

# Semi-Automated Curation Allows Causal Network Model Building for the Quantification of Age-Dependent Plaque Progression in ApoE<sup>-/-</sup> Mouse

Justyna Szostak, Florian Martin, Marja Talikka, Manuel C. Peitsch and Julia Hoeng

Philip Morris International R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland.

**ABSTRACT:** The cellular and molecular mechanisms behind the process of atherosclerotic plaque destabilization are complex, and molecular data from aortic plaques are difficult to interpret. Biological network models may overcome these difficulties and precisely quantify the molecular mechanisms impacted during disease progression. The atherosclerosis plaque destabilization biological network model was constructed with the semiautomated curation pipeline, BELIEF. Cellular and molecular mechanisms promoting plaque destabilization or rupture were captured in the network model. Public transcriptomic data sets were used to demonstrate the specificity of the network model and to capture the different mechanisms that were impacted in ApoE<sup>-/-</sup> mouse aorta at 6 and 32 weeks. We concluded that network models combined with the network perturbation amplitude algorithm provide a sensitive, quantitative method to follow disease progression at the molecular level. This approach can be used to investigate and quantify molecular mechanisms during plaque progression.

**KEYWORDS:** atherosclerosis, plaque destabilization, network model, ApoE<sup>-/-</sup> mice, network perturbation amplitude

**CITATION:** Szostak et al. Semi-Automated Curation Allows Causal Network Model Building for the Quantification of Age-Dependent Plaque Progression in ApoE<sup>-/-</sup> Mouse. *Gene Regulation and Systems Biology* 2016;10 95–103 doi: 10.4137/GRSB.S40031.

**TYPE:** Original Research

**RECEIVED:** April 26, 2016. **RESUBMITTED:** August 31, 2016. **ACCEPTED FOR PUBLICATION:** August 31, 2016.

**ACADEMIC EDITOR:** Li-Na Wei, Editorial Board Member

**PEER REVIEW:** Three peer reviewers contributed to the peer review report. Reviewers' reports totaled 564 words, excluding any confidential comments to the academic editor.

**FUNDING:** The study was funded by Philip Morris Products S.A.

**COMPETING INTERESTS:** Authors disclose no potential conflicts of interest.

**CORRESPONDENCE:** Justyna.Szostak@pmi.com

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal.

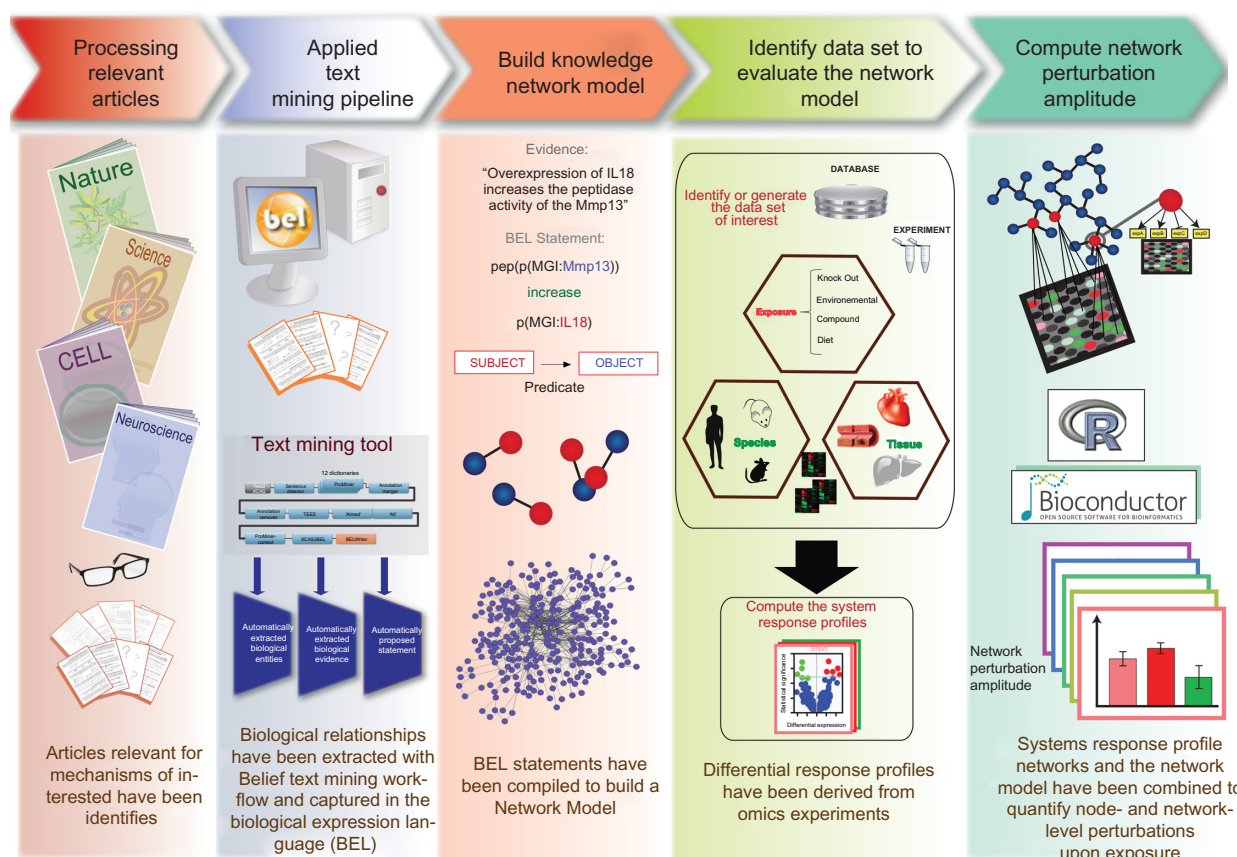
## Introduction

Quantitative methodologies to evaluate high-throughput omics data are indispensable for biological risk assessment. Frequently used methods to interpret the transcriptomic analysis and to understand the biological response upon different exposures include gene set enrichment analysis (GSEA)<sup>1</sup> or pathway analyses [eg, the Database for Annotation, Visualization and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>) and Ingenuity Pathways Analysis (IPA, <http://www.ingenuity.com/>)]. While these tools are relevant for finding the most affected biological processes or pathways in large-scale data sets,<sup>1–3</sup> the results obtained are typically qualitative and are not specific for a particular tissue type or disease. To quantify the impact of exposure or potentially affected biological processes in disease, biological network models (BNMs) have emerged as powerful and useful tools in the field of systems biology.<sup>4</sup> The major advantage of the BNM approach is to capture unstructured knowledge into interconnected and organized knowledge that describes biological processes precisely and accurately. There are several modeling languages, such as the Markup Language (SMBL: <http://sbml.org/>)<sup>5</sup> and Biological Pathways Exchange (BioPAX),<sup>6</sup> that can be employed to convert scientific results into a computable form. The Biological Expression Language (BEL) (<http://www.openbel.org/>)<sup>7,8</sup> represents scientific information

as statements with causal or noncausal relationships between biological entities.<sup>7</sup> BEL also addresses the common challenge in traditional pathway analyses, which is missing condition- and cell-specific information (Fig. 1).<sup>9</sup> Today, more than 80 commercial pharmaceutical and life science projects use BEL to build causal BNMs.<sup>10</sup>

