EviNet: a web platform for network enrichment analysis of flexibly defined gene sets

ABSTRACT

The purpose of the new web resource EviNet is to provide an easily run interface to network enrichment analysis on novel, experimentally defined gene sets. The major advantages of this version of enrichment analysis are 1) applicability to any genes found in the global network rather than to only those with pathway/GO term annotations, 2) ability to connect genes via different molecular mechanisms rather than within one high-throughput platform, and 3) a statistical power sufficient to detect enrichment of very small sets, down to individual genes. The users' gene sets are either defined prior to the upload or derived interactively from an uploaded file by adjusting filtering criteria on differential expression values. The pathways and networks used in the analysis are chosen from available collections. The calculation is typically done within seconds or minutes, while the stable URL is provided immediately. The results are presented in both visual and tabular formats using jQuery libraries (no Java and Flash plugins required). Uploaded data and analysis results are kept in separate project directories and are not accessible by other users. EviNet is available at https://www.evinet.org/.

INTRODUCTION

Interactions between genes, proteins, and other biological molecules enable the existence of living organisms. Using global maps of such interactions, called networks or interactomes, has become a crucial step in modern analysis of biological data.

We and others have developed Bayesian tools that allow reconstructing global networks of functional coupling via integration of evidence from multiple high-throughput platforms and literature (1), (2), (3), (4). Further, testing biological hypotheses requires methods of statistically adequate network analysis.

We proposed the method of network enrichment analysis (NEA) (5), (6), where network topology was employed to evaluate functional impact of experimentally determined genes and gene sets. NEA became a natural extension of the gene set enrichment analysis (GSEA) into the interactomics domain. The GSEA methodology (7) has been well developed during the last decade. It utilizes the abundance of known functional gene sets (FGS), such as pathways, to characterize novel, experimentally defined altered gene sets (AGS). This is done by finding overrepresentation (typically gene set overlap) of the FGS in the AGS and testing its statistical significance. Performance and applicability of GSEA have been limited by incomplete pathway annotation of gene space and by only considering alterations observable within one platform, such as a transcriptomics microarray. NEA largely overcomes these limitations of GSEA due to a key difference: while GSEA counts the number of genes shared between an experimental list and a pathway, NEA considers network edges between any genes of the both groups in the global network. To the best of our knowledge, this feature is surprisingly absent in all previously proposed algorithms for network enrichment analysis. Indeed, methods such as that of Ingenuity Pathway Analysis [Ingenuity® Systems, http://www.ingenuity.com], PheNetic (8), SteinerNet (9), ResponseNet (10), (11) identify, in various ways, network modules (clusters, sub-networks etc.) that appear relatively rich in altered genes. Then GSEA might be applied post-hoc to evaluate gene set overlap between the module and each of the tested pathways. The method of EnrichNet (12) is likely to be the closest to the idea of EviNet in the way it connects AGS and FGS via network paths (although the latter allowed here to be multi-step and thus might be imprecise in an imperfect network). We believe that the state-of-the-art global networks are dense enough (hence small-world) for a single-step analysis to be sufficient. This approach also makes the analysis much less noisy, transparent, and fast. Thus EviNet mades the analysis maximally similar to GSEA and at the same time statistically solid, straightforward, and transparent for the users.

Novel users of enrichment analysis tools are looking for 1) biological interpretability and transparency, 2) a clearly defined, sequential analytic procedure, 3) statistically rigorous hypothesis testing, and 4) high-quality visualization of findings. The purpose of the new resource EviNet (available at https://www.evinet.org/) is to provide a user-friendly interface for gene network analysis in both hypothesis-driven and hypothesis-free research. It has been employed in a number of research projects as well as a platform for teaching systems biology.

DATA INPUT

An algorithm of gene set enrichment analysis (GSEA) requires two components: set(s) of experimentally derived genes/ proteins that we term “altered gene sets” (AGS), and a collection of gene/protein sets with previously characterized common function (“functional gene sets”, FGS). Network enrichment analyses (NEA) in addition need a third component: a network in which edges are represented by functional couplings,

interactions, regulatory relationships etc. between genes and/or protein nodes (Fig. 1). In order to be unbiased, such a network should be global, that is include all available nodes and edges rather than e.g. only those relevant to the analyzed AGS. In the following sections we describe these three components and ways to define them for a particular analysis. Beyond that, new users of the web site can begin by initiating the animated demo analysis. Flowcharts under the tab *Help and download* explain the major analysis components and typical applications.

Altered gene sets

A typical user input to NEA should be one or multiple AGS, in the form of gene or protein lists. Such lists can be submitted in one of three major ways, represented with three horizontal sections at the tab *Altered gene sets*: 1) a single list to be typed in or pasted in the text box (a single gene ID in the simplest case) ; 2) an uploaded file with predefined single or multiple lists (multiple list IDs must be present in a dedicated column); and 3) an uploaded file with results of differential expression (DE) analysis that allows re-defining the lists by changing the DE criteria. For the latter option, the file must have a header in a standardized format. Upon the file upload, the header is rendered into a set of web form controls, which allows to simultaneously consider two, three, or four DE contrasts by watching overlaps between respective DE lists.

