MasterPATH: network analysis of functional genomics screening data

Natalia Rubanova[1,2*], Anna Polesskaya[3], Anna Campalans[4,5,6,7], Guillaume Pinna[8], Jeremie Kropp[9], Annick Harel-Bellan[9], and Nadya Morozova[9]

[1] Université Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France

[2] Skolkovo Institute of Science and Technology, Skolkovo, Russia

[3] Ecole Polytechnique, Université Paris-Saclay, CNRS UMR 7654, Palaiseau, France

[4] Institute of Molecular and Cellular Radiobiology, CEA, F-92265 Fontenay-aux-Roses, France.

[5] INSERM, U967, F-92265 Fontenay-aux-Roses, France.

[6] Université Paris Diderot, U967, F-92265 Fontenay-aux-Roses, France.

[7] Université Paris Sud, U967, F-92265 Fontenay-aux-Roses, France.

[8] Institute for Integrative Biology of the Cell (I2BC), IBITECS, CEA, CNRS, University Paris XI, University Paris-Saclay, 91198, Gif-sur-Yvette cedex, France

[9] Institut des Hautes Etudes Scientiques, 91440 Bures-sur-Yvette, France

Correspondence to Natalia Rubanova nrubanova@gmail.com

DRAFT

Abstract

Functional genomics employs several experimental techniques to investigate gene functions. These techniques such as loss-of-function screening and transcriptome profiling performed in a high-throughput manner give as result a list of genes involved in the biological process of interest. There exist several computational methods for analysis and interpretation of the list. The most widespread methods aim at investigation of biological processes significantly represented in the list or at extracting significantly represented subnetworks. Here we present a new exploratory network analysis method that employs the shortest path approach and centrality measure to uncover members of active molecular pathways leading to the studied phenotype based on the results of functional genomics screening data. We present the method and we demonstrate what data can be retrieved by its application to the terminal muscle differentiation miRNA loss-of-function screening and transcriptomic profiling data and to the 'druggable' loss-of-function RNAi screening data of the DNA repair process.

Introduction

The most wide-spread functional genomics techniques are loss-of-function screening and transcriptome profiling. The biological mechanism for repression of gene expression by causing the degradation of mRNA is exploited in the RNA interference(RNAi) based loss-of-function screening. Synthesized sequence specific RNAi reagents are introduced into the cells in either arrayed or pooled experiment(Mohr et al. 2014). Typically, small interfering RNAs (siRNAs) or short hairpin RNAs (shRNA) are used for mammalian cells(Mohr et al. 2014). Within the cell, shRNAs are recognized by the Dicer complex, which cleaves the shRNAs into the siRNAs, while siRNAs are directly incorporated into RISC (RNA induced silencing) complex(Sontheimer 2005). Next, RISC-siRNA complex binds to a complementary nucleotide sequence of a protein-coding mRNA transcript. This interaction allows a nuclease in the RISC complex to cleave and to destroy the protein-coding mRNA, therefore silencing expression of the gene in a relatively sequence-specific manner(Sontheimer 2005).

In contrast to RNAi technique that allows to perform only temporal gene knock down, permanent gene knockout can be archived using CRISPR/Cas9 technique. CRISPR-Cas system is a bacteria and archaea adaptive immunity mechanism(Wiedenheft, Sternberg, and Doudna 2012). In nature Cas9, crRNA and tracrRNA are transcribed from the CRISPR locus and form crRNA:tracrRNA:Cas9 complex, where crRNA:tracRNA hybrid acts as a guide which allows endonuclease Cas9 to cleave in a sequence specific manner foreign genetic material(Jinek et al. 2012). A synthesized CRISPR library consists of chimeric RNA produced by fusing crRNA and tracrRNA(Mali et al. 2013; Cho et al. 2013). These RNAs known as sgRNAs form sgRNA:Cas9 endonuclease complex that is transfected into the cell(Mali et al. 2013). Whatever technique is used for silencing the gene, the result of the silencing is assessed with a reporter assay specific to the studied phenotype(Mohr et al. 2014). The development of siRNA(Sui et al. 2002; Sen et al. 2004; Yu, DeRuiter, and Turner 2002; Chung et al. 2008; Yang et al. 2006), shRNA(Hu and Luo 2012; Yu, DeRuiter, and Turner 2002), CRISPR sgRNA(Shalem et al. 2015; Tim Wang, Lander, and Sabatini 2016; T Wang et al. 2014) libraries targeting whole genomes made it possible to perform genome scale loss-of function screening.

Transcriptome profiling allows to study a cellular system at the transcriptome level and aims at monitoring gene expression. The most widespread techniques for transcriptome profiling are microarrays and RNA-seq techniques (Lowe et al. 2017). mRNA abundance is measured by

hybridization of fluorescently labelled transcripts to short DNA fragments, known as "probes", in microarray-based experiment (Lockhart et al. 2000). The probes are arrayed on a solid substrate that is usually made from glass(Lockhart et al. 2000). RNA-seq refers to the combination of a high-throughput sequencing methodology with computational methods to measure the presence of transcripts in an RNA extract(Lowe et al. 2017).

Numerous computational methods to interpret functional genomics datasets and infer molecular machinery underlying a given biological process have been developed in the past 15 years. The most widespread methods can be grouped into two categories. The first category is the pathways analysis methods which can be considered as the gold standard approach for analysis of both transcriptome profiling and loss-of-function screening results. The classical representatives of the pathways analysis methods are over representation analysis (ORA) methods. An ORA method uses a statistical test to assess the enrichment of a list of genes in annotated gene sets or canonical pathways. ORA evaluates the number of the hit genes that belong to a pathway. Then, each pathway is assigned a p-value to show how significantly it is represented in the hit list by testing over- or under-representation. The most commonly used statistical tests are based on the hypergeometric, Fisher's exact, chi-square, or binomial tests (Khatri et al. 2012). Several improvements over the standard ORA were developed including functional class scoring approaches that aims at detecting coordinated changes in pathways (Khatri et al. 2012) and topology-based approaches that consider pathway topology, connectivity and interactome information(Bankhead et al. 2009; Draghici et al. 2007).

The second category is network analysis methods which use molecular interaction networks as a complementary information(Markowetz 2010). Such methods can help to find functionally related biological components in a functional genomics dataset. This can be done by introducing network-based scoring methods that can use "guilt by association" principle and information from both network topology and screening results(L. Wang, Tu, and Sun 2009); by introducing the use of the connectivity of subgraphs of protein-protein interaction networks(Kaplow et al. 2009); by using network neighbor information(Jiang et al. 2015); by performing networks functional analysis that relies on assessing the clustering of selected nodes on the network(Cornish and Markowetz 2014); by extracting the largest connected component of a subnetwork that is created from the optimal number of the top-ranked genes(Kairov et al. 2012).

Another way to use molecular interaction networks is to find significantly enriched subnetworks within a functional genomics dataset. Even manual investigation of such subnetworks can give biologically meaningful results(Krishnan et al. 2011; Warner et al. 2014). The molecular interactions networks can be integrated with other types of biological information to archive higher network specificity: with canonical pathways(Tu et al. 2009; De Maeyer et al. 2015); with regulatory interactions(Lan et al. 2011; De Maeyer et al. 2015); with phosphorylation(Nizard et al. 2014). Moreover, subnetworks can be analyzed for finding functional modules(Dittrich et al. 2008; Mitra et al. 2013).