To aid in the curation of biological knowledge into BEL, the Biological Expression Language Information Extraction Workflow (BELIEF) has an integrated text mining tool<sup>11,12</sup> and simplifies the conversion of biological entities and relationships into BEL.<sup>13–16</sup> The compiled BEL statements can be visualized in the form of a BNM that is a static representation of the extracted biological knowledge. This model can be used for the analysis and interpretation of large-scale data.<sup>17</sup> Algorithms were developed to score such network models employing the causal structure of the model as well as transcriptomic data.<sup>18–20</sup>

Here, we describe a particular network model constructed from literature knowledge focusing on atherosclerosis plaque destabilization in ApoE<sup>-/-</sup> mice. We further demonstrate that in combination with the scoring algorithm, the network model can be used to accurately quantify the impact of plaque development in aging ApoE<sup>-/-</sup> mice, which allows the interpretation of the precise mechanisms involved. Figure 1 shows the high-level workflow for the building and scoring of BNMs described in this article.



**Figure 1.** Five-step approach describing the use of the text mining tool for model network building and transcriptomics data for validation of the network.

**Notes:** *Step 1:* Relevant articles that summarize specific biological mechanisms involved in disease development were processed into a text file to allow text mining. *Step 2:* Biological relationships were extracted with BELIEF and captured in the BEL. *Step 3:* Automatically created BEL statements were assembled and compiled into knowledge network models. *Step 4:* Public data sets were identified to evaluate the network model. *Step 5:* Gene expression data were analyzed using the network model and the NPA algorithm.

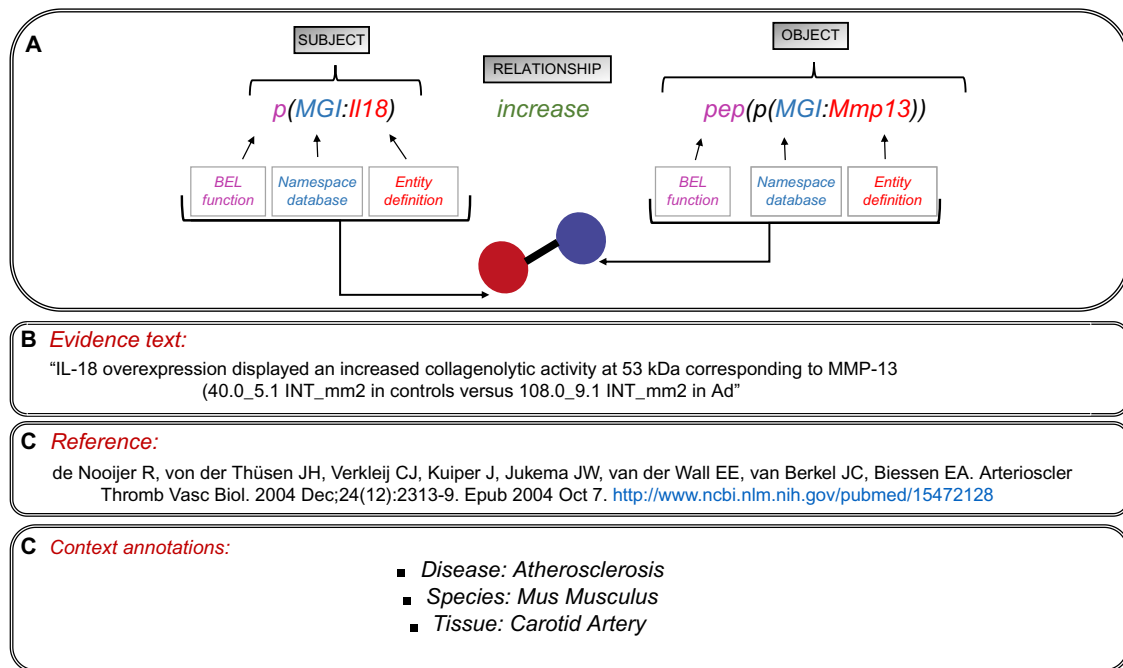
## Methods

**Biological Expression Language.** The BEL (<http://www.openbel.org>) was specially designed to represent causal relationships across a wide range of mechanistic events (eg, protein–protein functional or physical interactions, protein–diseases interactions, and protein–biological processes interactions) in the scientific literature. A BEL statement is a semantic triplet expressed in the form of subject–relationship–object, where the subject and object are coded by biological entity and where the relationship is a causal predicate (increase, decrease, directly decrease, directly increase; Fig. 2). Each BEL entity is defined by the BEL function (abundance chemical, abundance protein, gene, mRNA, abundance activity) and a referencing namespace (HUGO Gene Nomenclature Committee [HGNC], Mouse Genome Informatics [MGI], or Gene Ontology Consortium [GO]).<sup>15,21</sup> BEL also captures the experimental context and includes information about the biological and experimental system in which the relationships were observed (tissue type, species, cell line, or disease).

**The BELIEF workflow.** Thirty-three relevant articles (Supplementary File 1) derived from original experiments describing biological mechanisms involved in the plaque

destabilization process in atherosclerosis in mice with the ApoE<sup>-/-</sup> genetic background were identified using specific key words such as “atherosclerosis plaque destabilization”, “vulnerable lesion”, and “advanced lesions” in PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). After the conversion of PDF files (abstract, materials and methods, and results sections) into text files using ABBYY FineReader 11 (<http://www.abbyy.com/finereader/>), the articles were uploaded into BELIEF<sup>11,16</sup> and relevant biological terms in the text were recognized by the text mining software integrated in the BELIEF software, as described elsewhere (Fig. 2).<sup>16</sup> Each processed text document was opened in the curation interface, and the automatically assembled BEL triplets were reviewed, corrected, and annotated for the biological context (in this case, tissue, disease, cell, and species) to form a complete and accurate BEL document. To fulfill this process, the verified BEL statements in the BEL document were compiled to generate a knowledge assembly model, and exported into the eXtensible Graph Markup and Modeling Language (XGMML) format, to be visualized in Cytoscape.

