BEL Commons: an environment for exploration and analysis of networks encoded in Biological Expression Language

Charles Tapley Hoyt[1,2,*], Daniel Domingo-Fernández[1,2], and Martin Hofmann-Apitius[1,2]

1. Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin 53754, Germany
2. Bonn-Aachen International Center for IT, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn 53115, Germany

***Corresponding Author**: Hoyt, C. T., Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin 53754, Germany. Telephone details; +49 2241 14-2268

Abstract

The rapid accumulation of knowledge in the field of systems and networks biology during recent years requires complex, but user-friendly and accessible web services that allow from visualization to complex algorithmic analysis. While several web applications exist with various focus on creation, revision, curation, storage, aggregation, integration, collaboration, exploration, visualization, and analysis, many of these services remain disjoint and have yet to be packaged into a cohesive environment.

Here, we present BEL Commons; an integrative knowledge discovery environment for knowledge assembly networks encoded in the Biological Expression Language. Users can submit their BEL files to be parsed, validated, and stored with fine-granular permissions. After, users can summarize, explore, and optionally shared their networks with the scientific community. We have implemented a query builder wizard to help users find the relevant portions of increasingly large and complex knowledge assembly networks and a visualization interface that allows them to explore their resulting networks. Finally, we have included a dedicated analytical service for performing data-driven analysis of knowledge assemblies to support hypothesis generation.

This web application can be freely accessed at https://bel-commons.scai.fraunhofer.de.

Introduction

While there exist a variety of modeling languages, data formats, and analytical packages for knowledge assemblies in networks and systems biology, many require deep knowledge of computer programming to use and are generally inaccessible to biologists and clinicians. With the explosion of data and knowledge in the biomedical domain, it is paramount to develop tools that foster collaboration between groups of scientists with different backgrounds and skill sets who are working towards similar goals. Already, there are multiple freely available, web-based tools for

networks and systems biology with varying focus on creation, revision, curation, storage, aggregation, integration, collaboration, exploration, visualization, and analysis.

Knowledge assemblies in systems and networks biology comprise multiple relations between biological entities across multiple scales (e.g., from the genome via the cellular and tissue level to the entire organism or population). They provide an abstraction over the concepts of pathways and mechanisms in order to encode a variety of causal, correlative, and associative relations in network representations of complex biology. In this scope, knowledge assemblies directly correspond to networks and can be handled as such.

Because of the accelerating throughput of scientific publication in the biomedical domain, several workflows (e.g., BELIEF [1], REACH [2], TRIPS [3], etc.) have been developed to automate relation extraction. Limitations on the precision and recall of these techniques motivated the development of several semi-automated relation extraction pipelines (e.g., SBV Improver [4], BELIEF Dashboard [5]) embedded in web interfaces to allow users to manually revise and curate their results. When natural language processing is insufficient, manual curation interfaces (e.g., WikiPathways [6]) can drive crowdsourced creation of knowledge assemblies.

While some useful knowledge bases (e.g., miRTarBase [7], CTD [8], KEGG [9]) disseminate their data in non-standard formats, the use of standard formats (e.g., SBML [10], BioPAX [11], BEL [12]) has become much more common as in the case of the integration effort of Pathway Commons [13]. Other web tools (e.g., NDEx [14], GraphSpace [15] provide the ability to upload, store, share, and distribute networks while remaining schema-agnostic. Most of these resources also include simple network visualization, layout, and exploration with PathVisio [16], Cytoscape [17], Cytoscape.js [18], as well as in-browser navigators.

Numerous algorithms and analyses have been published for systems and networks biology, but most are bespoke due the heterogeneous nature of their underlying knowledge assemblies, data sets, and the scientific questions motivating their development. An exception lies in enrichment analysis, which is applicable across a variety of knowledge assemblies and data sets and is embedded in several web applications for user accessibility. Even in its simplicity, enrichment analysis has the ability to produce significant biological insight [19] as exemplified by the Enrichment Map Cytoscape Plugin [20, 21] with Pathway Commons as well as GSEA [22] with MSigDB [23].

Recently, BEL has been successfully used as a semantic and modeling framework for multi-scale and multi-modal knowledge in order to investigate the aetiology of complex disease as shown by Domingo-Fernández *et al*. [24] with the release of the NeuroMMSig Mechanism Enrichment Server. While the list of published BEL-specific algorithms is currently short (e.g., Reverse Causal Reasoning [25], Network Perturbation Amplitude [26], etc.), the advent of the modern PyBEL framework [27] has improved the accessibility and utility of BEL and motivates its wider adoption. Unfortunately, unlike many of the other functions that have been implemented in web applications, algorithms generally remain confined to programmatic use and are not accessible to potential users. Last, but not

least, the ecosystem of BEL-specific web applications is small, and does not include a service for parsing, validating, and converting BEL.

There are still several unmet needs for users that motivate the development of new of web applications. Generally, there is still the need to enable complex exploration and visualization as well as to to make algorithms and analyses generally accessible and reusable. Specifically to BEL, there is a need to make parsing, validating, and converting generally accessible. Finally, an integrative knowledge discovery environment that comprises many of the previously mentioned features would be greatly beneficial to the scientific community. Here, we present BEL Commons, a web application that addresses these unmet needs and is a first attempt at building such an environment.

Components

The user interface of BEL Commons contains several components that integrate the most useful features from the variety of the previously mentioned web applications for systems and networks biology (**Figure 1**). Below, we outline their functionalities and typical use cases. Implementation details can be found in the Supplementary Information.