Here we present a new network analysis method to analyze functional genomics datasets. The method aims at elucidating members of molecular pathways leading to the studied phenotype using the results of the functional screening data. The method works on a network that represent human interactome. The network is constructed from 8 databases: HIPPIE (Schaefer et al. 2012), Signor (Surdo et al. 2017), SignaLink (Fazekas et al. 2013), tFacts (Essaghir et al. 2010), KEGG Metabolic Pathways (Ogata et al. 1999), transMir (J. Wang et al. 2009), mirTarBase

(Hsu et al. 2011). The method extracts subnetwork built from the shortest paths between hit genes to so called “final implementers” - genes that are involved in molecular events responsible for final phenotypical realization (if known) or between hit genes (if “final implementers” are not known). The method calculates centrality score for each node and each linear path in the subnetwork as the number of paths found in the previous step that pass through the node and the linear path. Then the statistical significance of each centrality score is assessed by comparing it with centrality scores in subnetworks built from the shortest paths for randomly generated hit lists preserving the degree distribution of the initial hit list. We hypothesis that the nodes and the linear paths with statistically significant centrality score can be considered as putative members of active molecular pathways leading to the studied phenotype.

We illustrate the application of the method by analyzing the results of transcriptomic profiling and miRNA loss-of-function screening of terminal muscle differentiation and of ‘druggable’ loss-of-function screening of DNA repair processes.

Materials and Methods

Databases

The following databases were used to construct human integrated interactome of molecular interactions:

- Human Integrated Protein-Protein Interaction rEference (HIPPIE),
- Signor,
- SignaLink,
- tFacts,
- KEGG Metabolic Pathways,
- transMir,
- mirTarBase.

All databases contain experimentally validated interactions, except SignaLink database which contains a small number of predicted miRNA-mRNA interactions. We used only high confidence interactions from HIPPIE database. The confidence threshold was chosen according HIPPIE documentation(“HIPPIE Howto,” n.d.). Since all the databases use different types of gene ID, we converted the ids to the HUGO gene nomenclature and we used this nomenclature to construct human integrated network. Singnor database contains some interaction that involve phenotypes, protein families and stimuli, however we used only interactions between proteins, complexes and small molecules. The basic information about the databases is summarized in Table 1.

Table 1. Number of nodes, interactions and types of interactions in databases used to construct human integrated network. PPI stands for Protein-Protein Interactions, TF stands for Transcription Factor.

DRAFT

	Nodes	Interactions	Types of nodes	Types of interactions	Direction of interactions
HIPPIE (high confidence)	9368	41520	proteins	PPI	indirect
Signor	3977	13129	proteins, complexes, small molecules	PPI, enzymatic	direct, indirect
SignaLink	3285	27295	proteins, genes, miRNAs	PPI, miRNA-mRNA, TF-gene	direct, indirect
tFacts	2203	4312	TFs, genes	TF-gene	direct
KEGG metabolic pathways	2921	8231	proteins, small molecules	Enzymatic reactions	direct
transMir	324	647	TFs, miRNAs	TF-miRNA	direct
mirTarBase	2269	3511	miRNAs, genes	miRNA-mRNA	direct

Table 2 shows basic topological features of the direct integrated interactome constructed and indirect protein-protein interactome constructed only from. The exponent of the fitted power-law distribution in the degree distribution was calculated with powerlaw(Alstott, Bullmore, and Plenz 2014) Python package.

Table 2.

	Integrated network	PPI network
Nodes	13 419	9 365
Interactions	85 277	41 329

DRAFT

Average clustering coefficient	0.11	0.31
Number of connected components	126	121
Average number of neighbors	9.2	8.6
Density	0.0007	0.0009
Diameter	25	12
Average length of the shortest path	4.5	4.1
Exponent of fitted power-law distribution (total degree)	3.16	2.74
Exponent of fitted power-law distribution (in degree)	3.93	-
Exponent of fitted power-law distribution (out degree)	2.36	-

DRAFT

MasterPATH algorithm

The following notions are used in the mechanistic model of pathway construction: an unweighted graph $G=(V, E)$ represents a network of molecular interactions, where V are nodes that can be proteins, genes, small molecules or miRNAs; E are edges represent molecular interaction between nodes, interactions can be direct or undirect. List of hit genes of size n is as a set $H = \{h_i : h_i \in V \text{ for } \forall i \in [1..n]\}$. List of "final implementers" of size m as a set $F = \{f_i : f_i \in V \text{ for } \forall I \in [1..m]\}$. A simple linear paths p between a pair of nodes $(v,u) : v,u \in V$ is as set of pairs of nodes that represent existing edge in the graph $G: p(v,u)= (v,v_1),(v_1,v_2) \ldots (v_n,u)$ where $v_i \in V$ *for* $\forall i$ and each node vi is distinct. Length L of the path $p(v,u)$ is the number of edges in the path p. We distinguish 4 different types of paths:

- protein-protein paths if all edges represent protein-protein interactions;
- transcriptional paths if there exist at least one edge that represent transcriptional interaction;
- miRNA paths if there exist at least one edge that represent miRNA-mRNA interaction;
- metabolic paths if there exist at least one edge that represent enzymatic reaction.

The algorithm of the method is the following. For a given network G, hit list H, list of "final implementers" F the method finds for each pair of hit gene and "final implementer" (h_i, f_j) all the shortest paths $\{p_i\}$ of four abovementioned types of length less or equal the maximum length L_{max} (defined by the user) in the network G. The search is done using breadth-first algorithm. Then the centrality score which resembles centrality score c is calculated for each node v and each path q (of length of several interactions) as the number of the shortest paths from $\{p_i\}$ that pass through the node v and the path q : $c(v)= |p \in \{p_i\} : v \in p|$; $c(q) = |p \in \{p_i\}$:

$q \in p|$. After that, the statistical significance of each score is assessed. 10000 random hit lists are sampled from the set of nodes *N* preserving or not preserving the degree distribution of the initial hit lit. The probability (p-value^{Net}) of getting a node *v* or a path *q* with specific centrality score by chance is calculated as a proportion of sampled hit lists for which a node or a short path has the same or greater centrality score.

DRAFT

Application

Human muscle differentiation process transcriptomic profiling and functional screening

The screening data from the study by A. Polesskaya et al.(Polesskaya et al. 2013) was taken as the hit list for terminal human muscle differentiation process. In this study, genome-wide miRNA loss-of-function screening on late differentiating human muscle precursor cell line (LHCN) was performed in two-step approach. The primary screening was done in duplicate with LNA antisense inhibitors library targeting 870 miRNAs and a readout assay that detects multinucleated Myosin Heavy Chain (MHC) positive myotubes. Those miRNAs whose depletion resulted in differences to the negative control ≥ 2 SD were selected for the secondary screen which was done in triplicate. The total number of nuclei was checked in addition to the readout assay from the primary screen. A total of 63 miRNAs whose depletion resulted in differences to the negative control ≥ 2 SD were confirmed in the secondary screen.

The screening data from the study by J. Kropp et al.(Kropp et al. 2015) was taken as the second hit list for terminal human muscle differentiation process. Transcriptomic profiling for proliferation and late differentiation stages in human muscle precursor cell line (LHCN) was performed using Affymetrix Human Gene 1.1 ST arrays(Kropp et al. 2015). A total of 571 differentially expressed genes during late differentiation compared to proliferation stage were found with 2-fold change threshold.