**Transcriptomic data sets for verification of the atherosclerosis plaque destabilization network model.** To test the



**Figure 2.** Global representation of the BEL used in BELIEF.

**Notes:** (A) Example of a statement scripted in BEL. Biological entities (subject and object) are described as BEL functions and assigned a namespace identifier. The relationship represents the impact of the subject on the object, so the relationship has directionality. (B) Example of evidence automatically identified by the BELIEF text mining process serves as the basis for the construction of a BEL statement. (C) Example of metadata that are extracted by the text mining process.

ability of the plaque destabilization network model to provide a quantitative measure of plaque progression based on transcriptomic data from the aorta, we investigated publicly available data sets from studies describing aortic tissue affected by atherosclerosis or by nonatherosclerotic disease, as a negative control. The GSE10000 and GSE9371 data sets were obtained from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/gds>). The GSE1000 data set examined age-dependent gene expression in the aorta of ApoE<sup>-/-</sup> mice aged 6, 32, and 78 weeks versus that in C57BL/6J mice (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2626665/>). In this study, the ApoE<sup>-/-</sup> mice at 32 and 78 weeks represented a disease model with the high activation of mechanisms implicated in the atherosclerosis plaque destabilization process. The GSE9371 data set examined the role of alpha and beta estrogen receptors in the mouse aorta<sup>22</sup>; hence, such data are from the aorta of mice without the atherosclerotic disease.

#### Network perturbation amplitude (NPA) methodology.

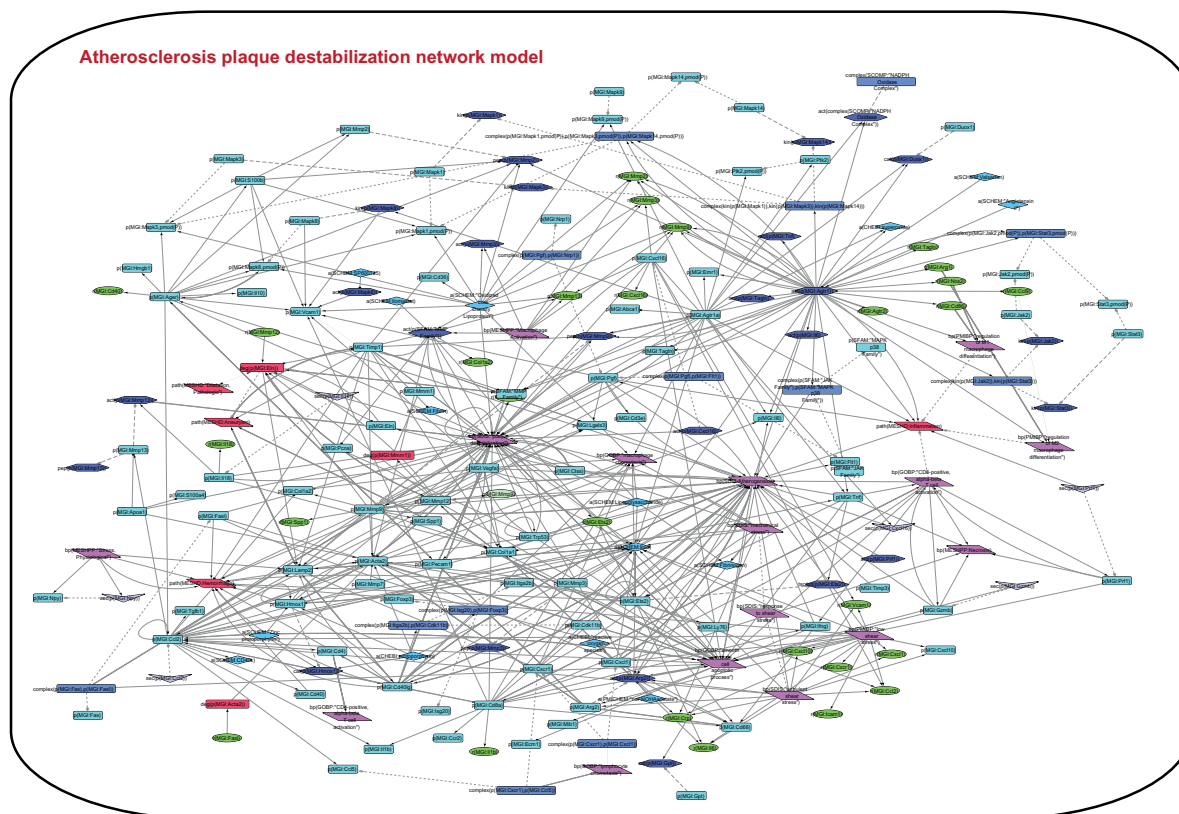
The NPA methodology is used to analyze the transcriptomics data sets in the context of the plaque destabilization network model. Briefly, network scoring exploits the functional layer described in the previous section. The signs of the relationships between the backbone and the functional layer represent the increase or decrease in the abundance of individual mRNAs and are used to infer the activation of a backbone node using transcriptomics data. Because not all backbone nodes have downstream mRNA nodes, the network models

need to be prepared for scoring to improve the specificity and relevancy of the overall NPA score. Only backbone nodes that are on a directed path that starts and ends with a node that has downstream mRNA nodes are considered. After removing the nodes that do not satisfy this criterion, the largest connected component is kept. Finally, the edges “causesNoChange” are disregarded for the scoring.

