



Grand Challenges for Artificial Intelligence in Molecular Medicine

Frank Emmert-Streib^{1,2*}

¹Predictive Society and Data Analytics Lab, Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland, ²Institute of Biosciences and Medical Technology, Tampere, Finland

Keywords: artificial intelligence, molecular medicine, machine learning, biomedicine, omics, data science, health data science, biomedical data science

1 INTRODUCTION

In recent years, novel methods from artificial intelligence (AI) and machine learning (ML) commonly referred to as data science (DS) enabled many advances in data-driven fields including computational biology, bioinformatics, network medicine, precision medicine and systems medicine (He et al., 2019; Rajkomar et al., 2019; Zou et al., 2019). Given the continuation of technological innovations that will further lead to new high-throughput measurements on all molecular levels, it can be expected that the importance of AI and ML for medicine and biomedicine will even increase in the future (Obermeyer and Emanuel, 2016; Emmert-Streib, 2021). For this reason, a scientific forum is needed for nurturing methodological developments and practical applications of AI, ML and general DS in molecular medicine allowing the community to disseminate and discuss recent results.

The Bioinformatics and AI Specialty Section aims to provide such a forum for publishing articles about the analysis of all types of Omics, clinical and health data for enhancing our understanding of molecular medicine. The emphasize is on either the application or the development of data-driven methods for diagnostic, prognostic, predictive or exploratory studies based on methods from AI or ML.

In **Figure 1**, we show an overview of the iterative process of scientific discovery utilizing artificial intelligence and machine learning to enhance our knowledge about molecular medicine. In the following, we discuss several of these topics that are in our opinion of particular relevance for the development of AI and ML in molecular medicine.

2 DATA: UTILIZATION AND INTEGRATION

In (Feldman et al., 2012), “data” have been called “fuel” because it is like oil for scientific discoveries. For this reason, it is not surprising that we start by discussing the importance of data for molecular medicine. In general, all methods from data science, regardless if they have their origin in artificial intelligence, machine learning or statistics, are based on data (Emmert-Streib and Dehmer, 2019). In other words, a method alone is not capable of contributing anything of meaning for molecular medicine but the combination with data is required.

Nowadays there are many big data resources available that can be utilized for developing and testing methods. Prominent examples thereof are The Cancer Genome Atlas Research Network (TCGA) (The Cancer Genome Atlas Research Network, 2008), Gene ontology (GO) (Ashburner et al., 2000), Gene Expression Omnibus (GEO) (Edgar et al., 2002) or Library of Integrated Network-based Cellular Signatures (LINCS) (Koleti et al., 2017). Interestingly, the idea that such data can also be used for making novel discoveries about the molecular understanding of disorders is so far largely underexplored.

A common problem encountered is how the diverse and often heterogeneous data can be integrated in a meaningful and sound way (Zitnik et al., 2019). Traditionally, one tried to accomplish this by the

OPEN ACCESS

Edited and reviewed by:

Masaru Katoh,
National Cancer Centre, Japan

***Correspondence:**

Frank Emmert-Streib
v@bio-complexity.com

Specialty section:

This article was submitted to
Bioinformatics and Artificial Intelligence
for Molecular Medicine,
a section of the journal
Frontiers in Molecular Medicine

