NETWORK ANALYSIS FOR IDENTIFYING POTENTIAL BIOMARKERS SUITABLE FOR MRD ASSESSMENT

Author: Kinga András[1]

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) will still experience relapse. Conventional morphologic detection for hematologic malignancies has a threshold limit of one tumor cell in 100 cells. Technology exists that can detect the presence of malignancy in orders of magnitude below the limit of conventional morphologic detection, a level of disease burden known as minimal residual disease (MRD). These 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 leukemic markers, but the list is still incomplete [1].

The clinical usefulness of MRD detection depends on the biomarker used. An ideal MRD marker helps discriminate between cells that would not cause relapse and those that hold the potential to cause relapse from the smallest clinically significant populations of leukemic cells. To assess known MRD markers and to identify new ones in one of the most common adult hematological 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 physical interactions within the cell, can be interpreted as a map and diseases as local perturbations. Our proposed network-based approach aims to identify the specific interactome neighborhood (called modules or communities) that is perturbed in AML, by constructing an interactome from public databases and using a set of highly specific AML and MRD associated genes (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 pathway databases, publications in PubMed, and knowledge of domain experts. The KEGG 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 Class II mutations) affecting prognosis and are standards for risk categorization. However many more genes are likely to contribute to leukemia pathogenesis as well as to potentially inform optimal therapeutics the final AML seed gene list was selected also based on the top genes mutated at frequencies > 5%: FLT3, NPM1, DNMT3A, IDH1, IDH2, TET2, KMT2A (AML1), TP53, NRAS, KRAS, ASXL1, GATA2, CEBPA, KIT, RUNX1 and WT1 [6-9] (see Figure 1.).

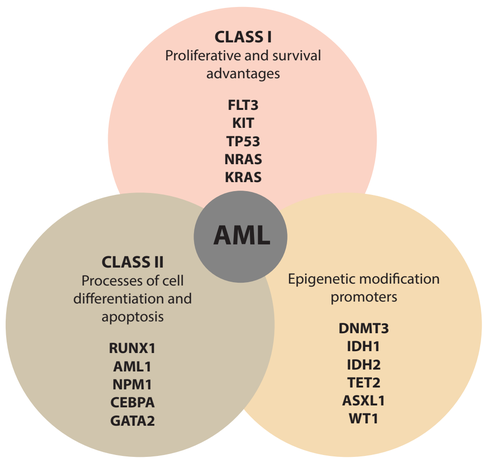

FIGURE 1 - AML associated seed gene collection classified based on key oncogenic events according to the two hits model [10]. Class I mutations confer proliferative and survival advantages, while Class II mutations affect the processes of cell differentiation and apoptosis. Mutations that do not conform to any of the two classes have not been classified; however, these mainly promote epigenetic modifications [5].

AML module detection within the interactome

According to the hypothesis, 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 module.

There are many different algorithms for module detection. Most of them use a basal network (interactome) and a set of genes/nodes of interest (already known disease associated genes), called the 'seeds' to generate a subnetwork possibly containing most of the seed genes and the added genes that are required in order to connect the seeds. The algorithm determines which node will be added to the subnetwork.

For module detection the DIAMoND algorithm [3] was used. DIAMoND introduces a new approach of community detection in the human interactome. Other community detection methods (weighted Steiner tree, Louvain, Link clustering, MCL) are able to find topologically dense communities, but when examining the number of already known disease associated genes in them, it shows that they only contain only a small fraction of these genes. The reason behind it is that disease associated genes probably do not exclusively form a locally dense community. In the DIAMOnD algorithm authors focused on this assumption and instead of only identifying dense communities they used a so-called connectivity significance as the most predictive quantity which was able to detect disease modules more precisely.

By exploring the local network neighborhood around known AML-associated seed genes we inferred likely new AML-associated gene candidates, ranking them according to their number of links to seed genes, these will be called added genes. The steps needed to be executed in order to achieve our goal can be seen in Figure 2.

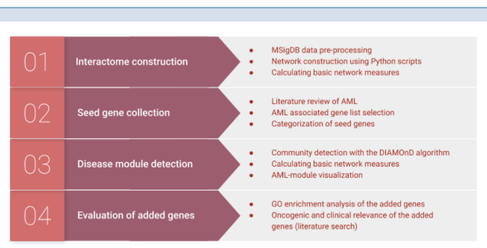

FIGURE 2 - Steps of the workflow.

Results

We were able to construct an interactome with 20.693 genes and more than 56 million interactions (edges) between them, with a high average degree and clustering coefficient, characterised by a relatively high density of edges. Biological networks are known to have the small-world property, characterized by a small diameter [11], as seen in the constructed interactome.

With the DIAMOnD algorithm we were able to identify several candidate genes with clinical or biological relevance to AML around a set of already known AML associated genes. More importantly, the DIAMOnD algorithm also identified highly reproducible marker genes and enriched functions from the interactome. The algorithm returned 100 genes that are added to the AML module, from which we selected the first 30 highest ranked genes for visualization (Figure 3.), and verified the first three genes if the literature confirms their AML associations.