The NPA methodology<sup>23</sup> aims at contextualizing the high-dimensional transcriptomics data by combining gene expression ( $\log_2$ ) fold changes,  $\beta$ , into fewer differential node values (one value for each node of the network),  $f$ . The differential node values are determined by a fitting procedure that infers the values that best satisfy the directionality of the causal relationships (positive or negative signs) contained in the network model, while being constrained by the experimental data (the gene  $\log_2$  fold changes, which are described as downstream effects of the network itself):

$$f = L_3^{-1} L_2^T \beta,$$

where  $L$  is the signed weighted Laplacian of the network and the extra edges and nodes describing the downstream effects (to expression gene nodes),  $L_3$  is the submatrix of  $L$  for the nodes in the network, and  $L_2$  is the submatrix corresponding to the edges connecting the network nodes to the downstream gene expression nodes. The differential node values are in turn



**Figure 3.** Atherosclerosis plaque destabilization network model containing 203 nodes (colored boxes) and 511 edges (gray lines connecting nodes). **Notes:** Biological entities or nodes: red parallelogram, pathology; pink parallelogram, biological process; cyan rectangle, protein abundance; blue triangle, chemical abundance; “V” shape, cell secretion; blue hexagon, activity of protein. Relationships or edges: lines with arrows indicate positive causal relationships; lines with dark “Ts” indicate negative causal relationships; gray dashed line indicates correlative and noncausal relationships.

summarized as a single positive number, which are referred to as the amplitude of perturbation (NPA scores):

$$NPA = \frac{1}{|E|} \sum_{e \in E} (f(e_0) + \sigma(e)f(e_1))^2,$$

where  $E$  is the set of edges in the network;  $|E|$  is its size;  $\sigma(e)$  is the sign of the edge  $e$ ; and  $e_0$  and  $e_1$  denote the start and the end, respectively, of an edge  $e$ . The sum computing the NPA score can be expressed as  $f^T Q f$ , where  $Q$  is the signed Laplacian of the network where all the edge signs have been reversed.

Details of the methodology are described in a previous publication.<sup>23</sup> For the NPA scores, a confidence interval accounting for experimental variation and associated  $P$ -values are computed. In addition, companion statistics, derived to determine the specificity of the NPA score with respect to the biology described in the network model, are shown as \*O and K\* if their  $P$ -values are below the significance level (0.05). A network is considered to be significantly impacted if the three values (the confidence interval, \*O, and K\* statistics) are below 0.05.

## Results

**The network model representing mechanisms involved in atherosclerosis plaque destabilization.** The atherosclerotic

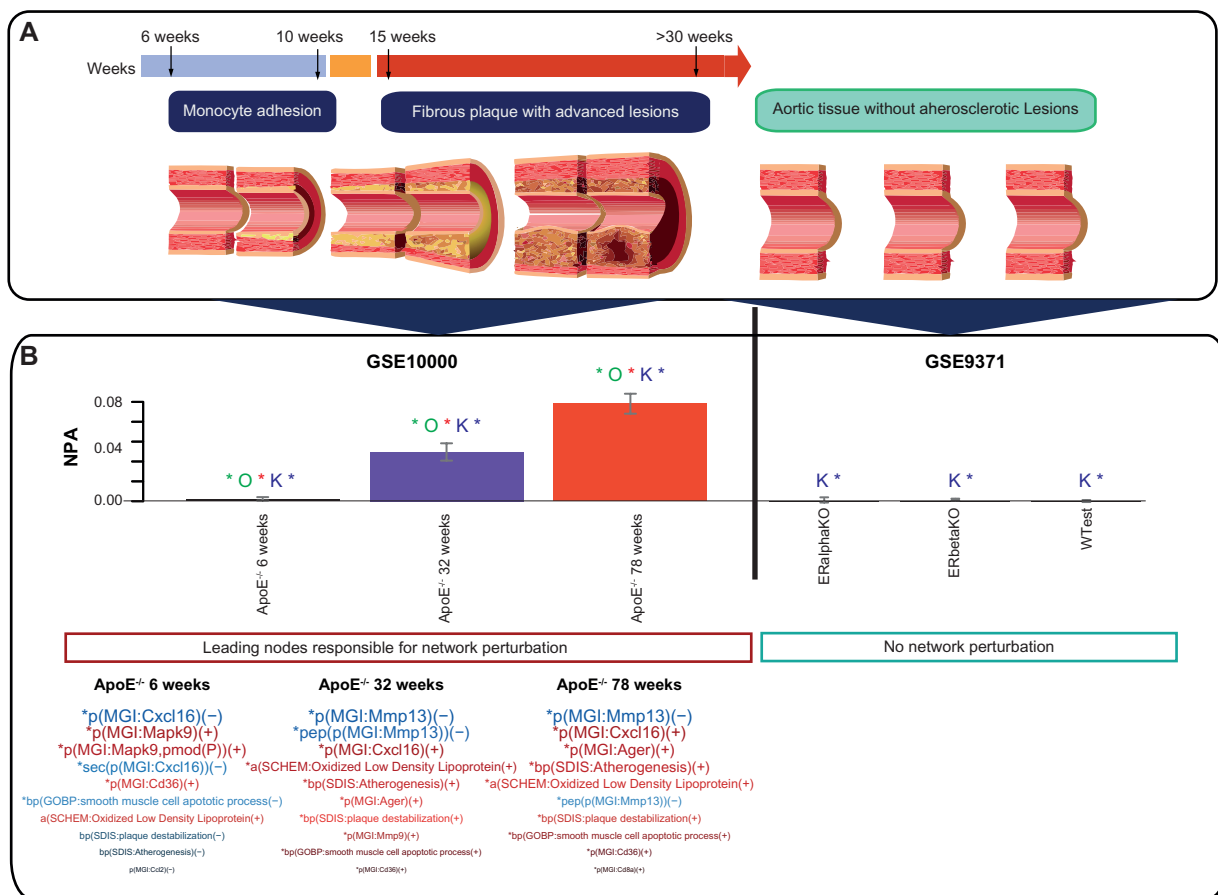
plaque network model represents a comprehensive collection of biological mechanisms that regulate the plaque destabilization process in the ApoE<sup>-/-</sup> mouse aorta.<sup>16</sup> A network model is a representation of the causal relationships between the biological entities in the determined cellular systems coded in the BEL (Fig. 2). The nodes of the network correspond to biological entities or processes and diseases; relationships are represented by edges that connect two nodes and represent the cause-and-effect relationship between the corresponding objects (Fig. 2). This first layer constitutes the backbone network. In addition to the backbone network model, the network contains a downstream layer, called the functional layer; this functional layer includes downstream nodes of a particular node from the backbone network. Downstream entities are in this case the differential expression of the genes causally affected by the considered nodes, which allows the integration of gene expression measurements into the model. This process is called “backward” reasoning and relies on the premise that the activity of the specific activities (eg, enzymes or transcription factors) is deduced from the changes in the expression of its downstream genes.<sup>23</sup> By definition, in the two-layered structure, there are no edges between genes in the transcript layer as we assume that the functional layer drives relationships between genes. The majority of these edges were

retrieved from the Selventa Knowledgebase,<sup>19</sup> and some new downstream nodes were also created via literature curation, as described above.