Proteins responsible for *inhibition*, *activation*, and *facilitating* of fusion of myotubes(Lee 2004; Alzhanov, McInerney, and Rotwein 2010; Gunning et al. 2001; Y. Wang et al. 2007; Bourmoum, Charles, and Claing 2016; Tachibana and Hemler 1999; Vasyutina et al. 2009; Doherty et al. 2008), namely MSTN, IGF2, ACTA1, MYH1, MYLPF, ARF6, CD81, CD9, CDC42, EHD2, MYOF, were used as a list of “final implementers”. The analysis was performed in the integrated human interactome.

It was found 2609 the shortest paths of 4 types of length from 2 to 5 interactions from each miRNA in the hit list from loss-of-function screening to each protein in the list of “final implementers”. The subnetwork constructed from these paths consists of 1063 nodes (384 of which are genes) and 2710 edges without duplicated edges. The centrality score and the p-

value were calculated for each node and path in the subnetwork according the procedure described in the *Method* section. 294 paths of length of 3 to 4 interactions, with centrality score ≥ 3 and p-value < 0.05; 240 nodes with centrality score ≥ 3 and with p-value < 0.05 were taken for further analysis. Analysis of the paths with high centrality scores had highlighted a possible role for a number of nuclear receptors (AR, ARDB1, NR3C1) in skeletal muscle differentiation, as well as suggested functions in myogenesis for such proteins as arrestin (ARRB1 and 2), intersectin (ITSN1), the Rho GTP echange factor VAV3, and the teratocarcinoma-derived growth factor (TDGF1). Interestingly, while the IGF1 regulatory role in myogenesis is very well studied, our approach allowed us to include the arrestin proteins in these pathways, and thus elaborate the known IGF1 network in skeletal muscle differentiation. The MEF2D, p300, CCND1 functions in differentiation have been abundantly demonstrated, and their presence among the results serves as a proof of efficiency of the analysis.

It was found 47714 the shortest paths of 4 types and of length from 1 to 5 interactions from each gene in the hit list from transcriptome profiling to each protein in the list of "final implementers". The subnetwork constructed from these paths consists of 2847 nodes and 13032 edges without duplicated edges. The centrality score and the p-value were calculated for each node and each path in the subnetwork. 1623 paths of length of 3 to 4 interactions and centrality score ≥ 3, with p-value < 0.05 and 328 nodes with centrality score ≥ 3 and p-value < 0.05 were taken for further analysis. There are 18 miRNAs among these 328 nodes. 4 of them (hsa-mir-125b, has-mir-133a, has-mir-133b, has-mir-145) are the hit miRNAs in the loss-of-function screen. 5 with the highest centrality score are hsa-mir-125b, hsa-mir-29a, hsa-mir-371, hsa-mir-216a, hsa-mir-1. These miRNAs except hsa-mir-371 were shown to be involved in muscle differentiation and/or proliferation(Callis, Chen, and Wang 2007; Winbanks et al. 2011; Meyer et al. 2015; Crist and Buckingham 2009). Moreover, almost half of 18 miRNAs are known to participate in terminal muscle differentiation, and potential roles in myogenesis could be predicted for other miRNAs in this list e.g that regulate cellular proliferation (such as miR-132, miR-145 or miR-224), as well as cardiac hypertrophy (miR-378). Interestingly, the majority of these miRNAs were not found in the original loss-of-function screen, most likely due to the redundancy of miRNA family members. Indeed, as the miRNAs of the same family share the seed sequence, an efficient loss-of-function screen should have contained not only individual miRNA inhibitors, but also the inactivators of whole miRNA families, in order to avoid false negative results. In this sense, our analysis of these data has been important in supplementing a group of miRNA targets that could have been overlooked. This possibility is highlighted by the presence of known myogenesis regulatory miRNAs (miR-1, miR-29a, miR-216a) in the list resulting from the analysis, whereas they have not been picked up by the original experimental screen.

The analysis of paths allowed to identify a potentially novel and interesting pathway in regulation of myogenesis, involving the Myc-associated factor X (MAX), the clathrin-coated pathway regulatory protein AP2M1, and the EH-domain protein EHD2, which links the clathrin coated transport to actin cytoskeleton, and also binds to myoferlin, a factor promoting myotube fusion. Together with integrin subunits ITGA4 and ITGB1, the extracellular matrix component fibronectin (FN1), and the protein chaperon HSP90, these proteins indicate a possible involvement of specific protein transport pathways in terminal myogenic differentiation. In addition, there is a possibility of involvement of beta-catenin (CTNNB1), C-KIT and PRKC in these processes. It should be noted that these three regulatory factors, while

extensively studied in a multitude of biological models, have never been shown to be specifically implicated in skeletal myogenesis. Two major regulatory molecules of skeletal myogenesis, MYOD and SMAD3, have been highlighted, together with their previously known muscle-related targets (TGFB1, CDC42, CTCF). Also, they are linked to a number of proteins that have not been previously studied in the context of muscle differentiation (MAX, KIT, AP2M1…).

Next, we compared the lists of nodes and the lists of paths from two experiments. We found 29 nodes and 16 paths common for both loss-of-function screening and transcriptomic profiling. Among the nodes with the highest centrality score, three - IGF1R, CBL, E2F1 - have been suggested to play key roles in the growth, development, and differentiation of skeletal muscle(Fernández et al. 2002; Nakao et al. 2009; Zappia and Frolov 2016). The path with the highest score consists of CDKN1A, MDM2, TCAP, MSTN proteins. The interaction between MDM2 and TCAP is known to be important for cardiac hypertrophy(Tian et al. 2006), it was also shown that TCAP controls secretion of MSTN(Nicholas et al. 2002). Our analysis shows that this path might be activated by the depletion of hsa-mir-17, hsa-mir-106a, hsa-mir-125a, hsa-mir-145, hsa-mir-93. I can also be noted, that not only androgen receptor (AR), but also the estrogen receptor ESR1 can play a role in human skeletal myogenesis. Interestingly, specific integrins (ITGB1, ITGA6) and adaptor proteins (CRKL) have also been found, confirming the importance of certain membrane/adherence structures. Strikingly, both the receptor of activated C kinase (RACK1), and the inhibitor of this kinase (YWHAB, a 14-3-3 protein), as well as multiple other protein-processing enzymes (peptidase inhibitor SERPIN1, casein kinase CSNK1A1, proprotein convertase FURIN, and activator or protein secretion CHRM3) were found by the analysis, attracting the attention to the role of protein metabolism in myogenesis. It was also very interesting to see the chromosome breakpoint generation factor FRA11B among these potential novel factors that might impact on the differentiation of human myoblasts. This comparison has shown potentially novel paths originating from well-known actors in muscle differentiation (such as IGF1R - RACK1 - CD81); and vice versa, has shown that previously unknown potential regulators of myogenesis, such as YWHAB or FRA11B, can act upon proteins that are well known to regulate myotube hypertrophy and/or fusion (IGFR1, CD81).

The fact that the comparison resulted only in a few number of paths might indicate, that although these experimental systems study one biological process, they interfere biological machinery on two different levels – on the level of translation (miRNAs) and on the level on transcription (transcriptome), which might account on two different regulation mechanisms.

