I am a computer scientist with more than two years of postdoctoral research experience in computational biology, genomics, and bioinformatics. My Ph.D. is in machine learning (**ML**) with emphasis on high dimensional statistics. During the last two years of graduate school, I was fighting Acute Myeloid Leukemia (**AML**) which motivated me to change direction from ML theory and made me personally invested in biology and health research. I joined the Mathematical Biosciences Institute (**MBI**) of The Ohio State University (**OSU**), which prepared me for working at the intersection of computational sciences and biology. **I have been developing methods in high-dimensional statistics, graphical models, causal inference, outlier detection, and submodular function optimization areas to address problems in gene regulation, high throughput drug screening, drug sensitivity analysis, tumorigenesis modeling, and social network analysis**. During my Ph.D., apart from my advisor Prof. Banerjee, I worked with Profs. Sapiro and Zhang and collaborated with a group of scientists at Stanford Research Institute (SRI). All of these collaborations resulted in articles published in prominent computer science conferences. While being a postdoctoral fellow at MBI working with my mentor Prof. Coombes, I have been collaborating with oncologists and cancer biology and genetics researchers at the OSU Comprehensive Cancer Center (**OSUCCC**)- James, such as Drs. Sampath, Toland, and Hertlein. In addition, I have ongoing collaborations with faculty in the Statistics Department Profs. Chkrebtii, Huling, Kurtek, and Paul on various projects. Outcomes of some of my postdoc projects have already been published or submitted and the rest are being prepared for submission.

1 Completed Research

1.1 Social Network Analysis

I worked on social network analysis during my M.Sc. and early years of my Ph.D. training. Specifically, I have investigated the "influence maximization" problem[1,2] and sentiment analysis.[3] In influence maximization, the goal is to increase the probability of success of an advertising campaign by effectively utilizing the word-of-mouth. The objective is to select the most influential nodes from the customers-network for targeted advertisements. I have studied influence maximization under both progressive[1] and non-progressive[2] influence propagation models. I have also worked on Twitter sentiment analysis where the goal is to extract the positive, negative, or neutral sentiments.[3]

1.2 High Dimensional Prediction and Inference

The goal of high-dimensional statistics is to perform inference in problems where the number of samples n is much smaller than the dimension p, i.e., measured features. My Ph.D. thesis focused on developing statistically sound algorithms for high dimensional regression problems.[4–6] I have designed and implemented an algorithm to consistently estimate the linear model parameter for noisy high dimensional data sets, i.e., estimation for the error-in-variables model.[6] In another project, **I studied multi-task learning**[4,5] **where we have G linked high-dimensional regression problems that have similarities and differences**. In the *data sharing* framework, samples of the G groups are modeled by a shared parameter $\boldsymbol{\beta}_0$ plus a per-group individual parameter $\boldsymbol{\beta}_g$, Fig. 1. I proposed the following estimator:

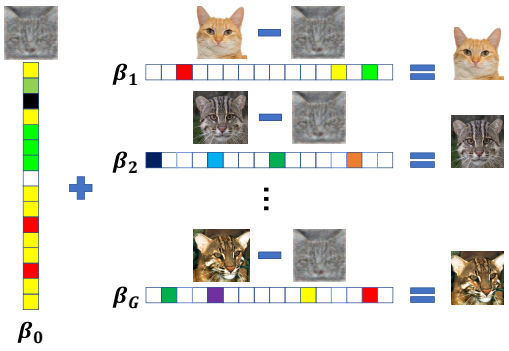


Figure 1: A conceptual illustration of data sharing model for learning representation of cat species. The common parameter $\boldsymbol{\beta}_0$ captures a *generic cat* which consists of shared features among all cats.

