networks (PPINs) systematically map physical and functional relationships between proteins, serving as powerful frameworks for both drug target identification and elucidation of disease pathways. Databases like STRING and BioGRID provide curated and experimental data for constructing these networks (Oughtred et al., 2019; Szklarczyk et al., 2021).

Gene regulatory networks (GRNs) model transcriptional and post-transcriptional regulatory relationships between genes, typically reconstructed from gene expression profiles using computational inference methods (Marbach et al., 2012; Stuart et al., 2003). Algorithmic tools including ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) and GENIE3 (Gene Network Inference with Ensemble of Trees) represent state-of-the-art methods for GRN inference, leveraging mutual information and machine learning approaches respectively (Huynh-Thu et al., 2010; Margolin et al., 2006).

Metabolic networks reconstruct biochemical reaction pathways through stoichiometric modeling of enzyme-substrate relationships (Orth et al., 2010), while signaling networks map information flow via protein phosphorylation cascades and second messenger dynamics (Kholodenko, 2006) – together providing systems-level insights into cellular regulation. The KEGG (Kyoto Encyclopedia of Genes and Genomes) and Reactome databases offer curated, hierarchical pathway representations that enable systematic modeling of biological interactions, from metabolic fluxes to signal transduction cascades (Kanehisa et al., 2021; Milacic et al., 2024).

Drug-target interaction (DTI) networks employ bipartite graph architectures to systematically connect pharmacological compounds with their protein targets, integrating both experimentally validated and computationally predicted binding data (Yamanishi et al., 2008). These networks rely on curated knowledge bases such as DrugBank (comprehensive drug-target annotations) and ChEMBL (bioactivity data from medicinal chemistry literature) as primary data sources (Gaulton et al., 2017; Wishart et al., 2018).

Disease-gene association networks systematically map relationships between pathological phenotypes and their underlying genetic variants, enabling the identification of disease-specific functional modules, and shared etiological mechanisms across clinically distinct disorders (Ghiassian et al., 2015; Goh et al., 2007; Menche et al., 2015).

Drug-disease networks structurally represent established and computationally inferred therapeutic relationships, facilitating drug repositioning through two primary analytical paradigms: (1) topological proximity measures between drug and disease modules, and (2) multi-scale similarity assessments of network signatures (Cheng et al., 2018; Guney et al., 2016).

1.2 Data Sources for Network Construction

The construction of biologically and pharmacologically relevant networks is enabled by multiple tiers of data resources: (1) manually curated knowledge bases (e.g., KEGG, Reactome), (2) experimentally derived interaction datasets (e.g., STRING, BioGRID), and (3) computationally predicted repositories (e.g., STITCH, DrugComb), each providing complementary levels of evidence (Kanehisa et al., 2021; Milacic et al., 2024; Oughtred et al., 2019; Szklarczyk et al., 2021; Szklarczyk et al., 2016; Zagidullin et al., 2019). For example, DrugBank integrates drug chemistry, pharmacology, and drug–target interaction data (Wishart et al., 2018) and STRING provides high-confidence protein–protein interactions aggregated from multiple sources, including computational and experimental predictions (Szklarczyk et al., 2019; Szklarczyk et al., 2021).

The Comparative Toxicogenomics Database (CTD) provides manually curated, evidence-based associations spanning chemical-gene interactions, chemical-disease relationships, and gene-disease linkages, systematically integrating toxicological and pharmacological data from published literature (Davis et al., 2023). KEGG and Reactome are authoritative sources for metabolic and signaling pathways (Kanehisa et al., 2021;

Milacic et al., 2024). The ChEMBL database systematically archives experimentally derived bioactivity measurements for structurally diverse small molecules, encompassing quantitative data from high-throughput screens, binding assays, and functional studies across multiple target classes (Gaulton et al., 2017; Mendez et al., 2019). The LINCS (Library of Integrated Network-Based Cellular Signatures) and Connectivity Map (CMap) projects generate comprehensive, standardized transcriptomic profiles characterizing dose- and time-dependent cellular responses to pharmacological perturbations, enabling pattern-based drug discovery through signature matching (Koleti et al., 2018; Lamb et al., 2006; Subramanian et al., 2017). DisGeNET systematically aggregates and standardizes gene-disease relationships through rigorous integration of heterogeneous data sources (Pinero et al., 2017; Pinero et al., 2020).

2 COMPUTATIONAL METHODS FOR NETWORK-BASED REPOSITIONING

Network-based drug repositioning employs similarity-based methods, network propagation, graph-based machine learning, matrix factorization, network clustering analysis, expression signature matching, and integrative multi-method frameworks to systematically predict novel therapeutic applications through computational analysis of biological networks (Himmelstein et al., 2017; Luo et al., 2017; Vanunu et al., 2010; Wang et al., 2013; Wu et al., 2013; Zitnik et al., 2018). Methods, their principle and example tools/algorithms are summarized in Table 2.

Table 2. Overview of Computational Methods in Network-Based Drug Repositioning

Method Category	Principle	Example Tools/ Algorithms	Typical Use Cases	References
Similarity-Based Methods	Compares drug-drug or disease-disease profiles using network topology or biological features	DrugNet, NBI	Identifying structurally similar drugs for repurposing	(Chen et al., 2016; Martinez et al., 2015)

Network Propagation	Propagates signals (e.g., disease genes) through networks via diffusion or random walks	Random Walk with Restart (RWR), HotNet, PRINCE	Prioritizing candidate drugs close to disease modules	(Valdeolivas et al., 2019; Vandin et al., 2011; Vanunu et al., 2010)
Graph-Based Machine Learning	Learns patterns from graph-structured data	deepDR, Graph Convolutional Neural Network (GCN), GraphSAGE	Predicting drug-disease edges in hetero-geneous networks	(Hamilton et al., 2017; Li et al., 2018; Zeng et al., 2019)
Matrix Factorization	Decomposes drug-target-disease matrices into latent features for prediction	Nonnegative Matrix Factorization (NMF), DTINet, Graph Regularized Matrix Factorization (GRMF)	Inferring unknown drug-target interactions	(Cui et al., 2019; Lee & Seung, 1999; Luo et al., 2017)
Network Clustering	Detects densely connected modules (communities) of drugs/targets/diseases	ClusterONE, Markov Cluster Algorithm (MCL), Infomap	Discovering polypharma-cology or shared-pathway drugs	(Dongen, 2000; Nepusz et al., 2012; Rosvall & Bergstrom, 2008)
Expression Signature Matching	Matches drug-induced gene expression changes to disease profiles	CMap, L1000CDS²	Repurposing based on transcripto-mic reversal	(Duan et al., 2016; Lamb et al., 2006)
Integrative Approaches	Combines multiple methods/data types (e.g., networks + omics + clinical data)	deepDR, Hetionet, PREDICT	Improving prediction robustness via multi-evidence integration	(Gottlieb et al., 2011; Himmelstein et al., 2017; Zeng et al., 2019)

2.1 Similarity-Based Methods

These methods hypothesize that similar drugs treat similar diseases.

Similarity can be defined in terms of chemical structure, gene expression profiles, side effects or therapeutic indications. For instance, Gottlieb et al. developed PREDICT, which integrates multiple similarity scores into a machine learning framework for drug–disease association prediction (Gottlieb et al., 2011).

2.2 Network Propagation Algorithms

Network propagation methods simulate the diffusion of information through biological networks. Random Walk with Restart (RWR) is a widely used technique that ranks nodes based on their network proximity to disease-related entities (Li & Patra, 2010). Advanced network propagation algorithms, including the Neighborhood Regularized Weighted Random Walk with Restart (NRWRH) (Luo et al., 2017) framework and diffusion kernel-based approaches (Taheri-Ledari et al., 2022), enable robust inference in heterogeneous biological networks by simultaneously modeling multiple interaction types (e.g., drug-target, disease-gene, and protein-protein interactions) through:

1. Edge-type-specific diffusion kernels that weight distinct biological relationships,

2. Topological regularization to preserve network structure during inference, and

3. Multi-layer network integration for cross-modal relationship prediction (Zitnik et al., 2018).

These methods are particularly powerful in identifying drugs that are not directly linked to a disease module but are in close topological proximity.

2.3 Graph-Based Machine Learning and Deep Learning

Recent breakthroughs in deep learning have revolutionized the analysis of biological networks through graph-structured representation learning. Graph Neural Networks (GNNs) now enable unified integration of topological connectivity patterns (e.g., protein-protein interactions) with node/edge attributes (e.g., drug chemical features, disease genomics), significantly enhancing drug-disease association predictions (Zhou et al., 2020). Notable implementations include:

1. Decagon (Zitnik et al., 2018): A multimodal graph convolution-

al framework that models polypharmacy side effects by learning from 1,048,576 drug-protein-disease triplets across 964 biological networks, achieving 0.91 AUROC in side-effect prediction.

2. SAveRUNNER (Fiscon et al., 2021): A hybrid approach combining network proximity metrics with machine learning, which identified 83% of clinically investigated COVID-19 repurposing candidates (including dexamethasone and baricitinib) by analyzing 17,000 drug-disease pairs in the human interactome.

These advances demonstrate how GNN architectures can capture both local network neighborhoods and global system properties for therapeutic discovery (Sanchez-Lengeling et al., 2021). These methods are further strengthened by multi-omic integration, allowing for the inclusion of transcriptomic, proteomic, and epigenetic information.

3 TOOLS AND DATABASES SUPPORTING REPOSITIONING

Cytoscape is a widely adopted open-source platform for visualizing and analyzing biological networks. Its extensible plugin architecture supports specialized analyses, including functional annotation through tools like ClueGO (gene-set enrichment) and BiNGO (Gene Ontology term visualization), enabling mechanistic interpretation of network modules (Bindea et al., 2009; Maere et al., 2005; Shannon et al., 2003).

DT-Web provides interactive visualization and prediction capabilities for drug–target and drug–disease networks (Alaimo et al., 2015). The platform allows real-time exploration of inferred therapeutic relationships, with user-adjustable parameters to prioritize candidates based on network proximity.

Transcriptomic databases—including GEO (gene expression omnibus), LINCS (perturbation signatures), and CMap (connectivity mapping)—offer large-scale gene expression datasets suitable for expression-based drug repositioning (Edgar et al., 2002; Koleti et al., 2018; Subramanian et al., 2017). These resources enable pattern-matching strategies to identify drugs that reverse disease-associated transcriptomic signatures.

PharmGKB curates pharmacogenomic knowledge, linking genetic variants to drug response variability. It serves as a critical resource for precision medicine by annotating clinically actionable variants in drug metabolism pathways (Hewett et al., 2002).

SIDER catalogs adverse drug reactions (ADRs) from clinical trials and post-marketing surveillance. ADR profiles facilitate side-effect-based drug similarity assessments, which can reveal repurposing opportunities (Kuhn et al., 2016).

DisGeNET and OMIM are authoritative resources for disease–gene annotation. DisGeNET integrates evidence from multiple sources (e.g., GWAS, animal models), while OMIM focuses on Mendelian disorders, together supporting genetic target identification (Hamosh et al., 2005; Pinero et al., 2020).

4 CASE STUDIES AND APPLICATIONS

Network-based drug repositioning has demonstrated clinical validation through multiple high-impact successes. The vasodilatory effects of sildenafil (originally an angina therapeutic) were mechanistically linked to phosphodiesterase-5 (PDE5) inhibition in erectile tissue and pulmonary vasculature through protein interaction network analysis, leading to its FDA approval for erectile dysfunction (1998) and pulmonary arterial hypertension (2005) (Galie et al., 2005; Goldstein et al., 1998). The immunomodulatory action of thalidomide—withdrawn in 1961 due to teratogenicity—was rediscovered via network analysis of its binding to cereblon (CRBN) and subsequent modulation of inflammatory cytokines. This underpinned its repurposing for multiple myeloma and leprosy (D'Amato et al., 1994; Ito et al., 2010).

During COVID-19, network tools like SAveRUNNER prioritized baricitinib (a JAK1/2 inhibitor) by identifying its overlap with SARS-CoV-2 host-entry pathways (e.g., ACE2 endocytosis regulation). Clinical trials confirmed its efficacy, resulting in emergency use authorization (EUA) (Fiscon et al., 2021; Stebbing et al., 2020). In neurodegenerative diseases, clozapine's dual action on dopamine D2 and serotonin 5-HT2A networks supported its use in Parkinson's psychosis (Meltzer et al., 2010).

5 CHALLENGES & LIMITATIONS

Despite its promise, network-based drug repositioning faces several challenges. Data quality remains a major concern, as incomplete or biased datasets can impair model reliability. Most current models are static and do not incorporate temporal or tissue-specific dynamics.