For example, a user might possess transcriptomics data from 3 experiments X, Y, Z, and a control condition C. A DE analysis has been done in advance using convenient software tools, given available replicates, which resulted in fold change values and (adjusted) p-values for each contrast of interest: X vs. C, Y vs. C, and Z vs. C. The DE lists can now be flexibly defined by choosing different criteria (fold change and/or p-value) and respective cut-off values. After setting the criteria, the user can generate a Venn diagram of the three DE sets. This reveals gene groups that might be of particular interest in the given experimental design, such as those differentially expressed e.g. in X vs. C and Y vs. C but not between Z vs. C (dubbed "++-"). Given that the criteria were set in a sensible way, there will likely be multiple genes in each group ("-+-", "-++" etc.), including the full overlap of DE genes between the three contrasts "+++". The respective Venn intersections are clickable, resulting in pop-up windows containing sortable and searchable tables of genes with respective DE values. These DE lists can be selected as AGS input for NEA. In our experience, there is a frequent need to find overlaps in complex DE analyses and accompany it with enrichment analysis. In section “Transcriptomic changes during stem cell differentiation” we present an analysis of cell differentiation stages compared to their original embryonic stem cell state.

Network

Research articles that present a network analysis typically use one or a few out of the broad range of previously published global networks. The most popular categories are networks of physically interacting proteins (PPI) as well as networks from extensive integration of published interactions and high-throughput experiments (3), (4), (2), (14), (15). Gene regulatory networks inferred via reverse engineering of transcriptomics data (16) are less popular due to their usually smaller size. In order to approach the network analysis from a specific angle or consider a particular molecular mechanism, one can use functional links between genes and/or proteins from curated databases of protein complexes (17), signaling and metabolic pathways (18), (19), protein phosphorylation (20), transcription factor binding (21) etc.

We provide a menu of such networks as well as NEA-based benchmark results as described by Merid et al.(6) (section “Sensitivity vs. Specificity...” in the *Network* tab). The benchmarking is presented as ROC curves of sensitivity versus specificity in finding member genes of known FGSs. Evidently, smaller networks have lower sensitivity due to having fewer pathway gene nodes and respective edges. On the other hand, larger networks do gain from data integration: e.g. a FunCoup network is superior to the pure PPI network iRefDB (22). At the same time, the STRING database, being largely based on prokaryotic evidence (2), performs best on the metabolic pathways but in the cancer domain it appears inferior to FunCoup (purely eukaryotic evidence). By considering these benchmarks the user can select the most suitable network, depending on the research question. In addition, it is possible to choose any combination of networks and merge them into a custom network union.

Functional gene sets

The usage of FGS is most similar to that in the GSEA methodology. Each FGS (typically a pathway or a Gene Ontology term) can be viewed as a dimension in a functional space. Upon choosing one of the collections (currently Reactome (19), KEGG (18), BioCarta (23), MetaCyc (24), WikiPathways (25), and Gene Ontology terms (26)), the resulting network enrichment scores will be used to place each analyzed AGS in the respective functional space. Thus the NEA output can be utilized for either an exploratory analysis (typically when there are few AGSs) or as a feature set potentially predictive of e.g. a clinical phenotype. In order to enable the latter, each patient in the cohort has to be represented with a specific AGS.

We also provide custom collections used in our previous research projects, such as cancer-relevant FGS as well those related to cytokine/chemokine signaling and inter-cellular communications. Alternatively, users can submit their own genes as a custom FGS via the text box in the same way as it is done for an AGS. Optionally, maximum and minimum gene set cardinality can be limited before the analysis.

ANALYSIS, ITS OUTPUT AND OPTIONS

In a practical analysis, AGS, FGS, and network options are selected in the tabs described above and can then be reviewed in the *Checklist* on the tab *Check and submit*. The results can be generated by pressing the *Submit* button. Importantly, this NEA implementation does not involve multiple network randomization runs and is, therefore, fast. A randomization step was needed in earlier algorithms (5), (27) in order to estimate errors for the ACG-FGS connectivity statistics. However, the NEA procedure at EviNet.org has been streamlined, i.e. it is focused only on direct connections between AGS and FGS nodes, which does not require calculating complex network statistics such as connectivity through shared neighbor nodes, shortest paths etc. This makes it possible to employ a simple binomial formula that calculates connectivity expected by chance between the given AGS and FGS on fly, from their cumulative global connectivities (the sums of individual node degrees, i.e. connectivity values of the member nodes). This expected ACG-FGS connectivity value is then compared to the actual observed value in the chi-squared test (28), (13). The NEA software used by EviNet.org is the same as in Merid et al. (6). Thus, the time needed for the whole analysis ranges from around 10 s (required to load a network file of a million edges) to around 90 s for an analysis of 10 large (>500 genes each) AGSs versus 60 KEGG FGSs. If the analysis takes longer, the user can either bookmark the permanent URL in order to access it later or submit an email address in order to receive the notification upon job completion.

The output appears in both graphical and tabular forms. The tabular output provides an overview of general and AGS-FGS specific connectivity values as well as chi-squared statistics and adjusted p-values. In fact, the applied adjustment for multiple testing by Benjamini and Hochberg (29) is only relevant on large numbers of AGS vs. FGS combinations. Otherwise the adjusted values should approach the original p-values. For comparison with GSEA, one can look in the second last column *Shared genes* which displays the number of genes belonging to both AGS and FGS, as well the overlap p-value estimated with the binomial GSEA (formally significant cases are labeled with asterisk). This reveals the much higher sensitivity of NEA compared to GSEA: a typical gene set overlap does not exceed 10 genes (mostly 0 or 1), while there are tens and hundreds of AGS-FGS network edges (column *linksAGS2FGS*). The last column of the table contains buttons to open sub-networks behind the given network enrichments, i.e. nodes and edges that connect the AGS-FGS pairs. Another way to see those genes and respective numbers of AGS-FGS edges is to mouse-over in the columns *genesAGS* and *genesFGS*.