We also found that 14 paths from the analysis of transcriptomic profiling have miRNAs hits from the loss-of-function screening and 60 paths from the analysis of loss-of-function screening have hit genes from the transcriptomic profiling on them. These are the paths from the analysis of transcriptomic profiling that pass through hsa-mir-125b which controls IGF2 gene and the paths that pass through has-mir-145 that control TRIP10 protein which, according to OMIM database, has highest expression in skeletal muscle(McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, n.d.) and interacts with CDC42 protein. Also, analyzing these paths one can notice the factors participating in at least three major cellular pathways, that, however, have not been extensively studied in skeletal muscle differentiation. These factors include beta-transduction repeat containing protein (BTRC), which has a strong impact on both beta-catenin and NF-kappa B signaling, as well as the p53-related protein TP73, and, finally, the protein LRIG1 that has a strong negative effect on the expression of epidermal

growth factor receptor. These pathways represent interesting new directions to follow, in order to further understand the mechanics of skeletal myogenesis.

Human DNA repair process functional screening

We used the data from the siRNA functional screening that identified 'druggable' genes involved in oxidative damaged DNA repair(Guyon et al. 2015; Robinson et al. 2017). The screening was performed on genetically engineered HeLa cells that express OGG1-GFP fusion protein. OGG1 is a DNA glycosylate protein that is recruited to chromatin to initiate the repair of oxidized chromatin(Guyon et al. 2015). Each 'druggable' gene was targeted by 3 siRNAs(Robinson et al. 2017). The intensity of chromatin-bound OGG1-GFP was detected after inducing DNA damage and 18 hit genes were identified.

All 18 hit genes were used as a list of "final implementers", since little is known about the proteins involved in the final steps of the DNA repair process. The analysis was performed in the protein-protein human interactome.

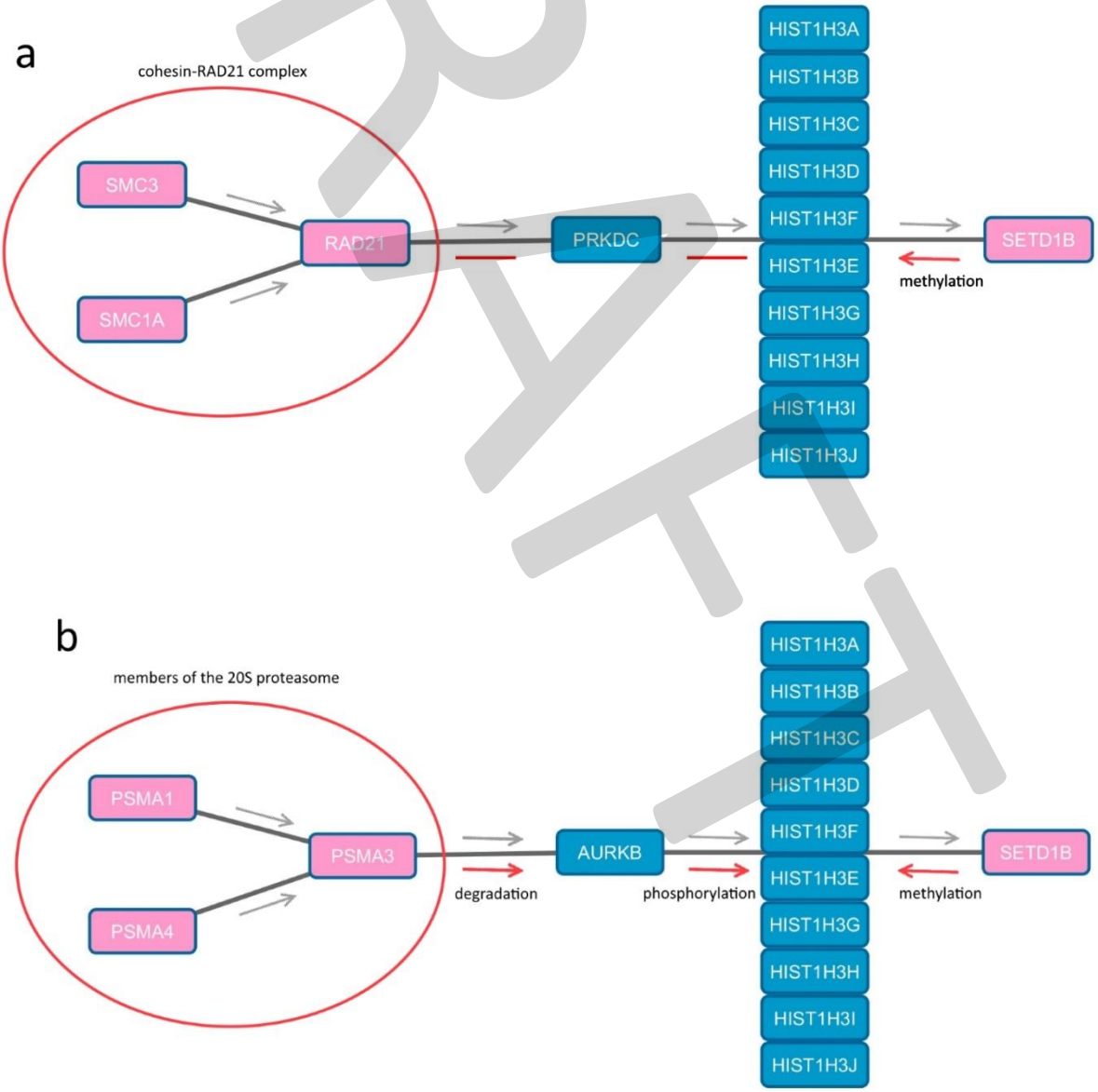
It was found 4876 the shortest protein-protein paths of length from 1 to 4 interactions from each gene product in the hit list to each gene product in the list of "final implementers". The subnetwork constructed from these paths consists of 381 nodes and 1764 edges without duplicated edges. The centrality score and the p-value were calculated for each node and path in the subnetwork according the procedure described in the *Method* section. 64 paths of length of 2 to 3 interactions, with centrality score ≥ 3 and p-value < 0.05; 28 nodes with centrality score ≥ 3 and with p-value < 0.05 were taken for further analysis.

The top 10 nodes with the highest centrality score are histones 3H proteins: HIST1H3A, HIST1H3B, HIST1H3C, HIST1H3D, HIST1H3E, HIST1H3F, HIST1H3G, HIST1H3H, HIST1H3I, HIST1H3J. It is known that the DNA damage is associated with higher level of chromatin mobility(Strzyz 2017; Dion et al. 2012; Miné-Hattab and Rothstein 2012) and it was shown recently that the increase in chromatin mobility is governed by the proteasome-mediated degradation of core histones(Hauer et al. 2017). Other proteins with high centrality score are SETDB1 protein – a member of the SET1 family of proteins; WDR5 protein - a core component of SET1 family complexes(Ruthenburg et al. 2016); TP53BP1 - a binding partner of the tumour suppressor protein p53. SETDB1 and WDR5 are associated with post-translational histone modification which allows recruitment of the chromatin-associated proteins and protein complexes(Schultz et al. 2002; Odho, Southall, and Wilson 2010). TP53BP1 protein is known to be an important regulator of the cellular response to DNA double-strand breaks(Panier and Boulton 2013).