Experimental validation remains a bottleneck, with in vitro and in

vivo testing requiring substantial time and resources. Generalized models may not account for inter-patient molecular heterogeneity, limiting their translational impact. The interpretability of deep learning models is another critical issue, particularly in clinical settings.

Finally, commercial and legal considerations—such as limited patent protection for repositioned drugs—pose barriers to investment and implementation.

6 FUTURE DIRECTIONS

Emerging technologies are poised to address many limitations in current repositioning pipelines. Single-cell and spatial transcriptomics provide high-resolution views of disease states, allowing for more targeted network modeling.

Dynamic and context-aware networks, which capture disease progression and environmental factors, will improve the specificity of predictions. The adoption of Explainable AI (XAI) techniques will enhance model interpretability and foster clinical trust.

Causal inference frameworks, such as directed acyclic graphs and counterfactual models, may better distinguish correlation from causality in drug–disease relationships. Personalized network medicine, integrating patient-specific omics data, will enable tailored repositioning strategies (Greene et al., 2015). Establishing standardized benchmarks and shared evaluation datasets is also critical for comparing and validating computational approaches.

7 CONCLUSION

Network-based drug repositioning represents a paradigm shift in drug discovery. By modeling the interconnected landscape of biological and pharmacological systems, these methods offer mechanistic insights and predictive capabilities beyond traditional approaches. As computational models evolve and omics data become richer, network-based frameworks are poised to play a central role in precision medicine and rational drug development.

8 REFERENCES

Alaimo, S., Bonnici, V., Cancemi, D., Ferro, A., Giugno, R., & Pulvirenti, A. (2015). DT-Web: a web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference. *BMC Syst Biol*, *9 Suppl 3*(Suppl 3), S4. https://doi.org/10.1186/1752-0509-9-S3-S4

Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov*, *3*(8), 673-683. https://doi.org/10.1038/nrd1468

Barabasi, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat Rev Genet*, *12*(1), 56-68. https://doi.org/10.1038/nrg2918

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W. H., Pages, F., Trajanoski, Z., & Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, *25*(8), 1091-1093. https://doi.org/10.1093/bioinformatics/btp101

Chen, X., Yan, C. C., Zhang, X., Zhang, X., Dai, F., Yin, J., & Zhang, Y. (2016). Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform*, *17*(4), 696-712. https://doi.org/10.1093/bib/bbv066

Cheng, F., Desai, R. J., Handy, D. E., Wang, R., Schneeweiss, S., Barabasi, A. L., & Loscalzo, J. (2018). Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun*, *9*(1), 2691. https://doi.org/10.1038/s41467-018-05116-5

Csermely, P., Korcsmaros, T., Kiss, H. J., London, G., & Nussinov, R. (2013). Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther*, *138*(3), 333-408. https://doi.org/10.1016/j.pharmthera.2013.01.016

Cui, Z., Gao, Y. L., Liu, J. X., Dai, L. Y., & Yuan, S. S. (2019). L(2,1)-GRMF: an improved graph regularized matrix factorization method to predict drug-target interactions. *BMC Bioinformatics*, *20*(Suppl 8), 287. https://doi.org/10.1186/s12859-019-2768-7

D'Amato, R. J., Loughnan, M. S., Flynn, E., & Folkman, J. (1994). Thalidomide is an inhibitor of angiogenesis. *Proc Natl Acad Sci U S A*, *91*(9), 4082-4085. https://doi.org/10.1073/pnas.91.9.4082

Davis, A. P., Wiegers, T. C., Johnson, R. J., Sciaky, D., Wiegers, J., & Mattingly, C. J. (2023). Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Res*, *51*(D1), D1257-D1262. https://doi.org/10.1093/nar/gkac833

DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *J Health Econ*, *22*(2), 151-185. https://doi.org/10.1016/S0167-6296(02)00126-1

Dongen, S. (2000). *A cluster algorithm for graphs*. CWI (Centre for Mathematics

and Computer Science).

Duan, Q., Reid, S. P., Clark, N. R., Wang, Z., Fernandez, N. F., Rouillard, A. D., Readhead, B., Tritsch, S. R., Hodos, R., Hafner, M., Niepel, M., Sorger, P. K., Dudley, J. T., Bavari, S., Panchal, R. G., & Ma'ayan, A. (2016). L1000CDS(2): LINCS L1000 characteristic direction signatures search engine. *NPJ Syst Biol Appl*, *2*, 16015-. https://doi.org/10.1038/npjsba.2016.15

Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, *30*(1), 207-210. https://doi.org/10.1093/nar/30.1.207

Fiscon, G., Conte, F., Farina, L., & Paci, P. (2021). SAveRUNNER: A network-based algorithm for drug repurposing and its application to COVID-19. *PLoS Comput Biol*, *17*(2), e1008686. https://doi.org/10.1371/journal.pcbi.1008686

Galie, N., Ghofrani, H. A., Torbicki, A., Barst, R. J., Rubin, L. J., Badesch, D., Fleming, T., Parpia, T., Burgess, G., Branzi, A., Grimminger, F., Kurzyna, M., Simonneau, G., & Sildenafil Use in Pulmonary Arterial Hypertension Study, G. (2005). Sildenafil citrate therapy for pulmonary arterial hypertension. *N Engl J Med*, *353*(20), 2148-2157. https://doi.org/10.1056/NEJMoa050010

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrian-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magarinos, M. P., Overington, J. P., Papadatos, G., Smit, I., & Leach, A. R. (2017). The ChEMBL database in 2017. *Nucleic Acids Res*, *45*(D1), D945-D954. https://doi.org/10.1093/nar/gkw1074

Ghiassian, S. D., Menche, J., & Barabasi, A. L. (2015). A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Comput Biol*, *11*(4), e1004120. https://doi.org/10.1371/journal.pcbi.1004120

Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabasi, A. L. (2007). The human disease network. *Proc Natl Acad Sci U S A*, *104*(21), 8685-8690. https://doi.org/10.1073/pnas.0701361104

Goldstein, I., Lue, T. F., Padma-Nathan, H., Rosen, R. C., Steers, W. D., & Wicker, P. A. (1998). Oral sildenafil in the treatment of erectile dysfunction. Sildenafil Study Group. *N Engl J Med*, *338*(20), 1397-1404. https://doi.org/10.1056/NEJM199805143382001

Gottlieb, A., Stein, G. Y., Ruppin, E., & Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol*, *7*, 496. https://doi.org/10.1038/msb.2011.26

Greene, C. S., Krishnan, A., Wong, A. K., Ricciotti, E., Zelaya, R. A., Himmelstein, D. S., Zhang, R., Hartmann, B. M., Zaslavsky, E., Sealfon, S. C., Chasman, D. I., FitzGerald, G. A., Dolinski, K., Grosser, T., & Troyanskaya, O. G. (2015). Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet*, *47*(6), 569-576. https://doi.org/10.1038/ng.3259

Guney, E., Menche, J., Vidal, M., & Barabasi, A. L. (2016). Network-based in silico drug efficacy screening. *Nat Commun*, *7*, 10331. https://doi.org/10.1038/ncomms10331

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems*, *30*.

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*, *33*(Database issue), D514-517. https://doi.org/10.1093/nar/gki033

Hewett, M., Oliver, D. E., Rubin, D. L., Easton, K. L., Stuart, J. M., Altman, R. B., & Klein, T. E. (2002). PharmGKB: the Pharmacogenetics Knowledge Base. *Nucleic Acids Res*, *30*(1), 163-165. https://doi.org/10.1093/nar/30.1.163

Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., Green, A., Khankhanian, P., & Baranzini, S. E. (2017). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife*, *6*. https://doi.org/10.7554/eLife.26726

Hodos, R. A., Kidd, B. A., Shameer, K., Readhead, B. P., & Dudley, J. T. (2016). In silico methods for drug repurposing and pharmacology. *Wiley Interdiscip Rev Syst Biol Med*, *8*(3), 186-210. https://doi.org/10.1002/wsbm.1337

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, *5*(9). https://doi.org/10.1371/journal.pone.0012776

Ito, T., Ando, H., Suzuki, T., Ogura, T., Hotta, K., Imamura, Y., Yamaguchi, Y., & Handa, H. (2010). Identification of a primary target of thalidomide teratogenicity. *Science*, *327*(5971), 1345-1350. https://doi.org/10.1126/science.1177319

Jeong, H., Mason, S. P., Barabasi, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41-42. https://doi.org/10.1038/35075138

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*, *49*(D1), D545-D551. https://doi.org/10.1093/nar/gkaa970

Kholodenko, B. N. (2006). Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol*, *7*(3), 165-176. https://doi.org/10.1038/nrm1838

Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D. J., Turner, J. P., Vidovic, D., Forlin, M., Kelley, T. T., D'Urso, A., Allen, B. K., Torre, D., Jagodnik, K. M., Wang, L., Jenkins, S. L., Mader, C., Niu, W., Fazel, M., Mahi, N.,… Schurer, S. C. (2018). Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res*, *46*(D1), D558-D566. https://doi.org/10.1093/nar/gkx1063

Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Res*, *44*(D1), D1075-1079. https://doi.org/10.1093/nar/gkv1075

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, *313*(5795), 1929-1935. https://doi.org/10.1126/science.1132939

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788-791. https://doi.org/10.1038/44565

Li, Q., Han, Z., & Wu, X.-M. (2018). Deeper insights into graph convolutional networks for semi-supervised learning. Proceedings of the AAAI conference on artificial intelligence,

Li, Y., & Patra, J. C. (2010). Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network. *Bioinformatics*, *26*(9), 1219-1224. https://doi.org/10.1093/bioinformatics/btq108

Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., Peng, J., Chen, L., & Zeng, J. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun*, *8*(1), 573. https://doi.org/10.1038/s41467-017-00680-8

Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, *21*(16), 3448-3449. https://doi.org/10.1093/bioinformatics/bti551

Marbach, D., Costello, J. C., Kuffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Consortium, D., Kellis, M., Collins, J. J., & Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. *Nat Methods*, *9*(8), 796-804. https://doi.org/10.1038/nmeth.2016

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., & Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, *7 Suppl 1*(Suppl 1), S7. https://doi.org/10.1186/1471-2105-7-S1-S7

Martinez, V., Navarro, C., Cano, C., Fajardo, W., & Blanco, A. (2015). DrugNet: network-based drug-disease prioritization by integrating heterogeneous data. *Artif Intell Med*, *63*(1), 41-49. https://doi.org/10.1016/j.artmed.2014.11.003

Meltzer, H. Y., Mills, R., Revell, S., Williams, H., Johnson, A., Bahr, D., & Friedman, J. H. (2010). Pimavanserin, a serotonin(2A) receptor inverse agonist, for the treatment of parkinson's disease psychosis. *Neuropsychopharmacology*, *35*(4), 881-892. https://doi.org/10.1038/npp.2009.176

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., & Barabasi, A. L. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science*, *347*(6224), 1257601. https://doi.org/10.1126/science.1257601

Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Felix, E., Magarinos, M. P., Mosquera, J. F., Mutowo, P., Nowotka, M., Gordillo-Maranon, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C. J., Segura-Cabrera, A.,…Leach, A. R. (2019). ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res*, *47*(D1), D930-D940. https://doi.org/10.1093/nar/gky1075

Milacic, M., Beavers, D., Conley, P., Gong, C., Gillespie, M., Griss, J., Haw, R., Jassal, B., Matthews, L., May, B., Petryszak, R., Ragueneau, E., Rothfels, K., Sevilla, C., Shamovsky, V., Stephan, R., Tiwari, K., Varusai, T., Weiser, J.,…D'Eustachio, P. (2024). The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res*, *52*(D1), D672-D678. https://doi.org/10.1093/nar/gkad1025

Morselli Gysi, D., do Valle, I., Zitnik, M., Ameli, A., Gan, X., Varol, O., Ghiassian, S. D., Patten, J. J., Davey, R. A., Loscalzo, J., & Barabasi, A. L. (2021). Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc Natl Acad Sci U S A*, *118*(19). https://doi.org/10.1073/pnas.2025581118

Nepusz, T., Yu, H., & Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods*, *9*(5), 471-472. https://doi.org/10.1038/nmeth.1938

Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? *Nat Biotechnol*, *28*(3), 245-248. https://doi.org/10.1038/nbt.1614

Oughtred, R., Stark, C., Breitkreutz, B. J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers, M. (2019). The BioGRID interaction database: 2019 update. *Nucleic Acids Res*, *47*(D1), D529-D541. https://doi.org/10.1093/nar/gky1079

