

船井情報科学振興財団 留学報告書

Stanford University  
School of Medicine  
[Biomedical Informatics Ph.D. program](#)  
[谷川洋介](#)

2019年6月30日



写真1 学会参加の際に立ち寄ったハワイ休暇の様子

## PhD 学生生活の折り返し目標地点を過ぎました

大学院留学の3年目も終わろうとしています。[前回の報告書](#)では、博士候補生になったことを報告しましたが、その後も無事に、博論審査委員となってくれる先生方を見つけることができました。今年の5月にはThesis pre-proposal talk という中間報告を兼ねたpublic talk を行い、審査委員の先生方と、同僚達からフィードバックをもらいました。毎年、卒業されていく先輩方を見送っているうちに、いつしか自分の学年も上がり、学位取得までの折返し地点を過ぎつつあるように思います。財団をはじめとして、皆様にご支援をいただき、恵まれた環境で留学できていることに感謝し、後悔することがないように研究に取り組みたいと思います。



© Yosuke Tanigawa 2019

この作品は、[クリエイティブ・コモンズの表示 - 非営利 - 継承 4.0 国際ライセンス](#)で提供されています。  
ただし、写真や図表はこの限りではありません。

## 研究活動の様子 (1) Thesis pre-proposal talk を無事に終えました

博士論文審査委員の先生方とミーティングを行い、今までの研究活動と今後の方向性についてフィードバックをいただきました。このミーティングにあわせて、45分間、Thesis pre-proposal talk という博論の概要とこれまでの取り組みをまとめた発表をしました。博士課程の修了に向けたマイルストーンがより明確になったように思います。

## 研究活動の様子 (2) バイオマーカーの遺伝学

健康診断などで、血液検査や尿検査を受けたことのあるかたは少なくないかと思えます。これらの検査では、血液や尿の中のタンパク質などの成分を調べ、それをもとに疾患になっていないか、あるいはその予兆となるような傾向が見られないか調べます。では、これらの検査の対象として用いられるバイオマーカーには、遺伝的な要因が効いているのでしょうか。もし、遺伝的要因によって、バイオマーカーが影響を受けることがあるとすれば、それらは疾患のリスクにも影響を与えるのでしょうか。

これらの疑問に答えるため、我々は35万人・38種類のバイオマーカーのデータを用いて大規模遺伝学解析を行い、[論文の草稿を公開しました](#)[1]。データセットには、以前から Rivas Lab で解析に取り組んでいる UK Biobank というイギリスのものをメインに使い、ゲノムワイド相関解析という手法や、因果推論、遺伝子情報によるリスクスコアの構築などを行いました。とくに、病気に対して計算されたリスクスコアと、バイオマーカーに対して構成されたりリスクスコアを組み合わせることで、疾患リスクの予測性能が向上することを報告しています。

## 研究活動の様子 (3) 緑内障の遺伝学

人類遺伝学の手法を用いた遺伝的要因の発見・解析は、実社会に何らかの影響を与えることができるのでしょうか。たとえば、遺伝学の手法を用いて創薬のターゲットを探すことはできるのでしょうか？我々が行っている Biobank の解析は、これらの応用にも有用となるのでしょうか。

バイオバンクを用いた創薬候補の探索の例として、我々は緑内障の遺伝学に関する[論文の草稿を公開しました](#)[2]。これは、イギリスの UK Biobank と、フィンランドの FinnGen を用いた国際共同研究となっています。まず、UK Biobank を用いた解析により、ANGPTL7 という遺伝子に変異を持った人々は、そうでない人々と比べて、眼圧と緑内障のリスクが低くなることを発見し、フィンランドの集団でもこれらの効果が認められることを報告しました。この遺伝子を標的とした薬を創ることができれば、緑内障の予防につながる可能性があります。

## 研究活動の様子 (4) 高次元・大規模データの解析の手法開発に関する共同研究

Biobank などの大規模データを扱う研究の現場には、新しい統計解析手法をつくる機会が広がっています。このような観点から統計学部の先生方やその学生さんたちも、Biobank 研究に興味を持っていただいているようで、我々といくつかの共同研究を行っています。

高次元のデータ（関係があるかもしれない変数の数—たとえば遺伝子変異の数—が多い状況）の中で、ごく一部の変数のみが結果（たとえば血液中のタンパク質の量）に影響を与えている状況で、どの変数が重要であるのか（変数の選択）とそれらの変数の効果量の両方を同時に推定したいという問題設定は、遺伝学解析に限らず、よくある状況です。これを調べるた