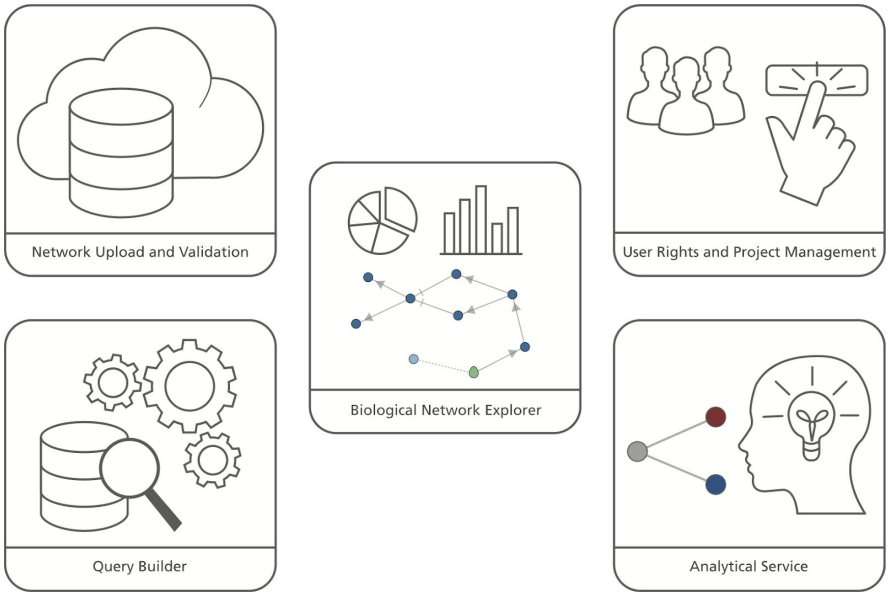


Figure 1: BEL Commons comprises several components: 1) the network uploader and validator, 2) user rights and project management, 3) the query builder, 4) the interactive biological network explorer, and 5) the analytical service.

Network Uploader

The first point of entry for many users of BEL Commons will be through its BEL uploader. Even with recent publication of a modern BEL parser (i.e., PyBEL), a web interface greatly improves accessibility for a more general

audience. The uploader contains a short form allowing users to choose a file to upload and to toggle common parsing and compilation parameters. After submitting, users' files are sent to an asynchronous task queue, implemented with RabbitMQ (https://www.rabbitmq.com) and Celery (http://www.celeryproject.org), that performs parsing, validation, and compilation in the background. The queue then performs several assessments that enumerate errors and warnings encountered during parsing, produce statistics over the resulting network, and identify biological network motifs. Finally, the queue notifies the submitter upon completion..

The parsing errors and warnings are categorized first as syntactic or semantic then with much more detail, as described in the PyBEL documentation (https://pybel.rtfd.io). Each is presented with provenance information including the line, line number, and position so curators can quickly make changes. Recurring errors and warnings are identified and grouped separately to allow curators to quickly make impactful improvements. Finally, a faceted search is presented for situations where an overwhelming number of errors and warnings are present.

The statistical summary (**Figure 2A**) of the network presents information about the contents of the network and also network-theoretic measurements of the full network. Several charts are generated depicting the types and number of nodes, edges, modifications, namespaces, annotations, and citations existing in the network. Furthermore, scalar values describing network properties such as network density and average node degree and also node overlap with other networks using the Szymkiewicz-Simpson coefficient in the database are listed.

The analysis of knowledge network motifs (**Figure 2B**) builds upon the ideas of transcriptional network motifs presented by Alon [28] and makes generalizations to apply them to knowledge networks. The simplest motifs that are informative to the robustness, correctness, and applicability of a knowledge network are based on the identification of biological contradictions. One type is the contradictory pair, where knowledge has been curated stating A increases B but also A decreases B. The ability to identify the contradiction that RB1 has been shown to increase the transcriptional activity of E2F4 by Li *et al.* [29] and that it decreases the transcriptional activity of E2F4 [30] is a significant step towards quality control during curation.