**Evaluation of the plaque destabilization network model.** The GSE10000 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10000>) published dataset was used to test the network model and the NPA methodology. The GSE10000 captures the gene expression changes in aortic tissue in response to aging. This study compared the aorta transcriptomes of C57BL/6/J and ApoE<sup>-/-</sup> knockout mice at different ages. In the ApoE<sup>-/-</sup> genetic background, age is considered a biological stimulus of atherogenesis and a strong inducer of the mechanisms implicated in the atherosclerotic destabilization process.<sup>24</sup> The 6-week-old ApoE<sup>-/-</sup> mice described in the GSE10000 data set had minor atherosclerotic lesions that advanced with time to become the advanced atherosclerotic lesions observed in the aorta of 32- and 78-week-old mice.<sup>24</sup> The GSE9371 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9371>) data set examined the implication of alpha and beta estrogen receptors in mouse arteries. This data set was used as a negative control to assess

the specificity of our atherosclerosis plaque destabilization network model.

Figure 4 shows the values for the perturbation amplitude for the network as a whole. The NPA score peaked at the age of 78 weeks, whereas at 6 weeks, only slight perturbation of the network was observed (Fig. 4). By contrast, evaluation of the data set from the study investigating estrogen receptors in the aorta (GSE9371) did not result in significant perturbation of the plaque destabilization network model. The key nodes contributing to the amplitude of network perturbation are referred to as the leading nodes, which are shown below the NPA bar graph (Fig. 4). Interestingly, while the NPA was substantially lower in ApoE<sup>-/-</sup> mice, at the age of 6 weeks, mitogen-activated protein kinase 9 (*Mapk9*), Cd36 antigen (*Cd36*), and “oxidation of low-density lipoprotein (LDL)” were inferred to be activated, whereas *Cxcl16* and the “smooth muscle cell apoptosis process” were inferred to be suppressed. These leading nodes were different from the ones obtained with the data from older mice in the study. The most primary controllers implicated in the atherosclerosis plaque destabilization process at 32 and 78 weeks were *Mmp13*, *Cxcl16*,



**Figure 4.** Perturbation of the atherosclerosis plaque destabilization network model as a whole.

**Notes:** (A) Quantification of the entire NPA in the aortic tissue of 6-, 32-, and 78-week-old ApoE<sup>-/-</sup> mice and in Era knock out estrogen (ERaKa) mice. \*O and K\* inform on the specificity of the NPA score with respect to the biology described in the network mode. A network is considered to be significantly impacted if the three values (the confidence interval, \*O, and K\* statistics) are below 0.05. (B) The top 10 leading nodes for each contrast are shown below the bar graph. The blue color and the negative sign mean that the activation of the leading node is decreased, and the red color indicates the increased activation.

advanced glycosylation end product-specific receptor *Ager*, *Mmp9*, and *Cd36* (Fig. 4).

The firm regulation of these biological entities was also associated with the robust activation of biological mechanisms such as “oxidation of LDL”, followed by “atherogenesis” and “plaque destabilization.”

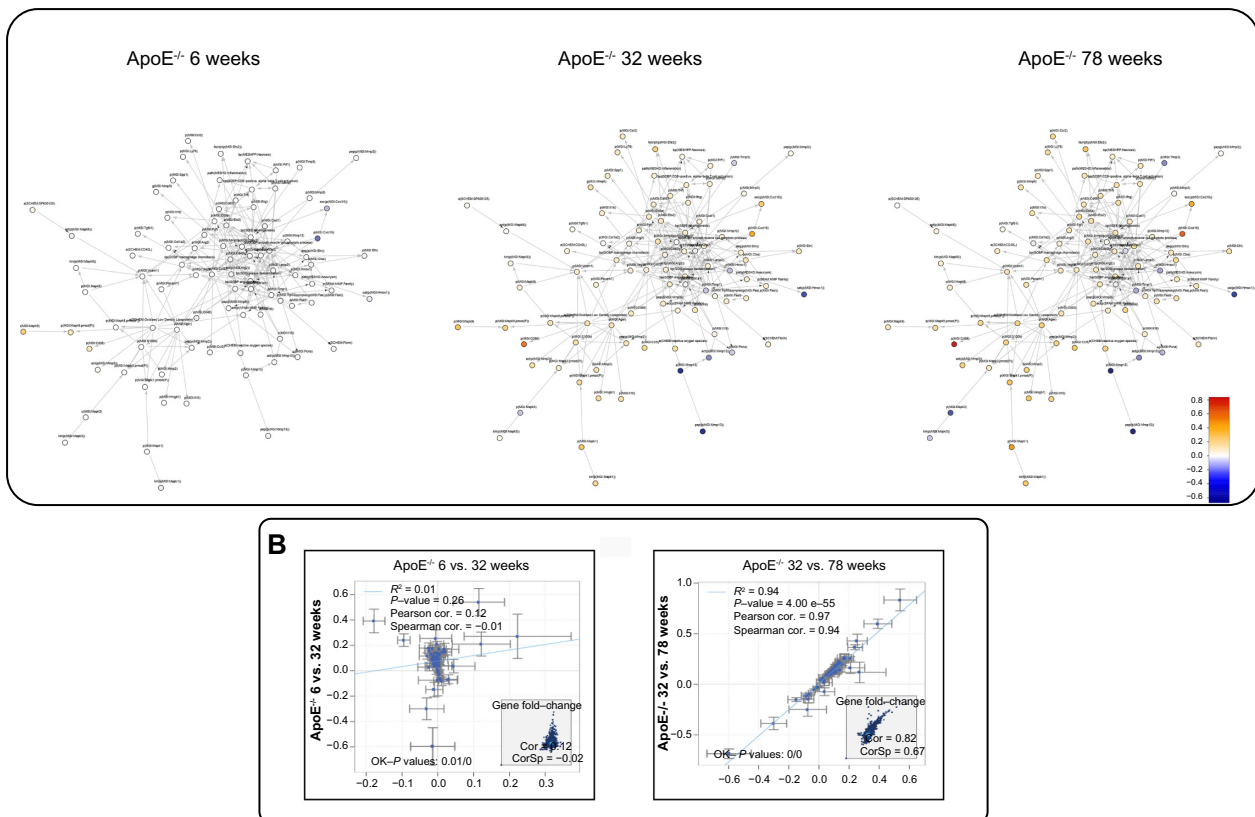
Figure 5A shows a graphic representation of the response of the network model to plaque progression in the ApoE<sup>-/-</sup> mouse aorta based on transcriptomics data. While the network model is marginally impacted when scoring the sample of 6-week-old mice, the number of affected nodes gradually increases as the mice age, and plaques develop in the ApoE<sup>-/-</sup> background. As shown in Figure 5B, the mechanisms regulated in the network at 6 and 32 weeks in ApoE<sup>-/-</sup> mice were considerably different (Pearson correlation: 0.12). This was in line with the fact that at 6 weeks, the ApoE<sup>-/-</sup> aorta does not yet show pathology, which was observed later, toward 32 weeks. By contrast, the mechanisms that were impacted at 32 and 78 weeks were quite similar, as demonstrated by a robust positive correlation (Pearson correlation: 0.97) between the activation of the backbone network model at these two ages.