$$\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_0^T, \dots, \hat{\boldsymbol{\beta}}_G^T) \in \underset{\boldsymbol{\beta}_0,\dots,\boldsymbol{\beta}_G}{\operatorname{argmin}} \frac{1}{n} \sum_{g=1}^{G} \|\mathbf{y}_g - \mathbf{X}_g(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_g)\|_2^2, \qquad (1)$$
$$\text{s.t. } \forall g \in [G] \cup \{0\} : f_g(\boldsymbol{\beta}_g) \leq f_g(\boldsymbol{\beta}_g^*),$$

where f_gs are arbitrary structure-inducing convex functions, e.g., l_1-norm for sparse $\boldsymbol{\beta}_g$s. I provided sample complexity, estimation error bounds, and convergence rate analysis for (1). The corresponding theory paper has been presented in a workshop[5] and the longer version is under review in the *SIAM Journal on Mathematics of Data Science*.[7] I also initiated a collaborative project between Profs. Banerjee and Coombes where the goal was to use (1) in the anti-cancer drug-response prediction where the G groups are different cancer (sub)types. I applied (1) to the Cancer Cell Line Encyclopedia dataset,[8] and presented the findings in two medical conferences.[9,10] The paper is available on arXiv.[4]

1.3 Data Science – General

I have been active in other areas of computer and data science. We have recently published a tutorial on the applications of statistical shape analysis in biology.[11] The tutorial begins with *differential and Riemannian geometry* and builds up to recent tools in *elastic shape analysis*. I have also worked on the problem of *malware domain detection*, where the goal is to identify malware from a sequence of DNS requests using *recurrent neural networks*.[12]

2 Ongoing Research: ML Applications in Biology and Medicine

2.1 Quantifying the Role of Transcription Factors and MicroRNAs in Transcription

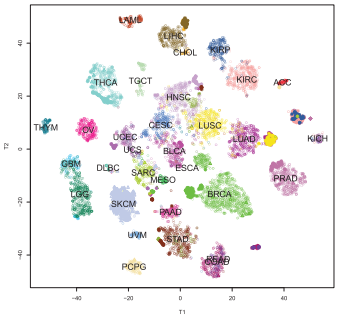
The amount of control that Transcription Factors (**TF**s) and microRNAs (**miR**s) exert on transcription is not quantified yet. In a series of papers, I investigated the amount of control of TFs and miRs on transcription using the TCGA dataset.[13] First, using Thresher,[14] the feature extraction technique that was developed in the Coombes lab, we extracted 30 biologically interpretable clusters of TFs and showed that they can distinguish normal samples from cancerous ones and also cancer types from each other, Fig.2a.[15] Then, we showed that the 21 miRs extracted by Thresher are sufficient to explain 31% of gene expression variability across tissues.[16] Finally, we showed that the combination of 51 regulatory components (**RC**s; selected miRs and TFs) can explain more than 50% variability of gene expression, Fig. 2b.[17] Analysis of genes that are well-explained by TFs or miRs and those that are poorly explained by both provide interesting biological insights regarding the similarities and differences of pathways that are controlled by each one of the RCs, Fig.2c. Future directions of this work are explained in Section 3.

2.2 Cancer Combination Drug Screening

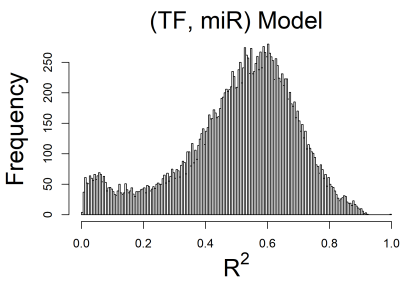
Combination therapy utilizes more than one drug to treat disease and is at the frontier of oncology.[18] For a wisely selected combination, the hope is to get the same efficacy of monotherapies with less toxicity and side effects. Also, combination therapy can attack multiple pathways of cancer and therefore overcome resistance induced by sub-clones.[19] Finally, combining already approved drugs (by the FDA) is a more cost-effective way of advancing cancer treatment. However, the search space of all combinations is huge ($O(n^2)$ for a two-drugs combination out of n approved drugs). This level of complexity is one of the primary impediments to the advance of combination therapy in oncology. To test the effect of combination therapy, first, we select two promising drugs based on their chemical properties. Then, we perform a grid search to find the most effective concentrations of the drugs by combining them in m various doses, which results in m^2 combinations.[20] This experiment is called a *full synergy experiment*.

In collaboration with leukemia researchers of the OSUCC, I have developed a novel pipeline in which by performing only a single test we can determine if the combination is promising and worth moving forward with the full synergy experiment.[21] **This approach allows us to search for effective combination therapies in a data-driven manner and combine many drugs and expand our search landscape**. I have designed and implemented a pipeline for outlier detection, batch effect removal, and synergy finding from scratch since none of the existing synergy scoring methods work in the single-dose regime. Two papers are under preparation based on this project.