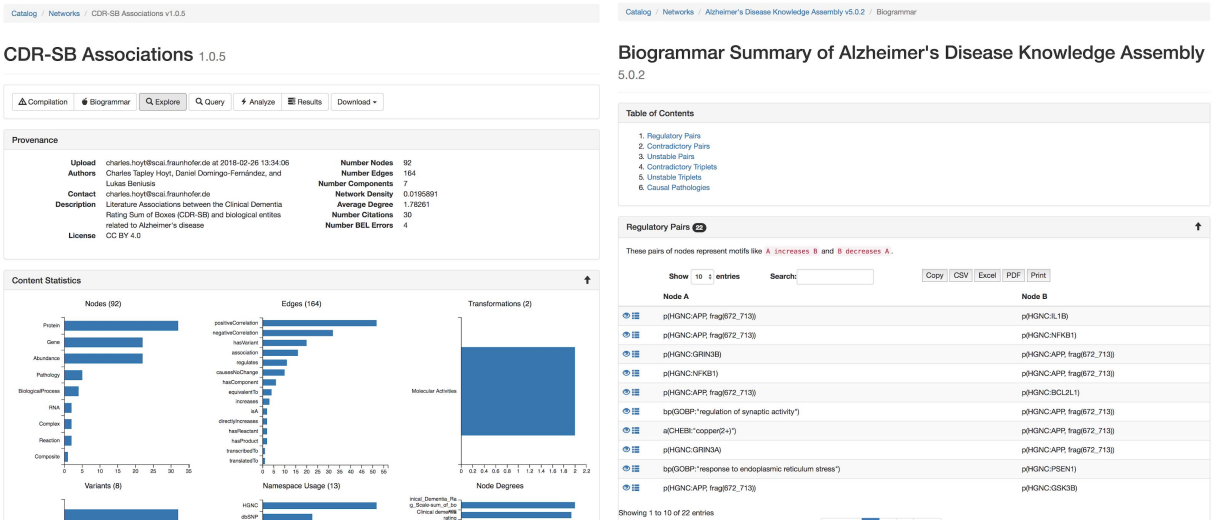


Figure 2: The statistical (A - left) and biogrammar (B - right) summary pages

User Rights Managements and Collaboration

Upon BEL upload, the user is presented with the option to make the resulting network either public or private. Networks can be uploaded privately for use during research then later released publicly to accompany a publication and share their work with the scientific community. The network catalog (**Figure 3A**) dynamically shows users only networks for which they have the appropriate permissions.

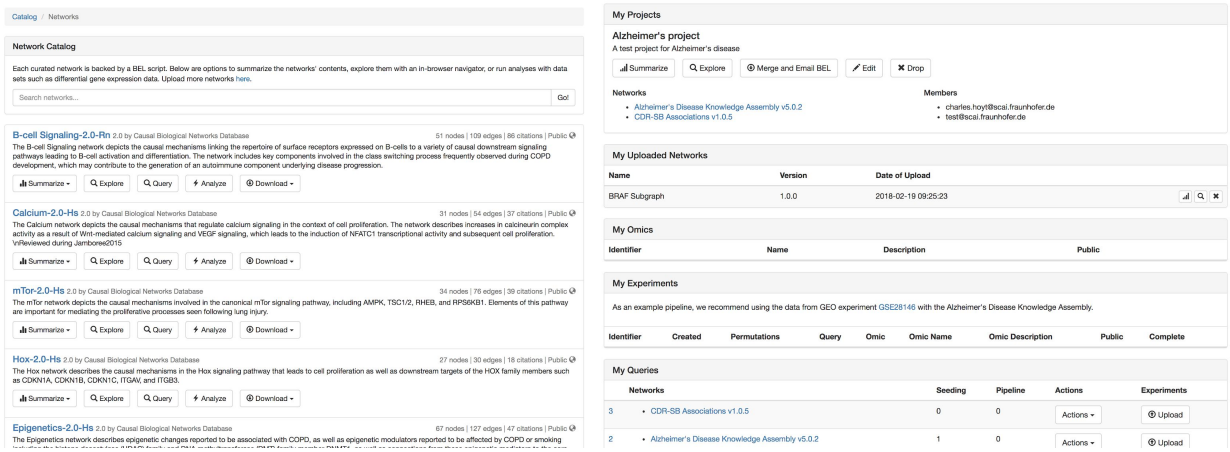


Figure 3: The network catalog (A - left) and user activity page (B -right)

Users can create projects that allow for multiple users to mutually share networks. For example, a curation project in a given disease area could contain networks generated from the efforts of multiple curators. Projects can generate a merged network that can be summarized, explored, analyzed, and exported with the same tools available for

stand-alone networks. Users can access their private activity page (**Figure 3B**), which provides a global summary over their projects, networks, queries, data sets, and experiments.

We do not presume all users plan to produce their own BEL script, especially with the growing number of both general and context-specific publicly-available knowledge assemblies. In light of this, we have included several of these resources with BEL Commons for these users (**Table 1**). The catalog of networks for which a user has the appropriate permissions can be accessed directly from the home page of BEL Commons.

Resource	**Networks**	**Nodes**	**Edges**	**Citations**
Selventa Example Corpora (http://resources.openbel.org)	5	16339	36971	5083
Causal Biological Networks Database [31]	138	5343	28766	4580
NeuroMMSig [24]	8	1411	3221	201

Table 1. Statistics over a selection of the resources publicly initially available in BEL Commons. These numbers are accurate to the best of our ability, but may not reflect nominal values from their sources depending on the ability of PyBEL to parse their contents.

Query Builder

While knowledge assemblies and their corresponding networks encoded in BEL have the benefit of being readily mergeable, it becomes increasingly difficult to visualize, explore, and interpret larger networks. Furthermore, most query interfaces are limited to exploration of the Nth neighbors of a biological entity of interest, which is often insufficient to capture complex biology. The second main point of entry for many users of BEL Commons will be through the query builder, which enables the generation of powerful, precise, and expressive queries. Each query contains three parts: assembly, seeding, and application of mutations.

The first facet of the query builder allows users to select knowledge assemblies relevant to their scientific question. While it still remains computationally feasible to query a merged view over the entire catalog of knowledge assemblies, users have the opportunity to apply their intuition and select the most relevant on the basis of their specificity towards target diseases areas, their curation methods, or any other appropriate criteria described in their metadata. Additionally, other structured knowledge resources such as protein families [32, 33], biochemical reactions [34, 35], gene orthologies (e.g., Entrez Gene [36], MGI [37], RGD [38], etc.) with high confidence can be included to enrich the curated knowledge assemblies. Later, we show how this novel feature can be used to enrich networks as a pre-processing step to connect disparate components before analysis.

The second facet of the query builder allows the user to choose from a combination of seeding methods to select nodes and edges (**Table 2**). Besides the more simple questions (e.g., which nodes are neighbor of my protein of interest?, which nodes are closely connected?), the query builder allows scientists to ask scientific questions like the

one proposed in the following scenario: the leukemia drug, nilotinib, triggers cells to remove of faulty components - including the ones associated with several brain diseases [39]. In 2015, the Georgetown University Medical Center published findings that the drug had a therapeutic effect on patients in Alzheimer's and Parkinson's diseases [40]. Though it is currently unknown, a search of the paths between this drug and these diseases could provide insight to the drug's mechanism of action.