Pinero, J., Bravo, A., Queralt-Rosinach, N., Gutierrez-Sacristan, A., Deu-Pons, J., Centeno, E., Garcia-Garcia, J., Sanz, F., & Furlong, L. I. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*, *45*(D1), D833-D839. https://doi.org/10.1093/nar/gkw943

Pinero, J., Ramirez-Anguita, J. M., Sauch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., & Furlong, L. I. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res*, *48*(D1), D845-D855. https://doi.org/10.1093/nar/gkz1021

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A*, *105*(4), 1118-1123. https://doi.org/10.1073/pnas.0706851105

Sanchez-Lengeling, B., Reif, E., Pearce, A., & Wiltschko, A. B. (2021). A gentle introduction to graph neural networks. *Distill*, *6*(9), e33.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, *13*(11), 2498-2504. https://doi.org/10.1101/gr.1239303

Stebbing, J., Phelan, A., Griffin, I., Tucker, C., Oechsle, O., Smith, D., & Richardson, P. (2020). COVID-19: combining antiviral and anti-inflammatory treatments. *Lancet Infect Dis*, *20*(4), 400-402. https://doi.org/10.1016/S1473-3099(20)30132-8

Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, *302*(5643), 249-255. https://doi.org/10.1126/science.1087447

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., Lahr, D. L., Hirschman, J. E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I. C., Lam, D.,…Golub, T. R. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, *171*(6), 1437-1452 e1417. https://doi.org/10.1016/j.cell.2017.10.049

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, *47*(D1), D607-D613. https://doi.org/10.1093/nar/gky1131

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*, *49*(D1), D605-D612. https://doi.org/10.1093/nar/gkaa1074

Szklarczyk, D., Santos, A., von Mering, C., Jensen, L. J., Bork, P., & Kuhn, M. (2016). STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res*, *44*(D1), D380-384. https://doi.org/10.1093/nar/gkv1277

Taheri-Ledari, M., Zandieh, A., Shariatpanahi, S. P., & Eslahchi, C. (2022). Assignment of structural domains in proteins using diffusion kernels on graphs. *BMC Bioinformatics*, *23*(1), 369. https://doi.org/10.1186/s12859-022-04902-9

Valdeolivas, A., Tichit, L., Navarro, C., Perrin, S., Odelin, G., Levy, N., Cau, P., Remy, E., & Baudot, A. (2019). Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics*, *35*(3), 497-505. https://doi.org/10.1093/bioinformatics/bty637

Vandin, F., Upfal, E., & Raphael, B. J. (2011). Algorithms for detecting significantly mutated pathways in cancer. *J Comput Biol*, *18*(3), 507-522. https://doi.org/10.1089/cmb.2010.0265

Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., & Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation. *PLoS Comput Biol*, *6*(1), e1000641. https://doi.org/10.1371/journal.pcbi.1000641

Wang, Y., Chen, S., Deng, N., & Wang, Y. (2013). Drug repositioning by kernel-based integration of molecular structure, molecular activity, and

phenotype data. *PLoS One*, *8*(11), e78518. https://doi.org/10.1371/journal.pone.0078518

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D.,…Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*, *46*(D1), D1074-D1082. https://doi.org/10.1093/nar/gkx1037

Wu, C., Gudivada, R. C., Aronow, B. J., & Jegga, A. G. (2013). Computational drug repositioning through heterogeneous network clustering. *BMC Syst Biol*, *7 Suppl 5*(Suppl 5), S6. https://doi.org/10.1186/1752-0509-7-S5-S6

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, *24*(13), i232-240. https://doi.org/10.1093/bioinformatics/btn162

Yildirim, M. A., Goh, K. I., Cusick, M. E., Barabasi, A. L., & Vidal, M. (2007). Drug-target network. *Nat Biotechnol*, *25*(10), 1119-1126. https://doi.org/10.1038/nbt1338

Zagidullin, B., Aldahdooh, J., Zheng, S., Wang, W., Wang, Y., Saad, J., Malyutina, A., Jafari, M., Tanoli, Z., Pessia, A., & Tang, J. (2019). DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res*, *47*(W1), W43-W51. https://doi.org/10.1093/nar/gkz337

Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R., & Cheng, F. (2019). deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics*, *35*(24), 5191-5198. https://doi.org/10.1093/bioinformatics/btz418

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI open*, *1*, 57-81.

Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, *34*(13), i457-i466. https://doi.org/10.1093/bioinformatics/bty294

99

Chapter 5

REGULATORY RNAS AND FUNCTIONAL RNAOME

Atakan DEMİR[1], Prof. Kazım Yalçın ARGA[2], Assoc. Prof. Beste TURANLI[3]

1 Marmara University, Faculty of Engineering, Department of Bioengineering, Istanbul, Türkiye

https://orcid.org/0009-0007-4771-056X
2 Marmara University, Faculty of Engineering, Department of Bioengineering, Istanbul, Türkiye

https://orcid.org/0000-0002-6036-1348
3 Marmara University, Faculty of Engineering, Department of Bioengineering, Istanbul, Türkiye

https://orcid.org/0000-0003-1330-9712

Introduction

Ribonucleic acids (RNAs) are the biomolecules that are essential to life, not only serving as carriers of genetic information encoded in DNA, but also actively participating in a broad spectrum of cellular processes (Cech & Steitz, 2014). Traditionally, RNAs are considered to be intermediate molecules that have messenger roles in the process of transferring genetic instructions from DNA to protein. However, with the advent of technology over the years, modern studies have shown that RNAs take part in an expanding array of biological functions, and they have now emerged as dynamic players involved in practically all aspects of cellular physiology, gene expression regulation, structural organization, catalysis, and signal transduction (Statello et al., 2020). Even though the classical central dogma, which the concept defines the directional flow of genetic information as DNA → RNA → Protein, is sufficient to emphasize the vital role of RNAs, it continues to be revealed that the significance of RNAs is not limited to that linear framework, but rather covers a complex and dynamic network of functional and regulatory actions requisite for viability (Crick, 1970; Statello et al., 2020).

With respect to the structure, RNA is a single-stranded polymer composed of ribonucleotides, each of which consists of a 5C ribose sugar, a phosphate group, and one of four nitrogenous bases: adenine (A), uracil (U), guanine (G), or cytosine (C). Unlike DNA molecules with stable and double-stranded nature, this structural arrangement of RNA allows RNA molecules to be more chemically active and conformationally flexible, granting them to adopt secondary and tertiary configurations through folding and base-complementarity for functional interactions (Tinoco & Bustamante, 1999). Due to this structural versatility, RNAs are capable of serving as not only a messenger, but also as a scaffold, regulator, enzyme (catalyst), and sensor (Doudna & Cech, 2002).

From an evolutionary perspective, the RNA World Hypothesis is one of the most persuasive theories for assessing the importance of RNA in life. This hypothesis was initially postulated by Carl Woese, Francis Crick, and Leslie Orgel, and later formalized by Walter Gilbert in 1986. What was posited by these scientists briefly is that RNA evolved before DNA and proteins. They proposed that early life on Earth may have relied only on RNA for both genetic information storage and catalytic activity prior to the emergence of DNA and proteins (Gilbert, 1986). This notion is supported by the exceptional dual-functioning ability of RNAs: they can carry genetic information, much like DNA, and catalyze biochemical reactions, akin to protein enzymes. In particular, the existence of ribozymes—RNA molecules with catalytic activity—is key evidence for the ancestral role of

RNA (Muñoz-Velasco et al., 2024). For instance, ribosomes are formed by ribosomal RNAs (rRNAs), and these rRNAs catalyze peptide bond formation between the amino acids that are brought by transfer RNAs (tRNAs) during protein synthesis without any involvement of proteins (Hoffer et al., 2020; Kirchner & Ignatova, 2015). Discovery of additional examples, such as auto-catalytic self-splicing introns and riboswitches, sides with the idea of RNA being an ideal candidate for having driven early cellular evolution due to its capacity to store information, adopt complex tertiary structures, catalyze biochemical reactions, and regulate gene expression (Garst et al., 2011).

Over the past two decades, the understanding of RNA has revolutionized with regard to the advancements in high-throughput sequencing technologies, functional genomics, transcriptomics, and RNA structural profiling. Genome-wide analyses, such as ENCODE, have revealed that only ~2% of the human genome encodes for protein while approximately 80% of the genome is transcribed into various RNA species. Subsequently, it was enlightened that the process of transcription does not exclusively serve protein-coding purposes and results in a coding RNA. The remaining ~78% of the genome produces a vast quantity of RNA molecules that possess different functional properties, which is referred to as non-coding RNA (ncRNA) (Dunham et al., 2012).

The whole RNA content present within a cell, tissue, or organism under time and space-dependent conditions is collectively called the RNAome. Both coding and non-coding RNAs together comprise this phenomenon, involving distinct RNA molecules with interrelated roles (Derks & Pothof, 2015). Although the literature lists many different kinds of RNA classification, the major classes of RNA are widely accepted as messenger RNA (mRNA), tRNA, and rRNA (Brosius & Raabe, 2016). The only coding RNA within organisms is mRNA, which forms a bridge between genetic information and functionality and serves as a template during the translation into proteins (Romão, 2022). On the other hand, ncRNAs encompass the vast majority of the transcriptome without any further translation into proteins (George et al., 2024). However, emerging research has exposed that there are also hybrid RNAs that are capable of both functioning and protein-coding, which blurred the boundaries between coding and non-coding transcripts. For example, some long non-coding RNAs (lncRNAs) have been found to be able to produce small peptides along with their regulatory function, which challenges the traditional idea that RNAs are either coding or non-coding (Choi et al., 2019). Similar to the regulatory elements of mRNA as promoters and terminators, untranslated regions (UTRs) of mRNAs have been identified as the hubs of regulatory elements that interact with RNA-binding proteins and

ncRNA in order to control translation efficiency, stability, and half-life of mRNA (Navarro et al., 2021).

Modern approaches for a better understanding of RNAome have been developed both computationally and experimentally. High-throughput RNA sequencing (RNA-Seq), crosslinking immunoprecipitation (CLIP), and high-resolution RNA structure mapping are several techniques that have been introduced to decipher the specific functions or proper categorizations of both familiar and novel RNA species. These recent technologies provide scientists with the ability to quantify RNA molecules, map RNA-RNA and RNA-protein interactions, and determine particular RNA functions in different biological contexts, including those of disease (Hwang et al., 2020; Stark et al., 2019; Van Nostrand et al., 2020). The abundance or functional dysregulations of RNAs have been linked to the signature of most human diseases, in which the RNA species with abnormal behaviors/amounts are considered to be biomarkers. Through RNAome analyses, we are able to achieve the differentially expressed gene (DEG) patterns of disease phenotypes in order to unveil the altered mechanisms at the transcriptional level. Dysfunctions in RNA biogenesis, processing, or decay may cause pathologies ranging from cancer and neurodegeneration on the one hand, to cardiovascular and autoimmune diseases on the other (Goodall & Wickramasinghe, 2020; Liu et al., 2017). To exemplify, disruptions in miRNA mechanism, which plays crucial roles in oncogenesis and tumor suppression, may lead to complex disease manifestations as well as abnormal long non-coding RNAs that have regulatory effects on cell growth, apoptosis, and metastasis (Ahmad et al., 2023; Otmani & Lewalle, 2021). Recent studies have also highlighted the significant contribution of improper RNA metabolism and processing, such as RNA-binding protein (RBP) dysfunctions, to neurodegeneration (Li & Sun, 2025). In this case, separating disease-relevant RNAs from transient or irrelevant species is the main challenge, which requires an integration of approaches utilizing transcriptomics, structural biology, functional genomics, and systems-level approaches. Nevertheless, the synthetic biology field enhances the controllability and modularity of RNA through engineering synthetic circuitry, RNA switches, and systems for the regulation of gene expression, with huge potential in future advances toward RNA-based therapeutic applications (Schmidt & Smolke, 2019). Concerning RNA therapeutics, antisense oligonucleotides, small interfering RNAs, RNAi mechanisms, and RNA vaccines are already utilized and being developed with the purpose of curing genetic illness, viral diseases, and cancer (Bajan & Hutvagner, 2020). Additionally, the CRISPR-Cas system has been utilized as one of the most recent therapeutic applications, in which the technology also relies on guide RNAs (gRNAs) for sequence-specific lo-

calization, although it aims to alter the DNA sequence itself (Xu et al., 2025). The RNAome, therefore, provides not only a window into disease mechanisms, but also a therapeutic target aspect, connecting genotype and phenotype with a high level of intricacy.