The graphical output is enabled with the jQuery plugin Cytoscape.js (https://github.com/cytoscape/cytoscape.js), which involves neither Java nor Abobe Flash engines and is thus sufficiently simple and robust. The accompanying menu provides a comprehensive control over the graph features (edge/node content, layout, naming, coloring, size), which allows preparing presentation-quality figures and saving them in the PNG format.

The *Archive* tab displays a list of previous analyses in the current project with applied parameters and an URL to restore them.

Analysis of individual genes

A unique feature of our NEA is that it is applicable to single node AGS or FGS. Indeed, an analysis based on estimating gene set overlap would not be applicable to such tasks. On the contrary, estimating significance of network connectivity of a single node against a multi-node set is perfectly possible. Of course, not every network node would have any network connections to an arbitrary node set. However, the expected number of network edges is usually a positive value which, when compared to the actual zero, would indicate a certain level of depletion (as opposed to enrichment). In reality almost any single node AGS is enriched against certain FGSs and vice versa. This feature can help in describing poorly characterized genes in the functional space, determining potential impact of cancer mutations etc. The NEA can be turned gene-specific with check boxes at *Check and submit* or, alternatively, by submitting single nodes as AGSs or FGSs in the first place.

TRANSCRIPTOMIC CHANGES DURING STEM CELL DIFFERENTIATION

In order to test and demonstrate the web server functionality, we used an RNA sequencing dataset generated from mouse embryonic stem cells (mESC) cultured in non-differentiating conditions and from mESC cultured for 1, 2, or 3 days in differentiating conditions (DDC) using a protocol that drives differentiation towards a ventral hindbrain neural progenitor identity (30). mESCs are pluripotent stem cells that can give rise to ectodermal, endodermal, and mesodermal cell lineages when cultured in appropriate differentiation conditions (31). mESCs treated with retinoic acid (RA) and a Shh agonist (SAG) recapitulate ventral hindbrain development and produce a population of neural progenitors with a minor presence of other cell types. Both the cell identities in this low abundant population and the signaling cascades that govern this process are unknown. However, this diversification occurs in a largely uniform mESC culture upon treatment with the same

morphogen cocktail. We aimed at identifying the cell types and signaling cascades that are induced in the mESCs upon RA and SAG treatment.

First, we analyzed the transcriptome differences between mESC in non-differentiating and differentiating conditions. We detected a dramatic and progressive downregulation of key regulators of pluripotency (Nanog, POU5F1) and upregulation of neuroectodermal (Sox1, Sox3) and endodermal (Sox17, HNF1β) markers but not of mesoderm-specific genes (32), (33), (34). In order to find genes involved in pluripotent state maintenance or differentiation progression we generated a Venn diagram that compared three DE lists: mESC vs. 1DDC, mESC vs. 2DDC, and mESC vs. 3DDC. This analysis identified genes that were either progressively downregulated or upregulated from 0 to 3DDC. At the Venn diagrams, the full intersection of the three DE contrasts delineated 65 genes putatively involved in the maintenance of pluripotent state (i.e. genes downregulated throughout 1, 2, 3 DDC (log2(FC) < -2; FDR<0.05) and 151 genes of the differentiation progression (i.e. genes upregulated throughout 1, 2, 3 DDC; log2(FC) > 2 ; FDR<0.05). These two AGSs were then used as input to NEA using the collection of signaling KEGG pathways as FGSs. This NEA (Fig. 2A) detected enrichment of signaling pathways previously associated with either the pluripotent state, such as JAK-STAT, TGF-beta, MAPK and insulin signaling pathways (35), (36), (37) or with the differentiation state, such as TGF-beta, NOTCH, WNT, hedgehog, MAPK signaling pathways (38), (39). Apparently, certain pathways were specific to only one of the two AGSs, such as the NOTCH pathway in differentiating stem cells. For comparison, the trivial GSEA detected only one pathway as enriched with a formally significant p-value (TGF-beta signaling pathway in the pluripotent state). Importantly, the signaling pathways used in the NEA often significantly overlapped with each other, so that it could be same genes behind enrichment of multiple FGS. As an example, Spry4 (sprouty homolog 4) (40) was downregulated 5-10 fold in the 1, 2, 3 DDC, as compared to mESC. This gene was connected in the global network to Cd37, Cyp2j5, Irak3, Klf5, Sgk2, and thus contributed, among other downregulated genes, to enrichment scores of respective FGSs. On the other hand, individual important FGS genes could also be connected to multiple AGS genes, thus being likely regulators of the up- or down-regulation processes. We identified such AGS genes by using the option “Analyze the FGS genes/proteins individually” at *Check and submit* tab (Fig. 2B). As expected, only a minor fraction (N=69) of the total set of individual signaling KEGG genes (N=3819) was highly enriched against the same two AGSs. Further, only one signaling KEGG gene (Pdgfra) was enriched against both the up- and down-regulated AGSs, whereas all the others had clearly distinct roles. Among the genes most connected to the downregulation were Fgfr1, Fgfr 2, Fgfr 4; Tgf1, Tgf 2, Tgf3, TgfA, Igf1r, and Pdgfrb. On the upregulation side, the most connected were Gria4, Grm3, Snap25, Camk2a, Shh, Ihh, Dhh and others, which partially explains the pathway pattern seen at Fig. 2A. Again, we emphasize that many FGS member genes that contributed to respective enrichment scores were not DE themselves.