Seed Method	**Description**
Nth Neighbors	Induces a subgraph over nodes in paths of length less than or equal to N from any query node, terminating at any node, including ones not included in the query
Upstream Subgraph	Induces a subgraph over nodes with causal edges targeting the query nodes then repeats a second time for that subgraph in order to include a second layer. Finally, induces all causal edges between resulting nodes.
Downstream Subgraph	Induces a subgraph over nodes with causal edges originating from the query nodes then repeats a second time for that subgraph in order to include a second layer. Finally, induces all causal edges between resulting nodes.
Shortest Paths	Induces a subgraph over all nodes in the shortest paths between any pair of query nodes, implemented by NetworkX [41]
All Paths	Induces a subgraph over all nodes in all of the paths (length less than seven) between any pair of query nodes, implemented by NetworkX [41]
Provenance	Builds a subgraph from all edges with provenance from articles with the given PubMed identifiers
Authors	Builds a subgraph from all edges from articles written by authors with the given names
Annotations	Builds a subgraph from all edges with the given biological or contextual annotations

Table 2. Seed methods available in the query builder.

The third and final facet of the query builder is the application of mutations (i.e., enrichments, selections, filters, and transformations) to the seed network. These vary from the enrichment of the central dogma, which extends protein nodes with their corresponding RNA and gene nodes; to the induction of purely causal subgraphs, which remove any non-causal edges; to the removal of pathology nodes, which are often uninformative in analyses; to pre-processing steps for analysis that might involve collapsing gene, RNA, and protein nodes. BEL Commons dynamically loads mutation functions from the PyBEL Tools package (https://github.com/pybel/pybel-tools), such that new functions, pipelines, and workflows for processing networks can be written quickly and made available to users. A full list is available through the PyBEL Tools documentation (https://pybel-tools.rtfd.io), or through the query builder itself.

Each query is saved with a unique identifier such that queries can be rerun, shared, merged, or compared. Rather than storing the results of queries, the assembly, seeding, and mutations are stored as a "transaction" so that they can

be applied to new assemblies, for example, when a network is updated. Effectively, queries correspond to an experimental protocol for processing raw knowledge assemblies before visualization, exploration, analysis, and interpretation. However, the construction of a query is not the end of its life. The next section summarizing the biological network explorer describes how queries can be extended and evolve during the process of scientific inquiry.

Biological Network Explorer

The third main point of entry for many users, the core feature of BEL Commons, will be its biological network explorer (**Figure 4**), which uses D3.js (https://d3js.org) to render networks with a force-directed layout algorithm that can be panned and zoomed. Because the complexity of biological networks often limits the utility of automated layouts [42], users can also manually drag and reposition nodes. Furthermore, users can adjust the edge length parameter of the algorithm to rarefy densely grouped nodes and improve readability. The networks are styled with minimum visual clutter and make use of easily-distinguishable colors rather than obtrusive shapes for nodes as well as patterns and colors for different types of edges.

The explorer has several contextual actions registered to the nodes and edges. Users can left-click a node to populate the information box below the explorer with information from external data sources (e.g., Entrez Gene, ChEBI [43], ExPASy [33], GO [44] gathered from Bio2BEL services (https://github.com/bio2bel) and the EMBL Ontology Lookup Service (OLS) [45]. Alternatively, users can right-click a node to open a contextual menu that enables further modification to the network (e.g., delete the node, add the neighbors of the node to the network) that are interactively appended to the original query used to render the visualization. The contents of the network can also be further modified by the inline query builder, which allows additional mutations (including enrichments) to be applied interactively. The query history is displayed at the bottom of the explorer, new changes are highlighted in red, and because queries are stored as transactions, changes can be reverted with an "undo" button.

When an edge is clicked, the information box is populated with relevant citations, evidences, and annotations. Each edge is linked to a voting and commenting system so domain-specific experts, curators, and bioinformaticians can engage in discussion on the correctness and robustness of the chosen representation of knowledge.

To the right of the explorer is the filter tool box, which incorporates a novel approach to filtering and exploring networks using a linked hierarchical explorer of the terminologies/ontologies annotated to the edges in the currently displayed networks. Users can search and select groups of annotations to filter the network. For example, this could be useful to exclude edges asserted from research on cell lines that are not relevant. The filter tool box has three additional tabs; nodes, edges and highlight. Users can either search for specific nodes and edges in their corresponding tabs or use the highlight tab to select nodes and edges with specific properties to highlight in the network.

Above the explorer is the general tool box that includes several additional interactions for exploration, analysis, and export of the network using the protocols described by Hoyt, *et al.* [27]. Notably, it contains a path mining tool that enables path searches between given nodes with fine-granular, configurable settings (e.g., directedness, path search algorithm, application of filters for pathologies, etc.). Hence, it can immediately be used to identify the causal root affecting two nodes, or generate hypothetical links across modes and scales.

The visualization can be further modified by resizing the nodes corresponding to the results of topological or data-driven analyses, such as their degree, betweenness centrality, or by the results of an experiment (e.g. Unbiased Candidate Mechanism Perturbation Amplitude; see below) in order to identify novel biological entities.

