**NETWORK ANALYSIS FOR IDENTIFYING POTENTIAL BIOMARKERS SUITABLE FOR MRD ASSESSMENT**

**Author:** Kinga András

1. Omixon Biocomputing Kft, Budapest, Hungary

**Introduction**

Despite improvements in the treatment of hematologic malignancies, in some cases, patients who have achieved complete remission or complete response (CR) still experience relapse. Conventional morphologic detection for hematologic malignancies has a threshold limit of one tumor cell in 10^6 cells. Technology exists that can detect the presence of myeloid malignancies in lower numbers than the limit of conventional morphologic detection, a level of disease burden known as minimal residual disease (MRD). Those technologies measure cell characteristics, such as genetic mutations, cell surface markers, or specific DNA gene rearrangements. In recent years, the number of biomarkers has increased and there are several already known specific hematologic markers, but the list is still incomplete.

The clinical usefulness of MRD detection depends on the biomarker used. An ideal MRD marker can distinguish between cells that would not cause relapse and those that hold the potential to cause relapse from the smallest clinically significant population of leukemic cells. As we know MRD markers and to identify new ones is one of the most common adult hematologic malignancy, acute myeloid leukemia (AML), we used a network analysis approach.

This work utilizes the interactome-based approach to human disease [2]. The interactome, i.e., the integrated network of all cellular interactions with the cell, can be interpreted as a map and disease as localizations. Our proposed network-based approach aims to identify the specific interactome neighborhoods of biomarkers or communities that are present in AML, to be validated from public databases and using a set of highly specific AML and MRD associated gene (seed genes) and a module detection algorithm called DIAMoND [3]. In order to search for new prospective biomarkers, which could be also potentially valuable for MRD monitoring.

**Methods**

**Interactome construction**

The first step of this work was a data-driven network construction of the so called interactome with Python scripts using experimental biological data from the Molecular Signatures Database (MSigDB) [4]. The MSigDB collection includes known gene interactions from various sources such as online protein databases, publications in PubMed, and knowledge of domain experts. The REGG pathway collection, the immunologic and the oncogenic collection was used for interactome construction. The interactome integrates gene interactions, nodes representing genes and they are connected with edges if they are involved in the same molecular pathway or process.

**AML associated seed gene collection**

In this work we chose AML for testing our network-based approach for MRD marker identification. The molecular heterogeneity of AML poses major challenges to find MRD biomarkers [5]. No single prevalent mutation is present in all or even in the majority of patients. The AML associated seed genes were collected from the scientific literature, including key molecular markers with implications for clinical practice (both class I and II mutations) affecting prognosis and are standards for risk categorization. However, many more genes are likely to contribute to leukemia pathogenesis as well as potentially inform optimal therapeutic. The final AML seed gene list was also selected also based on the AML-1 subtypes – Vh, ET, TET2, NPM1, CALR, IDH1, IDH2, TET2, ASXL1, RUNX1, TET2, KIT, WT1 [6-9] (see Figure 1). For module detection the DIAMoND algorithm [3] was used. DIAMoND introduces a new approach of community detection in the human interactome. The first step of this work was a data-driven network construction of the so called interactome with Python scripts using experimental biological data from the Molecular Signatures Database (MSigDB) [4]. The MSigDB collection includes known gene interactions from various sources such as online protein databases, publications in PubMed, and knowledge of domain experts. The REGG pathway collection, the immunologic and the oncogenic collection was used for interactome construction. The interactome integrates gene interactions, nodes representing genes and they are connected with edges if they are involved in the same molecular pathway or process.

**AML module detection within the interactome**

According to the literature, disease associated genes tend to cluster within so-called disease modules. Such disease modules are connected subgraphs within the interactome that contain all molecular determinants of a certain disease. To unfold the biological mechanisms of a disease in a network based framework is therefore to identify the respective disease modules.

**Results**

We were able to construct an interactome with 20,693 genes and more than 56 million interactions (edges) between them, with high average degree and clustering coefficient, characterized by a relatively high density of edges. Biological networks are known to have the small-world property, characterized by a relatively high density of edges. Biological networks are known to have the small-world property, characterized by a relatively high density of edges. Biological networks are known to have the small-world property, characterized by a relatively high density of edges. Biological networks are known to have the small-world property, characterized by a relatively high density of edges.

**Discussion**

We were able to identify several candidate markers, but their clinical relevance and molecular role in AML pathophysiology need to be evaluated one by one. The highest ranked added gene was the humanized hydrosoluble (HIV) involved in the regulation of stem cell metabolism, essential for tissue functions and tumor suppression. According to the literature, the TGF gene plays a role in the mesenchymal-to-epithelial transition (MET), which results in the shedding of epithelial cells from the tumor, which is necessary for the progression of carcinogenesis in progression. This gene is known to be involved in the pathogenesis of acute myeloid leukemia (AML), which is characterized by the presence of leukemic cells in the bone marrow. The second ranked added gene was the GST gene, a functional polymorphism encoding the glutathione S-transferase (GST) reactivating enzyme. GSTs deconjugate intracellular drug metabolites, including metabolites of several chemotherapeutic agents, some of which are known to increase the risk of acute leukemia. The third ranked added gene was the TGF gene involved in the regulation of cell adhesion and cell-cell interactions, which is known to be involved in the pathogenesis of acute myeloid leukemia (AML), which is characterized by the presence of leukemic cells in the bone marrow.

The results show that community detection analysis and a network-based approach can be an efficient strategy for biomarker detection and to explore the underlying connectivity patterns of disease. It can be used as a first step in the targeted search for MRD biomarkers, but it requires further investigations and refinements of methods and data selection.

**References**