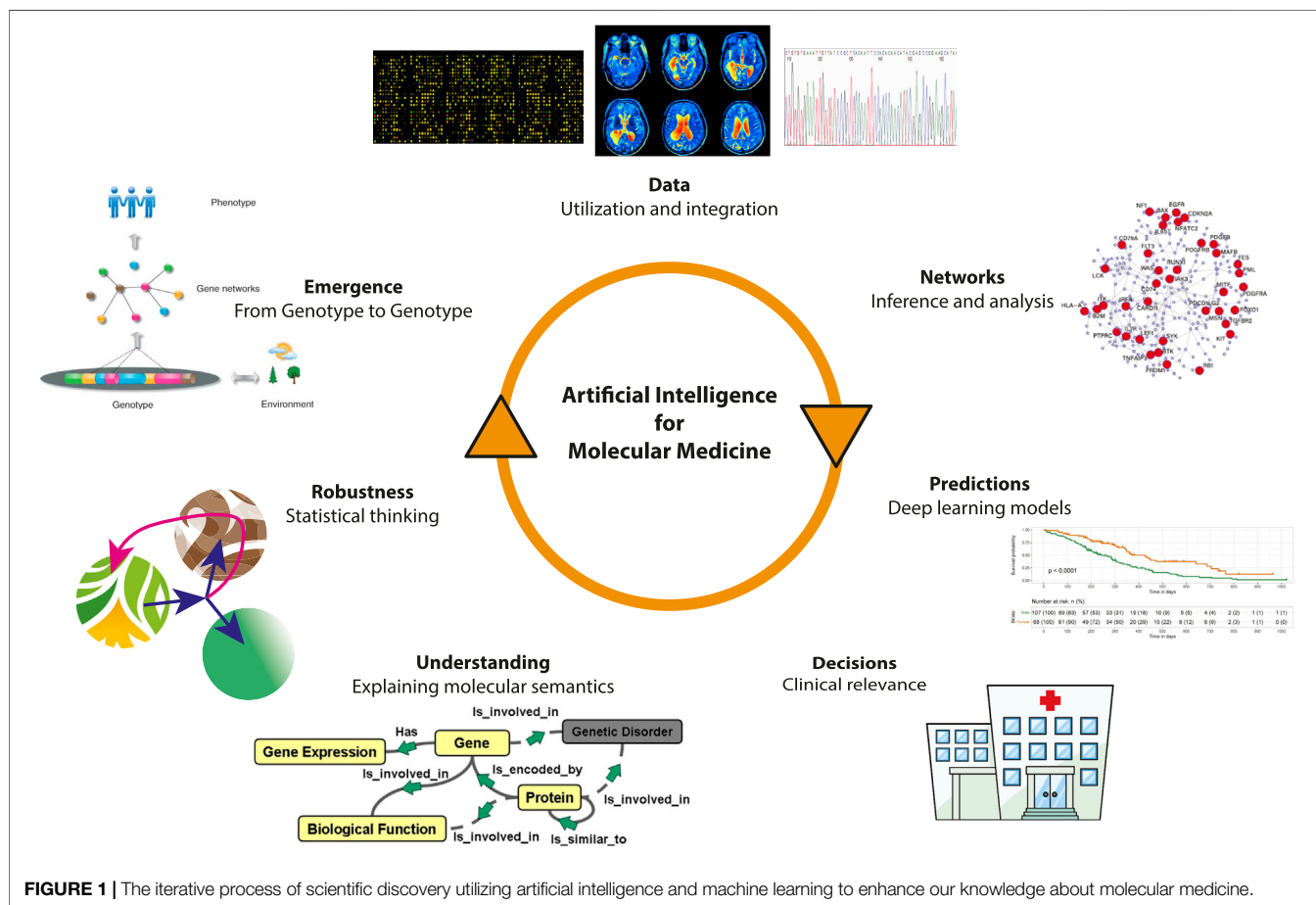
Received: 01 July 2021

Accepted: 08 July 2021

Published: 22 July 2021

Citation:

Emmert-Streib F (2021) Grand
Challenges for Artificial Intelligence in
Molecular Medicine.
Front. Mol. Med. 1:734659.
doi: 10.3389/fmmed.2021.734659



normalization of data with the hope that this allows the pooling of data, i.e., two or more data sets can be combined, from different sources. While this approach is applicable in certain situations it does not offer a generic solution. Instead, a conceptual approach that could be of great practical relevance in this context is provided by transfer learning (Pan and Yang, 2009). The basic idea of transfer learning is to utilize data from two different domains and to use both for learning a so called target task. Importantly, the underlying feature spaces of both domains can be different. Hence, this framework allows to utilize data from different domains without actually combining them. For instance, data from DNA microarrays can be used to improve tasks for RNA-seq data or even to utilize imaging data, e.g., from X-Rays or fMRIs, or text data from electronic health records (eHR) for the same target task. Other machine learning paradigms that could be of relevance are multi-task learning or semi-supervised learning (Chapelle et al., 2006; Zhang and Yang, 2018).