To sum up, the RNAome is the total coding and non-coding RNA network that occupies a central, yet multifaceted, position in biological systems and operates in harmony to identify the general structure and function of an organism by acting as regulators, catalysts, and scaffolds in addition to being informational intermediates. In this review, we provide a comprehensive overview of the general classes of RNA, emphasize the distinction between coding and non-coding RNA species, and highlight the biological significance of the functional RNAome in diverse cellular and disease-related contexts.

Classes of RNA and their Regulatory Roles

mRNA: Structure, synthesis, and messenger role in translation

Messenger ribonucleic acid (mRNA) is the main type of RNA that is an intermediate molecule formed through transcription of the DNA sequence of interest into a message that is suitable for accurate ribosomal entry and precise codon-anticodon pairing for the synthesis of a functional protein. It is a single-stranded ribonucleic acid that is composed of nucleotides bound by a phosphodiester bond, transcribed by RNA polymerase II from protein-coding genes. The mRNA displays a dynamic nature that allows for rapid cellular responses to environmental, stress-related conditions and developmental processes besides from its primary duty in protein translation (Romão, 2022). It functions not only as a protein coding tool but also as a precise regulator of cellular processes. With regards to development, translational mechanisms at certain times and asymmetric localization of mRNAs are critical aspects for cell fate and embryonic axis determination. For example, bicoid mRNA determines the anteroposterior axis in the Drosophila embryo while cyclin B mRNA regulates early mitotic divisions (Cai et al., 2021; Clemm von Hohenberg et al., 2022). The localization of specific RNAs within the cells is known to orchestrate asymmetric cell division and be an important mechanism for the establishment of cellular asymmetry during embryonic development. In many organisms, a large proportion of the oocyte cytoplasm is maternal origin. Tightly regulated transcriptional programs ensure that certain mRNAs, such as bicoid, nanos, and oskar, are specifically localized to the destined regions of the egg during oogenesis. After fertilization, each of these maternal mRNAs is distinctly involved in the regulation of cellular

processes according to its localization, even before the activation of genomic transcription in embryos. This asymmetric distribution provides uneven transferring of mRNAs to daughter cells, resulting in the initiation of cell lineage-specific gene expression programs; thus, differentiation (Kugler & Lasko, 2009).

rRNA: Structural and functional roles in ribosomes and protein synthesis

rRNA (ribosomal RNA) forms the structural skeleton of ribosomes and performs catalytic functions throughout the translation process. In both prokaryotic and eukaryotic organisms, ribosomes consist of rRNA and ribosomal proteins; however, their ribosomal structures are quite different. Prokaryotes have 70S ribosomes which consist of 50S (containing 23S and 5S) large subunits and 30S (containing 16S) small subunits, whereas eukaryotes have 80S ribosomes composed of 60S (containing 28S, 5.8S, and 5S) large subunits and 40S (containing 18S) small subunits (Yusupova & Yusupov, 2021). rRNA is vital for the cell because it forms the structural and catalytic basis of ribosomes. Hence, rRNA synthesis is a strictly regulated process and is tightly linked to cellular metabolism. However, rRNAs do not only determine the structure of the ribosomes with proteins, but they also have an active role in different steps of the translation. In particular, 16S (in prokaryotes) and 18S (in eukaryotes) rRNAs are found in the small subunit of the ribosome and play a critical role in the binding of mRNA to the ribosome and the correct codon-anticodon pairing (Hoffer et al., 2020). Interestingly, the direct role of ribosomal proteins in this catalytic function in protein synthesis is rather limited; the main catalytic activity is mediated by the three-dimensional conformation of rRNAs. This enzymatic activity is carried out by the rRNAs which are found in the ribosomal subunits, not by proteins. This finding revealed that ribosome is composed of ribozymes, the RNA molecules with catalytic activity (Gray & Gopalan, 2020). Additionally, rRNA also functions as the structural platform for the binding of ribosomal subunits, positioning of tRNA and mRNA, and initiation and termination of translation. In prokaryotes, the anti-Shine-Dalgarno sequence at the 3' end of the 16S rRNA pairs with the Shine-Dalgarno sequence in the mRNA, allowing the recognition of the correct start site with an additional level of regulation (Li et al., 2023).

tRNA: Functions as an adaptor in translation

tRNA (transfer RNA) carries out the transition phase where the genetic information encoded in codons is converted into the exact amino acid sequence as the primary structure of a protein. Within this phase, tRNA

acts as an adaptor molecule in protein synthesis and takes over the job of transporting amino acids to ribosomes, making the decoding of the genetic code possible (Kirchner & Ignatova, 2015). The length of tRNAs is around 70-95 nucleotides. They exist in an L-shaped three-dimensional structure formed through the folding of the classic "cloverleaf" secondary structure that consists of four main strands and an anticodon loop (acceptor stem, D stem-loop, anticodon stem-loop, variable arm, and T-stem-loop). The site for amino acid binding is located at the 3' end of tRNA, marked by a CCA nucleotide sequence. On the other hand, the anticodon loop base pairs with a specific codon on the mRNA. By this means, tRNA both transports the amino acid and recognizes its location by binding to the correct codon corresponding to the amino acid that is being carried. (Li et al., 2024). Recognition and the attachment of the correct amino acid are accomplished by the specific enzyme called aminoacyl-tRNA synthetase. Aminoacyl-tRNA synthetases are the family of enzymes that manage the amino acid attachments to the 3' end of the tRNA, generating aminoacyl-tRNA through a two-step ATP-driven reaction (Angel et al., 2020). The specificity of amino acid-tRNA pairing is controlled by proofreading mechanisms that hydrolyze incorrectly formed aminoacyl adenylates and aminoacyl-tRNAs in the editing site (Gupta et al., 2023).

miRNA: Post-transcriptional gene regulation

miRNAs (MicroRNAs) are single-stranded small non-coding RNAs, approximately 20-24 nucleotides in length, which act in post-transcriptional gene expression regulation. Mainly, they target the 3' UTR of an mRNA to prevent the translation or to cause the degradation of its target mRNA. miRNAs play a fundamental role in fine-tuning gene expression in many eukaryotes, including mammals. Nearly all events in cells are affected by miRNA's action, providing fine-tuning of gene expression at the post-transcriptional level. For example, in vertebrates, let-7 miRNAs are key regulators in determining cell fate, prominently expressed in differentiated tissues during late embryonic development (Reynolds, 2024). Another example is miR-21, which is known to act as an oncogene by targeting tumor suppressors such as PTEN and PDCD4 to activate the PI3K/AKT pathway. miR-21 is highly expressed in breast, lung, and colorectal cancers and contributes to tumorigenesis (Liu et al., 2019). Therefore, it can be said that miRNAs act as key regulators of gene expression at the post-transcriptional level, contributing to the fine-tuning of cellular functions and the maintenance of homeostasis. Their ability to coordinate the repression of multiple target mRNAs in a time and space-dependent manner further advances the intricacy of gene regulatory networks to a new level.

siRNA: RNA interference and gene silencing

siRNAs (Small interfering RNAs) are 21-23 nucleotide long double-stranded RNA molecules that have starter roles in RNA interference (RNAi)-mediated post-transcriptional gene silencing (Alshaer et al., 2021). RNAi is a highly conserved and sequence-specific post-transcriptional gene silencing mechanism that was first discovered in 1998 by Fire and Mello in Caenorhabditis elegans as double stranded RNAs, and later described in higher eukaryotes, including plants, animals and fungi (Cornec & Poirier, 2023; Fire et al., 1998). In addition to gene expression regulation, RNAi mechanism was found to be a natural defense mechanism against viruses and transposons (Hung & Slotkin, 2021). The central players of this intricate network of RNA regimentation are shown to be siRNAs as they guide the sequence-dependent endonucleolytic cleavage of mRNAs (Hamilton & Baulcombe, 1999). siRNAs are typically generated from long double-stranded RNAs (dsRNAs), which could arise exogenously, such as through viral infection or in vitro introduction, or endogenously from transcription of repetitive genomic sequences. siRNA serves as a tool in molecular biology applications, owing to its ability to repress gene expression with high level of specificity at the post-transcriptional level. The applications of siRNA extend widely across basic research, functional genomics, and clinical therapeutics. Hopefully, siRNAs can give us new opportunities for silencing protein targets that are considered "undruggable". Nevertheless, the applications of siRNA therapeutics display some limitations, such as off-target effects, immune system activation and toxicity, stability and degradation of siRNA, delivery challenges of siRNA, oversaturation of RNAi mechanism (Ali Zaidi et al., 2023).

piRNA: Transposon silencing in germ cells

piRNAs (PIWI-interacting RNAs) are 2'-O-methylated 3'-end small non-coding RNAs with an approximate length of 24-35 nucleotides that form complexes with PIWI proteins. Their main role, which is evolutionary conserved, is to maintain genome integrity in germline cells by silencing mobile genetic elements called transposable elements (TE) (Hirakata & Siomi, 2016). The majority of piRNAs contain sequences that are antisense to transposon RNAs and thus can act as oligonucleotides to repress the corresponding mobile DNA elements. These RNAs were named "PIWI-interacting RNA (piRNA)" and were shown to work specifically with PIWI proteins such as Aubergine and AGO3 in Drosophila to suppress transposons. This discovery revealed the importance of RNA-mediated defense mechanisms in the maintenance of genome integrity in germ cells (Ozata et al., 2019). Although the roles of piRNAs were initially thought to be limited merely to transposable element activity suppression in germ

cells, recent studies have enlightened that piRNAs are involved in processes such as gene regulation, tumor suppression, synaptic plasticity, and memory formation in both neuronal and somatic tissue (Hegde & Smith, 2019).

snRNA: Spliceosome components and RNA processing

Small nuclear RNAs (snRNAs) are an essential class of non-coding RNAs, which function primarily in the cell nucleus and participate in the pre-mRNA splicing machinery of post-transcriptional gene regulation (Morais et al., 2021). snRNAs constitute catalytic and structural basis of spliceosome ribonucleoprotein complex responsible for removing non-coding intronic region of pre-mRNA transcript. Each snRNA assembles with a set of core proteins to form a small nuclear ribonucleoprotein (snRNP) as the fundamental functional unit of the spliceosome. These snRNPs not only serve as structural stability but also have crucial roles in recognizing splice sites and catalyzing the stepwise removal of introns. snRNAs enable precise and efficient removal of introns although they have different transcriptional origins, biogenetic pathways and functional roles according to spliceosome type and RNA polymerase specificity.

snoRNA: rRNA modification and maturation

Small nucleolar RNAs (snoRNAs) are a class of small non-coding RNA, typically 60-300 nucleotides in length, which are primarily localized in the nucleolus of the eukaryotic cells, where they play fundamental roles in performing to guide the chemical modifications of other RNA types, especially ribosomal RNAs (rRNAs), transfer RNA (tRNA), and small nuclear RNAs (snRNAs) (Xiao et al., 2023). The best-known function of snoRNA is directing the chemical modifications on different RNA moieties, such as 2'-O-methylation and pseudouridylation on rRNAs, which are critical for proper ribosome assembly, stability, and translation efficiency (Li et al., 2023). Generally, snoRNAs are transcribed from introns of protein-coding or non-coding host genes by RNA polymerase II, although some are synthesized by RNA polymerase III (Scott & Ono, 2011; Zimta et al., 2020). Beyond this, snoRNAs have the potential to be used both as diagnostic biomarkers and are being evaluated as therapeutic targets for many diseases. Synthetic snoRNAs are also being developed as a molecular biology tool to direct specific modifications to desired RNA regions. Therefore, the importance of snoRNAs in basic biology and medicine is on a rise, due to not only rRNA maturation and ribosome biogenesis, but also a wide range of regulatory roles associated with cellular physiology and pathology (Grützmann et al., 2024).

Ribozymes: Catalytic RNA molecules

Ribozymes are RNA molecules that have the ability to catalyze certain biochemical reactions, and they act as protein enzymes. The discovery of ribozymes thoroughly changed the idea that only proteins could be biological catalysts (Scott, 2007). In the early 1980s, T. R. Cech identified a self-cutting and self-assembling intron in the rRNA precursor of Tetrahymena thermophila, and S. Altman showed that the RNA component of ribonuclease P (RNase P) performs the catalytic activity that cuts the 5' leader sequences of precursor tRNAs (Guerrier-Takada et al., 1983; Kruger et al., 1982). These discoveries have provided strong support for the RNA World Hypothesis by demonstrating that RNA molecules can function as carriers of genetic information and as catalytic agents.