Figure 1 presents subnetworks visualized with Cytoscape software(Shannon et al. 2003) for the paths of length 3 interactions with centrality score 3. Figure 1a shows that the method identified that two cohesin proteins SMC3 and SMC1A can act by interacting with RAD21 protein, known cohesin-RAD21 complex(Losada 2014) enriched at DNA double-strand break sites and facilitates recombinational DNA repair(Watrin and Peters 2009). Figure 1b shows possible mechanism of involvement of PSMA1, PSMA3, PSMA4 proteins, all members of the

20S proteasome(Coux, Tanaka, and Goldberg 1996), through interaction with AURKB, Aurora Kinase B(Shu et al. 2003), which in turn interacts with histones H3(Crosio et al. 2002). The path ends with histones H3 – SETDB1 interaction. SETDB1 is a histone methyltransferase that specifically methylates histone H3(Schultz et al. 2002) and is also a member of the hit list. The red arrow shows the direction of this interaction on Figure 1a. Considering this, histones H3 are the proteins where the signal converges from different members of the hit list and we hypothesize that histone H3 can be a "final implementer" for this system.

Figure 1. Subnetworks for two paths with centrality score 3. The hit proteins are colored pink, the intermediate proteins are colored blue. The grey arrows show the direction of the paths when it is built by the method. The red arrows show the direction and the effect of the interaction found in literature. **a** RAD21-PRKDC-histone H3-SETDB1 path **b** PSMA3-AURKB-histone H3- SETDB1 path



DRAFT

Java implementation

The method is implemented in the Java code. The code and full tables for the results presented in the paper available at GitHub page https://github.com/daggoo/masterPath.

Discussion

We used two different types of networks in our work. The first one was a mixed direct and indirect network constructed from PPI, transcriptional, post- transcriptional and metabolic data. The second network was indirect PPI network. PPI networks are the most common networks used in network analysis, although they are known to be incomplete and biased towards the well-studied proteins. Incorporating transcriptional, post-transcriptional and metabolic data, does not solve the issues associated with PPI networks, but adds information on directionality and positive or negative effect of the interactions and gives the ability to build heterogeneous paths.

The bias in the results towards highly connected nodes or paths that pass by highly connected nodes is removed by generating random hit lists. We used the threshold of 0.05 for p-value. However, the nodes/paths with higher p-value can also be considered, it is just not guaranteed that the value of the centrality score is due to the specificity of the node/path to the phenotype or due to its high connectivity.

Other general problems in network analysis is network specificity for the biological system and lack of information relating to some members of the hit list. We used networks that represent human interactome in our work. However, the nodes that are not active can be excluded from the network for some biological systems, based for example on the transcriptomic data, to create a smaller but more specific network. On the other hand, the hit nodes from the functional screening might be poorly studied or might even not be present in the network. Low confidence or predicted interactions for hit nodes might be added to the network in this case.

We introduced a notion of “final implementer” in this work. We denoted a “final implementer” as a molecule that is involved in events responsible for final phenotypical realization of the biological process. Modern molecular biology accumulated vast amount of knowledge. A part of such molecules is known for some processes, e.g. for apoptosis caspase 3, caspase 6 and caspase 7 could be considered as “final implementers”. In case such molecules are unknown, the members of the hit list can be used as a list of “final implementers” in the analysis on the PPI network and by studying directionality of the paths candidates for “final implementers” could be found as we demonstrated for the DNA repair process.

DRAFT

Conclusion

We presented a new exploratory network analysis method that employs the shortest path approach and centrality measure for uncovering members of active molecular pathways leading to the studied phenotype based on the results of functional screening and in this work. The source code was provided on the GitHub page. We illustrated the application of the method to the analysis of the results of transcriptomic profiling and miRNA loss-of-function screening of human muscle proliferation process and of 'druggable' loss-of-function screening of human DNA repair process. We also presented the basic topological features of the integrated and the protein-protein human interactome.

References

Alstott, Jeff, Ed Bullmore, and Dietmar Plenz. 2014. "Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions." *PLoS ONE* 9 (1). https://doi.org/10.1371/journal.pone.0085777.

Alzhanov, Damir T., Stephanie F. McInerney, and Peter Rotwein. 2010. "Long Range Interactions Regulate Igf2 Gene Transcription during Skeletal Muscle Differentiation." *Journal of Biological Chemistry* 285 (50):38969–77. https://doi.org/10.1074/jbc.M110.160986.

Bankhead, Armand, Iliana Sach, Chester Ni, Nolwenn LeMeur, Mark Kruger, Marc Ferrer, Robert Gentleman, and Carol Rohl. 2009. "Knowledge Based Identification of Essential Signaling from Genome-Scale siRNA Experiments." *BMC Systems Biology* 3:80. https://doi.org/10.1186/1752-0509-3-80.

Bourmoum, Mohamed, Ricardo Charles, and Audrey Claing. 2016. "The GTPase ARF6 Controls ROS Production to Mediate Angiotensin II-Induced Vascular Smooth Muscle Cell Proliferation." *PLoS ONE* 11 (1):1–17. https://doi.org/10.1371/journal.pone.0148097.

Callis, Thomas E, Jian-Fu Chen, and Da-Zhi Wang. 2007. "MicroRNAs in Skeletal and Cardiac Muscle Development." *DNA and Cell Biology* 26 (4):219–25. https://doi.org/10.1089/dna.2006.0556.

Cho, Seung Woo, Sojung Kim, Jong Min Kim, and Jin-Soo Kim. 2013. "Targeted Genome Engineering in Human Cells with the Cas9 RNA-Guided Endonuclease." *Nature Biotechnology* 31 (3). Nature Publishing Group:230–32. https://doi.org/10.1038/nbt.2507.

Chung, N., L. Locco, K. W. Huff, S. Bartz, P. S. Linsley, M. Ferrer, and B. Strulovici. 2008. "An Efficient and Fully Automated High-Throughput Transfection Method for

DRAFT

Genome-Scale siRNA Screens." *Journal of Biomolecular Screening* 13 (2):142–48. https://doi.org/10.1177/1087057107312032.

Cornish, Alex J., and Florian Markowetz. 2014. "SANTA: Quantifying the Functional Content of Molecular Networks." *PLoS Computational Biology* 10 (9). https://doi.org/10.1371/journal.pcbi.1003808.

Coux, O, K Tanaka, and a L Goldberg. 1996. "Structure and Functions of the 20S and 26S Proteasomes." *Annual Review of Biochemistry* 65:801–47. https://doi.org/10.1146/annurev.biochem.65.1.801.

Crist, Colin G., and Margaret Buckingham. 2009. "microRNas Gain Magnitude in Muscle." *Cell Cycle* 8 (22):3627–28. https://doi.org/10.4161/cc.8.22.9960.

Crosio, Claudia, Gian Maria Fimia, Romain Loury, Yukio Okano, Hongyi Zhou, Subrata Sen, C David, et al. 2002. "Mitotic Phosphorylation of Histone H3 : Spatio-Temporal Regulation by Mammalian Aurora Kinases Mitotic Phosphorylation of Histone H3 : Spatio-Temporal Regulation by Mammalian Aurora Kinases." *Molecular and Cellular Biology* 22 (3):874–85. https://doi.org/10.1128/MCB.22.3.874.

Dion, Vincent, Véronique Kalck, Chihiro Horigome, Benjamin D. Towbin, and Susan M. Gasser. 2012. "Increased Mobility of Double-Strand Breaks Requires Mec1, Rad9 and the Homologous Recombination Machinery." *Nature Cell Biology* 14 (5). Nature Publishing Group:502–9. https://doi.org/10.1038/ncb2465.