Future Direction. In an ongoing collaboration, I am extending this framework to analyze the data of the follow-up full synergy experiments and find the minimum effective doses. To compare the efficacy of various drug combinations, I proposed a new measure based on *IC50 level set*, which is the contour line of all concentrations that kill half of the cells. To estimate the IC50 level set, I propose to use *additive isotonic model* (**AIM**) $y = f_1(x_1) + f_2(x_2) + \epsilon$ where y is the efficacy of combination and f_is are monotone univariate functions corresponding to each drug. Since increasing the concentration of each drug should not decrease efficacy, a linear combination of monotonic functions is a proper model for such a surface. AIM can be solved for f_is using cyclic pooled adjacent violators algorithm.[22]



(a)



(b)

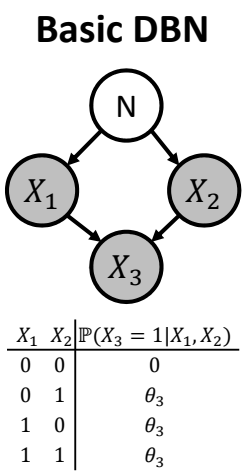
Pathways	miR	TF
Tight junction interactions	1.26E-03	NS
Cell junction organization	7.59E-03	NS
Cell-Cell communication	2.02E-02	NS
Cell Cycle	NS	4.17E-31
Mitotic Prometaphase	NS	7.36E-25
M Phase	NS	2.01E-16

(c)

Figure 2: (a) Plot of the non-linear t-SNE map of TCGA samples resulted from 30 biological components derived from 486 transcription factors by Thresher produced a clear separation between most TCGA cancer types and also normal cases.[15] (b) Distribution of the percentage of the variance of mRNA expression explained (R^2) by the 51 regulatory components on an independent test set for ∼20,000 genes.[17] (c) Gene enrichment for well-explained genes ($R^2 > 0.8$). P-values are calculated by a false discovery rate (FDR) correction (Benjamini-Hochberg). NS indicates non-significant.[17]

2.3 Inference of Cancer Progression Network: Learning Structure of Bayesian Networks

Cancer is an evolutionary process that can be modeled as a sequence of fixation of genetic alterations throughout the tumor cell population.[23,24] Each new driver alteration confers a selective growth advantage to the cell and sweeps through the population, which results in clonal expansion.[25] But the order in which accumulating alterations fixate in tumors is not arbitrary. This observation leads to the first model of tumorigenesis of colon cancer as a *chain* of aberrations.[26] Inferring the order of alterations has been shown to have diagnostic and prognostic importance[23,24] but is challenging because we often have a single observation from the tumor at the time of diagnosis. Chain progression models have been extended to *trees*,[27] mixture of trees,[28,29] and *Directed Acyclic Graphs*[30,31] (DAGs). Bayesian networks (BN), which are DAGs equipped with a joint probability distribution,[32] lend themselves naturally to such models. Perhaps the most famous BN model of cancer progression is Conjunctive Bayesian Network (CBN)[30,31] where the assumption is that all parent alterations must be present in order for a child aberration to occur.



X_1	X_2	$\mathbb{P}(X_3 = 1 \mid X_1, X_2)$
0	0	0
0	1	θ_3
1	0	θ_3
1	1	θ_3

Figure 3: The basic DBN model where the progression starts from normal state, N, and moves to accumulate X_i alterations according to the probabilistic OR rule. The table shows the progression rule for the node X_3. The goal is to learn both network and the node parameters θ_is from cross-sectional data.

The assumption of CBNs is restrictive because a single advantageous hit is usually enough for progression.[25] **I proposed the Disjunctive Bayesian Network (DBN) in which each alteration can occur if at least one of its parents has happened before**, Fig. 3. DBN generalizes CBN and therefore has a larger search space but we have designed a scalable algorithm to infer DBN. Last summer, I mentored two math undergraduate students (with Prof. Chkrebtii) and implemented this algorithm. We are finalizing a methodological paper which is my first paper as senior author.[34] We are writing another paper on our specific findings for the progression of melanoma.