Ribozymes are divided into large and small ribozymes according to their structure and catalytic mechanisms. Large ribozymes are the naturally occurring ribozymes involved in essential cellular processes, with the models of Group I introns, Group II introns, RNase P, and ribosomal RNA (rRNA). Small ribozymes have much simpler secondary structures and typically catalyze site-specific phosphodiester cleavage, including hammerhead ribozyme, hairpin ribozyme, hepatitis delta virus (HDV) ribozyme, and Varkud Satellite (VS) ribozyme (Tanner, 1999). Group I and Group II introns are capable of self-splicing by two-step transesterification reactions. Group II introns are thought to be evolutionary precursors for the spliceosome found in eukaryotes(Gomes et al., 2024; Monachello et al., 2021). RNase P is a ribonucleoprotein complex involved in the maturation of precursor tRNAs by cleaving the 5' leader sequence (Shaukat et al., 2021). Another important example of a ribozyme is the large subunit ribosomal RNA (23S or 28S), which forms the peptidyl transferase center of the ribosome that catalyzes peptide bond formation during protein synthesis (Gray & Gopalan, 2020). In contrast, small ribozymes are often found in viral genomes and can perform self-cut to regulate replication (Ren et al., 2017).

Ribozymes have been engineered and repurposed in therapeutics and synthetic biology. Using in vitro evolution techniques, such as SELEX, ribozymes with altered substrate specificity and increased catalysis rate have been developed, such as riboswitch systems (Ge et al., 2021). Chemical modifications, such as 2′ -O-methylation and locked nucleic acids (LNA), and nanoparticle-based delivery systems are current strategies to improve ribozyme stability and therapeutic efficacy. Moreover, integrating ribozymes with modern gene editing technologies such as CRISPR could pave the way for more precise and programmable RNA-level interventions in the future (Zhou et al., 2019).

lncRNA: A different world of regulation

The non-coding RNA molecules that are longer than 200 nucleotides are termed as long non-coding RNAs (lncRNAs), which comprise various functional RNA species as the vast majority of total transcripts (Djebali et al., 2012). As the name implies, they do not encode for and are translated into a protein but instead participate in numerous regulatory processes with different functions, including regulation of gene expression and structural functions both in the nucleus and the cytoplasm (Wilusz et al., 2009). There is a diverse catalog of lncRNAs in terms of their biogenesis and function, which we discussed in this section while underlining the biological significance and variety of their existence.

Even the recent literature is not sufficient to propose a clear categorization that is well-defined and universally accepted for lncRNAs. Thus far, scientists have suggested a wide variety of classifications within the scope of genomic location or functional mechanisms of lncRNAs. Most recently, from a functional aspect, Wang & Chang propounded a function-oriented model that includes 4 archetypes: signals, decoys, guides, and scaffolds (Wang & Chang, 2011). Despite the fact that this classification is deeply detailed and distills the myriad functions of lncRNAs, it was not considered to be comprehensive enough to illustrate the commonalities and diversities of the lncRNA world clearly, and to unveil the functional complexity to be accepted universally. Herein, we consider the classification of Yao et al to be the most comprehensive and contextually appropriate model to date; thus, we accepted and summed that model in this section: (1) large intervening/intergenic non-coding RNAs (lincRNAs), (2) natural antisense transcripts (NATs), (3) RNase P processed lncRNAs, (4) snoRNA-ended lncRNAs (sno-lncRNAs), (5) 5' snoRNA-ended and 3'-polyadenylated lncRNAs (SPAs), (6) circular intronic RNAs (ciRNAs), (7) circular RNAs (circRNAs) (Yao et al., 2019).

In addition to the functions in gene expression regulation, lncRNAs also have structural roles both in the nucleus and the cytoplasm. After being localized to nucleus, lncRNAs can regulate nuclear organization and function as they take part in chromosome territory, chromatin structure, transcription, and post-transcriptional regulations. One of the most well-known examples is the X-inactive-specific transcript (Xist), which coats one of the X chromosome from which it is transcribed and recruits chromatin silencing complexes such as PRC1 and PRC2 to inactivate the chromosome and ensure dosage compensation in female embryos. This interaction results in widespread deposition of repressive histone modifications, such as H3K27me3, leading to stable silencing of one X chromosome in female mammals (Brockdorff, 2013). HOTAIR, as another example, is a prominent scaffold lncRNA that links PRC2 and LSD1/CoREST complexes, coordinating histone methylation and demethylation at

distant genomic sites such as the HOXD locus (Rinn et al., 2007). Some lncRNAs may also bind DNA directly through triplex-forming sequences, aiding in the targeting of chromatin modifiers (Mondal et al., 2015). Hence, lncRNAs display a great clinical potential as biomarkers and therapeutic targets. Their expression is often cell- or tissue-specific, and many are detectable in biofluids like blood, where they are protected from degradation through exosomal encapsulation or unique secondary structures (Hosseini et al., 2022). This stability enhances their utility in non-invasive diagnostics. Furthermore, manipulating lncRNA expressions or functions through antisense oligonucleotides, small compounds, or CRISPR-based methods reflects an appealing approach for improved therapeutic interventions.

Extracellular and Circulating RNAs

Extracellular or circulating RNAs (often termed exRNAs or cell-free RNAs, cfRNAs) are RNA molecules detected in nearly all body fluids, including blood (plasma/serum), saliva, urine, milk, and cerebrospinal fluid as a result of their release from various cell types (Etheridge et al., 2013). Interestingly, exRNAs display remarkable persistence in circulation although the nature of naked RNAs is known to be inherently unstable. Early assumptions that free RNA would be rapidly degraded were overturned by findings that circulating RNAs remain stable and incorporate in various biological functions like cell-to-cell communication and gene regulation, which might lead to cell proliferation, apoptosis, differentiation, metabolic alterations, and immune responses in distant target cells (Zhong et al., 2024).

The stability of exRNAs in extracellular environments is explained by the way exRNAs are released, packaged and protected. In fact, they are not completely free, but mostly are bound to or encapsulated in protective carriers that provide RNase avoidance, which is referred to as active release. Viable cells can actively secrete RNA by loading it into those protective carriers, whereas passive release occurs when cells die or are injured. Therefore, the biogenesis of an exRNA is the key factor in determining its fate and characteristics (Sohail et al., 2022). The analysis of plasma exRNAs has attracted a great interest in its clinical applications as liquid biopsy biomarkers since they offer early diagnosis, monitoring treatment responses, and understanding disease progression with reduced invasiveness and increased reproducibility. Additionally, because circulating RNAs reflects tissue-specific or pathology-specific expression patterns that can sensitively flag disease processes, they can serve

as non-invasive indicators of diseases (Zhong et al., 2024). In contrast to genomic DNA markers, exRNAs are accepted to be highly functional as a biomarker since they report dynamic functional information about which genes are active or suppressed in diseased tissue rather than indicating possible risks arising from mutations in the genome. For instance, tumor cells release exRNAs into the bloodstream; profiling these circulating exRNA signatures can reveal tumor-specific expression changes without needing a tissue biopsy, which shows a great promise for early diagnosis in cancer (Metzenmacher et al., 2020). Scientists even suggest that exRNA biomarkers may be more sensitive than circulating cell-free DNA for reflecting tumor presence due to the fact that exRNAs include signals from multiple cell types and regulatory processes (Kan et al., 2023). Since their initial discovery in 2008, specific exRNAs have already been linked to various disorders. For example, miR-21 was found to be elevated in breast, lung, colon, and pancreatic tumors, while plasma miR-155 is associated with lymphoma progression and poor prognosis, particularly in diffuse large B-cell lymphoma (DLBCL) (Bautista-Sánchez et al., 2020; Mazan-Mamczarz & Gartenhaus, 2013). In liver disease, circulating miR-122 levels rise significantly in drug-, viral-, alcohol-, or chemically induced liver injury, including paracetamol overdose where capillary miR-122 predicts biopsy-proven injury (Cirronis et al., 2024; Ding et al., 2012; Zhang et al., 2010).

Despite these encouraging findings, challenges that limit the routine use of exRNA biomarkers remain. Sensitivity can be addressed as a key issue since exRNA levels in blood are relatively low and the molecules are usually fragmented. This makes the precise detection of desired biomarkers in serum/plasma difficult. To avoid degradation, samples must be handled carefully with a proper follow-up of experimental protocols (e.g. EV separation, precipitation, immunocapture). However, there is not a golden standard for experimental protocols as it varies widely, and no single method works universally. Even from bioinformatics aspect, quantification and normalization lack consensus, and techniques may favor certain RNA types over others, which results in inconsistent findings across laboratories. In addition to the lack of standardization, the need for large sample volumes to detect rare transcripts and the complexity of deep sequencing also hinder clinical implementation. Thus, all of these drawbacks must be overcome in order for the transition in the use of circulating RNA biomarkers from research to clinic (Zhong et al., 2024).

Current Research Approaches and Databases

Over the past two decades, the area of RNA research has remarkably expanded with the rapid development of high-throughput sequencing techniques and precise chemical probing methods. These advancements have enabled scientists to screen the whole transcriptome with high resolution along with unraveling novel RNA molecules and complex pathways of RNA regulation across various contexts, including diseases (D'Agostino et al., 2022). With respect to RNAs in disease, the main focus of contemporary research is to elucidate the molecular basis of pathological malfunctions through profiling and characterizing a broad range of both coding and non-coding transcripts that are altered in disease conditions. Since differentially expressed genes as transcripts can point out disrupted pathways and serve as biomarkers, researchers aim to identify transcripts whose expression shifts in disease versus healthy conditions and to suggest an expression pattern that represents the disease phenotype (Hasin et al., 2017). Therefore, it can be said that the modern research of RNA mostly relies on a range of experimental approaches that yield structural and functional insights into both coding and non-coding RNA species.

With respect to experimental techniques in RNA research, there are numerous methods that approach RNA from different facets. These include RNA sequencing (RNA-seq), ribosome profiling (Ribo-seq), cross-linking and immunoprecipitation sequencing (CLIP-seq) with advanced versions as RNA immunoprecipitation (RIP), capture hybridization analysis of RNA targets (CHART), chromatin isolation by RNA purification (ChIRP), and RNA antisense purification (RAP), selective 2′ -hydroxyl acylation analyzed by primer extension sequencing (SHAPE-seq), dimethyl sulfate–mutational profiling with sequencing (DMS-MaPseq), and CRISPR-based RNA targeting systems such as CRISPR-Cas13. RNA-seq allows for comprehensive profiling of RNA expression within bulk samples. Its methodology comprises two steps as conversion of RNA molecules into complementary DNA (cDNA) molecules, which is then followed by deep sequencing. RNA-seq provides not only quantification of transcripts, but also sequence analysis of many types of RNA species, involving mRNAs, lncRNAs, and small non-coding RNAs (Hrdlickova et al., 2016). The capabilities of RNA-seq have improved after the development of its further variants, such as strand-specific RNA-seq (ssRNA-seq), which allows determination of transcriptional direction, RNA-seq with rRNA depletion, and single-cell RNA-seq (scRNA-seq), which analyzes heterogeneity and rare transcriptomic states under cellular resolution (Tang et al., 2009; Ziegenhain et al., 2017). Moreover, recent adaptations of RNA-seq have enabled indirect detection of RNA modifications through misincorporation signals or the use of chemical libraries such as pseudo-UTP analogs (Safra

et al., 2017). However, detecting transcripts with relatively low abundance like lncRNAs and distinguishing RNA molecules that display structural similarities, such as tRNAs and snRNAs, are current limitations. Another method is Ribo-seq, which involves sequencing ribosome-protected fragments of mRNAs, resulting in a snapshot of translational activity at sub-codon resolution (Ingolia et al., 2009). By this way, it is possible to detect which RNAs and ORFs are actively translated at a certain moment, including those embedded within non-coding regions such as lncRNAs and upstream ORFs (uORFs) (Brar & Weissman, 2015). Through integrating with RNA-seq, Ribo-seq bridges the gap between transcriptome and proteome by highlighting functional translation rather than mere transcript presence. CLIP-seq, on the other hand, provides scientists with the exploration of RNA-protein interactions in nucleotide resolution. This family of methods utilizes UV crosslinking to covalently link RNA-binding proteins (RBPs) to their RNA targets with a subsequent immunoprecipitation and deep sequencing of the bound RNA fragments (Stork & Zheng, 2016). CLIP-seq has been widely utilized in mapping the binding regions of RBPs (e.g. Argonaute), splicing regulators, RNA helicases, miRNA target sites, and lncRNA-interacting protein partners (Hafner et al., 2021). For structural characterization of RNA molecules, techniques such as SHAPE-seq and DMS-MaPseq provide high-resolution screening of RNA secondary structures in vitro and in vivo (Spitale et al., 2013; Zubradt et al., 2016). CRISPR-based screening platforms are another emerging frontier in RNA research. In contrast to the well-established Cas9 system that acts on DNA, CRISPR-Cas13 systems have been adapted for RNA-guided knockdown and RNA editing without altering the genome (Abudayyeh et al., 2017).