Dittrich, Marcus T., Gunnar W. Klau, Andreas Rosenwald, Thomas Dandekar, and Tobias Müller. 2008. "Identifying Functional Modules in Protein-Protein Interaction Networks: An Integrated Exact Approach." *Bioinformatics* 24 (13):223–31. https://doi.org/10.1093/bioinformatics/btn161.

Doherty, Katherine R., Alexis R. Demonbreun, Gregory Q. Wallace, Andrew Cave, Avery D. Posey, Konstantina Heretis, Peter Pytel, and Elizabeth M. McNally. 2008. "The Endocytic Recycling Protein EHD2 Interacts with Myoferlin to Regulate Myoblast Fusion." *Journal of Biological Chemistry* 283 (29):20252–60. https://doi.org/10.1074/jbc.M802306200.

Draghici, Sorin, Purvesh Khatri, Adi Laurentiu Tarca, Kashyap Amin, Arina Done, Calin Voichita, Constantin Georgescu, and Roberto Romero. 2007. "A Systems Biology Approach for Pathway Level Analysis." *Genome Research*, 1537–45. https://doi.org/10.1101/gr.6202607.1.

Essaghir, Ahmed, Federica Toffalini, Laurent Knoops, Anders Kallin, Jacques van Helden, and Jean Baptiste Demoulin. 2010. "Transcription Factor Regulation Can Be Accurately Predicted from the Presence of Target Gene Signatures in Microarray Gene Expression Data." *Nucleic Acids Research* 38 (11). https://doi.org/10.1093/nar/gkq149.

Fazekas, Dávid, Mihály Koltai, Dénes Türei, Dezső Módos, Máté Pálfy, Zoltán Dúl, Lilian Zsákai, et al. 2013. “SignaLink 2 - a Signaling Pathway Resource with Multi-Layered Regulatory Networks.” *BMC Systems Biology* 7:7. https://doi.org/10.1186/1752-0509-7-7.

Fernández, Ana M., Joëlle Dupont, Roger P. Farrar, Sukho Lee, Bethel Stannard, and Derek Le Roith. 2002. “Muscle-Specific Inactivation of the IGF-I Receptor Induces Compensatory Hyperplasia in Skeletal Muscle.” *Journal of Clinical Investigation* 109 (3):347–55. https://doi.org/10.1172/JCI200213503.

Gunning, P W, V Ferguson, K J Brennan, and E C Hardeman. 2001. “Alpha-Skeletal Actin Induces a Subset of Muscle Genes Independently of Muscle Differentiation and Withdrawal from the Cell Cycle.” *Journal of Cell Science* 114 (Pt 3):513–24. http://www.ncbi.nlm.nih.gov/pubmed/11171321.

Guyon, Laurent, Christian Lajaunie, Frédéric Fer, Ricky Bhajun, Eric Sulpice, Guillaume Pinna, Anna Campalans, et al. 2015. “Φ-Score: A Cell-to-Cell Phenotypic Scoring Method for Sensitive and Selective Hit Discovery in Cell-Based Assays.” *Scientific Reports* 5 (1). Nature Publishing Group:14221. https://doi.org/10.1038/srep14221.

Hauer, Michael H, Andrew Seeber, Vijender Singh, Raphael Thierry, Ragna Sack, Assaf Amitai, Mariya Kryzhanovska, et al. 2017. “Histone Degradation in Response to DNA Damage Enhances Chromatin Dynamics and Recombination Rates.” *Nature Structural & Molecular Biology* 24 (2). Nature Publishing Group:99–107. https://doi.org/10.1038/nsmb.3347.

“HIPPIE Howto.” n.d. http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/information.php.

Hsu, Sheng Da, Feng Mao Lin, Wei Yun Wu, Chao Liang, Wei Chih Huang, Wen Ling Chan, Wen Ting Tsai, et al. 2011. “MiRTarBase: A Database Curates Experimentally Validated microRNA-Target Interactions.” *Nucleic Acids Research* 39 (SUPPL. 1):163–69. https://doi.org/10.1093/nar/gkq1107.

Hu, Guang, and Ji Luo. 2012. “A Primer on Using Pooled shRNA Libraries for Functional Genomic Screens.” *Acta Biochimica et Biophysica Sinica* 44 (2):103–12. https://doi.org/10.1093/abbs/gmr116.

Jiang, Peng, Hongfang Wang, Wei Li, Chongzhi Zang, Bo Li, Yinling J. Wong, Cliff Meyer, Jun S. Liu, Jon C. Aster, and X. Shirley Liu. 2015. “Network Analysis of Gene Essentiality in Functional Genomics Experiments.” *Genome Biology* 16 (1):239. https://doi.org/10.1186/s13059-015-0808-9.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer a Doudna, and Emmanuelle Charpentier. 2012. “A Programmable Dual-RNA–Guided DNA

Endonuclease in Adaptive Bacterial Immunity" 337 (August):816–22. https://doi.org/10.1126/science.1225829.

Kairov, Ulykbek, Tatyana Karpenyuk, Erlan Ramanculov, and Andrei Zinovyev. 2012. "Network Analysis of Gene Lists for Finding Reproducible Prognostic Breast Cancer Gene Signatures." *Bioinformation* 8 (16):773–76. https://doi.org/10.6026/97320630008773.

Kaplow, Irene M, Rohit Singh, Adam Friedman, Chris Bakal, Norbert Perrimon, and Bonnie Berger. 2009. "RNAiCut: Automated Detection of Significant Genes from Functional Genomic Screens." *Nature Methods* 6 (7). Nature Publishing Group:476–77. https://doi.org/10.1038/nmeth0709-476.

Khatri, Purvesh, Marina Sirota, Atul J Butte, Lucile Packard, Welch Road, and Palo Alto. 2012. "10 Years of Pathway Analysis : Current Approaches and Outstanding Challenges - Table S2," 10.

Krishnan, Manoj N, Aylwin Ng, Bindu Sukumaran, Felicia D Gilfoy, Pradeep D Uchil, Hameeda Sultana, Abraham L Brass, et al. 2011. "RNA Interference Screen for Human Genes Associated with West Nile Virus Infection." *Nature* 455 (7210):242–45. https://doi.org/10.1038/nature07207.RNA.

Kropp, Jeremie, Cindy Degerny, Nadezda Morozova, Julien Pontis, Annick Harel-Bellan, and Anna Polesskaya. 2015. "miR-98 Delays Skeletal Muscle Differentiation by down-Regulating E2F5." *Biochemical Journal* 466 (1):85–93. https://doi.org/10.1042/BJ20141175.

Lan, Alex, Ilan Y. Smoly, Guy Rapaport, Susan Lindquist, Ernest Fraenkel, and Esti Yeger-Lotem. 2011. "ResponseNet: Revealing Signaling and Regulatory Networks Linking Genetic and Transcriptomic Screening Data." *Nucleic Acids Research* 39 (SUPPL. 2):424–29. https://doi.org/10.1093/nar/gkr359.

Lee, Se-Jin. 2004. "Regulation of Muscle Mass By Myostatin." *Annual Review of Cell and Developmental Biology* 20 (1):61–86. https://doi.org/10.1146/annurev.cellbio.20.012103.135836.

Lockhart, D J, D J Lockhart, E a Winzeler, and E a Winzeler. 2000. "Genomics, Gene Expression and DNA Arrays." *Nature* 405 (6788):827–36. https://doi.org/10.1038/35015701.