Future Directions. Exploring the relation of DBN with population genetic models of cancer,[36] inference in the continuous-time version of DBN,[37] explicit modeling of mutual exclusivity of alterations in pathways for DBN,[38] and using causal discovery methods[39] for progression network inference are among the future directions of this work. **These are all suitable projects for graduate and senior undergraduate students**.

3 Future Directions

In addition to *immediate* future directions mentioned before, here, I will elaborate on my *longer-term* research goals.

3.1 Improving Prediction of Gene Expression Profile from Regulatory Molecules

This topic is a future direction of Section 2.1.

3.1.1 Tissue-specific Modeling

We have shown that the Regulatory Molecules (TFs + miRs = **RM**) explain a large amount of transcriptome variation and expect the rest to be explained by factors such as methylation and histone modification. Remodeling the methylome and histones are hallmarks of development and cell differentiation.[40,41] Therefore, one would expect that parts of these signals are captured by the tissue type. **I hypothesize that considering tissue type in prediction will increase the accuracy**. I will consider three ways of incorporating the tissue type. First, I will treat it as a categorical variable in the regression. Next, I will perform the prediction in each tissue separately. Finally, I will use the *data sharing* framework (1).

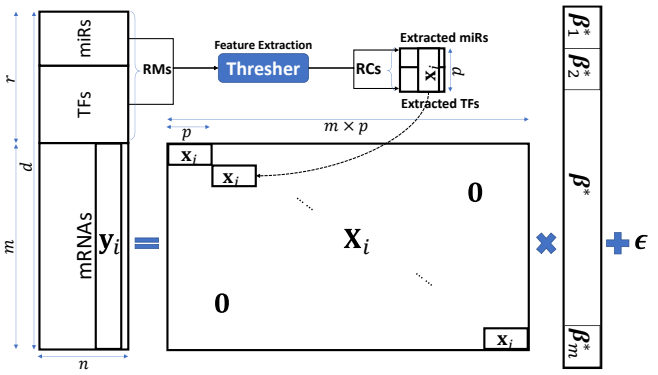


Figure 4: Extracting RCs from RMs by Thresher and formulating the multi-response regression problem where the goal is to predict the rest of mRNA expressions $\mathbf{y}_i$ using extracted features, RCs, by learning $\boldsymbol{\beta}^*$ under the assumption $\boldsymbol{\epsilon} \sim N(0, \Sigma^*)$.

3.1.2 Multi-response Regression

A naive approach treats the prediction of each of the $\sim 20,000$ genes as an independent task. **I hypothesize that learning the correlation structure of tasks will improve accuracy and I propose a variant of the multi-response prediction to do so**.[42–44] Consider a multi-response model with m outputs, $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}^* + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \Sigma^*)$ where $\mathbf{y}_i \in \mathbb{R}^m$ is the response vector (expression of $\sim 20,000$ genes for the ith sample), $\mathbf{X}_i \in \mathbb{R}^{m \times pm}$ contains the RC feature vector $\mathbf{x}_i$ replicated on diagonal, and $\boldsymbol{\epsilon}_i \in \mathbb{R}^m$ is noise, Fig. 4. Each response has a specific parameter $\boldsymbol{\beta}_i^*$ which are stacked into one "long" vector $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_m^*)$.[43] Both Σ^* and $\boldsymbol{\beta}^*$ should be estimated

by maximum likelihood estimation which is non-trivial because the number of *outputs* is more than the number of samples ($m > n$) and $\hat{\Sigma}$ is rank deficient. I assume that the precision matrix $\Omega = \Sigma^{-1}$ is s-sparse, i.e., outputs have limited interactions.[46–49] **I propose the following alternate minimization estimator to boost the prediction accuracy by simultaneously learning parameters β^* and the relation of outputs Σ^*:**

$$\hat{\boldsymbol{\beta}}_t \in \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \|\hat{\Omega}_{(t-1)}^{\frac{1}{2}}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\|_2^2, \quad \hat{\Omega}_t \in \underset{\Omega\succcurlyeq 0}{\operatorname{argmin}} -\log|\Omega| + \operatorname{tr}(\hat{S}_t\Omega) + \lambda\|\Omega\|_1, \tag{2}$$

where $\hat{S}_t$ is the empirical covariance matrix. I will explore *response envelopes*[52] as an alternative approach.