At the second stage, we searched for FGSs that might regulate the selection between neuroectoderm or endoderm differentiation lineages. In order to do that, we used transcriptomics data for the neuroectodern population. The latter was derived by sorting 3 DDC culture using the cell surface marker Prominin-1 (30). In this case the DE values reflected a technical difference, i.e., a mixture of two (or even multiple) fractions versus one filtered fraction, rather than a transcriptional shift due to a biological transformation. Genes with expression values different between the non-sorted and sorted populations were used as a fourth contrast in addition to the ones described above (i.e., those involved in the differentiation progression). We separately retrieved gene lists with positive and negative fold changes in all the 4 contrasts, which provided us with AGSs

specific for neuroectodermal (N=79) and endodermal (N=43) lineages (Fig. 3A). There were FGSs both commonly and specifically enriched against the neuroectodermal and endodermal AGSs. Some of these pathways have been well-known, while others appeared novel and potentially interesting. For example, NOTCH, WNT, SHH, MAPK, the calcium signaling and dorso-ventral axis specification pathways were highly enriched against the neuroectoderm AGS. JAK-STAT, insulin, toll-receptor, TGF-beta, WNT, and calcium signaling pathways were connected to the endoderm specification.

CONSLUSION

We introduced a novel web implementation of our method for network enrichment analysis. We intentionally streamlined the algorithm by 1) considering only single-step, direct links between AGS and FGS genes, 2) ignoring intra-FGS and intra-AGS edges, 3) disregarding edge confidence weights and directionality (although inexistent in e.g. protein complexes), and 4) replacing the network randomization step with the unbiased binomial calculation of expected connectivity.

Compared to similar resources, the data flow and output of EviNet appears the most similar to the classical GSEA. On the other hand, due to a much higher statistical power it can be used for purposes other than an exploratory analysis, such as evaluation of candidate disease genes, testing cancer mutations for being drivers, or building prognostic and predictive statistical models on patient cohorts.

We incorporated ancillary functionality for flexible re-definition of DE gene list to be submitted to NEA. The Venn diagram tool facilitated understanding and analysis of the complex high-throughput experimental design. In summary, the significant functional connections identified by EviNet in the differentiating mESC suggested potentially important roles in the integration of signaling for specification between neuroectoderm and endoderm cell fates.

Analysis of differential expression

In order to compensate for heteroscedasticity (the correlation between mean and variance of gene expression profiles), the sample-specific count values were processed in R package limma with function voom. Further after the empirical Bayesian variance interpretation, the DE and adjusted p-values were calculated with function topFC.

Data availability

The file with count, voom, and differential expression values for this experimental series is available at EviNet.org in the project “venn” and in the tab *Download and help*.

VennDiagrams

The Venn diagrams allowed overlapping pairwise comparisons of different experimental contrasts. Such cross-comparisons allow the user to focus on gene sets which might characterize specific processes. The server-side script re-reads the file with DE values and generates a list of all genes that satisfy each contrast-specific set of filtering conditions. Each list corresponds to one circle on the Venn diagram (when the latter has been generated with R package Vennerable). Further, all possible overlaps at Venn diagrams (3, 7, and 15 in 2, 3, and 4-contrast analyses, respectively) are accompanied by corresponding gene lists. The user can test different filtering criteria, which leads to re-generation of the Venn diagram and the gene lists. The latter pop up on the screen upon mouse clicks at the intersection areas. The lists are accompanied with DE values from the user-submitted file and can be investigated by sorting, gene ID search etc. Finally, the user chooses (with checkboxes) an arbitrary number of intersection gene lists as AGSs and proceeds to tabs *Network*, *Functional Gene Sets*, and *Check and submit* in order to execute NEA.

Figure 1. Data flow of network enrichment analysis on EviNet.org.

Formula for the binomial calculation is from (27).