The added genes with the 16 AML seed genes formed an also highly interconnected subnetwork within the interactome with 46 genes and 871 interactions between. The subnetwork measures show small-world properties as well: with the average degree of 37, RUNX1 and JAK1 acting as hubs with the highest degrees. The average clustering coefficient of the module is 0.894 and has a small diameter.

Based on GO enrichment analysis all 30 added genes are enriched in the Wnt signaling pathway and are enriched in several molecular functions (Figure 4.).

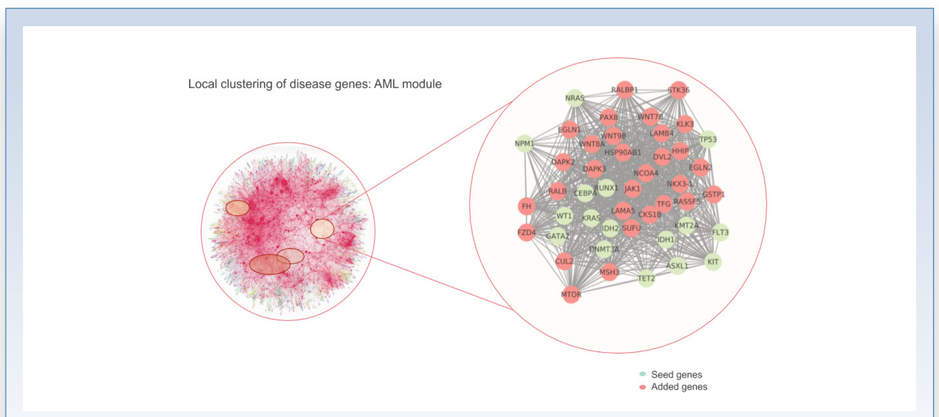

FIGURE 3 - Schematic view of disease modules within the interactome and the identified AML module using the DIAMOnD community detection algorithm. Molecular components that form a topological module are functionally closely related, thus they correspond to a function module, and a disease is result of the breakdown of this functional module [3].

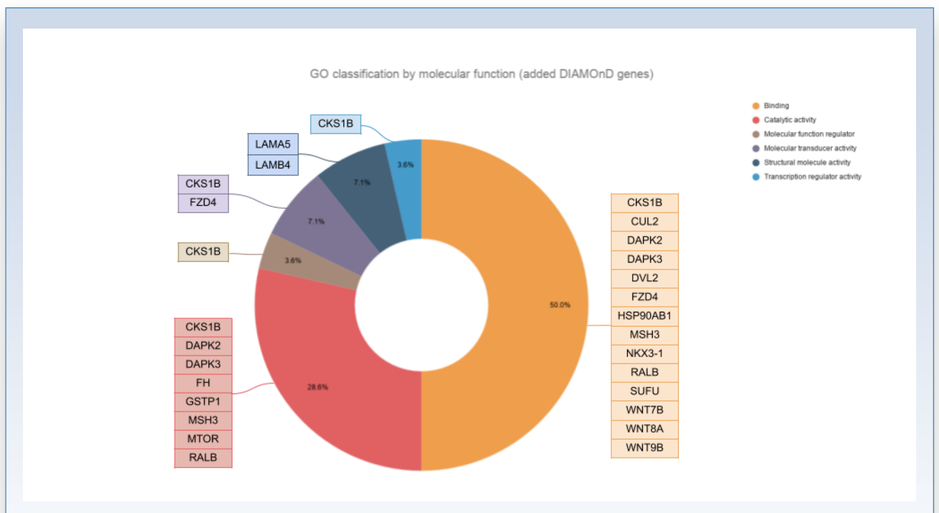

FIGURE 4 - GO enrichment analysis of the added AML candidate genes by the DIAMOnD algorithm. The genes in the chart showed enrichment in six molecular function categories. Genes are classified according to their function in several different ways using the PANTHER (protein annotation through evolutionary relationship) classification system [12].

Discussion

We were able to identify several candidate markers, but their clinical relevance and molecular roles in AML pathophenotype need to be evaluated one by one.

The highest ranked added gene was the fumarate hydratase gene (FH) involved in the regulation of stem cell metabolism, essential for tissue functions and tumor suppression. According to the literature, the FH gene plays a key role in the mitochondrial tricarboxylic acid (TCA) cycle and cytosolic fumarate metabolism, in normal and leukemic hematopoiesis [13,14].

The second candidate gene was the GSTP1 gene, a functional polymorphism encoding the glutathione S-transferases (GSTs) metabolizing enzyme. GSTs detoxify potentially mutagenic and toxic DNA-reactive electrophiles, including metabolites of several chemotherapeutic agents, some of which are suspected human carcinogens. Evidence shows that polymorphisms in genes that encode GSTs alter susceptibility to chemotherapy-induced carcinogenesis, specifically to therapy-related acute myeloid leukemia [15-18].