3.2 Causal Effect Estimation of Regulatory Molecules

Associations do not necessarily lead to actionable knowledge about the biological mechanisms whereas significant causal effects will. Causal inference will play a major role in data science and biology as we move from observational studies to large scale experimental studies such as gene perturbation. In particular, it has a tremendous potential to explain gene regulation mechanisms, which is a key link in the translation chain from genotypes to phenotypes. Motivated by the importance of causality in biology and genomics, in the summer of 2018 I started to examine the causal inference literature by attending conferences and talking to OSU's experts in the field. **I have initiated and am organizing a weekly causality reading group with faculty of Computer Science and Statistics Departments where we discuss recent papers**. Here, I will discuss a few directions that I am interested in and working on.

3.2.1 Invariant Causal Prediction: Inferring Gene Regulation from the Cancer Genome

It is infeasible to measure the causal effect of *all* RMs by perturbation experiments such as TF knockdown[54] and miR transfection.[55] Due to various aberrations, cancer cells can serve as a *natural experiment* from which the causal effect of RMs can be estimated. **To infer the causal effect of RMs, I propose to apply *Invariant Causal Prediction*[53] (ICP) to TCGA[56] and GTEx[57] data**. ICP argues that the set of direct causes can predict the outcome across different *environments* with invariant accuracy.[53] Environments can be created by precise experimental design (e.g., gene knockdowns) or generated by unknown and uncontrolled interventions, Figure 5. Given different environments, one can collect all models that show prediction invariance and the true causal model will be a member of this set with high probability.[58] ICP can leverage multiple heterogeneous data sets such as normal vs. cancer, different tissue types, or sub-types of cancer for causal inference. In contrast to other methods, ICP excels in inference if provided with more sources and thus is suitable for integrating multiple heterogeneous data sets.

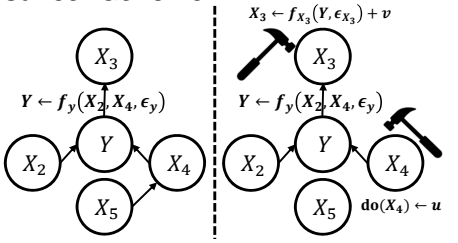


Figure 5: Example of observational (left) and interventional/experimental (right) distributions.[53] Mechanism $f_y(\cdot)$ that generates Y from its causes remains invariant.

3.2.2 Double/Debiased Machine Learning: Scaling Causal Effect Estimation

Multi-output prediction (Section 3.1.2) and ICP (Section 3.2) assume a linear relationship between RMs and outcome mRNAs expressions and are hard to extend theoretically for capturing non-linear interactions.[59] On the other hand, the prediction accuracy of more complex ML models such as deep neural networks surpasses the linear model's performance but leads to biased estimates of the causal effect.[60] *Double/Debiased machine learning*[60] (**DML**) is a new framework for causal inference from observational data introduced in economics that corrects for the regularization bias incurred from ML prediction methods. **To use the prediction ability and scalability of modern ML methods while estimating the causal effects of RMs on transcription rigorously, I will build upon DML.** DML uses arbitrary ML methods for two prediction tasks (estimating the effect of confounders on RM and mRNA, Fig. 6) while overcoming the induced bias on causal effect estimation by *orthogonalization*.[60] Since DML better controls the confounding factors, it can squeeze more information out of the observational data for causal effect estimation, i.e., DML is *sample efficient*. As DML uses advanced ML methods in its core, it is more scalable, i.e., *computationally efficient*.

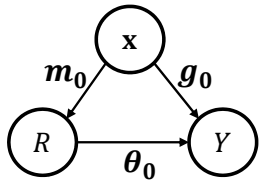


Figure 6: The DAG representing DML. DML estimates the nuisance parameters $m_0(\mathbf{x})$ and $g_0(\mathbf{x})$ with ML prediction methods to infer the causal effect θ_0.