3 NETWORKS: INFERENCE AND ANALYSIS

Another type of approach that is of crucial relevance for molecular medicine is network-based approaches (Vidal, 2009). Specifically, there have been many studies inferring various types of gene

regulatory networks (GRNs), including transcription regulation networks, protein interaction networks, metabolic networks or signalling networks (Emmert-Streib and Dehmer, 2018). Each of these provide useful information about molecular interactions on the cellular level (Emmert-Streib et al., 2014). However, in order to obtain a full systems biology understanding an integration of such networks is needed. Hence, multi-scale network studies are needed to provide us with comprehensive blue-prints about the hierarchical molecular organization pattern (Yu and Gerstein, 2006; Ravasz, 2009).

A field that is dedicated for utilizing such approaches is network medicine (Barabási, 2007). A particular example for the utility of networks is to study the relations between disorders and genes (Goh et al., 2007; Emmert-Streib et al., 2013). Importantly, instead of focusing on individual disorders or genes at a time, network medicine aims at providing insights into the intricate interrelations among all such entities. This allows not only the exploitation of common biological processes or pathways but also to make predictions, e.g., about the drug repurposing (AY et al., 2007; Pushpakom et al., 2019). Hence, networks provide efficient means for studying basic molecular biological questions of disorders and pharmacogenomic problems to gain insights into treatment options for patients.

4 PREDICTIONS: DEEP LEARNING MODELS

A good example to show that machine learning and artificial intelligence are dynamical fields with constant innovations is deep learning (LeCun et al., 2015). Methods of this type came to the awareness of the general community around 2012 and have since then contributed to enhance our understanding in many domains. One particular reason contributing to the success of deep learning methods is the flexibility they offer for building neural networks of different tasks. As a result, nowadays a large number of network architectures is known, e.g., Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM) or Deep Belief Networks, that have been applied in a large variety of application domains (Schmidhuber, 2015; Emmert-Streib et al., 2020a).

One particular deep learning model for the analysis of text data that received considerable attention is BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019). BERT is an autoencoding language model trained using stacked encoder blocks from transformers with a masked language modeling (MLM) to learn word-embeddings bidirectionally. Part of the success of this model is its flexibility to be utilized for a number of different prediction tasks, including named entity recognition, question answering and relation detection (Perera et al., 2020). Hence, this model is of great relevance for analyzing, e.g., electronic Health Records (eHR) from hospitals (Lee et al., 2019; Li et al., 2019).

5 DECISIONS: CLINICAL RELEVANCE

Of particular practical relevance for molecular medicine are studies investigating diagnostic, predictive, prognostic or therapeutic signatures of biomarkers. The reason for this is that such studies have the potential to inform clinical decision making by influencing the diagnosis or treatment of patients in profound ways. The surge of genomics data provides ample opportunities for such studies and one key issue of these is feature selection. Specifically, while the number of molecular entities, e.g., about genes or mRNAs, is in the tens of thousands, interpretable models aim to limit this number to the smallest possible number.

Another interesting topic in this context is the utility of network biomarkers. In contrast to traditional approaches that are based on, e.g., sets of genes or proteins, network biomarkers utilize structural features from gene regulatory networks (Chen et al., 2012; Zeng et al., 2013). This converts a structureless set of genes (sometimes called gene bag) into a complex entity conveying more predictive and interpretable information. As a side-note we would like to mention that this could be also beneficial for the visualization of results and the doctor-patient communication in order to explain therapeutic measures.

6 UNDERSTANDING: EXPLAINING MOLECULAR SEMANTICS

A common goal of all above approaches is to enhance our understanding of the molecular bases of disorders. In order to

see that this is a non-trivial endeavour let's discuss some examples. Deep learning models have been criticized for being black-box models (Adadi and Berrada, 2018). That means such models are good for making predictions but defy a straightforward interpretation making the models non-explainable (Emmert-Streib et al., 2020b). This is particularly problematic in a medical context involving humans because this ultimately means that clinical decisions, e.g., based on the analysis of personal genomics data, cannot be explained to the patient.