The third ranked added gene was the TFG gene encoding a protein that plays a role in the normal dynamic function of the endoplasmic reticulum and its associated microtubules. TFG is involved in the secretory pathway, that can be actively involved in sensing stress stimuli and possibly even initiating and propagating cell death signaling. The TFG gene has been observed to form fusion partners with several other genes (RARA,TEK, ALK and NTRK1) and it was reported in patients with leukemia and lymphoma. The TFG-RARA fusion gene confers sensitivity to ATRA treatment, and shows that it can also be used for minimal residual disease (MRD) monitoring using RT-PCR [19,20].

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 investigation and refinements of methods and data selection.

References

1. US Food and Drug Administration. (2019). Hematologic malignancies: regulatory considerations for use of minimal residual disease in development of drug and biological product for treatment.
2. Caldera, M., Buphamalai, P., Müller, F., & Menche, J. (2017). Interactome-based approaches to human disease. Current Opinion in Systems Biology, 3, 88-94.
3. Ghiassian, S. D., Menche, J., & Barabási, A. L. (2015). A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. PLoS computational biology, 11(4).
4. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics, 27(12), 1739-1740.
5. Lagunas-Rangel, F. A., Chávez-Valencia, V., Gómez-Guijosa, M. Á., & Cortes-Penagos, C. (2017). Acute myeloid leukemia—genetic alterations and their clinical prognosis. International journal of hematology-oncology and stem cell research, 11(4), 328.
6. Gruszka, A. M., Valli, D., & Alcalay, M. (2017). Understanding the molecular basis of acute myeloid leukemias: where are we now?. International journal of hematologic oncology, 6(2), 43-53.
7. DiNardo, C. D., & Cortes, J. E. (2016). Mutations in AML: prognostic and therapeutic implications. Hematology 2014, the American Society of Hematology Education Program Book, 2016(1), 348-355.
8. Ilyas, A. M., Ahmad, S., Faheem, M., Naseer, M. I., Kumosani, T. A., Al-Qahtani, M. H., ... & Ahmed, F. (2015). Next generation sequencing of acute myeloid leukemia: influencing prognosis. BMC genomics, 16(S1), S5.
9. Roloff, G. W., & Griffiths, E. A. (2018). When to obtain genomic data in acute myeloid leukemia (AML) and which mutations matter. Hematology 2014, the American Society of Hematology Education Program Book, 2018(1), 35-44.
10. Kelly, L. M., & Gilliland, D. G. (2002). Genetics of myeloid leukemias. Annual review of genomics and human genetics, 3(1), 179-198.
11. Zhang, Z., & Zhang, J. (2009). A big world inside small-world networks. PloS one, 4(5).
12. Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nature protocols, 8(8), 1551.
13. Guitart, A. V., Panagopoulou, T. I., Villacreces, A., Vukovic, M., Sepulveda, C., Allen, L., ... & Sas, Z. (2017). Fumarate hydratase is a critical metabolic regulator of hematopoietic stem cell functions. Journal of Experimental Medicine, 214(3), 719-735.
14. Laukka, T., Mariani, C. J., Ihantola, T., Cao, J. Z., Hokkanen, J., Kaelin, W. G., ... & Koivunen, P. (2016). Fumarate and succinate regulate expression of hypoxia-inducible genes via TET enzymes. Journal of Biological Chemistry, 291(8), 4256-4265.
15. Allan, J. M., Wild, C. P., Rollinson, S., Willett, E. V., Moorman, A. V., Dovey, G. J., ... & Morgan, G. J. (2001). Polymorphism in glutathione S-transferase P1 is associated with susceptibility to chemotherapy-induced leukemia. Proceedings of the National Academy of Sciences, 98(20), 11592-11597.
16. Das, P., Shaik, A. P., & Bammidi, V. K. (2009). Meta-analysis study of glutathione-S-transferases (GSTM1, GSTP1, and GSTT1) gene polymorphisms and risk of acute myeloid leukemia. Leukemia & lymphoma, 50(8), 1345-1351.
17. Das, P., Shaik, A. P., & Bammidi, V. K. (2009). Meta-analysis study of glutathione-S-transferases (GSTM1, GSTP1, and GSTT1) gene polymorphisms and risk of acute myeloid leukemia. Leukemia & lymphoma, 50(8), 1345-1351.
18. Xi, Y. M., Shi, X. E., Zhang, H., Jia, M. F., Li, M., Li, P., ... & Yao, X. J. (2011). Relation of GSTP1 and CYP2E1 polymorphisms with susceptibility to acute leukemia. Zhongguo shi yan xue ye xue za zhi, 19(3), 589-593.
19. Wlodkowic, D., Skommer, J., McGuinness, D., Hillier, C., & Darzynkiewicz, Z. (2009). ER–Golgi network—A future target for anti-cancer therapy. Leukemia research, 33(11), 1440-1447.
20. Chong, M. L., Cheng, H., Xu, P., You, H., Wang, M., Wang, L., & Ho, H. H. (2018). TFG-RARA: A novel fusion gene in acute promyelocytic leukemia that is responsive to all-trans retinoic acid. Leukemia research, 74, 51.