Summary. Causal inference is finding its place in data science at a level between purely correlation-based models and physical models based on differential equations.[58] These methods can generate testable biological hypotheses and have a lower false discovery rate compared to other computational approaches.[61] I expect causal inference methods to improve our understanding of mechanisms of gene regulation and move us closer to learning the overall effect of targeted therapies. I believe that by bringing the best of causal and ML methods into biology and genomics, the proposed algorithms will have a broad impact and add a new class of methods to the computational biologist's toolbox.

References

[1] **Amir Asiaee T.**, Mohammad Afshar, and Masoud Asadpour. Influence maximization for informed agents in collective behavior. In *Distributed Autonomous Robotic Systems*, pages 389–402. Springer, 2013.

[2] Golshan Golnari*, **Amir Asiaee T.***, Arindam Banerjee, and Zhi-Li Zhang. Revisiting non-progressive influence models: Scalable influence maximization in social networks. In *31st Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 316–325, 2015 (*Equal contribution).

[3] **Amir Asiaee T.**, Mariano Tepper, Arindam Banerjee, and Guillermo Sapiro. If you are happy and you know it... tweet. In *21st ACM international conference on Information and knowledge management (CIKM)*, pages 1602–1606. ACM, 2012.

[4] **Amir Asiaee**, Samet Oymak, Arindam Banerjee, and Kevin R. Coombes. High dimensional data enrichment: Interpretable, fast, and data-efficient. *arXiv:1806.04047*.

[5] **Amir Asiaee**, Samet Oymak, Kevin R. Coombes, and Arindam Banerjee. Data enrichment: Multi-task learning in high dimension with theoretical guarantees. In *Adaptive and Multi-Task Learning Workshop at ICML*, 2019.

[6] **Amir Asiaee T.**, Soumyadeep Chaterjee, and Arindam Banerjee. High dimensional structured estimation with noisy designs. In *16th SIAM International Conference on Data Mining (SDM)*, pages 801–809. SIAM, 2016.

[7] **Amir Asiaee**, Samet Oymak, Arindam Banerjee, and Kevin R. Coombes. High dimensional data sharing: Multi-task learning with theoretical guarantee. Under review in SIAM Journal on Mathematics of Data Science.

[8] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A Margolin, Sungjoon Kim, Christopher J Wilson, Joseph Lehár, Gregory V Kryukov, Dmitriy Sonkin, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603, 2012.

[9] **Amir Asiaee**, Samet Oymak, Arindam Banerjee, and Kevin R. Coombes. Precision cancer treatment with high dimensional data enrichment. *Nature Conference on Big Data and Cancer Precision Medicine*, 2018. Boston, MA.

[10] **Amir Asiaee**, Samet Oymak, Arindam Banerjee, and Kevin R. Coombes. Precision cancer treatment with high dimensional data sharing model. *Stanford Conference on Big Data in Precision Health*, 2018. Stanford, CA.

[11] Min Ho* Cho, **Amir Asiaee***, and Sebastian Kurtek. Elastic statistical shape analysis of biological structures with case studies: A tutorial. *Bulletin of Mathematical Biology*, 81(7):2052–2073, 2019.

[12] **Amir Asiaee T.**, Hardik Goel, Shalini Ghosh, Vinod Yegneswaran, and Arindam Banerjee. Time series deinterleaving of dns traffic. In *1st Deep Learning and Security Workshop*, 2018.

[13] Cancer Genome Atlas Research Network, John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas Pan-Cancer analysis project. *Nature genetics*, 45(10):1113–1120, October 2013.

[14] Min Wang, Zachary B Abrams, Steven M Kornblau, and Kevin R Coombes. Thresher: determining the number of clusters while removing outliers. *BMC bioinformatics*, 19(1):9, January 2018.

[15] Zachary B Abrams, Mark Zucker, Min Wang, Amir Asiaee Taheri, Lynne V Abruzzo, and Kevin R Coombes. Thirty biologically interpretable clusters of transcription factors distinguish cancer type. *BMC genomics*, 19(1):738, 2018.

[16] **Amir Asiaee***, Zachary B. Abrams*, Samantha Nakayiza, Deepa Sampath, and Kevin R. Coombes. Explaining gene expression using twenty-one microrna s. *Journal of Computational Biology*, Forthcoming, 2020 (*Equal contribution).