Indeed, the big data generated by the aforementioned high-throughput RNA profiling techniques demand just as advanced computational infrastructures to ensure accurate analysis and interpretation. This complexity has led scientists to improve and establish bioinformatic pipelines and tools, which are critical for aligning sequencing reads, quantifying expression, predicting structural and functional attributes, and integrating biological pathways and molecular functions arising from data. Alignment and quantification are fundamental processes in RNA-seq analysis. For that, tools such as HISAT2 enable rapid and accurate mapping of short reads to reference genomes (Kim et al., 2019). Once raw counts are obtained, differential expression analysis is typically performed using mostly R/Bioconductor software packages for data manipulation and statistical analysis, such as DESeq2 and limma (Love et al., 2014; R Core Team, 2025; Ritchie et al., 2015). RNA structural analysis also benefits from bioinformatic modeling tools as programs, such as RNAfold, Mfold,

and RNAstructure, are built to predict RNA secondary structures based on thermodynamic stability and chemical calculations (Kazanskii et al., 2024; Mathews et al., 2004; Zuker, 2003). Proceeding from individual gene analysis to overall data interpretation, functional enrichment tools, such as GSEA (Gene Set Enrichment Analysis) and DAVID, are commonly used in order for the identification of biological pathways and molecular functions associated with RNA expression profiles (Chen et al., 2013; Huang et al., 2009; Subramanian et al., 2005; Yu et al., 2012).

While these analytic tools bring enormously large RNA data in terms of their sequences, annotations, interactions, and experimental validations, the need for curated databases that provide structured repositories of the generated data has emerged inevitably. Today, there are hundreds of periodically updated RNA databases covering all acknowledged types of RNAs, their interactions, and even their associations with diseases, in which the most popular ones, such as miRTarBase, NONCODE, and starBase are obtained from Oxford Academic and listed in the **Table 1** (Cui et al., 2025; J. H. Li et al., 2013; Nucleic Acids Research | Oxford Academic, n.d.; Zhao et al., 2021).

Table 1. Popular RNA databases.

Database	*Description*	*URL*
miRBase	*Repository of miRNA sequences, annotation and nomenclature.*	https://www.mirbase.org/
miRTarBase	*Experimentally validated miRNA–target interactions.*	https://miRTarBase.cuhk.edu.cn/
miRDB	*miRNA target prediction and functional annotation database.*	https://mirdb.org/
mirGeneDB	*Manually curated microRNA gene database across metazoan species.*	https://www.mirgenedb.org/
RepTar	*Predicted cellular targets of host and viral miRNAs.*	http://reptar.ekmd.huji.ac.il/

DIANA-LncBase v3	*Experimentally validated miRNA–lncRNA interactions.*	https://diana.e-ce.uth.gr/lncbasev3
LNCipedia	*Compendium of human lncRNAs with expression and functional data.*	https://lncipedia.org/
LncExpDB	*Database of human lncRNA expression across biological contexts.*	https://bigd.big.ac.cn/lncexpdb/
LncACTdb	*Curated database of experimentally supported ceRNA (lncRNA–miRNA–mRNA).*	http://bio-bigdata.hrbmu.edu.cn/LncACTdb/
circBase	*Repository of experimentally validated circular RNAs.*	http://www.circbase.org/
circRNADb	*Manually curated database of human circRNAs.*	http://reprod.njmu.edu.cn/circrnadb
circBank	*Comprehensive circRNA database compiling human and various model organisms.*	https://www.circbank.cn/
CIRCpedia v2	*circRNA annotations from >180 RNA-seq datasets across six species.*	http://yang-laboratory.com/circpedia/
starBase (ENCORI)	*miRNA–mRNA, miRNA–lncRNA, RBP–RNA interactions from CLIP-seq data.*	http://starbase.sysu.edu.cn/
RNAcentral	*Aggregates ncRNA sequences from multiple sources.*	https://rnacentral.org/

NONCODE	*Integrated knowledgebase of lncRNAs and circRNAs across species.*	http://www.noncode.org/
snoDB	*Database of human snoRNAs with expression and interaction data.*	https://bioinfo-scottgroup.med.usherbrooke.ca/snoDB/
RBPDB	*Database of RNA-binding protein specificities.*	http://rbpdb.ccbr.utoronto.ca/
doRiNA	*Post-transcriptional RNA interaction repository from CLIP-seq.*	https://dorina.mdc-berlin.de/
Rfam	*Catalog of RNA families based on conserved sequence and structure.*	https://rfam.org/
SILVA rRNA	*Quality-checked aligned small and large subunit rRNA sequences.*	https://www.arb-silva.de/
tRNADB-CE	*Manually curated tRNA gene database including environmental sequences.*	http://trna.ie.niigata-u.ac.jp/
GtRNAdb	*Genomic tRNA gene predictions (via tRNAscan-SE).*	http://gtrnadb.ucsc.edu/
RNAxs	*Tool/database for predicting target sites of small RNAs.*	http://rna.tbi.univie.ac.at/cgi-bin/RNAxs/RNAxs.cgi
TargetScan	*Predicts biological targets of miRNAs across multiple species.*	http://www.targetscan.org/
RNAfold	*Predicts RNA secondary structure via thermodynamic models (Vienna).*	https://www.tbi.univie.ac.at/RNA/

RNAhybrid	*Predicts miRNA binding sites and hybridization energies.*	https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid
GEO	*Archive of gene expression and sequencing datasets.*	https://www.ncbi.nlm.nih.gov/geo/
SRA	*Nucleotide sequence read archive (sequencing data repository).*	https://www.ncbi.nlm.nih.gov/sra/
RNADisease	*Curated RNA–disease associations across multiple RNA types and species.*	https://www.rnadisease.org/
LncRNADisease	*Experimentally supported lncRNA and circRNA disease associations with confidence scoring.*	http://www.rnanut.net/lncrnadisease/
circRNADisease	*High-quality experimentally supported circRNA–disease relationships across multiple species.*	http://cgga.org.cn/circRNADisease/
CircR2Disease	*Curated database of experimentally validated circRNA–disease associations.*	https://github.com/bioinforlab/CircR2Disease-v2.0
circad	*Manually curated database of circular RNAs associated with disease phenotypes.*	https://clingen.igib.res.in/circad/
Circ2Disease	*Curated human circRNA–disease associations with integrated miRNA and RBP data.*	https://bioinformatics.zju.edu.cn/Circ2Disease/

DisGeNET	*Broad gene–disease association database (includes RNA gene data).*	https://www.disgenet.org/
Expression Atlas	*Gene expression (RNA and protein) across tissues, conditions, including disease contexts.*	https://www.ebi.ac.uk/gxa/home

The full potential of RNA research is exposed when computational and experimental data are combined through integrative and systems-level approaches. Multi-omics integration connects transcriptomic data with genomic, epigenomic, proteomic, and metabolomic datasets in order to understand how biological components interact and function together as a system, and how mechanisms are affected under disease conditions, which is referred to as the field of systems biology (Hasin et al., 2017). For instance, co-expression networks and gene regulatory networks can map interactions among RNAs, proteins, and chromatin modifiers, shedding light on hierarchical gene regulation of the omics pyramid, where transcriptome act as a bridge between the information and the function. Machine learning and artificial intelligence further enhance RNA analysis by enabling the classification of ncRNA subtypes, prediction of disease associations, and identification of novel functional modules. Supervised and unsupervised learning models trained on sequence, structure, and expression features can predict lncRNA functions or prioritize miRNA–mRNA target pairs (Zeng et al., 2021). These databases, integrative frameworks, and computational tools collaborate in order to create a robust groundwork for deciphering the complexity of RNAome and how RNAs contribute cellular and organismal phenotypes.

Conclusion and Future Perspectives

This review strongly emphasizes the fact that RNAs are much more than simple intermediate molecules in protein synthesis as they construct a dynamic network of gene expression regulation, structural assembly, catalytic activity, and signal transduction. In addition to an overview of the major RNAs with their biogenesis, structure, and classical function, we discussed the specific properties and the emerging roles of non-coding RNAs in fine-tuning transcriptional regulation and maintaining cellular homeostasis. Additionally, the ancient evolutionary role of RNA molecules was described by the RNA World Hypothesis, which also forms

the conceptual basis for modern functional genomics and high-throughput sequencing studies. Currently, alongside experimental methods that reveal diverse RNA functions and structures, the development of comprehensive databases and advanced bioinformatics platforms with multi-omics data integration methods have revolutionized the concepts of molecular cell biology and clinical applications. Yet, there are plentiful non-coding RNAs whose functions are incognita. Looking forward, there is a necessity of more detailed mapping of RNA's structural and functional diversity, especially to clarify the roles of obscure non-coding RNAs. It is also anticipated that the previously identified limitations on the clinical applications of RNAs will be transcended soon with the rapid advancements in technology and computational science. We do not envision the routine use of RNA-based treatment strategies in clinical applications as a distant future prospect.

References

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., Verdine, V., Cox, D. B. T., Kellner, M. J., Regev, A., Lander, E. S., Voytas, D. F., Ting, A. Y., & Zhang, F. (2017). RNA targeting with CRISPR–Cas13. Nature 2017 550:7675, 550(7675), 280–284. https://doi.org/10.1038/nature24049

Ahmad, M., Weiswald, L. B., Poulain, L., Denoyelle, C., & Meryet-Figuiere, M. (2023). Involvement of lncRNAs in cancer cells migration, invasion and metastasis: cytoskeleton and ECM crosstalk. Journal of Experimental & Clinical Cancer Research 2023 42:1, 42(1), 1–14. https://doi.org/10.1186/S13046-023-02741-X

Ali Zaidi, S. S., Fatima, F., Ali Zaidi, S. A., Zhou, D., Deng, W., & Liu, S. (2023). Engineering siRNA therapeutics: challenges and strategies. In Journal of Nanobiotechnology (Vol. 21, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12951-023-02147-z

Alshaer, W., Zureigat, H., Al Karaki, A., Al-Kadash, A., Gharaibeh, L., Hatmal, M. M., Aljabali, A. A. A., & Awidi, A. (2021). siRNA: Mechanism of action, challenges, and therapeutic approaches. European Journal of Pharmacology, 905. https://doi.org/10.1016/j.ejphar.2021.174178

Angel, M., Gomez, R., & Ibba, M. (2020). Aminoacyl-tRNA synthetases. https://doi.org/10.1261/rna

Bajan, S., & Hutvagner, G. (2020). RNA-Based Therapeutics: From Antisense Oligonucleotides to miRNAs. Cells 2020, Vol. 9, Page 137, 9(1), 137. https://doi.org/10.3390/CELLS9010137

Bautista-Sánchez, D., Arriaga-Canon, C., Pedroza-Torres, A., De La Rosa-Velázquez, I. A., González-Barrios, R., Contreras-Espinosa, L., Montiel-Manríquez, R., Castro-Hernández, C., Fragoso-Ontiveros, V., Álvarez-Gómez, R. M., & Herrera, L. A. (2020). The Promising Role of miR-21 as a Cancer Biomarker and Its Importance in RNA-Based Therapeutics. Molecular Therapy. Nucleic Acids, 20, 409. https://doi.org/10.1016/J.OMTN.2020.03.003

Bernstein, E., & Allis, C. D. (2005). RNA meets chromatin. Genes and Development, 19(14), 1635–1655. https://doi.org/10.1101/GAD.1324305,

Brar, G. A., & Weissman, J. S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. Nature Reviews Molecular Cell Biology 2015 16:11, 16(11), 651–664. https://doi.org/10.1038/nrm4069

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA, 19(4), 429–442. https://doi.org/10.1261/RNA.037598.112

Brosius, J., & Raabe, C. A. (2016). What is an RNA? A top layer for RNA classification. RNA Biology, 13(2), 140–144.

Burnett, J. C., & Rossi, J. J. (2012). RNA-based therapeutics: Current progress and future prospects. Chemistry and Biology, 19(1), 60–71.

Cai, X., Rondeel, I., & Baumgartner, S. (2021). Modulating the bicoid gradient in space and time. Hereditas, 158(1). https://doi.org/10.1186/s41065-021-

00192-y

Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution - Trashing old rules to forge new ones. Cell, 157(1), 77–94.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics, 14(1), 1–14.