Another example is given by biomarkers. In general, biomarkers are used for diagnostic, prognostic, predictive or therapeutic purposes to make decisions about the care of a patient (Califf, 2018). It is widely believed that aside from this clinical utility based on the predictive capabilities of such signatures, biomarkers are also offering insights into the molecular functioning of biological processes and their causal involvement in disorders (Van De Vijver et al., 2002; Cuzick et al., 2011). However, for prognostic signatures of breast cancer it has been demonstrated that this is not the case (Venet et al., 2011; Manjang et al., 2021). This implies that also the prognostic signatures are black-box models with sensible predictions of breast cancer outcome but no value for revealing causal connections. Hence, such models have a predictive utility, e.g., for applications in the clinical practice but no biological utility for enhancing our understanding of breast cancer biology. If similar results are observed for other cancer types or different disorders remains to be seen.

From these examples one can see that establishing a good prediction model does not imply that we also obtain immediately an understanding of the molecular semantics offered by disorders. Hence, ideally, causal prediction models are required that provide prediction capabilities along with an interpretable structure for giving causal explanations of molecular activities (Holzinger et al., 2019). In case such ideal models are unachievable one needs measures for quantifying these deficiencies.

7 ROBUSTNESS: STATISTICAL THINKING

An aspect that does not receive enough appreciation is the fact that any type of the analysis of data from molecular medicine requires statistical considerations. That means even modern developments in AI and ML do not make a statistical understanding obsolete but are built upon it. This includes, for instance, ensuring the reproducibility of studies (Peng, 2011; Begley and Ioannidis, 2015), multiple testing corrections of hypotheses (Noble, 2009) or the regularization of regression models (Tibshirani, 1996). Of particular interest are studies that clarify the understanding of problems of widely used methods or approaches (Ioannidis, 2005; Tripathi et al., 2013; Wasserstein and Lazar, 2016). Hence, investigations that enhance our understanding of molecular medicine by applying any form of *statistical thinking* are welcome to advance bioinformatics because only such approaches lead to the robustness of findings that are of biological and clinical significance (Vingron, 2001).

8 EMERGENCE: FROM GENOTYPE TO PHENOTYPE

Finally, we would like to emphasize that molecular medicine aims to study the connection between genotype and phenotype (Ginsburg and Willard, 2009; Collins and Varmus, 2015). That means, while aberrant molecular processes give rise to various forms of disorders, those molecular processes should not be studied in isolation but their phenotypic consequences need to be systematically documented. However, this requires to bridge from genotype to phenotype (Noble, 2008a; Gjuvslund et al., 2013; Ritchie et al., 2015). For practical approaches 'networks' have been suggested to capture relevant information as an intermediate layer (Emmert-Streib and Glazko, 2011; Carter et al., 2013; Kim and Przytycka, 2013), however, further investigations are needed, e.g., to merge such approaches with predictive models.

On a theoretical note, we would like to highlight that the above problem is actually severe because it requires an understanding of *emergence* (Noble, 2008b; Pigliucci, 2010). Hence, reductionist approaches are prone to fail in molecular medicine which possesses major challenges for conceptual approaches provided by AI or ML to overcome such limitations (Mazzocchi, 2012). Hence, even personalized medicine or precision medicine depend on our theoretical understanding of the biological complexity

of emergent features arising from the transition between the genotype to the phenotype.