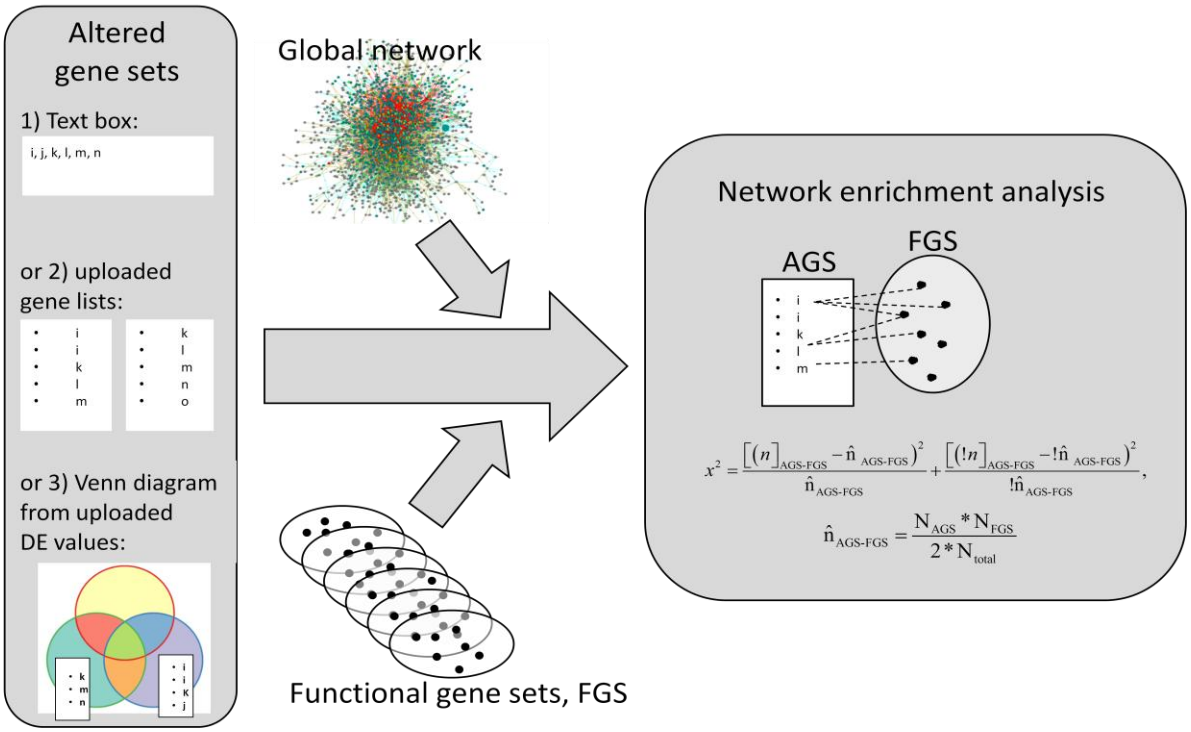


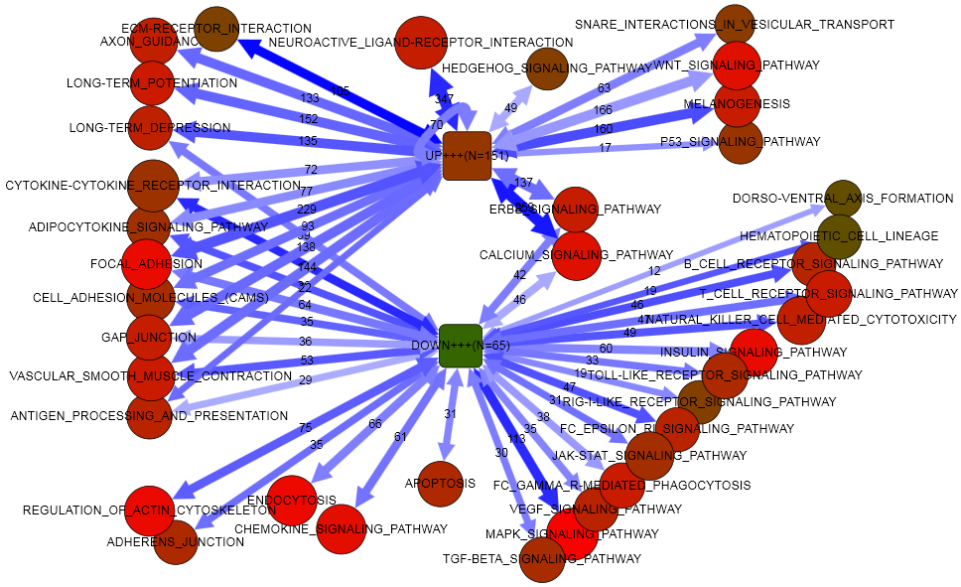
Figure 2. Network enrichment of KEGG signaling pathways against gene sets that showed ongoing up- and down-regulation during differentiation toward ventral hindbrain neural progenitor identity.

A: AGSs of 151 and 65 genes that were respectively up- and down-regulated at least 4-fold (absolute value of log2(fold change) > 2; FDR < 0.05) at each of the 1, 2, and 3 DDC compared to mESC and analyzed versus KEGG pathways as whole gene sets.

B: The same AGSs analyzed versus KEGG pathway members as individual gene nodes.

Boxes: AGSs; circles in A: KEGG signaling pathways as provided by the EviNet collection; circles in B: member genes of KEGG pathways; node coloring: cumulative "network activity" of nodes (red and green correspond to high/low sum of member genes' connectivity values in the global network); two-headed arrows: summaries of individual gene-gene links from the global network that connected AGS and FGS (gene-gene links could be undirected or have an arbitrary direction); edge labels: the number of individual gene-gene links behind each enrichment; edge transparency: confidence of network enrichment score (maximal allowed NEA FDR = 0.1 corresponds to the highest transparency).