Figure 6 shows a graphic representation of the leading nodes in the plaque destabilization model, when scored with transcriptomics data from the ApoE<sup>-/-</sup> mouse aorta. In 6-week-old mice, the activation of nodes, such as Cd36 and oxidized LDL, was predicted to be principally responsible for macrophages chemotaxis. At 32 weeks, the macrophages chemotaxis was accompanied by the development of inflammation through macrophages activation, which could lead to smooth muscle cell apoptosis and plaque destabilization. The mechanisms that were predicted to be impacted at 32 weeks were also the most important contributors to the effect at 78 weeks. In addition, inflammatory players, such as the cytotoxic CD8<sup>+</sup> T-lymphocytes, were predicted to be activated, which indicated progression of the plaque destabilization process.

## Discussion

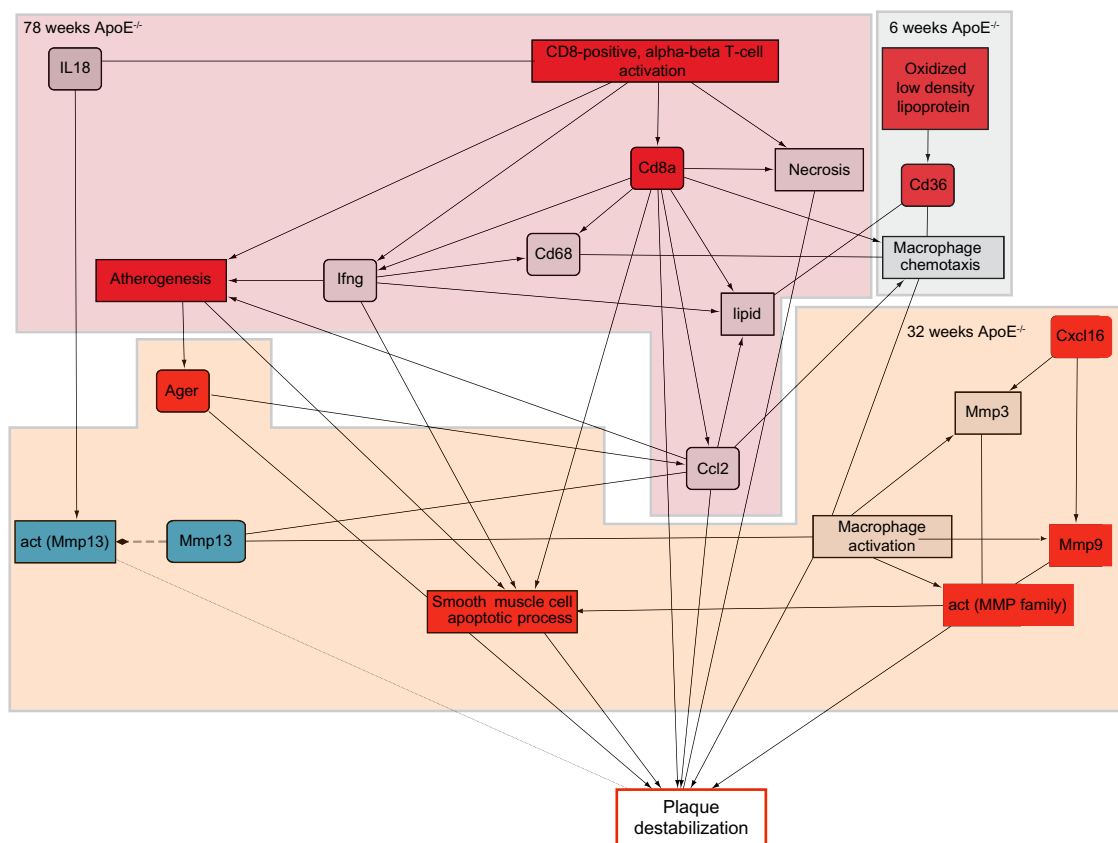
The cellular and molecular mechanisms behind the process of atherosclerosis plaque destabilization are complex and still relatively poorly understood.<sup>25–27</sup> Here, we demonstrated how a five-step approach combining semiautomated curation with transcriptomics data and a perturbation network algorithm could be used to study the process of atherosclerosis plaque

### A Backbone network value



**Figure 5.** Comparability of aortic tissue response between 6, 32, and 78 weeks.

**Notes:** (A) Differential network backbone values were generated from the GSE10000 data set, showing the impact of age on progression of the atherosclerosis plaque destabilization process illustrated by the model backbone. The blue color indicates the decreased activation of the biological entity; orange and red indicate its increased activation. The white color indicates that the biological entity is not activated. Relationships are represented by lines with arrows. (B) Correlation of different network model backbone values generated from the aortic tissue extracted from 6-, 32-, and 78-week-old mice.