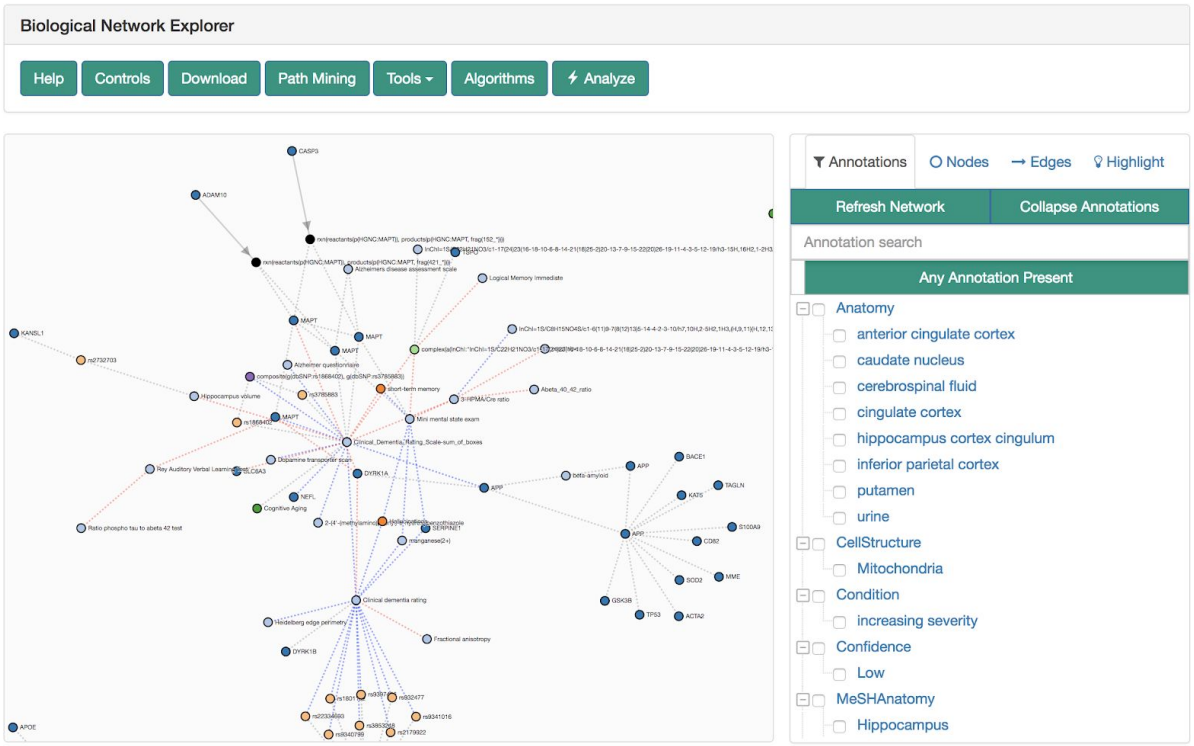


Figure 4: The biological network explorer and related navigation components. Truncated from this image are the node information and query information boxes.

Finally, there are several alternatives to exploring networks that are too large to render in-browser due to the limits of Javascript-based graphics. First, the network catalog opts to present users with a random subsampling of large networks. If a network in the explorer becomes too big, then the explorer prompts the usage of the filter tool box to identify a more relevant, smaller network. Otherwise, users can export the current network to multiple formats for use in desktop visualization applications.

Analytical Service

Algorithms for analyzing pathways and networks were categorized by Kathri in 2012 into three main categories: over-representation analysis, functional class scoring, and pathway topology [19]. These algorithms have been developed for a wide variety of applications, data formats, and graph types, but few are specific to BEL.

BEL Commons includes the newly developed Unbiased Candidate Mechanism Perturbation Amplitude (UCMPA) workflow, which begins by generating unbiased candidate mechanisms (with respect to their boundaries) based on the upstream controllers of biological process nodes in BEL networks and including their induced causal edges. It addresses issues with graph topologies (i.e., cycles, contradictory edges) and other limits (e.g., need for pre-defined subgraphs, usage of correlative relationships) posed by previous BEL-specific algorithms [26, 46] with more complex randomization and sampling approaches. Users can map pre-processed high-throughput experimental data (e.g., differential gene expression) to the network then execute a classical schema-free analytical method inspired by other heat diffusion analyses in networks biology [47, 48]. The algorithm outputs scores for each biological process based on the propagated effect of its upstream controllers, dictated by the input data. A technical description of this method can be found in the PyBEL Tools documentation (https://pybel-tools.rtfd.io). Finally, users are presented with a statistics page and are able to overlay the results on the original networks.

In the following section, the query builder, interactive explorer, and analytical service are used to assess several differential gene expression experiments representing Alzheimer's disease patients at different disease states using NeuroMMSig networks.

Application Scenario

This section describes a typical use case, where a multiple networks are queried, explored, then analyzed.

First, a sampling of networks from NeuroMMSig that represent well-known mechanisms with different progression patterns in the context of Alzheimer's disease (i.e., Low Density Lipoprotein Subgraph, GABA Subgraph, Notch Signaling Subgraph, and Reactive Oxygen Species Subgraph) were assembled with the query builder. No seeding methods were used since the analysis is general across mechanisms annotated by NeuroMMSig.

Because underlying knowledge assemblies generated by multiple curators often manifest as disjoint networks that contain related nodes, automated methods can be used to connect them. First, the central dogma was inferred for all proteins and RNAs. For example, this connected the GABRA4 gene and GABR4 mRNA from disjoint components. Further modifications (i.e., collapsing of variants, collapsing on the central dogma, enriching unqualified edges) were also applied.