9 CONCLUSION

The Bioinformatics and AI Specialty Section of *Frontiers in Molecular Medicine* will provide a venue for world-class interdisciplinary research addressing the above, and many more challenges arising in the future. In order to provide a forum for the exchange of ideas and growth of innovations for a multi-disciplinary research community, the journal does not only publish Original Research and Review articles but a number of additional paper types. For instance, the journal welcomes submissions for the following article types: Hypothesis and Theory, Perspective, Opinion and General Commentary. This will allow to express the perspectives and opinions of the community and to discuss recent developments critically. Furthermore, the journal publishes also Technology and Code articles which present 1) new software, 2) new applications of software or 3) implementations of existing algorithms under novel settings.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- Adadi, A., and Berrada, M. (2018). Peeking inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. doi:10.1109/access.2018.2870052
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Michael Cherry, J., et al. (2000). Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi:10.1038/75556
- Ay, M., Goh, K. I., Cusick, M. E., Barabasi, A. L., and Vidal, M. (2007). Drug–target Network. *Nat. Biotechnol.* 25, 1119–1127. doi:10.1038/nbt1338
- Barabási, A.-L. (2007). Network Medicine – from Obesity to the “Diseasome”. *N. Engl. J. Med.* 357, 404–407. doi:10.1056/nejme078114
- Begley, C. G., and Ioannidis, J. P. (2015). Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. *Circ. Res.* 116, 116–126. doi:10.1161/circresaha.114.303819
- Califf, R. M. (2018). Biomarker Definitions and Their Applications. *Exp. Biol. Med.* 243, 213–221. doi:10.1177/1535370217750088
- Carter, H., Hofree, M., and Ideker, T. (2013). Genotype to Phenotype via Network Analysis. *Curr. Opin. Genet. Dev.* 23, 611–621. doi:10.1016/j.gde.2013.10.003
- Chapelle, O., Schölkopf, B., and Zien, A. (2006). *Semi-Supervised Learning. Adaptive Computation and Machine Learning*. The MIT Press.
- Chen, L., Liu, R., Liu, Z.-P., Li, M., and Aihara, K. (2012). Detecting Early-Warning Signals for Sudden Deterioration of Complex Diseases by Dynamical Network Biomarkers. *Scientific Rep.* 2, 1–8. doi:10.1038/srep00342
- Collins, F. S., and Varmus, H. (2015). A New Initiative on Precision Medicine. *New Engl. J. Med.* 372, 793–795. doi:10.1056/nejmp1500523
- Cuzick, J., Swanson, G. P., Fisher, G., Brothman, A. R., Berney, D. M., Reid, J. E., et al. (2011). Prognostic Value of an Rna Expression Signature Derived from Cell Cycle Proliferation Genes in Patients with Prostate Cancer: a Retrospective Study. *Lancet Oncol.* 12, 245–255. doi:10.1016/s1470-2045(10)70295-3
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). “Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota. 4171–4186.
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucleic Acids Res.* 30, 207–210. doi:10.1093/nar/30.1.207
- Emmert-Streib, F., and Dehmer, M. (2018). Inference of Genome-Scale Gene Regulatory Networks: Are There Differences in Biological and Clinical Validations? *Mach. Learn. Knowl. Extr.* 1, 138–148. doi:10.3390/make1010008
- Emmert-Streib, F., and Dehmer, M. (2019). Defining Data Science by a Data-Driven Quantification of the Community. *Mach. Learn. Knowl. Extr.* 1, 235–251. doi:10.3390/make1030058
- Emmert-Streib, F., Dehmer, M., and Haibe-Kains, B. (2014). Gene Regulatory Networks and Their Applications: Understanding Biological and Medical Problems in Terms of Networks. *Front. Cel Dev. Biol.* 2, 38. doi:10.3389/fcell.2014.00038
- Emmert-Streib, F. (2021). From the Digital Data Revolution toward a Digital Society: Pervasiveness of Artificial Intelligence. *Mach. Learn. Knowl. Extr.* 3, 284–298. doi:10.3390/make3010014
- Emmert-Streib, F., and Glazko, G. (2011). Network Biology: A Direct Approach to Study Biological Function. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3, 379–391. doi:10.1002/wsbm.134
- Emmert-Streib, F., Tripathi, S., de Matos Simoes, R., Hawwa, A., and Dehmer, M. (2013). The Human Disease Network: Opportunities for Classification, Diagnosis and Prediction of Disorders and Disease Genes. *Syst. Biomed.* 1, 1–8. doi:10.4161/sysb.22816
- Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., and Dehmer, M. (2020a). An Introductory Review of Deep Learning for Prediction Models with Big Data. *Front. Artif. Intelligence* 3, 4. doi:10.3389/frai.2020.00004
- Emmert-Streib, F., Yli-Harja, O., and Dehmer, M. (2020b). Explainable Artificial Intelligence and Machine Learning: A Reality Rooted Perspective. *WIREs Data Mining Knowledge Discov.* 10, e1368. doi:10.1002/widm.1368
- Feldman, B., Martin, E. M., and Skotnes, T. (2012). Big Data in Healthcare Hype and hope. *Dr. Bonnie.* 360, 122–125. doi:10.1109/TKDE.2009.191