**Figure 6.** Mechanistic representation of the leading nodes at 6, 32, and 78 weeks in ApoE<sup>-/-</sup> mice.

**Notes:** Blue indicates predicted downregulation, and red indicates predicted activation of the biological entity. Nodes labeled in gray did not appear as leading nodes.

destabilization in great detail at the molecular level. The central component of our analysis is the causal BNM describing molecular relationships that take place during plaque instability in ApoE<sup>-/-</sup> mouse aorta. While being a static representation, this computable causal model could serve as the basis for developing a dynamical model describing the dynamics of its molecular entities, in line with the perspectives discussed in Ref. 28. Dynamical models describing, for example, the initiation of atherosclerosis through a small number of molecular species have been previously published.<sup>29</sup> With the use of two public datasets, we have demonstrated that the atherosclerosis plaque destabilization network model is accurate to the atherosclerotic disease and is very sensitive to data from advanced lesions. However, the network model was extracted in propriety from ApoE<sup>-/-</sup> model; the mechanism of atherosclerotic disease could be used in another disease model to investigate the severity of atherosclerotic plaque. High connected nodes were captured as the backbone network model (Fig. 3) to allow the NPA scoring (Fig. 3). The results suggest that lesions observed at 78 weeks present the most elevated risk for the plaque destabilization process in ApoE<sup>-/-</sup> mice. Interestingly, the network was already slightly impacted in ApoE<sup>-/-</sup> mouse aorta at 6 weeks, when no plaques should be present. However, the significant molecular contributors to the perturbation differed from those observed for more advanced lesions,

as demonstrated by the leading node and correlation analyses. The inferred downregulation of CXCL16 in the younger mice could be related to enhanced macrophage recruitment to the aortic arch, which could promote the acceleration of the inflammatory response and atherosclerosis, as has been previously proposed.<sup>30</sup> This enhanced macrophage recruitment could be additionally enhanced through CD36 activation, as suggested by the NPA analysis. CD36 is an essential scavenger receptor strongly implicated in the capture of lipids and macrophages in the subendothelial layer; overexpression of CD36 has been shown to lead to the development of an inflammatory response mediated through lipid oxidation.<sup>31</sup> Together, these findings suggest that in ApoE<sup>-/-</sup> mice aged 6 weeks, mechanisms related to the recruitment of lipids and inflammatory cells are predominantly activated and that these mechanisms strongly involve CXCL16 and CD36 (Fig. 6). Our network model further integrated the crucial role of MMPs in the plaque destabilization process and showed that MMP13 was strongly deactivated, whereas MMP9 and AGER were strongly activated in the network model scored with 32-week ApoE<sup>-/-</sup> mouse data. Together, these results suggest that pro- and antiplaque destabilization mechanisms are regulated at 32 weeks. While MMP13 inhibition promotes collagen accumulation in atherosclerotic lesions, even in established plaques, and participates in preventing plaque rupture,<sup>32,33</sup>



the pro-rupture mechanisms related to MMP9 and AGER degrade the extracellular matrix and could promote smooth muscle cell apoptosis. However, antiplaque destabilization mechanisms seemed to be regulated at 32 weeks; the imbalance between matrix accumulation and degradation reduces plaque stability, as suggested by the activation of “smooth muscle cell apoptosis” and the “plaque destabilization process,” which occurs despite MMP13 deactivation.<sup>34</sup> At 78 weeks, the atherosclerosis plaque destabilization process had progressed further, as indicated by the reinforcement of mechanisms already inferred to be activated at 32 weeks. Despite the substantial agreement between the leading nodes activated at 32 and 78 weeks, we observed that additional mechanisms mediated by CD8 were increasingly activated at 78 weeks (Fig. 6). Cd8a is a biomarker for cytotoxic CD8<sup>+</sup> T-lymphocytes. The activation of such lymphocytes suggests that the advanced lesions at 78 weeks present infiltration of not only macrophages but also cytotoxic Cd8a<sup>+</sup> T-lymphocytes. As demonstrated in the literature, CD8<sup>+</sup> T-lymphocytes play a critical role in the plaque destabilization process; the infiltration of CD8<sup>+</sup> T-lymphocytes promotes the induction of macrophages, endothelial cells, and smooth muscle cell apoptosis, leading to plaque instability.<sup>32,33</sup> Our conclusion demonstrating the additional effects of CD8<sup>+</sup> T-lymphocytes on the plaque destabilization process is in accordance with literature results that revealed that the advanced atherosclerotic lesions promote the infiltration of CD8<sup>+</sup> T-lymphocytes, which account for plaque rupture.<sup>27,33</sup>

## Conclusion

As demonstrated in this study, BNMs in combination with the NPA algorithm provide a sensitive, quantitative method to follow disease progression at the molecular level.<sup>20</sup> In this study, we additionally demonstrated that the mechanisms playing important roles in atherosclerotic lesions at 6, 32, and 78 weeks differed markedly. While the mechanisms at 6 weeks focused on lipids and macrophage adhesion, the mechanisms in advanced lesions were strongly associated with the promotion of macrophages and Cd8<sup>+</sup> T-cell activation, maintaining the essential role of inflammatory cells in advanced lesions. The causal BNMs describing the atherosclerotic plaques destabilization could be utilized for scoring of different data sets to determine the severity of atherosclerotic lesions and could help in the discrimination of vulnerable and nonvulnerable lesions. Although we focused on the atherosclerosis plaque destabilization process, the approach described here could be used in various contexts or diseases; the NPA methodology is not restricted to a particular domain and can be widely used in pharmaceutical and toxicological contexts. In fact, semiautomated curation using BELIEF tools is a flexible approach; BELIEF integrates the gene, protein, chemical, and disease database namespaces and allows the recognition of biological entities and their associated relationships in various biological and toxicological contexts. Although BELIEF strongly supports the creation of BNMs,

direct visualization of the structured network knowledge in BELIEF is not currently possible. The BELIEF tool could benefit from a dynamic visualization tool, which is suggested as the next step to advance this approach. Finally, a key advantage of network models built using a text mining tool is that they can be regularly updated and maintained to keep the knowledge up to date.

## Acknowledgments

The authors acknowledge and thank Stephanie Boué, Alain Sewer, and Samantha Elmhurst for their contributions to this project.

## Author Contributions

Carried curation process and network building: JS. Drafted the manuscript: JS, MT. Scored the network model with transcriptomics data: FM. Participated in the design and conception of the study: JH, MCP. All the authors read and approved the final manuscript.

## Abbreviations

- AGER: advanced glycosylation end product-specific receptor
- BEL: Biological Expression Language
- BNMs: biological network models
- CXCXL16: C-X-C motif chemokine ligand 16
- CD36: CD36 antigen (collagen type I receptor, thrombospondin receptor)
- MMP13: matrix metalloproteinase 13 (collagenase 3)
- MMP9: matrix metalloproteinase 9
- MAPK9: mitogen-activated protein kinase 9
- CD8a: T-cell surface glycoprotein CD8 alpha chain
- GSEA: gene set enrichment analysis
- IPA: Ingenuity Pathways Analysis
- DAVID: Database for Annotation, Visualization and Integrated Discovery
- BELIEF: Biological Expression Language Information Extraction Workflow
- BioPAX: Biological Pathways Exchange
- NPA: network perturbation amplitude
- HGNC: HUGO Gene Nomenclature Committee
- MGI: Mouse Genome Informatics
- GO: Gene Ontology Consortium
- GEO: Gene Expression Omnibus.