As an example from the NeuroMMSig GABA Subgraph, knowledge-based approaches can be used to connect the disconnected GABBR2 to GABBR1 gene nodes to their common families using resources like PFAM [49] or InterPro [32]. Hierarchical knowledge sources like these can then be used to reason over the network, like using the

knowledge that GABBR2 decreases the cAMP catabolic process [50] to assert that GABBR1, the other member of the GPCR family 3, GABA-B receptor (IPR001828), shares the same activity. While this knowledge does not exist in the assembly, literature search also notes several connections between GABBR1 and cAMP signalling [51, 52].

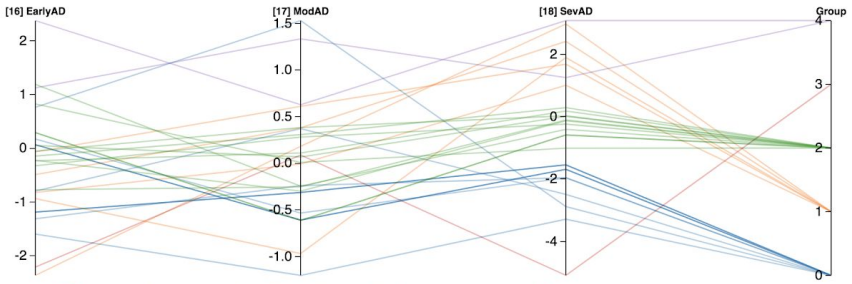
Unbiased candidate mechanisms were generated from the assembly from the upstream controllers of each biological process present and the UCMPA workflow was applied to each to assist in the interpretation of the differential gene expression experiments from GSE28146 [53] that were preprocessed with GEO2R [54]. This trial classified patients into three disease progression stages: incipient, moderate, and severe. BEL Commons includes a parallel coordinate display (**Figure 5**) to assist in comparison of multiple UCMPA experiments across different data sets and clustering their results with K-Means clustering. While BEL has inherent limits in its temporal expressivity, interpreting data that has an inherent temporal ordering helps overcome this limit.

Summary of Experiments

Identifier	Query	Permutations	Omic	Name	Description	
16	14	250	13	EarlyAD.csv	Patients from GSE28146 with early Alzheimer's disease	Delete
17	14	250	14	ModAD.csv	Patients from GSE28146 with moderate Alzheimer's disease	Delete
18	14	250	15	SevAD.csv	Patients from GSE28146 with severe Alzheimer's disease	Delete

Comparison of Experiments

Download Data as TSV



K-Means clustering with K=5 is enabled. Be aware that this is a randomized algorithm and results may not be consistent.

Type	Namespace	Name	[16] EarlyAD	[17] ModAD	[18] SevAD	Group
BiologicalProcess	GOBP	glial cell differentiation	-0.7989	0.371	-2.4916	0
BiologicalProcess	GOBP	Notch signaling pathway	-2.3671	0.1815	2.9763	1
BiologicalProcess	GOBP	lipid metabolic process	0.291	-0.6136	-0.5847	2
BiologicalProcess	GOBP	response to oxidative stress	-2.2128	0.088	-5.0908	3
BiologicalProcess	GOBP	apoptotic process	1.1241	1.3339	1.2571	4

Figure 5: The resulting UCMPA scores of biological processes. After multiple several applications of K-means clustering with *K=5*, the following consensuses emerged: Group 0 (blue) represented biological processes that decreased only at severe onset of disease (e.g., glial cell differentiation). On the other hand, Group 1 (orange) consistently increased throughout progression of disease (e.g., Notch signaling pathway). Finally, Group 2 (green) processes remained relatively unchanged (e.g. lipid metabolic process) through progression of disease and Group 4 (purple) remained consistently elevated (e.g., apoptotic process). The unabridged results of this experiment as well as a tutorial for reproduction can be found in the supplementary data.

While Alzheimer's disease must be studied with respect to its progression over time, this analysis can provide insight directly to measurements performed on a single time series. Those results provide a ranking that prioritizes the most up- and down-regulated biological processes as a function of the observed data.

Discussion

No web application, however feature-rich, will ever satiate the desire and creativity of researchers for generate novel solutions to scientific problems. While BEL Commons has taken inspiration from many well-constructed services to advance towards a more feature-rich knowledge discovery environment that enables researchers to explore knowledge and data in new ways, it still has this limit. However, we are not discouraged, and we hope to make several improvements to BEL Commons in the future:

We would like to improve the interoperability of BEL Commons and the platform build on BEL itself by integrating open authentication systems like ORCID in order to harmonize identification of users across multiple web services and provide reliable provenance for networks, queries, and analyses. We would also like to integrate further tools for converting BEL to RDF in order to connect BEL with other linked data. Further, we would like to improve exporters to other services, notably, NDEx, which have brighter outlooks on sharing and feedback systems. Recent developments in integrating INDRA [55] with PyBEL enables conversion from BioPAX documents to BEL. A future update to BEL Commons will include an option to upload these documents as well.

We would like to integrate BEL Commons with other BEL-specific systems developed with different underlying technologies. First, integrating the BELIEF Dashboard to use the underlying network and edge store from PyBEL would enable a more thorough feedback and curation interface so users could not only vote on the correctness of edges, but fix them directly. Second, the NeuroMMSig Mechanism Enrichment Server will be re-implemented completely with reusable PyBEL code and BEL Commons components in order to advance its goals of achieving patient-stratification by using common algorithms and tools.