A



B

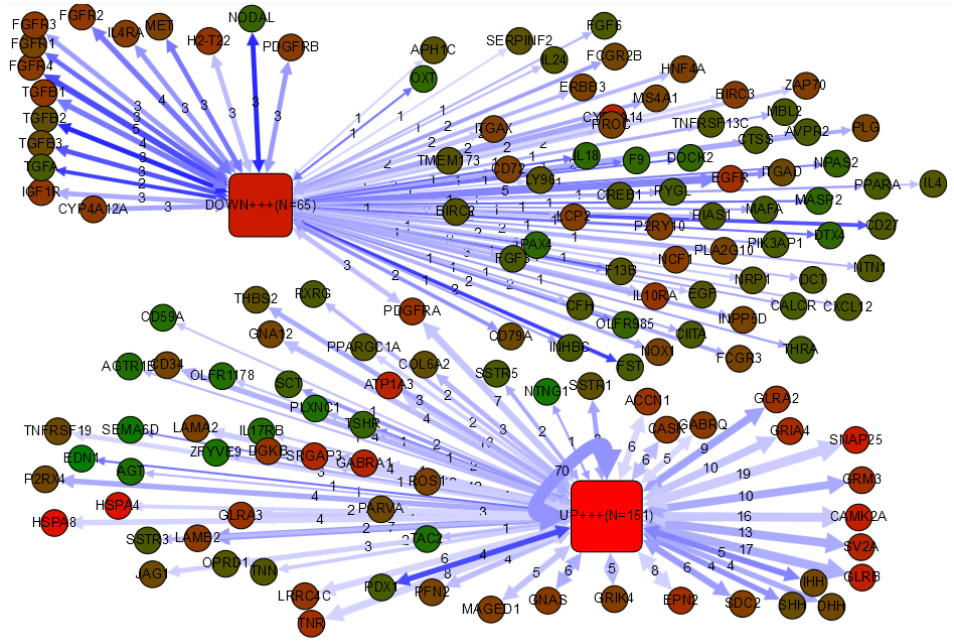


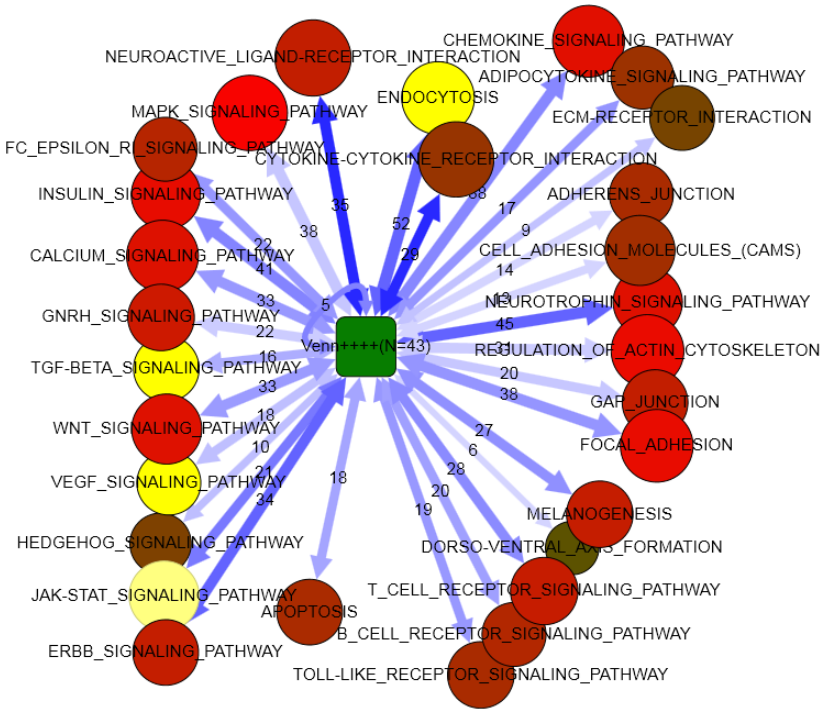
Figure 3. Network enrichment of genes characterized by up/down-regulation during differentiation and by enrichment after sorting between neuroectoderm or endoderm.

A: AGS of 43 genes upregulated at least 2, 4, 4, and 2-fold at 1 DDC vs. mESC, 2 DDC vs. mESC, 3 DDC vs. mESC, and at 3 DDC (sorted) vs. 3 DDC, respectively.

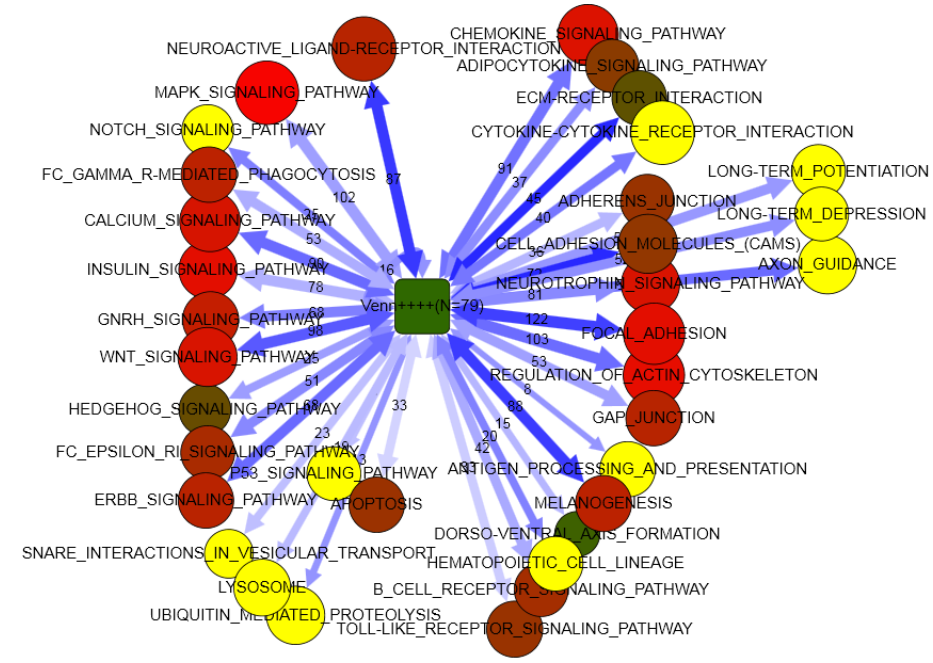
B: AGS of DE 79 genes under the same fold change thresholds as in A but with negative sign, i.e. downregulated.