- Ginsburg, G. S., and Willard, H. F. (2009). Genomic and Personalized Medicine: Foundations and Applications. *Translational Res.* 154, 277–287. doi:10.1016/j.trsl.2009.09.005
- Gjuvsland, A. B., Vik, J. O., Beard, D. A., Hunter, P. J., and Omholt, S. W. (2013). Bridging the Genotype–Phenotype gap: what Does it Take? *J. Physiol.* 591, 2055–2066. doi:10.1113/jphysiol.2012.248864
- Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabasi, A. (2007). The Human Disease Network. *Proc. Natl. Acad. Sci.* 104, 8685–8690. doi:10.1073/pnas.0701361104
- He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., and Zhang, K. (2019). The Practical Implementation of Artificial Intelligence Technologies in Medicine. *Nat. Med.* 25, 30–36. doi:10.1038/s41591-018-0307-0
- Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. (2019). Causability and Explainability of Artificial Intelligence in Medicine. *Wiley Interdiscip. Rev. Data Mining Knowledge Discov.* 9, e1312. doi:10.1002/widm.1312
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *Plos Med.* (New York: Springer) 2, e124. doi:10.1371/journal.pmed.0020124
- Kim, Y.-A., and Przytycka, T. M. (2013). Bridging the gap between Genotype and Phenotype via Network Approaches. *Front. Genet.* 3, 227. doi:10.3389/fgene.2012.00227
- Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D. J., Turner, J. P., et al. (2017). Data portal for the Library of Integrated Network-Based Cellular Signatures (Lincs) Program: Integrated Access to Diverse Large-Scale Cellular Perturbation Response Data. *Nucleic Acids Res.* 46, D558–D566. doi:10.1093/nar/gkx1063
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature* 521, 436. doi:10.1038/nature14539
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., et al. (2019). BioBERT: a Pre-trained Biomedical Language Representation Model for Biomedical Text Mining. *Bioinformatics* 36. doi:10.1093/bioinformatics/btz682
- Li, F., Jin, Y., Liu, W., Rawat, B. P. S., Cai, P., and Yu, H. (2019). Fine-tuning Bidirectional Encoder Representations from Transformers (BERT)-based Models on Large-Scale Electronic Health Record Notes: an Empirical Study. *JMIR Med. Inform.* 7, e14830. doi:10.2196/14830
- Manjang, K., Tripathi, S., Yli-Harja, O., Dehmer, M., Glazko, G., and Emmert-Streib, F. (2021). Prognostic Gene Expression Signatures of Breast Cancer Are Lacking a Sensible Biological Meaning. *Scientific Rep.* 11, 1–18. doi:10.1038/s41598-020-79375-y
- Mazzocchi, F. (2012). Complexity and the Reductionism–Holism Debate in Systems Biology. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 4, 413–427. doi:10.1002/wsbm.1181
- Noble, D. (2008a). Genes and Causation. *Phil. Trans. R. Soc. A.* 366, 3001–3015. doi:10.1098/rsta.2008.0086
- Noble, D. (2008b). *The Music of Life: Biology beyond Genes*. Oxford University Press.
- Noble, W. S. (2009). How Does Multiple Testing Correction Work?. *Nat. Biotechnol.* 27, 1135. doi:10.1038/nbt1209-1135
- Obermeyer, Z., and Emanuel, E. J. (2016). Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. *New Engl. J. Med.* 375, 1216. doi:10.1056/nejmp1606181
- Pan, S. J., and Yang, Q. (2009). A Survey on Transfer Learning. *IEEE Trans. Knowledge Data Eng.* 22, 1345–1359.
- Peng, R. D. (2011). Reproducible Research in Computational Science. *Science* 334, 1226–1227. doi:10.1126/science.1213847
- Perera, N., Dehmer, M., and Emmert-Streib, F. (2020). Named Entity Recognition and Relation Detection for Biomedical Information Extraction. *Front. Cell Dev. Biol.* 8, 673. doi:10.3389/fcell.2020.00673