[17] **Amir Asiaee***, Zachary B. Abrams*, Samantha Nakayiza, Deepa Sampath, and Kevin R Coombes. Identification and comparison of genes differentially regulated by transcription factors and miRNAs. *bioRxiv*, page 803643, 2019 (*Equal contribution). https://doi.org/10.1101/803643.

[18] Reza Bayat Mokhtari, Tina S Homayouni, Narges Baluch, Evgeniya Morgatskaya, Sushil Kumar, Bikul Das, and Herman Yeger. Combination therapy in combating cancer. *Oncotarget*, 8(23):38022–38043, June 2017.

[19] Bissan Al-Lazikani, Udai Banerji, and Paul Workman. Combinatorial drug therapy for cancer in the post-genomic era. *Nature biotechnology*, 30(7):679–692, July 2012.

[20] Aleksandr Ianevski, Liye He, Tero Aittokallio, and Jing Tang. SynergyFinder: a web application for analyzing drug combination dose-response matrix data. *Bioinformatics*, 33(15):2413–2415, August 2017.

[21] **Amir Asiaee**, Kevin R. Coombes, Melanie Davis, and Erin Hertlein. Large scale screening of drug combination with single doses. *In Preparation*.

[22] Peter Bacchetti. Additive isotonic model. *Journal of the American Statistical Association*, 84(405):289–294, 1989.

[23] Philipp M Altrock, Lin L Liu, and Franziska Michor. The mathematics of cancer: integrating quantitative models. *Nature reviews. Cancer*, 15(12):730–745, December 2015.

[24] Niko Beerenwinkel, Roland F Schwarz, Moritz Gerstung, and Florian Markowetz. Cancer evolution: mathematical models and computational inference. *Systematic biology*, 64(1):e1–25, January 2015.

[25] Johannes G Reiter, Marina Baretti, Jeffrey M Gerold, Alvin P Makohon-Moore, Adil Daud, Christine A Iacobuzio-Donahue, Nilofer S Azad, Kenneth W Kinzler, Martin A Nowak, and Bert Vogelstein. An analysis of genetic heterogeneity in untreated cancers. *Nature reviews. Cancer*, August 2019.

[26] E R Fearon and B Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, 61(5):759–767, June 1990.

[27] R Desper, F Jiang, O P Kallioniemi, H Moch, C H Papadimitriou, and A A Schäffer. Inferring tree models for oncogenesis from comparative genome hybridization data. *Journal of computational biology: a journal of computational molecular cell biology*, 6(1):37–51, 1999.

[28] Niko Beerenwinkel, Jörg Rahnenführer, Rolf Kaiser, Daniel Hoffmann, Joachim Selbig, and Thomas Lengauer. Mtreemix: a software package for learning and using mixture models of mutagenetic trees. *Bioinformatics*, 21(9):2106–2107, May 2005.

[29] Niko Beerenwinkel, Jörg Rahnenführer, Martin Däumer, Daniel Hoffmann, Rolf Kaiser, Joachim Selbig, and Thomas Lengauer. Learning multiple evolutionary pathways from cross-sectional data. *Journal of computational biology: a journal of computational molecular cell biology*, 12(6):584–598, July 2005.

[30] Niko Beerenwinkel, Nicholas Eriksson, and Bernd Sturmfels. Conjunctive bayesian networks. *Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability*, 13(4):893–909, November 2007.

[31] Moritz Gerstung, Michael Baudis, Holger Moch, and Niko Beerenwinkel. Quantifying cancer progression with conjunctive bayesian networks. *Bioinformatics*, 25(21):2809–2815, November 2009.

[32] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[33] Hesam Montazeri, Jack Kuipers, Roger Kouyos, Jürg Böni, Sabine Yerly, Thomas Klimkait, Vincent Aubert, Huldrych F Günthard, Niko Beerenwinkel, and Swiss HIV Cohort Study. Large-scale inference of conjunctive bayesian networks. *Bioinformatics*, 32(17):i727–i735, September 2016.

[34] Phillip B. Nicol, Courtney Deaver, Oksana Chkrebtii, Kevin R. Coombes, Subhadeep Paul, and **Amir Asiaee**. Disjunctive bayesian network infers cancer progression. *In Preparation*.