Losada, Ana. 2014. "Cohesin in Cancer: Chromosome Segregation and beyond." *Nature Reviews Cancer* 14 (6). Nature Publishing Group:389–93. https://doi.org/10.1038/nrc3743.

Lowe, Rohan, Neil Shirley, Mark Bleackley, Stephen Dolan, and Thomas Shafee. 2017. "Transcriptomics Technologies." *PLoS Computational Biology* 13 (5):1–23. https://doi.org/10.1371/journal.pcbi.1005457.

Maeyer, Dries De, Bram Weytjens, Joris Renkens, Luc De Raedt, and Kathleen Marchal. 2015. “PheNetic: Network-Based Interpretation of Molecular Profiling Data.” *Nucleic Acids Research* 43 (W1):W244–50. https://doi.org/10.1093/nar/gkv347.

Mali, Prashant, Luhan Yang, Kevin M Esvelt, John Aach, Marc Guell, James E DiCarlo, Julie E Norville, and George M Church. 2013. “RNA-Guided Human Genome Engineering via Cas9 Prashant.” *Science* 339 (6121):823–26. https://doi.org/10.1126/science.1232033.RNA-Guided.

Markowetz, Florian. 2010. “How to Understand the Cell by Breaking It: Network Analysis of Gene Perturbation Screens.” *PLoS Computational Biology* 6 (2). https://doi.org/10.1371/journal.pcbi.1000655.

McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). n.d. “Online Mendelian Inheritance in Man, OMIM®.” https://omim.org/.

Meyer, Swanhild U, Christian Thirion, Anna Polesskaya, Stefan Bauersachs, Sebastian Kaiser, Sabine Krause, and Michael W Pfaffl. 2015. “TNF-A and IGF1 Modify the microRNA Signature in Skeletal Muscle Cell Differentiation.” *Cell Communication and Signaling* 13 (1):1–14. https://doi.org/10.1186/s12964-015-0083-0.

Miné-Hattab, Judith, and Rodney Rothstein. 2012. “Increased Chromosome Mobility Facilitates Homology Search during Recombination.” *Nature Cell Biology* 14 (5). Nature Publishing Group:510–17. https://doi.org/10.1038/ncb2472.

Mitra, K, Ar Carvunis, Sk Ramesh, and T Ideker. 2013. “Integrative Approaches for Finding Modular Structure in Biological Networks.” *Nature Reviews Genetics* 14 (10):719–32. https://doi.org/10.1038/nrg3552.Integrative.

Mohr, Stephanie E., Jennifer A. Smith, Caroline E. Shamu, Ralph a. Neumüller, and Norbert Perrimon. 2014. “RNAi Screening Comes of Age: Improved Techniques and Complementary Approaches.” *Nature Reviews Molecular Cell Biology* 15 (9). Nature Publishing Group:591–600. https://doi.org/10.1038/nrm3860.

Nakao, R., K. Hirasaka, J. Goto, K. Ishidoh, C. Yamada, a. Ohno, Y. Okumura, et al. 2009. “Ubiquitin Ligase Cbl-B Is a Negative Regulator for Insulin-Like Growth Factor 1 Signaling during Muscle Atrophy Caused by Unloading.” *Molecular and Cellular Biology* 29 (17):4798–4811. https://doi.org/10.1128/MCB.01347-08.

Nicholas, Gina, Mark Thomas, Brett Langley, Wayne Somers, Ketan Patel, C. Fred Kemp, Mridula Sharma, and Ravi Kambadur. 2002. “Titin-Cap Associates With, and Regulates Secretion Of, Myostatin.” *Journal of Cellular Physiology* 193 (1):120–31. https://doi.org/10.1002/jcp.10158.

Nizard, Philippe, Frédéric Ezan, Dominique Bonnier, Nolwenn Le Meur, Sophie Langouët, Georges Baffet, Yannick Arlot-Bonnemains, and Nathalie Théret. 2014. “Integrative Analysis of High-Throughput RNAi Screen Data Identifies the FER and CRKL Tyrosine

Kinases as New Regulators of the Mitogenic ERK-Dependent Pathways in Transformed Cells." *BMC Genomics* 15 (1):1169. https://doi.org/10.1186/1471-2164-15-1169.

Odho, Zain, Stacey M. Southall, and Jon R. Wilson. 2010. "Characterization of a Novel WDR5-Binding Site That Recruits RbBP5 through a Conserved Motif to Enhance Methylation of Histone H3 Lysine 4 by Mixed Lineage Leukemia Protein-1." *Journal of Biological Chemistry* 285 (43):32967–76. https://doi.org/10.1074/jbc.M110.159921.

Ogata, Hiroyuki, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, and Minoru Kanehisa. 1999. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 27 (1):29–34. https://doi.org/10.1093/nar/27.1.29.

Panier, Stephanie, and Simon J. Boulton. 2013. "Double-Strand Break Repair: 53BP1 Comes into Focus." *Nature Reviews Molecular Cell Biology* 15 (1). Nature Publishing Group:7–18. https://doi.org/10.1038/nrm3719.

Polesskaya, Anna, Cindy Degerny, Guillaume Pinna, Yves Maury, Gueorgui Kratassiouk, Vincent Mouly, Nadya Morozova, Jeremie Kropp, Niels Frandsen, and Annick Harel-Bellan. 2013. "Genome-Wide Exploration of miRNA Function in Mammalian Muscle Cell Differentiation." *PLoS ONE* 8 (8):1–10. https://doi.org/10.1371/journal.pone.0071927.

Robinson, Sean, Jaakko Nevalainen, Guillaume Pinna, Anna Campalans, J. Pablo Radicella, and Laurent Guyon. 2017. "Incorporating Interaction Networks into the Determination of Functionally Related Hit Genes in Genomic Experiments with Markov Random Fields." *Bioinformatics* 33 (14):i170–79. https://doi.org/10.1093/bioinformatics/btx244.

Ruthenburg, Alexander J, Wooikoon Wang, Daina M Graybosch, Haitao Li, C David Allis, Dinshaw J Patel, and Gregory L Verdine. 2016. "Histone H3 Recognition and Presentation by the WDR5 Module of the MLL1 Complex" 13 (8):704–12. https://doi.org/10.1038/nsmb1119.Histone.

Schaefer, Martin H., Jean Fred Fontaine, Arunachalam Vinayagam, Pablo Porras, Erich E. Wanker, and Miguel a. Andrade-Navarro. 2012. "Hippie: Integrating Protein Interaction Networks with Experiment Based Quality Scores." *PLoS ONE* 7 (2):1–8. https://doi.org/10.1371/journal.pone.0031826.

Schultz, David C, Kasirajan Ayyanathan, Dmitri Negorev, Gerd G Maul, and Frank J Rauscher Iii. 2002. "SETDB1 : A Novel KAP - 1 - Associated Histone H3 , Lysine 9 - Specific Methyltransferase That Contributes to HP1 - Mediated Silencing of Euchromatic Genes by KRAB Zinc - Finger Proteins." *Genes and Development* 16:919–32. https://doi.org/10.1101/gad.973302.

Sen, George, Tom S Wehrman, Jason W Myers, and Helen M Blau. 2004. “Restriction Enzyme-Generated siRNA (REGS) Vectors and Libraries.” *Nature Genetics* 36 (2):183–89. https://doi.org/10.1038/ng1288.

