The signalling networks regulation and exploration in different tumours

PhD thesis

Dr. Dezső Módos

Semmelweis University

Pathology Ph.D. School

Supervisor: Földvári-Nagy Lászlóné Katalin Lenti, Ph.D. college professor

Consultant: Tamás Korcsmáros, Ph.D., fellow

Official Opponents: Orsolya Kapuy, Ph.D., adjunct professor

Balázs Egyed, Ph.D., adjunct professor

Head of Examination board:

Tóth Sára, PhD, habilitated university docent

Members of Examination board:

Sipeki Szabolcs, MD, PhD, adjunct professor

Hegedűs Tamás, PhD, research fellow

Budapest

2017

Introduction

For the complex function of the human body myriad process need to work with optimal regulation. The tissues, organs and the smallest working component of living systems the cell, form an entangled, communicating, system.

The fate of a cell is determined by the numerous incoming signals and the cell can answer to them numerous ways depending on its internal state at the moment. To understand such a complex system, we need to look on not just on one particular biochemical reaction, but also to investigate the whole system, the signalling processing machinery of the cell. To understand the signalling systems we need collections, signalling databases which contain the signalling and the transcriptional, regulatory interactions.

From these databases, we can build the interactional model of signalling. One such simple model is the proteins and their interactions containing protein-protein interaction network.

To understand and change signalling we need to find the key points in the signalling networks. During my Ph.D. I have investigated the role of during evolution duplicated genes which are in the same species -a.k.a. paralogues in signalling. These genes can be affected in diseases especially in cancer.

If we can see the differences in specific diseases, then we can visualize their effects in the network, then we can now where and what have been changed in the signalling network. Most of the time in signalling our research focus on the changed proteins, but the signalling neighbourhood of

Aims

I have aimed during my Ph.D. the following:

1. Developing the SignLink 2.0 transcriptional and regulatory layers, which contain transcription factor-target gene and miRNA-target gene interactions
2. Developing an algorithm which can be used for TF-target gene prediction in other cellular processes like in autophagy and NRF2 stress regulation
3. Defining the critical paralogues in singling and analyse their network importance
4. Finding the cancer related signalling proteins and their first neighbours in four high prevalence cancers (colorectal carcinoma, breast cancer, non-small cell lung cancer, hepatocellular carcinoma) with the first neighbours illustrating cancer rewiring strategy
5. Showing the critical paralogues and the first neighbours importance in drug discovery

Methods

The SignaLink 2.0, the ARN and the NRF2Ome database regulatory layers

For the transcription factor target-gene layers, I downloaded the transcription-factor target gene position weighted matrixes from the JASPAR database. The transcription factor biding sites were searched in the genes and upstream or downstream form them 2000 base pairs. For details see the dissertation.

For the micro-RNA (miRNA) target gene layers the miRNAs were translated to MIRBASE ID from miR2Disease, doRiNA, DianaMicroT, miRDeathDB, miRanda, PicTar, TargetScan and TarBase databases. In the case of PuTmiR, Transmir and Encode TF-miRNA data we used similar protocol. The target genes were translated to UniProt IDs.

Finding critical paralogue groups within the signalling networks:

A critical paralogue group is a group of proteins, which

1. Have paralogues in signalling (evolutionary criterion)
2. At least one member goes through a tissue specific signalling path (tissue specific graph based criterion)
3. At least one member is involved in a cross-talk (signalling based biological criterion)

To fulfil this definition, I searched the paralogues of the SignaLink 2 signalling proteins in Inparanoid and in OrthoDB database. I considered a signalling path as a cascade from ligand to transcription factor. The tissue specificity was downloaded from ENSMEBL database, then I merged for

larger organ systems. The cross-talk information came from the SignaLink 2 database. The network parameters were measured by iGraph program. For the specificity analysis, I used the GenOntology, OMIM, ChEMBL, and the Cancer Gene Census (CGC) databases. For the continuous values statistical analysis, I used Kolmogorov-Smirnov and Wilcoxon rank sum test. I used Chi-squere test to test the independence of discrete variables.

Determination of cancer related first neighbours and their investigation

During my work, I searched in four solid cancers the mutated genes according to CGC and the differently expressed genes according to altogether 1557 Affymetrix HGU 133plus 2 microarrays. The microarray chips were normalized to each other by RMA method therefore they became comparable to each other. I projected these genes to different signalling networks (SingnaLink 2, Reactome, Cui *et al*) and protein interaction networks (HPRD, BIOGRID-DIP-INTACT) to find their first-neighbour interactor partners. Then, I measured different network parameters (degree, betweenness, clustering coefficient) by iGraph. For statistical tests, I used Kolmogorov Simonov and Wilcoxon rank sum tests. During my work, I also looked for potential novel drug targets between the first neighbours and their interaction partners. To do this I used the ChEMBL database. I considered a molecule bound to a protein if the biding affinity was lower than 10 μmol/l.