[35] **Amir Asiaee**, Phillip B. Nicol, Kevin R. Coombes, Courtney Deaver, Oksana Chkrebtii, Subhadeep Paul, and Amanda E. Toland. Inferring evolution trajectory of melanoma. *In Preparation*.

[36] Camille Stephan-Otto Attolini, Yu-Kang Cheng, Rameen Beroukhim, Gad Getz, Omar Abdel-Wahab, Ross L Levine, Ingo K Mellinghoff, and Franziska Michor. A mathematical framework to determine the temporal sequence of somatic genetic events in cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41):17604–17609, October 2010.

[37] N Beerenwinkel and S Sullivant. Markov models for accumulating mutations. *Biometrika*, 96(3):645–661, September 2009.

[38] Benjamin J Raphael and Fabio Vandin. Simultaneous inference of cancer pathways and tumor progression from cross-sectional mutation data. *Journal of computational biology: a journal of computational molecular cell biology*, 22(6):510–527, June 2015.

[39] Peter Spirtes, Clark N Glymour, Richard Scheines, David Heckerman, Christopher Meek, Gregory Cooper, and Thomas Richardson. *Causation, prediction, and search*. MIT press, 2000.

[40] Roderick C Slieker, Matthias S Roost, Liesbeth van Iperen, H Eka D Suchiman, Elmar W Tobi, Francoise Carlotti, Eelco J P de Koning, P Eline Slagboom, Bastiaan T Heijmans, and Susana M Chuva de Sousa Lopes. DNA methylation landscapes of human fetal development. *PLoS genetics*, 11(10):e1005583, October 2015.

[41] Howard Cedar and Yehudit Bergman. Linking DNA methylation and histone modification: patterns and paradigms. *Nature reviews. Genetics*, 10(5):295–304, May 2009.

[42] Leo Breiman and Jerome H Friedman. Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):3–54, 1997.

[43] Sheng Chen and Arindam Banerjee. Alternating estimation for structured high-dimensional multi-response models. In *Advances in Neural Information Processing Systems*, pages 2838–2848, 2017.

[44] Seyoung Kim and Eric P Xing. Tree-guided group lasso for multi-response regression with structured sparsity, with an application to eqtl mapping. *The Annals of Applied Statistics*, 6(3):1095–1117, 2012.

[45] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

[46] Tony Cai, Weidong Liu, and Xi Luo. A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

[47] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine learning research*, 9(Mar):485–516, 2008.

[48] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[49] Pradeep Ravikumar, Martin J Wainwright, Garvesh Raskutti, and Bin Yu. High-dimensional covariance estimation by minimizing l_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011.

[50] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[51] Patrick Breheny and Jian Huang. Penalized methods for bi-level variable selection. *Statistics and its interface*, 2(3):369, 2009.

[52] R Dennis Cook. *An introduction to envelopes: dimension reduction for efficient estimation in multivariate statistics*, volume 401. John Wiley & Sons, 2018.

[53] Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):947–1012, 2016.

[54] Saskia Trescher and Ulf Leser. Estimation of transcription factor activity in knockdown studies. *Scientific reports*, 9(1):9593, July 2019.

[55] Hyun Yong Jin, Alicia Gonzalez-Martin, Ana V Miletic, Maoyi Lai, Sarah Knight, Mohsen Sabouri-Ghomi, Steven R Head, Matthew S Macauley, Robert C Rickert, and Changchun Xiao. Transfection of microRNA mimics should be used with caution. *Frontiers in genetics*, 6:340, December 2015.

[56] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M Stuart, and Cancer Genome Atlas Research Network. The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10):1113, 2013.

[57] GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*, 550(7675):204, 2017.

[58] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT Press, 2017.

[59] Christina Heinze-Deml, Jonas Peters, and Nicolai Meinshausen. Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6(2), 2018.

[60] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The econometrics journal*, 21(1):C1–C68, February 2018.

[61] Nicolai Meinshausen, Alain Hauser, Joris M Mooij, Jonas Peters, Philip Versteeg, and Peter Bühlmann. Methods for causal inference from gene perturbation experiments and validation. *Proceedings of the National Academy of Sciences*, 113(27):7361–7368, 2016.