Map legend: same as in Fig. 3A.

A


NEUROACTIVE_LIGAND-RECEPTOR_INTERACTION
CHEMOKINE_SIGNALING_PATHWAY
ADIPOCYTOKINE_SIGNALING_PATHWAY
ENDOCYTOSIS
MAPK_SIGNALING_PATHWAY
ECM-RECEPTOR_INTERACTION
FC_EPSILON_RI_SIGNALING_PATHWAY
CYTOKINE-CYTOKINE_RECEPTOR_INTERACTION
INSULIN_SIGNALING_PATHWAY
ADHERENS_JUNCTION
CALCIUM_SIGNALING_PATHWAY
CELL_ADHESION_MOLECULES_(CAMS)
GNRH_SIGNALING_PATHWAY
NEUROTROPHIN_SIGNALING_PATHWAY
Venn++++(N=43)
REGULATION_OF_ACTIN_CYTOSKELETON
TGF-BETA_SIGNALING_PATHWAY
GAP_JUNCTION
WNT_SIGNALING_PATHWAY
FOCAL_ADHESION
VEGF_SIGNALING_PATHWAY
MELANOGENESIS
HEDGEHOG_SIGNALING_PATHWAY
DORSO-VENTRAL_AXIS_FORMATION
JAK-STAT_SIGNALING_PATHWAY
T_CELL_RECEPTOR_SIGNALING_PATHWAY
APOPTOSIS
B_CELL_RECEPTOR_SIGNALING_PATHWAY
ERBB_SIGNALING_PATHWAY
TOLL-LIKE_RECEPTOR_SIGNALING_PATHWAY

B


NEUROACTIVE_LIGAND-RECEPTOR_INTERACTION
CHEMOKINE_SIGNALING_PATHWAY
ADIPOCYTOKINE_SIGNALING_PATHWAY
MAPK_SIGNALING_PATHWAY
ECM-RECEPTOR_INTERACTION
NOTCH_SIGNALING_PATHWAY
CYTOKINE-CYTOKINE_RECEPTOR_INTERACTION
FC_GAMMA_R-MEDIATED_PHAGOCYTOSIS
LONG-TERM_POTENTIATION
ADHERENS_JUNCTION
CALCIUM_SIGNALING_PATHWAY
LONG-TERM_DEPRESSION
CELL_ADHESION_MOLECULES_(CAMS)
INSULIN_SIGNALING_PATHWAY
AXON_GUIDANCE
NEUROTROPHIN_SIGNALING_PATHWAY
Venn++++(N=79)
GNRH_SIGNALING_PATHWAY
FOCAL_ADHESION
WNT_SIGNALING_PATHWAY
REGULATION_OF_ACTIN_CYTOSKELETON
HEDGEHOG_SIGNALING_PATHWAY
GAP_JUNCTION
FC_EPSILON_RI_SIGNALING_PATHWAY
ANTIGEN_PROCESSING_AND_PRESENTATION
P53_SIGNALING_PATHWAY
ERBB_SIGNALING_PATHWAY
APOPTOSIS
MELANOGENESIS
DORSO-VENTRAL_AXIS_FORMATION
SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT
HEMATOPOIETIC_CELL_LINEAGE
LYSOSOME
B_CELL_RECEPTOR_SIGNALING_PATHWAY
UBIQUITIN_MEDIATED_PROTEOLYSIS
TOLL-LIKE_RECEPTOR_SIGNALING_PATHWAY

REFERENCES

1. Rhodes DR, Tomlins SA, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, et al. Probabilistic model of the human protein-protein interaction network. Nat Biotechnol. 2005 Aug;23(8):951–9.

2. von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Krüger B, et al. STRING 7--recent developments in the integration and prediction of protein interactions. Nucleic Acids Res. 2007 Jan;35(Database issue):D358–62.

3. Alexeyenko A, Sonnhammer ELL. Global networks of functional coupling in eukaryotes from comprehensive data integration. Genome Res. 2009 Jun;19(6):1107–16.

4. Guan Y, Myers CL, Lu R, Lemischka IR, Bult CJ, Troyanskaya OG. A genomewide functional network for the laboratory mouse. PLoS Comput Biol. 2008;4(9):e1000165.

5. De Maeyer D, Weytjens B, Renkens J, De Raedt L, Marchal K. PheNetic: network-based interpretation of molecular profiling data. Nucleic Acids Res. 2015 Jul 1;43(W1):W244–50.