Choi, S. W., Kim, H. W., & Nam, J. W. (2019). The small peptide world in long noncoding RNAs. Briefings in Bioinformatics, 20(5), 1853–1864. https://doi.org/10.1093/BIB/BBY055

Cirronis, M., Schneemann, S., Pettie, J., Mannaioni, G., & Dear, J. W. (2024). Evaluation of capillary miR-122 as a prognostic biomarker of paracetamol-induced liver toxicity. Molecular Biology Reports, 51(1), 1–6.

Clemm von Hohenberg, K., Müller, S., Schleich, S., Meister, M., Bohlen, J., Hofmann, T. G., & Teleman, A. A. (2022). Cyclin B/CDK1 and Cyclin A/CDK2 phosphorylate DENR to promote mitotic protein translation and faithful cell division. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-28265-0

Cornec, A., & Poirier, E. Z. (2023). Interplay between RNA interference and transposable elements in mammals. In Frontiers in Immunology (Vol. 14). Frontiers Media SA. https://doi.org/10.3389/fimmu.2023.1212086

Crick, F. (1970). Central dogma of molecular biology. Nature, 227(5258), 561–563. https://doi.org/10.1038/227561A0;KWRD=SCIENCE

Cui, S., Yu, S., Huang, H. Y., Lin, Y. C. D., Huang, Y., Zhang, B., Xiao, J., Zuo, H., Wang, J., Li, Z., Li, G., Ma, J., Chen, B., Zhang, H., Fu, J., Wang, L., & Huang, H. Da. (2025). miRTarBase 2025: updates to the collection of experimentally validated microRNA–target interactions. Nucleic Acids Research, 53(D1), D147–D156. https://doi.org/10.1093/NAR/GKAE1072

D'Agostino, N., Li, W., & Wang, D. (2022). High-throughput transcriptomics. Scientific Reports, 12(1), 1–3.

Derks, K. W. J., & Pothof, J. (2015). RNAome sequencing delineates the complete RNA landscape. Genomics Data, 5, 381–384. https://doi.org/10.1016/J.GDATA.2015.07.002

Ding, X., Ding, J., Ning, J., Yi, F., Chen, J., Zhao, D., Zheng, J., Liang, Z., Hu, Z., & Du, Q. (2012). Circulating microRNA-122 as a potential biomarker for liver injury. Molecular Medicine Reports, 5(6), 1428–1432. https://doi.org/10.3892/MMR.2012.838/HTML

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., … Gingeras, T. R. (2012). Landscape of transcription in human cells. Nature, 489(7414), 101–108.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. https://doi.org/10.1093/

BIOINFORMATICS/BTS635

Doudna, J. A., & Cech, T. R. (2002). The chemical repertoire of natural ribozymes. Nature, 418(6894), 222–228. https://doi.org/10.1038/418222A;KWRD=SCIENCE

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., … Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 2012 489:7414, 489(7414), 57–74. https://doi.org/10.1038/nature11247

Earls, L. R., Westmoreland, J. J., & Zakharenko, S. S. (2014). Non-coding RNA regulation of synaptic plasticity and memory: implications for aging. Ageing Research Reviews, 0, 34. https://doi.org/10.1016/J.ARR.2014.03.004

Esteller, M. (2011). Non-coding RNAs in human disease. Nature Reviews Genetics 2011 12:12, 12(12), 861–874. https://doi.org/10.1038/nrg3074

Etheridge, A., Gomes, C. P. C., Pereira, R. W., Galas, D., & Wang, K. (2013). The complexity, function, and applications of RNA in circulation. Frontiers in Genetics, 4(JUN), 53745. https://doi.org/10.3389/FGENE.2013.00115/XML/NLM

Faghihi, M. A., & Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nature Reviews Molecular Cell Biology, 10(9), 637–643. https://doi.org/10.1038/NRM2738,

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. Nature, 391(6669), 806–811. https://doi.org/10.1038/35888;KWRD=SCIENCE

Garst, A. D., Edwards, A. L., & Batey, R. T. (2011). Riboswitches: Structures and Mechanisms. Cold Spring Harbor Perspectives in Biology, 3(6), a003533. https://doi.org/10.1101/CSHPERSPECT.A003533

Ge, H., Marchisio, M. A., Kato, Y., & Lou, C. (2021). Aptamers, Riboswitches, and Ribozymes in S. cerevisiae Synthetic Biology. Life 2021, Vol. 11, Page 248, 11(3), 248. https://doi.org/10.3390/LIFE11030248

George, T. P., Subramanian, S., & Supriya, M. H. (2024). A brief review of noncoding RNA. Egyptian Journal of Medical Human Genetics 2024 25:1, 25(1), 1–16. https://doi.org/10.1186/S43042-024-00553-Y

Gilbert, W. (1986). Origin of life: The RNA world. Nature, 319(6055), 618. https://doi.org/10.1038/319618A0;KWRD=SCIENCE

Gomes, R. M. O. da S., da Silva, K. J. G., & Theodoro, R. C. (2024). Group I introns: Structure, splicing and their applications in medical mycology. Genetics and Molecular Biology, 47, e20230228. https://doi.org/10.1590/1678-4685-GMB-2023-0228

Goodall, G. J., & Wickramasinghe, V. O. (2020). RNA in cancer. Nature Reviews Cancer 2020 21:1, 21(1), 22–36. https://doi.org/10.1038/s41568-020-00306-0

Gray, M. W., & Gopalan, V. (2020). Piece by piece: Building a ribozyme. Journal of Biological Chemistry, 295(8), 2313–2323. https://doi.org/10.1074/jbc.REV119.009929

Grützmann, K., Salomo, K., Krüger, A., Lohse-Fischer, A., Erdmann, K., Seifert, M., Baretton, G., Aust, D., William, D., Schröck, E., Thomas, C., & Füssel, S. (2024). Identification of novel snoRNA-based biomarkers for clear cell renal cell carcinoma from urine-derived extracellular vesicles. Biology Direct, 19(1), 1–15.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell, 35(3 PART 2), 849–857. https://doi.org/10.1016/0092-8674(83)90117-4

Gupta, S., Jani, J., Vijayasurya, Mochi, J., Tabasum, S., Sabarwal, A., & Pappachan, A. (2023). Aminoacyl-tRNA synthetase – a molecular multitasker. In FASEB Journal (Vol. 37, Issue 11). John Wiley and Sons Inc. https://doi.org/10.1096/fj.202202024RR

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., & Zavolan, M. (2021). CLIP and complementary methods. Nature Reviews Methods Primers, 1(1), 1–23.

Hamilton, A. J., & Baulcombe, D. C. (1999). A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. In New Series (Vol. 286, Issue 5441).

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. Genome Biology 2017 18:1, 18(1), 1–15. https://doi.org/10.1186/S13059-017-1215-1

Hegde, A. N., & Smith, S. G. (2019). Recent developments in transcriptional and translational regulation underlying long-term synaptic plasticity and memory. Learning & Memory, 26(9), 307. https://doi.org/10.1101/LM.048769.118

Hirakata, S., & Siomi, M. C. (2016). piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. In Biochimica et Biophysica Acta - Gene Regulatory Mechanisms (Vol. 1859, Issue 1, pp. 82–92). Elsevier. https://doi.org/10.1016/j.bbagrm.2015.09.002

Hoffer, E. D., Hong, S., Sunita, S., Maehigashi, T., Gonzalez Jnr, R. L., Whitford, P. C., & Dunham, C. M. (2020). Structural insights into mRNA reading frame regulation by tRNA modification and slippery codon–anticodon pairing. ELife, 9, 1–20. https://doi.org/10.7554/eLife.51898

Hosseini, K., Ranjbar, M., Pirpour Tazehkand, A., Asgharian, P., Montazersaheb, S., Tarhriz, V., & Ghasemnejad, T. (2022). Evaluation of exosomal non-coding RNAs in cancer using high-throughput sequencing. Journal of Translational Medicine, 20(1), 30. https://doi.org/10.1186/S12967-022-03231-Y

Hrdlickova, R., Toloue, M., & Tian, B. (2016). RNA-Seq methods for transcriptome analysis. Wiley Interdisciplinary Reviews. RNA, 8(1), 10.1002/wrna.1364. https://doi.org/10.1002/WRNA.1364

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57. https://doi.org/10.1038/NPROT.2008.211;KWRD=LIFE+SCIENCES

Hung, Y. H., & Slotkin, R. K. (2021). The initiation of RNA interference (RNAi) in plants. In Current Opinion in Plant Biology (Vol. 61). Elsevier Ltd. https://doi.org/10.1016/j.pbi.2021.102014

Hwang, J. Y., Jung, S., Kook, T. L., Rouchka, E. C., Bok, J., & Park, J. W. (2020). rMAPS2: an update of the RNA map analysis and plotting server for alternative splicing regulation. Nucleic Acids Research, 48(W1), W300–W306. https://doi.org/10.1093/NAR/GKAA237

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science (New York, N.Y.), 324(5924), 218. https://doi.org/10.1126/SCIENCE.1168978

Jacobs, D. I., Qin, Q., Fu, A., Chen, Z., Zhou, J., Zhu, Y., Jacobs, D. I., Qin, Q., Fu, A., Chen, Z., Zhou, J., & Zhu, Y. (2018). piRNA-8041 is downregulated in human glioblastoma and suppresses tumor growth in vitro and in vivo. Oncotarget, 9(102), 37616–37626. https://doi.org/10.18632/ONCOTARGET.26331

Kan, C. M., Pei, X. M., Yeung, M. H. Y., Jin, N., Ng, S. S. M., Tsang, H. F., Cho, W. C. S., Yim, A. K. Y., Yu, A. C. S., & Wong, S. C. C. (2023). Exploring the Role of Circulating Cell-Free RNA in the Development of Colorectal Cancer. International Journal of Molecular Sciences, 24(13). https://doi.org/10.3390/IJMS241311026,

Kazanskii, M. A., Uroshlev, L., Zatylkin, F., Pospelova, I., Kantidze, O., & Gankin, Y. (2024). RNAfold: RNA tertiary structure prediction using variational autoencoder. BioRxiv, 2024.06.18.599511. https://doi.org/10.1101/2024.06.18.599511

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology, 37(8), 907–915.

Kirchner, S., & Ignatova, Z. (2015). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. In Nature Reviews Genetics (Vol. 16, Issue 2, pp. 98–112). Nature Publishing Group. https://doi.org/10.1038/nrg3861

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell, 31(1), 147–157. https://doi.org/10.1016/0092-8674(82)90414-7

Kugler, J. M., & Lasko, P. (2009). Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during drosophila oogenesis. In Fly (Vol. 3, Issue 1, pp. 15–28). Taylor and Francis Inc. https://doi.org/10.4161/fly.3.1.7751

Lehmann, K., & Schmidt, U. (2003). Group II introns: Structure and catalytic

versatility of large natural ribozymes. Critical Reviews in Biochemistry and Molecular Biology, 38(3), 249–303.

Li, B., Qu, L., & Yang, J. (2023). RNA-Guided RNA Modifications: Biogenesis, Functions, and Applications. Accounts of Chemical Research, 56(22), 3198–3210.

Li, J. H., Liu, S., Zhou, H., Qu, L. H., & Yang, J. H. (2013). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Research, 42(Database issue), D92. https://doi.org/10.1093/NAR/GKT1248

Li, X., Wang, H., Chen, Y., Zhang, Y., Liu, J., Zhou, D., Wu, Z., & Sun, M. (2023). Upstream ribosome impediments activate roles of internal Shine-Dalgarno sequence for translation initiation in E. coli. https://doi.org/10.1101/2023.07.05.547755

Li, Y., & Sun, S. (2025). RNA dysregulation in neurodegenerative diseases. The EMBO Journal, 44(3), 613–638.

Li, Y., Yu, Z., Jiang, W., Lyu, X., Guo, A., Sun, X., Yang, Y., & Zhang, Y. (2024). tRNA and tsRNA: From Heterogeneity to Multifaceted Regulators. In Biomolecules (Vol. 14, Issue 10). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/biom14101340

Liao, J., Yu, L., Mei, Y., Guarnera, M., Shen, J., Li, R., Liu, Z., & Jiang, F. (2010). Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Molecular Cancer, 9(1), 1–10. https://doi.org/10.1186/1476-4598-9-198

Liu, E. Y., Cali, C. P., & Lee, E. B. (2017). RNA metabolism in neurodegenerative disease. Disease Models & Mechanisms, 10(5), 509–518. https://doi.org/10.1242/DMM.028613

Liu, H.-Y., Zhang, Y.-Y., Zhu, B.-L., Feng, F.-Z., Yan, H., Zhang, H.-Y., & Zhou, B. (2019). miR-21 regulates the proliferation and apoptosis of ovarian cancer cells through PTEN/PI3K/AKT.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., & Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proceedings of the National Academy of Sciences of the United States of America, 101(19), 7287–7292.