- Pigliucci, M. (2010). Genotype–phenotype Mapping and the End of the ‘genes as Blueprint’ metaphor. *Philos. Trans. R. Soc. B: Biol. Sci.* 365, 557–566. doi:10.1098/rstb.2009.0241
- Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug Repurposing: Progress, Challenges and Recommendations. *Nat. Rev. Drug Discov.* 18, 41–58. doi:10.1038/nrd.2018.168
- Rajkumar, A., Dean, J., and Kohane, I. (2019). Machine Learning in Medicine. *New Engl. J. Med.* 380, 1347–1358. doi:10.1056/nejmra1814259
- Ravasz, E. (2009). “Detecting Hierarchical Modularity in Biological Networks,” in *Computational Systems Biology*. Springer, 1–16. doi:10.1007/978-1-59745-243-4_7
- Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., and Kim, D. (2015). Methods of Integrating Data to Uncover Genotype–Phenotype Interactions. *Nat. Rev. Genet.* 16, 85–97. doi:10.1038/nrg3868
- Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural networks*. 61, 85–117. doi:10.1016/j.neunet.2014.09.003
- The Cancer Genome Atlas Research Network (2008). Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways. *Nature* 455, 1061–1068. doi:10.1038/nature07385
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B.* 58, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x
- Tripathi, S., Glazko, G., and Emmert-Streib, F. (2013). Ensuring the Statistical Soundness of Competitive Gene Set Approaches: Gene Filtering and Genome-Scale Coverage Are Essential. *Nucleic Acids Res.* 6, e53354. doi:10.1093/nar/gkt054
- Van De Vijver, M. J., He, Y. D., Van’t Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., et al. (2002). A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. *New Engl. J. Med.* 347, 1999–2009. doi:10.1056/nejmoa021967
- Venet, D., Dumont, J. E., and Detours, V. (2011). Most Random Gene Expression Signatures Are Significantly Associated with Breast Cancer Outcome. *PLoS Comput. Biol.* 7, e1002240. doi:10.1371/journal.pcbi.1002240
- Vidal, M. (2009). A Unifying View of 21st century Systems Biology. *FEBS Lett.* 583, 3891–3894. doi:10.1016/j.febslet.2009.11.024
- Vingron, M. (2001). [Dataset], Bioinformatics Needs to Adopt Statistical Thinking. *Bioinformatics* 17 (5), 389–390. doi:10.1093/bioinformatics/17.5.389
- Wasserstein, R. L., and Lazar, N. A. (2016). The ASA’s Statement on P-Values: Context, Process, and Purpose. *The Am. Stat.* 70, 129–133. doi:10.1080/00031305.2016.1154108
- Yu, H., and Gerstein, M. (2006). Genomic Analysis of the Hierarchical Structure of Regulatory Networks. *Proc. Natl. Acad. Sci. USA.* 103, 14724–14731. doi:10.1073/pnas.0508637103
- Zeng, T., Sun, S.-y., Wang, Y., Zhu, H., and Chen, L. (2013). Network Biomarkers Reveal Dysfunctional Gene Regulations during Disease Progression. *FEBS J.* 280, 5682–5695. doi:10.1111/febs.12536
- Zhang, Y., and Yang, Q. (2018). An Overview of Multi-Task Learning. *Natl. Sci. Rev.* 5, 30–43. doi:10.1093/nsr/nwx105
- Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., and Hoffman, M. M. (2019). Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. *Inf. Fusion* 50, 71–91. doi:10.1016/j.inffus.2018.09.012
- Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., and Telenti, A. (2019). A Primer on Deep Learning in Genomics. *Nat. Genet.* 51, 12–18. doi:10.1038/s41588-018-0295-5

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Emmert-Streib. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.