Results

The SignaLink 2, the ARN and the NRF2Ome database regulatory layers

The interactions which were integrated by me at each database are in Table 1. From the transcription-factor target gene data I mentioned only the predicted interactions, because those are which I were predicted the methods developed by myself.

Table 1 The regulatory interactions of each databases

	SignaLink 2.0	NRF2ome	ARN
Predicted TF-target gene interactions	23098	8217	2911
miRNA-target gen interactions	7381553	8157	325680
TF-miRNA interactions	5227	4982	2192

Finding critical paralogue groups within the signalling networks:

According to the definition mentioned in the methods I have found 75 critical paralogue groups and 265 critical proteins. My analysis showed, that these proteins equally important compared to the other signalling proteins ($p \leq 0,05$ Kolmogorov Smirnov test and Wilcoxon rank sum test), but more important compared to other proteins in paralogue groups. They have higher betweenness and degree compared to other paralogue proteins ($p \leq 0,05$ Kolmogorov Smirnov test and Wilcoxon rank sum test). The critical paralogues are more often involved in cancer and in other diseases compared to other paralogues according to CGC and OMIM database (Chi square test

$p<0.05$). The critical paralogues within a critical paralogue groups have more different functions compared to other paralogues in their groups according to Gene Ontology biological processes ($p< 0.05$ Kolmogorov Smirnov test, Wilcoxon rank sum test). Moreover, the critical prologues are more often drug targets compared to other paralogues (Chi square test $p<0.05$).

During my work, I could distinguish two kind of paralogue proteins in the human signalling. The first group is the group of critical paralogues, whose are distinguishable, meanwhile the other paralogues are interchangeable to each other

Determination of cancer related first neighbours and their investigation

During my work, I found in four cancers (colorectal carcinoma, breast cancer, hepatocellular carcinoma, non-small cell lung cancer) and in three signalling (SignaLink 2.0, Reactome, Cui *et al*) and two protein-protein interaction networks (HPRD, BioGRID-DIP-IntAct) independently that the first neighbours are at least as important as cancer related proteins themselves. They have similar or higher betweenness, degree and clustering coefficient ($p\leq 0,05$ Kolmogorov Smirnov test and Wilcoxon rank sum test).

The first neighbours form a more connected component with each other, then expected in all five networks of all four cancers (giant component percentage Z score analysis $p<0.001$).

Moreover, I have found that the differently expressed proteins first neighbours have higher centralities ($p\leq 0,05$ Kolmogorov Smirnov test and

Wilcoxon rank sum test), but the mutated ones have similar centralities ($p > 0,05$ Kolmogorov-Smirnov test). This means the cancers rewire the signalling by two different ways. In one hand mutation affects the central nodes of signalling directly in the other hand differential expression affects the central nodes through the first neighbours.

The drug target space increase significantly if we consider the first neighbours of cancer related proteins. There are 122 drugs and 3585 not druggable compounds targeting the 105 cancer related proteins. Proportionally significantly less drugs (233) and compounds (6699) target the 295 first neighbours (Bernoulli test $p<0.05$). It shows an opportunity to use the first neighbours to increase the drug target space.

Conclusion

During my Ph.D. work, I have successfully built the SignaLink 2, Autophagy Regulatory Network and NRF2ome regulatory layers. The above-mentioned databases became able to study regulatory circuits through my contribution. I am showing examples in the discussion of my dissertation.

From the three databases, I used the SignaLink 2 for signalling networks analysis. I have found and analysed the critical prologue groups and critical paralogues in the human signalling 7 most important signalling pathways. I used two complementary paralogue finding methods. My method even capable for finding unknown signalling proteins to categorize them to specific pathway.

After finding the paralogues I developed a workflow to determine which paralogue proteins are critical for signalling. I used for that whether a protein is involved in cross-talk and whether without a protein a tissue specific signalling path disappear. Only these two criteria were enough to find the critical paralogue groups. My method is capable to distinguish the druggable prologue groups, because the critical paralogues are involved in more different functions, and more diseases. My results showed the critical paralogues are not interchangeable within a paralogue group, meanwhile in other paralogue groups the paralogue proteins are.

My work about the cancer related first neighbours showed how can we increase the drug targets using the first neighbours of cancer related

Publications

2. Publications belong to the topic of the dissertation

The SignaLink 2.0, the ARN and the NRF2Ome database regulatory layers

Papp D, Lenti K, **Módos D**, Fazekas D, Dúl Z, Türei D, Földvári-Nagy L, Nussinov R, Csermely P, Korcsmáros T: The NRF2-related interactome and regulome contain multifunctional proteins and fine-tuned autoregulatory loops. (2012) FEBS Letters 586, 13, 1795–1802.