© Yosuke Tanigawa 2019

この作品は、[クリエイティブ・コモンズの表示 - 非営利 - 継承 4.0 国際ライセンス](#)で提供されています。

ただし、写真や図表はこの限りではありません。

めの手法の一つに Lasso regression という手法があり、広くつかわれています[3,4]。しかし、Biobank スケールのデータセットは、既存の方法を適用するには大きすぎることもあり、Lasso などの手法はあまり使われていませんでした。

我々は、Lasso を考案した統計学部の先生方・その学生さんたちと共同研究を行い、高次元かつデータ点の数も大きいときにも適用可能な、batch screening iterative lasso (BASIL) という手法を開発しました。遺伝学解析のために、この手法を snpnet という統計ソフト R のパッケージとして実装し、UK Biobank を用いて BASIL の有用性を論文の草稿中で報告しました[5]。また、(2)の「バイオマーカーの遺伝学」で報告した論文の草稿にも、この BASIL-snpnet を用いています。遺伝学の問題設定をスタートとして、他の分野にも大きな影響を与えうる手法開発につながったことを嬉しく思います。

## Teaching assistant として知識表現論を教えました

冬学期には、BIOMEDIN210/CS270 “Modeling Biomedical Systems: Ontology, Terminology, Problem Solving” というクラスの Teaching assistant (TA) として教育に携わる機会を得ました。このクラスでは、人間が持っている知識をどのようにコンピュータに表現するかそしてそれをどのように問題解決に応用するか、ということをお教えます。現在、人工知能などの手法が盛り上がっていますが、知識表現に関する研究は、一つ前のブームのときから行われており、現在の機械学習などの手法と補完的なものであると思います。私はこの授業を履修したことはなかったのですが、他の TA や先生の助けを得ながら、なんとか乗り切ったことに安堵しています。とくに、自分の専門分野をもとに、“Ontologies in bioinformatics” というコマの授業をゼロから準備して講義できたときには、大きな達成感を感じました。

## 生活の様子

今年の1月に学会発表でハワイを訪れました（写真1、[前回の報告書](#)にレポートさせていただいた[授業プロジェクトの発表](#)[6]）が、それ以外ではキャンパス付近での生活が続いています。週末に、ハイキングやスキー、カヤックといったアウトドア活動や、ボードゲームなどのインドア活動をするのが息抜きになっています。最近では、船井奨学生の先輩である野田さんが卒業されるということで、タホ湖でのカヤックやフェアウェルパーティーなど、卒業をお祝いするイベントを企画しました。先輩方にいろいろとお世話になったことは、後輩たちにも良い形で受け継いでいければと思っています。

最後になりましたが、常日頃から支援を頂いている[船井情報科学振興財団](#)に感謝して、この報告書の結びとさせていただきます。素晴らしい環境に恵まれていることに感謝し、在学中の残された時間内にできる限りの研究成果をあげられるよう引き続き頑張ります。



© Yosuke Tanigawa 2019

この作品は、[クリエイティブ・コモンズの表示 - 非営利 - 継承 4.0 国際ライセンス](#)で提供されています。ただし、写真や図表はこの限りではありません。

## 参考文献

1. Sinnott-Armstrong, N.\*, **Tanigawa, Y.\*** et al. Genetics of 38 blood and urine biomarkers in the UK Biobank. bioRxiv 660506 (2019). [doi:10.1101/660506](https://doi.org/10.1101/660506)
2. **Tanigawa, Y.**, et al. Rare protein-altering variants in *ANGPTL7* lower intraocular pressure and protect against glaucoma. bioRxiv 677443 (2019). [doi:10.1101/677443](https://doi.org/10.1101/677443)
3. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267–288 (1996).
4. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33, 1–22 (2010).
5. Qian, J., Du, W., **Tanigawa, Y.** et al. A Fast and Flexible Algorithm for Solving the Lasso in Large-scale and Ultrahigh-dimensional Problems. bioRxiv 630079 (2019). [doi:10.1101/630079](https://doi.org/10.1101/630079)
6. Anand, S.\*, Kalesinskas, L.\*, Smail, C\*. & **Tanigawa, Y.\*** SNPs2ChIP: Latent Factors of ChIP-seq to infer functions of non-coding SNPs. in Biocomputing 2019 184–195 (WORLD SCIENTIFIC, 2018). [doi:10.1142/9789813279827\\_0017](https://doi.org/10.1142/9789813279827_0017).



© Yosuke Tanigawa 2019

この作品は、[クリエイティブ・コモンズの表示 - 非営利 - 継承 4.0 国際ライセンス](https://creativecommons.org/licenses/by-nc-sa/4.0/)で提供されています。  
ただし、写真や図表はこの限りではありません。