Conclusion

Along with recent improvements in generation of BEL content through text mining (INDRA) and serialization of resources (Bio2BEL project (https://github.com/bio2bel)), we believe that BEL Commons will make BEL more accessible to both academic and industrial users. We have made this application freely available at https://bel-commons.scai.fraunhofer.de.

Authors' Contributions

C.T.H. and D.D.F. conceived the web application, implemented it, and wrote this manuscript. M.H.A. reviewed the content.

Acknowledgements

We would like to thank all colleges that assist in testing and provide feedback in order to improve this work, especially to André Gemünd for his technical assistance. We would also like to thank Scott Colby for making our logos.

Funding

This work was supported by the EU/EFPIA Innovative Medicines Initiative Joint Undertaking under AETIONOMY [grant number 115568], resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution.

Conflict of Interest

None declared.

References

1. Madan, S., Hodapp, S., Senger, P., *et al.* (2016). The BEL information extraction workflow (BELIEF): evaluation in the BioCreative V BEL and IAT track. *Database*, *2016*.
2. Valenzuela-Escárcega, M. A., Hahn-Powell, G., Hicks, T., & Surdeanu, M. (2015). A Domain-independent Rule-based Framework for Event Extraction. Proceedings of ACL-IJCNLP 2015 System Demonstrations, 127–132.
3. Allen, J. F., Swift, M., & De Beaumont, W. (2008). Deep semantic analysis of text. Proceedings of the 2008 Conference on Semantics in Text Processing STEP 08, 1, 343–354.
4. Meyer, P., Hoeng, J., Rice, J. J. , *et al*. (2012). Industrial methodology for process verification in research (IMPROVER): toward systems biology verification. *Bioinformatics*, *28*(9), 1193-1201.
5. Madan, S., Hodapp, S., & Fluck, J. (2015). BELIEF dashboard—a web-based curation interface to support generation of BEL networks. In *Proceedings of the Fifth BioCreative Challenge Evaluation Workshop* (pp. 409-417).
6. Slenter, D. N., Kutmon, M., Hanspers, K., *et al.* (2017). WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids research*, *46*(D1), D661-D667.
7. Chou, C. H., Shrestha, S., Yang, C. D., *et al*. (2017). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic acids research*, *46*(D1), D296-D302.
8. Davis, A. P., Grondin, C. J., Johnson, R. J., *et al*. (2016). The comparative toxicogenomics database: update 2017. *Nucleic acids research*, *45*(D1), D972-D978.
9. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, *45*(D1), D353-D361.
10. Hucka, M., Bergmann, F. T., Hoops, S., *et al*. (2015). The Systems Biology Markup Language (SBML): language specification for level 3 version 1 core. *Journal of integrative bioinformatics*, *12*(2), 382-549.

11. Demir, E., Cary, M. P., Paley, S., Fukuda, K., Lemer, C., Vastrik, I., … Bader, G. D. (2010). The BioPAX community standard for pathway data sharing. Nature Biotechnology, 28(12), 1308–1308.
12. Slater, T. (2014). Recent advances in modeling languages for pathway maps and computable biological networks. Drug Discovery Today, 19(2), 193–198.
13. Cerami, E. G., Gross, B. E., Demir, E., *et al*. (2010). Pathway Commons, a web resource for biological pathway data. *Nucleic acids research*, *39*(suppl_1), D685-D690.
14. Pratt, D., Chen, J., Welker, D., *et al*. (2015). NDEx, the network data exchange. *Cell systems*, *1*(4), 302-305.
15. Bharadwaj, A., Singh, D. P., Ritz, A., *et al*. (2017). GraphSpace: stimulating interdisciplinary collaborations in network biology. *Bioinformatics*, *33*(19), 3134-3136.
16. Kutmon, M., van Iersel, M. P., Bohler, A., *et al*. (2015). PathVisio 3: an extendable pathway analysis toolbox. *PLoS computational biology*, *11*(2), e1004085.
17. Shannon, P., Markiel, A., Owen Ozier, 2, Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, (13), 2498–2504.
18. Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O., & Bader, G. D. (2015). Cytoscape.js: A graph theory library for visualisation and analysis. Bioinformatics, 32(2), 309–311.
19. Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: Current approaches and outstanding challenges. PLoS Computational Biology, 8(2).
20. Isserlin, R., Merico, D., Voisin, V., & Bader, G. D. (2014). Enrichment Map–a Cytoscape app to visualize and explore OMICs pathway enrichment results. *F1000Research*, *3*.
21. Cavalli, F. M. G., *et al*. (2017). Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell, 31(6), 737–754.e6.
22. Subramanian, A., Tamayo, P., Mootha, V. K., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545-15550.
23. Liberzon, A., Subramanian, A., Pinchback, R., *et al*. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, *27*(12), 1739-1740.
24. Domingo-Fernández, D., *et al*. (2017). Multimodal Mechanistic Signatures for Neurodegenerative Diseases (NeuroMMSig): a web server for mechanism enrichment. *Bioinformatics*, *33*(22), 3679-3681.
25. Catlett, N. L., Bargnesi, A. J., Ungerer, S., *et al*.. (2013). Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data. *BMC bioinformatics*, *14*(1), 340.
26. Martin, F., Thomson, T. M., Sewer, A.,*et al*. (2012). Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks. *BMC systems biology*, *6*(1), 54.
27. Hoyt, C. T., Konotopez, A., & Ebeling, C. (2018). PyBEL: a computational framework for Biological Expression Language. Bioinformatics (Oxford, England), 34(4), 703–704.