IF:3,58

Fazekas D.*, Koltai M.*, Türei D.*, **Módos D**, Pálfy M, Dúl Z, Zsákai L, Szalay-Bekő M, Lenti K, Farkas I, Vellai T, Csermely P és Korcsmáros T (2013). SignaLink 2 – a signaling pathway resource with multi-layered regulatory networks. BMC Systems Biology 7:7.

IF: 3,26

Türei D., Papp D., Fazekas D., Földvári-Nagy L., **Módos D.,** Lenti K., Csermely P. és Korcsmáros T. (2013). NRF2-ome: an integrated web resource to discover protein interaction and regulatory networks of NRF2. Oxidative Medicine and Cellular Longevity, 2013: 737591.

IF: 3,36

Türei D, Földvári-Nagy L, Fazekas D, **Módos D**, Kubisch J, Kadlecsik T, Demeter A, Lenti K, Csermely P, Vellai T és Korcsmáros T. Autophagy

Regulatory Network – a systems-level bioinformatics resource for studying autophagy components and their regulation (2015). Autophagy, 11(1):155-165.

IF: 9,11

Finding critical paralogue groups within the signalling networks:
Modos D., Brooks J., Fazekas D., Ari E., Vellai T., Csermely P., Korcsmaros T és Lenti K. Identification of critical paralog groups with indispensable role in the regulation of signaling flow (2016) Scientific Reports

IF: 5,23

Determination of cancer related first neighbours and their investigation
Módos D., Bulusu K. C., Fazekas D, Kubisch J., Brooks J., Marczell I, Szabó P. M., Vellai T., Csermely P., Lenti K., Bender A. és Korcsmáros T. Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anti-cancer therapies (2017) npj Systems Biology and Applications,3 Article number: 2

2. Other publications

Csenge Földvári-Nagy, Kincső Csepke Földvári-Nagy, Dezső Módos, Katalin Lenti: Comparative study of hygiene habits in three different groups in Hungary 2016 New Medicine 20(4) 141-147

Csályi K, Fazekas D, Kadlecsik T, Türei D, Gul L, Horváth B, **Módos D**, Demeter A, Pápai N, Lenti K, Csermely P, Vellai T, Korcsmáros T és Varga M (2016) SignaFish: A Zebrafish-Specific Signaling Pathway Resource. *Zebrafish* April 2016,

IF:2,17

Földvári-Nagy Cs, **Módos D**., Feith H. J., Lenti K. Quantitative study of the generational changes among relationship habits in highly educated Hungarian population 2015 New Medicine

Perez-Lopez Á. R.. Szalay K. Z., Türei D., **Módos D**, Lenti K., Korcsmáros T, és Csermely P.: Targets of drugs are generally, and targets of drugs having side effects are specifically good spreaders of human interactome perturbations. 2015 Scientific Reports 5

Acknowledgments

I would like to thank for Katalin Lenti, my supervisor, who I can ask for not just professional but personal help. Tamás Kormcsaros, my consultant with whom we worked during the night to form the presentations and papers which are the backbone of my dissertation. Péter Csermely, who gave me his network viewpoint and through him I got to know Tamas. Gabreilla Dörnyei who gave me the opportunity as a Ph.D. student to teach the physiology practical and lecture at the Morphological and Physiological Department of Faculty of Health Sciences, through it I got better lecturer skills. Tibor Vellai who gave the opportunity to work at Eötvös Loránd University Department of Genetics. David Fazekas who guided my first steps at the terrain of programming and gave me lots of good ideas. Dénes Türei, who made the core of the databases SQL design and webpages. Gabor Szuromi, who helped me the generation of TF-target gene sequences. István Marczell and Péter Szabó, who helped renormalizing the enormous amount of microarray chips. János Kubisch, with whom we searched for the microarray data on GEO. Eszter Ari for the useful advices and the nice meals. László Földvári-Nagy, who checked my dissertation autophagy chapter and his literature knowledge made possible the Autophagy Regulatory Database. Johanne Brooks who checked my English in my publications. Krishna Bulusu, who taught me the

ChEMBL and became a friend under the way, Andreas Bender who hired me as a Post-doc without a Ph.D. I am grateful for every advice, idea and kind words from every colleague at the Morphology and Physiology Department and every member of the NetBiol group and the Department of Genetics at Eötvös Loránd University. I am also grateful for my family and friends, who accept when I am working on my research and not with them and at the end they pushed me to finish my dissertation.