Shalem, Ophir, Neville E Sanjana, Feng Zhang, and Cognitive Sciences. 2015. “High-Throughput Functional Genomics Using CRISPR-Cas9” 16 (5):299–311. https://doi.org/10.1038/nrg3899.High-throughput.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Beno Schwikowski, and Trey Ideker. 2003. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” *Genome Research* 13 (11):2498–2504. https://doi.org/10.1101/gr.1239303.

Shu, Fengjue, Shuguang Guo, Yongjun Dang, Meiyan Qi, Guangjin Zhou, Zekun Guo, Ying Zhang, Chaoqun Wu, Shouyuan Zhao, and Long Yu. 2003. “Human Aurora-B Binds to a Proteasome A-Subunit HC8 and Undergoes Degradation in a Proteasome-Dependent Manner.” *Molecular and Cellular Biochemistry* 254 (1-2):157–62. https://doi.org/10.1023/A:1027317014159.

Sontheimer, Erik J. 2005. “Assembly and Function of RNA Silencing Complexes.” *Nature Reviews Molecular Cell Biology* 6 (2):127–38. https://doi.org/10.1038/nrm1568.

Strzyz, Paulina. 2017. “DNA Repair: Histones Have Got to Go.” *Nature Reviews Molecular Cell Biology* 18 (3). Nature Publishing Group:138–39. https://doi.org/10.1038/nrm.2017.2.

Sui, G., C. Soohoo, E. B. Affar, F. Gay, Y. Shi, W. C. Forrester, and Y. Shi. 2002. “A DNA Vector-Based RNAi Technology to Suppress Gene Expression in Mammalian Cells.” *Proceedings of the National Academy of Sciences* 99 (8):5515–20. https://doi.org/10.1073/pnas.082117599.

Surdo, Prisca Lo, Alberto Calderone, Gianni Cesareni, and Livia Perfetto. 2017. “SIGNOR: A Database of Causal Relationships between Biological Entities-a Short Guide to Searching and Browsing.” *Current Protocols in Bioinformatics* 2017 (June):8.23.1–8.23.16. https://doi.org/10.1002/cpbi.28.

Tachibana, Isao, and Martin E. Hemler. 1999. “Role of Transmembrane 4 Superfamily (TM4SF) Proteins CD9 and CD81 in Muscle Cell Fusion and Myotube Maintenance.” *Journal of Cell Biology* 146 (4):893–904. https://doi.org/10.1083/jcb.146.4.893.

Tian, Li Feng, Hui Yan Li, Bao Feng Jin, Xin Pan, Jiang Hong Man, Pei Jing Zhang, Wei Hua Li, et al. 2006. “MDM2 Interacts with and Downregulates a Sarcomeric Protein, TCAP.” *Biochemical and Biophysical Research Communications* 345 (1):355–61. https://doi.org/10.1016/j.bbrc.2006.04.108.

Tu, Zhidong, Carmen Argmann, Kenny K Wong, Lyndon J Mitnaul, Stephen Edwards, Iliana C Sach, Jun Zhu, and Eric E Schadt. 2009. “Integrating siRNA and Protein – Protein Interaction Data to Identify an Expanded Insulin Signaling Network,” 1057–67. https://doi.org/10.1101/gr.087890.108.dependent.

Vasyutina, Elena, Benedetta Martarelli, Cord Brakebusch, Hagen Wende, and Carmen Birchmeier. 2009. “The Small G-Proteins Rac1 and Cdc42 Are Essential for Myoblast Fusion in the Mouse.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (22):8935–40. https://doi.org/10.1073/pnas.0902501106.

Wang, Juan, Ming Lu, Chengxiang Qiu, and Qinghua Cui. 2009. “TransmiR: A Transcription Factor microRNA Regulation Database.” *Nucleic Acids Research* 38 (SUPPL.1):119–22. https://doi.org/10.1093/nar/gkp803.

Wang, Li, Zhidong Tu, and Fengzhu Sun. 2009. “A Network-Based Integrative Approach to Prioritize Reliable Hits from Multiple Genome-Wide RNAi Screens in Drosophila.” *BMC Genomics* 10:220. https://doi.org/10.1186/1471-2164-10-220.

Wang, T, Jj Wei, Dm Sabatini, and Es Lander. 2014. “Genetic Screens in Human Cells Using the CRISPR/Cas9 System.” *Science* 343 (6166):80–84. https://doi.org/10.1126/science.1246981.Genetic.

Wang, Tim, Eric S Lander, and David M Sabatini. 2016. “Large-Scale Single-Guide RNA Library Construction and Use for Genetic Screens: CRISPR/Cas9-Based Genetic Screens” 2016 (3). https://doi.org/10.1101/pdb.top086892.Large-Scale.

Wang, Y., D. Szczesna-Cordary, R. Craig, Z. Diaz-Perez, G. Guzman, T. Miller, and J. D. Potter. 2007. “Fast Skeletal Muscle Regulatory Light Chain Is Required for Fast and Slow Skeletal Muscle Development.” *The FASEB Journal* 21 (9):2205–14. https://doi.org/10.1096/fj.06-7538com.

Warner, Neil, Aaron Burberry, Maria Pliakas, Christine McDonald, and Gabriel N????ez. 2014. “A Genome-Wide Small Interfering RNA (siRNA) Screen Reveals Nuclear Factor-??B (NF-??B)-Independent Regulators of NOD2-Induced Interleukin-8 (IL-8) Secretion.” *Journal of Biological Chemistry* 289 (41):28213–24. https://doi.org/10.1074/jbc.M114.574756.

Watrin, Erwan, and Jan-Michael Peters. 2009. “The Cohesin Complex Is Required for the DNA Damage-Induced G2/M Checkpoint in Mammalian Cells.” *The EMBO Journal* 28 (17):2625–35. https://doi.org/10.1038/emboj.2009.202.

Wiedenheft, Blake, Samuel H. Sternberg, and Jennifer a. Doudna. 2012. “RNA-Guided Genetic Silencing Systems in Bacteria and Archaea.” *Nature* 482 (7385):331–38. https://doi.org/10.1038/nature10886.

Winbanks, Catherine E., Bo Wang, Claudia Beyer, Phillip Koh, Lloyd White, Phillip Kantharidis, and Paul Gregorevic. 2011. "TGF-B Regulates miR-206 and miR-29 to Control Myogenic Differentiation through Regulation of HDAC4." *Journal of Biological Chemistry* 286 (16):13805–14. https://doi.org/10.1074/jbc.M110.192625.

Yang, Jian Ping, Wufang Fan, Cheryl Rogers, Jon E. Chatterton, Joshua Bliesath, Guohong Liu, Ning Ke, et al. 2006. "A Novel RNAi Library Based on Partially Randomized Consensus Sequences of Nuclear Receptors: Identifying the Receptors Involved in Amyloid ?? Degradation." *Genomics* 88 (3):282–92. https://doi.org/10.1016/j.ygeno.2006.03.010.

Yu, J.-Y., S. L. DeRuiter, and D. L. Turner. 2002. "RNA Interference by Expression of Short-Interfering RNAs and Hairpin RNAs in Mammalian Cells." *Proceedings of the National Academy of Sciences* 99 (9):6047–52. https://doi.org/10.1073/pnas.092143499.

Zappia, Maria Paula, and Maxim V Frolov. 2016. "E2F Function in Muscle Growth Is Necessary and Sufficient for Viability in Drosophila." *Nat Commun* 7:10509. Nature Publishing Group:1–16. https://doi.org/10.1038/ncomms10509.