6. Tuncbag N, McCallum S, Huang S-SC, Fraenkel E. SteinerNet: a web server for integrating “omic” data to discover hidden components of response pathways. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W505–9.

7. Lan A, Smoly IY, Rapaport G, Lindquist S, Fraenkel E, Yeger-Lotem E. ResponseNet: revealing signaling and regulatory networks linking genetic and transcriptomic screening data. Nucleic Acids Res. 2011 Jul 1;39(suppl):W424–9.

8. Basha O, Tirman S, Eluk A, Yeger-Lotem E. ResponseNet2.0: revealing signaling and regulatory pathways connecting your proteins and genes--now with human data. Nucleic Acids Res. 2013 Jul 1;41(W1):W198–203.

9. Glaab E, Baudot A, Krasnogor N, Schneider R, Valencia A. EnrichNet: network-based gene set enrichment analysis. Bioinforma Oxf Engl. 2012 Sep 15;28(18):i451–7.

10. Ideker T, Sharan R. Protein networks in disease. Genome Res. 2008 Apr;18(4):644–52.

11. Myers CL, Chiriac C, Troyanskaya OG. Discovering biological networks from diverse functional genomic data. Methods Mol Biol Clifton NJ. 2009;563:157–75.

12. Zhang S-Q, Ching W-K, Tsing N-K, Leung H-Y, Guo D. A new multiple regression approach for the construction of genetic regulatory networks. Artif Intell Med. 2010 Mar;48(2-3):153–60.

13. Ruepp A, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Stransky M, et al. CORUM: the comprehensive resource of mammalian protein complexes. Nucleic Acids Res. 2007 Dec 23;36(Database):D646–50.

14. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. Nucleic Acids Res. 2002 Jan 1;30(1):42–6.

15. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014 Jan;42(Database issue):D472–7.

16. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 2012 Jan 1;40(D1):D261–70.

17. Bovolenta LA, Acencio ML, Lemke N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. BMC Genomics. 2012;13:405.

18. Merid SK, Goranskaya D, Alexeyenko A. Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis. BMC Bioinformatics. 2014;15:308.

19. Razick S, Magklaras G, Donaldson IM. iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics. 2008;9:405.

20. Nishimura D. BioCarta. Biotech Softw Internet Rep. 2001 Jun;2(3):117–20.

21. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 2014 Jan;42(D1):D459–71.

22. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: Pathway Editing for the People. PLoS Biol. 2008 Jul 22;6(7):e184.

23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25–9.

24. Alexeyenko A, Lee W, Pernemalm M, Guegan J, Dessen P, Lazar V, et al. Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. BMC Bioinformatics. 2012;13:226.

25. McCormack T, Frings O, Alexeyenko A, Sonnhammer ELL. Statistical assessment of crosstalk enrichment between gene groups in biological networks. PloS One. 2013;8(1):e54945.

26. Alexeyenko A, Wassenberg DM, Lobenhofer EK, Yen J, Linney E, Sonnhammer ELL, et al. Dynamic zebrafish interactome reveals transcriptional mechanisms of dioxin toxicity. PloS One. 2010;5(5):e10465.

27. Alexeyenko A, Alkasalias T, Pavlova T, Szekely L, Kashuba V, Rundqvist H, et al. Confrontation of fibroblasts with cancer cells in vitro: gene network analysis of transcriptome changes and differential capacity to inhibit tumor growth. J Exp Clin Cancer Res CR. 2015;34(1):62.

28. Yosef Hochberg YB. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc, Series B. 1995;1:289–300.

29. Dias JM, Alekseenko Z, Applequist JM, Ericson J. Tgfβ signaling regulates temporal neurogenesis and potency of neural stem cells in the CNS. Neuron. 2014 Dec 3;84(5):927–39.

30. Keller G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. Genes Dev. 2005 May 15;19(10):1129–55.

31. Zhao S, Nichols J, Smith AG, Li M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Mol Cell Neurosci. 2004 Nov;27(3):332–42.

32. Yasunaga M, Tada S, Torikai-Nishikawa S, Nakano Y, Okada M, Jakt LM, et al. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. Nat Biotechnol. 2005 Dec;23(12):1542–50.

33. Loh Y-H, Wu Q, Chew J-L, Vega VB, Zhang W, Chen X, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet. 2006 Apr;38(4):431–40.

34. Pan G, Thomson JA. Nanog and transcriptional networks in embryonic stem cell pluripotency. Cell Res. 2007 Jan;17(1):42–9.

35. Niwa H, Ogawa K, Shimosato D, Adachi K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature. 2009 Jul 2;460(7251):118–22.

36. Dalton S. Signaling networks in human pluripotent stem cells. Curr Opin Cell Biol. 2013 Apr;25(2):241–6.

37. Hitoshi S, Seaberg RM, Koscik C, Alexson T, Kusunoki S, Kanazawa I, et al. Primitive neural stem cells from the mammalian epiblast differentiate to definitive neural stem cells under the control of Notch signaling. Genes Dev. 2004 Aug 1;18(15):1806–11.

38. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008 Jun 13;133(6):1106–17.

39. Ikeda M, Inoue F, Ohkoshi K, Yokoyama S, Tatemizo A, Tokunaga T, et al. B-box and SPRY domain containing protein (BSPRY) is associated with the maintenance of mouse embryonic stem cell pluripotency and early embryonic development. J Reprod Dev. 2012;58(6):691–9.