## REFERENCES

1. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
2. Kutmon M, Evelo CT, Coort SL. A network biology workflow to study transcriptomics data of the diabetic liver. *BMC Genomics*. 2014;15:971.
3. Huang DW, Sherman BT, Tan Q, et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res*. 2007;35(Web Server issue):W169–75.
4. Cho DY, Kim YA, Przytycka TM. Chapter 5: network biology approach to complex diseases. *PLoS Comput Biol*. 2012;8(12):e1002820.





5. Hucka M, Finney A, Sauro HM, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. 2003;19(4):524–31.
6. Bader GD, Cary MP, Sander C. BioPAX – biological pathway data exchange format. *Structuring and Integrating Data Introductory Review*. John Wiley & Sons, Inc; 2006.
7. Slater T. Recent advances in modeling languages for pathway maps and computable biological networks. *Drug Discov Today*. 2014;19(2):193–8.
8. Slater T, Song DH. Saved by the BEL: *Ringling in a Common Language for the Life Sciences*. Vol 75. Drug Discovery World Fall; 2012:80.
9. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*. 2012;8(2):e1002375.
10. Juliane F, Marc Z. Data management and processing in toxicoinformatics – from chemical databases to automatic extraction of unstructured resources. In: Hoeng J, ed. *Computational Systems Toxicology*. New York: Springer; 2015: 109–43.
11. Fluck J, Madan S, Ansari S, et al. BELIEF – a semiautomatic workflow for BEL network creation. Proceedings of the 6th International Symposium on Semantic Mining in Biomedicine (SMBM). Aveiro, Portugal: 2014.
12. Fluck J, Mevissen HT, Dach H, Oster M, Hofmann-Apitius M. ProMiner: recognition of human gene and protein names using regularly updated dictionaries. Proceedings of the Second BioCreative Challenge Evaluation Workshop, Critical Assessment of Information Extraction in Molecular Biology. Madrid, Spain: 2007.
13. Gebel S, Lichtner RB, Frushour B, et al. Construction of a computable network model for DNA damage, autophagy, cell death, and senescence. *Bioinform Biol Insights*. 2013;7:97–117.
14. Westra JW, Schlage WK, Frushour BP, et al. Construction of a computable cell proliferation network focused on non-diseased lung cells. *BMC Syst Biol*. 2011;5:105.
15. Boue S, Talikka M, Westra JW, et al. Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. *Database*. 2015;2015:bav030.
16. Szostak J, Ansari S, Madan S, et al. Construction of biological networks from unstructured information based on a semi-automated curation workflow. *Database*. 2015;2015:bav057.
17. Aderem A. Systems biology: its practice and challenges. *Cell*. 2005;121(4):511–3.
18. Martin F, Thomson TM, Sewer A, et al. Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks. *BMC Syst Biol*. 2012;6:54.
19. Sewer A, Martin F. Quantifying the Biological Impact of Active Substances Using Causal Network Models. *Computational Systems Toxicology*. 2015: 223–56.
20. Talikka M, Boué S, Schlage WK. Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. In: Hoeng J, Peitsch MC, eds. *Computational Systems Toxicology*. New York: Springer; 2015:65–93.
21. O’Lone R, Knorr K, Jaffe IZ, et al. Estrogen receptors alpha and beta mediate distinct pathways of vascular gene expression, including genes involved in mitochondrial electron transport and generation of reactive oxygen species. *Mol Endocrinol*. 2007;21(6):1281–96.
22. Martin F, Sewer A, Talikka M, Xiang Y, Hoeng J, Peitsch MC. Quantification of biological network perturbations for mechanistic insight and diagnostics using two-layer causal models. *BMC Bioinformatics*. 2014;15:238.
23. Nakashima Y, Plump AS, Raines EW, Breslow JL, Ross R. ApoE-deficient mice develop lesions of all phases of atherosclerosis throughout the arterial tree. *Arterioscler Thromb*. 1994;14(1):133–40.
24. Hartwig H, Silvestre-Roig C, Hendrikse J, et al. Atherosclerotic plaque destabilization in mice: a comparative study. *PLoS One*. 2015;10(10):e0141019.
25. Dickson BC, Gotlieb AI. Towards understanding acute destabilization of vulnerable atherosclerotic plaques. *Cardiovasc Patol*. 2003;12(5):237–48.
26. Silvestre-Roig C, de Winther MP, Weber C, Daemen MJ, Lutgens E, Soehnlein O. Atherosclerotic plaque destabilization: mechanisms, models, and therapeutic strategies. *Circ Res*. 2014;114(1):214–26.
27. Aslanian AM, Charo IF. Targeted disruption of the scavenger receptor and chemokine CXCL16 accelerates atherosclerosis. *Circulation*. 2006;114(6):583–90.
28. Clarke MC, Figg N, Maguire JJ, et al. Apoptosis of vascular smooth muscle cells induces features of plaque vulnerability in atherosclerosis. *Nat Med*. 2006;12(9):1075–80.
29. Quillard T, Araujo HA, Franck G, Tesmenitsky Y, Libby P. Matrix metalloproteinase-13 predominates over matrix metalloproteinase-8 as the functional interstitial collagenase in mouse atheromata. *Arterioscler Thromb Vasc Biol*. 2014;34(6):1179–86.
30. Quillard T, Tesmenitsky Y, Croce K, et al. Selective inhibition of matrix metalloproteinase-13 increases collagen content of established mouse atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2011;31(11):2464–72.
31. Rao VH, Kansal V, Stoupa S, Agrawal DK. MMP-1 and MMP-9 regulate epidermal growth factor-dependent collagen loss in human carotid plaque smooth muscle cells. *Physiol Rep*. 2014;2(2):e00224.
32. Sato K, Niessner A, Kopecky SL, Frye RL, Goronzy JJ, Weyand CM. TRAIL-expressing T cells induce apoptosis of vascular smooth muscle cells in the atherosclerotic plaque. *J Exp Med*. 2006;203(1):239–50.
33. Kyaw T, Winship A, Tay C, et al. Cytotoxic and proinflammatory CD8+ T lymphocytes promote development of vulnerable atherosclerotic plaques in apoE-deficient mice. *Circulation*. 2013;127(9):1028–39.
34. Gewaltig J, Kummer M, Koella C, Cathomas G, Biedermann BC. Requirements for CD8 T-cell migration into the human arterial wall. *Hum Pathol*. 2008; 39(12):1756–62.