Mazan-Mamczarz, K., & Gartenhaus, R. B. (2013). Role of microRNA deregulation in the pathogenesis of diffuse large B-cell lymphoma (DLBCL). Leukemia Research, 37(11), 10.1016/j.leukres.2013.08.020.

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: Insights into functions. Nature Reviews Genetics, 10(3), 155–159. https://doi.org/10.1038/NRG2521,

Metzenmacher, M., Váraljai, R., Hegedüs, B., Cima, I., Forster, J., Schramm, A.,

Scheffler, B., Horn, P. A., Klein, C. A., Szarvas, T., Reis, H., Bielefeld, N., Roesch, A., Aigner, C., Kunzmann, V., Wiesweg, M., Siveke, J. T., Schuler, M., & Lueong, S. S. (2020). Plasma Next Generation Sequencing and Droplet Digital-qPCR-Based Quantification of Circulating Cell-Free RNA for Noninvasive Early Detection of Cancer. Cancers 2020, Vol. 12, Page 353, 12(2), 353. https://doi.org/10.3390/CANCERS12020353

Monachello, D., Lauraine, M., Gillot, S., Michel, F., & Costa, M. (2021). A new RNA–DNA interaction required for integration of group II intron retrotransposons into DNA targets. Nucleic Acids Research, 49(21), 12394–12410. https://doi.org/10.1093/NAR/GKAB1031

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A. R., Hoberg, E., Moustakas, A., Gyllensten, U., Jones, S. J. M., Gustafsson, C. M., Sims, A. H., Westerlund, F., Gorab, E., & Kanduri, C. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA–DNA triplex structures. Nature Communications 2015 6:1, 6(1), 1–17. https://doi.org/10.1038/ncomms8743

Morais, P., Adachi, H., & Yu, Y. T. (2021). Spliceosomal snRNA Epitranscriptomics. Frontiers in Genetics, 12, 652129. https://doi.org/10.3389/FGENE.2021.652129

Muñoz-Velasco, I., Cruz-González, A., Hernández-Morales, R., Campillo-Balderas, J. A., Cottom-Salas, W., Jácome, R., & Vázquez-Salazar, A. (2024). Pioneering role of RNA in the early evolution of life. Genetics and Molecular Biology, 47(Suppl 1), e20240028. https://doi.org/10.1590/1678-4685-GMB-2024-0028

Navarro, E., Mallén, A., & Hueso, M. (2021). Dynamic Variations of 3′UTR Length Reprogram the mRNA Regulatory Landscape. Biomedicines 2021, Vol. 9, Page 1560, 9(11), 1560. https://doi.org/10.3390/BIOMEDICINES9111560

Nucleic Acids Research | Oxford Academic. (n.d.). Retrieved October 6, 2025, from https://academic.oup.com/nar/

Otmani, K., & Lewalle, P. (2021). Tumor Suppressor miRNA in Cancer Cells and the Tumor Microenvironment: Mechanism of Deregulation and Clinical Implications. Frontiers in Oncology, 11, 708765. https://doi.org/10.3389/FONC.2021.708765/XML/NLM

Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D., & Zamore, P. D. (2019). PIWI-interacting RNAs: small RNAs with big functions. In Nature Reviews Genetics (Vol. 20, Issue 2, pp. 89–108). Nature Publishing Group. https://doi.org/10.1038/s41576-018-0073-3

Palazzo, A. F., & Lee, E. S. (2015). Non-coding RNA: What is functional and what is junk? Frontiers in Genetics, 5(JAN), 127231. https://doi.org/10.3389/FGENE.2015.00002

Panwar, B., Arora, A., & Raghava, G. P. S. (2014). Prediction and classification of ncRNAs using structural information. BMC Genomics, 15(1), 1–13. https://doi.org/10.1186/1471-2164-15-127

Park, S. V., Yang, J. S., Jo, H., Kang, B., Oh, S. S., & Jung, G. Y. (2019). Catalytic RNA, ribozyme, and its applications in synthetic biology.

Biotechnology Advances, 37(8), 107452. https://doi.org/10.1016/J.BIOTECHADV.2019.107452

Prathiba, J., & Malathi, R. (2008). Group I introns and gnra tetraloops: Remnants of "The RNA world"? Molecular Biology Reports, 35(2), 239–249. https://doi.org/10.1007/S11033-007-9076-4/

R Core Team. (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ren, A., Micura, R., & Patel, D. J. (2017). Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. Current Opinion in Chemical Biology, 41, 71–83. https://doi.org/10.1016/J.CBPA.2017.09.017

Reynolds, J. A. (2024). MicroRNAs in the developmental toolbox — a comparative approach to understanding their role in regulating insect development. In Current Opinion in Insect Science (Vol. 66). Elsevier Inc. https://doi.org/10.1016/j.cois.2024.101256

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., & Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Non-Coding RNAs. Cell, 129(7), 1311. https://doi.org/10.1016/J.CELL.2007.05.022

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43(7), e47–e47. https://doi.org/10.1093/NAR/GKV007

Robinton, D. A., Chal, J., Lummertz da Rocha, E., Han, A., Yermalovich, A. V., Oginuma, M., Schlaeger, T. M., Sousa, P., Rodriguez, A., Urbach, A., Pourquié, O., & Daley, G. Q. (2019). The Lin28/let-7 Pathway Regulates the Mammalian Caudal Body Axis Elongation Program. Developmental Cell, 48(3), 396-405.e3. https://doi.org/10.1016/j.devcel.2018.12.016

Romão, L. (2022). mRNA Metabolism in Health and Disease. In Biomedicines (Vol. 10, Issue 9). MDPI. https://doi.org/10.3390/biomedicines10092262

Safra, M., Nir, R., Farouq, D., Slutzkin, I. V., & Schwartz, S. (2017). TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. Genome Research, 27(3), 393–406. https://doi.org/10.1101/GR.207613.116/-/DC1

Schmidt, C. M., & Smolke, C. D. (2019). RNA Switches for Synthetic Biology. Cold Spring Harbor Perspectives in Biology, 11(1), a032532. https://doi.org/10.1101/CSHPERSPECT.A032532

Scott, M. S., & Ono, M. (2011). From snoRNA to miRNA: Dual function regulatory non-coding RNAs. Biochimie, 93(11), 1987–1992. https://doi.org/10.1016/J.BIOCHI.2011.05.026

Scott, W. G. (2007). Ribozymes. Current Opinion in Structural Biology, 17(3), 280–286. https://doi.org/10.1016/J.SBI.2007.05.003

Shaukat, A. N., Kaliatsi, E. G., Skeparnias, I., & Stathopoulos, C. (2021). The

Dynamic Network of RNP RNase P Subunits. International Journal of Molecular Sciences 2021, Vol. 22, Page 10307, 22(19), 10307. https://doi.org/10.3390/IJMS221910307

Sohail, A. M., Khawar, M. B., Afzal, A., Hassan, A., Shahzaman, S., & Ali, A. (2022). Multifaceted roles of extracellular RNAs in different diseases. Military Medical Research 2022 9:1, 9(1), 1–9. https://doi.org/10.1186/S40779-022-00405-Z

Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., & Chang, H. Y. (2013). RNA SHAPE analysis in living cells. Nature Chemical Biology, 9(1), 18–20.

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. Nature Reviews Genetics 2019 20:11, 20(11), 631–656. https://doi.org/10.1038/s41576-019-0150-2

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2020). Gene regulation by long non-coding RNAs and its biological functions. Nature Reviews Molecular Cell Biology 2020 22:2, 22(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Stork, C., & Zheng, S. (2016). Genome-Wide Profiling of RNA–Protein Interactions Using CLIP-Seq. Methods in Molecular Biology (Clifton, N.J.), 1421, 137. https://doi.org/10.1007/978-1-4939-3591-8_12

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102(43), 15545–15550.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods, 6(5), 377–382.

Tanner, N. K. (1999). Ribozymes: the characteristics and properties of catalytic RNAs. FEMS Microbiology Reviews, 23(3), 257–275. https://doi.org/10.1111/j.1574-6976.1999.tb00399.x

Tinoco, I., & Bustamante, C. (1999). How RNA folds. Journal of Molecular Biology, 293(2), 271–281. https://doi.org/10.1006/JMBI.1999.3001

Van Nostrand, E. L., Pratt, G. A., Yee, B. A., Wheeler, E. C., Blue, S. M., Mueller, J., Park, S. S., Garcia, K. E., Gelboin-Burkhart, C., Nguyen, T. B., Rabano, I., Stanton, R., Sundararaman, B., Wang, R., Fu, X. D., Graveley, B. R., & Yeo, G. W. (2020). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. Genome Biology, 21(1), 1–26.

Wang, K. C., & Chang, H. Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. Molecular Cell, 43(6), 904–914.

Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs:

Functional surprises from the RNA world. Genes and Development, 23(13), 1494–1504. https://doi.org/10.1101/GAD.1800909,

Xiao, H., Feng, X., Liu, M., Gong, H., & Zhou, X. (2023). SnoRNA and lncSNHG: Advances of nucleolar small RNA host gene transcripts in anti-tumor immunity. Frontiers in Immunology, 14, 1143980. https://doi.org/10.3389/FIMMU.2023.1143980/XML/NLM

Xu, Y., Le, H., Wu, Q., Wang, N., & Gong, C. (2025). Advancements in CRISPR/Cas systems for disease treatment. Acta Pharmaceutica Sinica B. https://doi.org/10.1016/J.APSB.2025.05.007

Yao, R. W., Wang, Y., & Chen, L. L. (2019). Cellular functions of long noncoding RNAs. Nature Cell Biology, 21(5), 542–551.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. OMICS A Journal of Integrative Biology, 16(5), 284–287.

Yusupova, G., & Yusupov, M. (2021). A Path to the Atomic-Resolution Structures of Prokaryotic and Eukaryotic Ribosomes. In Biochemistry (Moscow) (Vol. 86, Issue 8, pp. 926–941). Pleiades journals. https://doi.org/10.1134/S0006297921080046

Zeng, M., Lu, C., Fei, Z., Wu, F. X., Li, Y., Wang, J., & Li, M. (2021). DMFLDA: A Deep Learning Framework for Predicting lncRNA-Disease Associations. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 18(6), 2353–2363. https://doi.org/10.1109/TCBB.2020.2983958

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology, 4(1).

Zhang, Y., Jia, Y., Zheng, R., Guo, Y., Wang, Y., Guo, H., Fei, M., & Sun, S. (2010). Plasma MicroRNA-122 as a Biomarker for Viral-, Alcohol-, and Chemical-Related Hepatic Diseases. Clinical Chemistry, 56(12), 1830–1838. https://doi.org/10.1373/CLINCHEM.2010.147850

Zhao, L., Wang, J., Li, Y., Song, T., Wu, Y., Fang, S., Bu, D., Li, H., Sun, L., Pei, D., Zheng, Y., Huang, J., Xu, M., Chen, R., Zhao, Y., & He, S. (2021). NONCODEV6: an updated database dedicated to long non-coding RNA annotation in both animals and plants. Nucleic Acids Research, 49(D1), D165–D171. https://doi.org/10.1093/NAR/GKAA1046

Zhong, P., Bai, L., Hong, M., Ouyang, J., Wang, R., Zhang, X., & Chen, P. (2024). A Comprehensive Review on Circulating cfRNA in Plasma: Implications for Disease Diagnosis and Beyond. Diagnostics 2024, Vol. 14, Page 1045, 14(10), 1045. https://doi.org/10.3390/DIAGNOSTICS14101045

Zhou, L.-Y., Qin, Z., Zhu, Y.-H., He, Z.-Y., & Xu, T. (2019). Current RNA-based Therapeutics in Clinical Trials. Current Gene Therapy, 19(3), 172–196.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., & Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Molecular Cell, 65(4), 631-643.e4.

Zimta, A. A., Tigu, A. B., Braicu, C., Stefan, C., Ionescu, C., & Berindan-Neagoe, I. (2020). An Emerging Class of Long Non-coding RNA With Oncogenic Role Arises From the snoRNA Host Genes. Frontiers in Oncology, 10, 516945.

Zubradt, M., Gupta, P., Persad, S., Lambowitz, A. M., Weissman, J. S., & Rouskin, S. (2016). DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nature Methods, 14(1), 75–82.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31(13), 3406. https://doi.org/10.1093/NAR/GKG595