28. Alon, U. (2007). Network motifs: theory and experimental approaches. Nature Reviews. Genetics, 8(6), 450–61.
29. Li, J. M., *et al*. (1997). E2F4-RB and E2F4-p107 complexes suppress gene expression by transforming growth factor beta through E2F binding sites. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 4948–53.
30. Kohn, K. W. (1999). Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular Biology of the Cell*, *10*(8), 2703–34.
31. Boué, S., Talikka, M., Westra, J. *et al*. (2015). Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. *Database*, *2015*, bav030.
32. Finn, R. D., Attwood, T. K., Babbitt, *et al*. (2016). InterPro in 2017—beyond protein family and domain annotations. *Nucleic acids research*, *45*(D1), D190-D199.
33. Gasteiger, E., Gattiker, A., Hoogland, *et al*. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, *31*(13), 3784-3788.
34. Placzek, S., Schomburg, I., Chang, A.,*et al*. (2016). BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic acids research*, gkw952.
35. Morgat, A., Lombardot, T., Axelsen, K. B., Aimo, L., Niknejad, A., Hyka-Nouspikel, N., … Bridge, A. (2017). Updates in Rhea – an expert curated resource of biochemical reactions. *Nucleic Acids Research*, *45*(D1), D415–D418.
36. Maglott, D., Ostell, J., Pruitt, K. D., & Tatusova, T. (2011). Entrez gene: Gene-centered information at NCBI. Nucleic Acids Research, 39(SUPPL. 1), 52–57.
37. Blake, J. A., Eppig, J. T., Kadin, J. A., Richardson, J. E., Smith, C. L., Bult, C. J., … Zhu, Y. (2017). Mouse Genome Database (MGD)-2017: Community knowledge resource for the laboratory mouse. Nucleic Acids Research, 45(D1), D723–D729.
38. Shimoyama, M., De Pons, J., Hayman, G. T., Laulederkind, S. J. F., Liu, W., Nigam, R., … Jacob, H. (2015). The Rat Genome Database 2015: Genomic, phenotypic and environmental variations and disease. Nucleic Acids Research, 43(D1), D743–D750.
39. Hebron ML, Lonskaya I, & Moussa CE (2013) Nilotinib reverses loss of dopamine neurons and improves motor behavior via autophagic degradation of alpha-synuclein in Parkinson's disease models. Hum Mol Genet, 22, 3315- 3328.
40. Pagan, F., Hebron, M., Valadez, E. H., Torres-Yaghi, Y., Huang, X., Mills, R. R., … Moussa, C. (2016). Nilotinib effects in Parkinson's disease and dementia with lewy bodies. Journal of Parkinson's Disease, 6(3), 503–517.
41. Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, (SciPy), 11–15.

42. Schreiber, F., Dwyer, T., Marriott, K., & Wybrow, M. (2009). A generic algorithm for layout of biological networks. *BMC Bioinformatics*, *10*, 1–12.
43. Hastings, J., De Matos, P., Dekker, *et al*. (2013). The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013. *Nucleic Acids Research*, *41*(D1), 456–463.
44. Carbon, S., Dietze, H., Lewis, S. E., Mungall, C. J., Munoz-Torres, M. C., Basu, S., … Westerfield, M. (2017). Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. Nucleic Acids Research, 45(D1), D331–D338.
45. Cote, R., Jones, P., Apweiler, R., & Hermjakob, H. (2006). The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics, 7, 1–7.
46. Vasilyev, D. M., Thomson, T. M., Frushour, B. P., Martin, F., & Sewer, A. (2014). An algorithm for score aggregation over causal biological networks based on random walk sampling. *BMC research notes*, *7*(1), 516.
47. Bernabò N., Barboni, B., & Maccarrone, M. (2014). The biological networks in studying cell signal transduction complexity: The examples of sperm capacitation and of endocannabinoid system. *Computational and Structural Biotechnology Journal*, *11*(18), 11–21.
48. Leiserson, M. D. M., Vandin, F., *et al*. (2015). Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature Genetics*, *47*(2), 106–14.
49. Finn, R. D., Coggill, P., Eberhardt, R. Y., *et al.* (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic acids research*, *44*(D1), D279-D285.
50. Massone, S., Vassallo, I., Fiorino, G., Castelnuovo, M., Barbieri, F., Borghi, R., … Pagano, A. (2011). 17A, a novel non-coding RNA, regulates GABA B alternative splicing and signaling in response to inflammatory stimuli and in Alzheimer disease. *Neurobiology of Disease*, *41*(2), 308–317.
51. Frère, S. G. a, & Lüthi, A. (2004). Pacemaker channels in mouse thalamocortical neurons are regulated by distinct pathways of cAMP synthesis. *The Journal of Physiology*, *554*(Pt 1), 111–125.
52. Palmer, A. A., Verbitsky, M., Suresh, R., Kamens, H. M., Reed, C. L., Li, N., … Phillips, T. J. (2005). Gene expression differences in mice divergently selected for methamphetamine sensitivity. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, *16*(5), 291–305.
53. Blalock, E. M., Buechel, H. M., Popovic, J., Geddes, J. W., & Landfield, P. W. (2011). Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease, *37*(1), 62–70.
54. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., … Soboleva, A. (2013). NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Research, 41(D1), D991–D995.
55. Gyori, B. M., Bachman, J. A., Subramanian, K., Muhlich, J. L., Galescu, L., & Sorger, P. K. (2017). From word models to executable models of signaling networks using automated assembly. *Molecular Systems Biology*, *13*(11), 954.