

## II. 分担研究報告

### ヒト ES/iPS 細胞遺伝子発現プロフィールのバイオインフォマティクス解析

研究分担者 水口 賢司

独立行政法人 医薬基盤研究所 創薬基盤研究部  
バイオインフォマティクスプロジェクト プロジェクトリーダー

**研究要旨：**現在は、多くのヒト ES/iPS 細胞が国内外の様々な研究所において樹立され、様々な培養方法により培養され、研究に使用されている。これらの細胞を治療目的あるいは創薬研究目的のために利用するには、詳細な細胞特性分析や培養手順の標準化が必要である。これらの細胞株の差異、及び培養条件の差異を調査するために、我々は、NIBIO ヒト幹細胞応用開発室において出されたヒト ES/iPS 細胞の遺伝子発現解析結果について、バイオインフォマティクス分析を行った。

#### A. 研究目的

ゲノム解析技術や各種ハイスループット実験技術の進展に伴い、医学生物学のいずれの分野においても大規模データの取り扱いが日常的なものとなり、コンピュータ解析が必須となっている。扱われるデータの種類は、遺伝子発現、相互作用ネットワーク、タンパク質立体構造など多種多様だが、それらの解析には、データベース技術や統計学、機械学習(コンピュー

タがデータから自動的にルールを抽出し学習する技術の総称)などの共通技術が用いられ、これらの基本技術に支えられ、生物情報からの知識抽出を目指す分野一般を広い意味でバイオインフォマティクスと呼ぶことが多い。

本研究では、各種 ES/iPS 細胞を特徴付けるために遺伝子発現情報を用いるが、その最初の段階では、膨大なデータ点(例えば、細胞株数×遺伝子数×実験条件)から何らかのパターン

を見いだせるかどうかという探索的な解析が重要な役割を果たす。階層的クラスタリングは、その目的のために一般的に広く用いられる手法であり、データ点にある共通の特徴を持つ部分集合(クラスタ)に分割する。例えば、各サンプルについて、特定の一群の遺伝子の発現量(測定値の組)を「遺伝子発現プロフィール」と呼ぶことにすると、2つのサンプル間でどの程度遺伝子発現プロフィールが似ているかを定義できるので、近い発現プロフィールを持つサンプルから順番に繋げていくことで、クラスタを構築することができる。

これらのクラスタリング結果の可視化のために、本研究ではヒートマップを用いる。これは、データを行列の形に整理し(例えば、行が遺伝子で列がサンプル)、データ値を色で表したものである。クラスタリングの結果に従って行と列の順序を入れ替えることにより、似た値を持つデータ点が近接し、視覚的に部分集合を識別することができる。

未分化状態で維持培養されたヒト ES/iPS 細胞は、Oct3/4、Nanog、Tra1-60、Tra1-81、Tra2-54、SSEA3、SSEA4 等の未分化マーカーを発現し、SSEA1 等の分化マーカーを発現しない。このことを判定する免疫染色やフ

ローサイトメトリー(FCM)解析は、細胞の培養状態、すなわち未分化/分化状態の比較的解釈が簡単な評価方法であり、世界中で採用されている。NIBIO ヒト幹細胞応用開発室においても、この評価方法が細胞の品質管理のひとつとして採用され、改良・改善を加えつつ、継続的に運用されている。しかし、免疫染色や FCM 解析では、基本的には蛋白質や糖脂質など抗体の抗原となりうる遺伝子の発現解析が可能であるが、mRNA レベルの遺伝子発現の解析が困難である。また、アフィニティーの高い抗体がある場合にのみ解析が可能であり、一度に解析できる蛋白質や糖脂質の数はある程度限られているため、網羅的に解析をすることはできない。しかし、ヒト ES/iPS 細胞を治療目的あるいは創薬研究目的のために利用するには、詳細な細胞特性分析や培養手順の標準化が求められている。そこで、NIBIO ヒト幹細胞応用開発室では、免疫染色、FCM 解析と平行して、未分化状態で維持培養されたヒト ES/iPS 細胞の RNA レベルの遺伝子発現を網羅的・定量的に測定するために、細胞株毎、培養条件ごとに RNA サンプルを採取し、PCR アレイをおこなってきた。我々は、NIBIO ヒト幹細胞応用開発室において蓄積された PCR アレイ測

定結果をより詳細に解析し、細胞株の差異、及び培養条件の差異を調査することを第一の目的とし、バイオインフォマティクス解析を進めた。さらに、より迅速且つ正確で詳細な遺伝子発現プロフィール解析を進めるための解析手順を策定することを目的とし、アレイデータの取り扱い方法についての検討を行った。

## B. 研究方法

### 未分化状態でのヒト ES/iPS 細胞の遺伝子発現プロフィール解析

NIBIO ヒト幹細胞応用開発室において、未分化状態でのヒト ES/iPS 細胞の特性を明らかにするため、未分化維持培養をおこなったヒト ES/iPS 細胞から RNA を抽出し、Human Stem Cell Pluripotency Array (Applied Biosystems) を用い、定量 RT-PCR による遺伝子発現プロフィール解析が行われた。その検査方法は、H23 年度に策定を行った国際幹細胞イニシャティブ (ISCI) にて策定された方法と同様のものがあった。

この PCR アレイ (Human Stem Cell Pluripotency Array) において発現量が測定される遺伝子は 96 遺伝子

であり、このうち 6 遺伝子 (ATCB、RAF1、CTNNB1、GAPD、EEFA1、18S) はハウスキーピング遺伝子と呼ばれる一般に発現量が変化しにくいとされる遺伝子群、残る 90 遺伝子が本来の解析対象の遺伝子群である (表 2)。各遺伝子の発現量の測定結果は CT 値 (ある一定の核酸量が測定されるまでに必要な PCR の増幅回数; 遺伝子発現量が高いと CT 値は低い) として表される。ハウスキーピング遺伝子をコントロールとしてこの CT 値に補正をかけ、サンプル間で遺伝子発現量を直接比較できるように算出したものが CT 値である。我々は、まず、CT 値を算出する際のコントロールとして適切なハウスキーピング遺伝子 3 遺伝子 (ATCB、RAF1、GAPD) を選抜し、これらをもとに算出された CT 値を用いて解析を進めた。

階層的クラスタリングは、R 統計ソフトウェア (<http://www.r-project.org>) の hclust 関数で実行し、距離としてユークリッド距離を使用、average-linkage のクラスタリングを行った。ヒートマップは、R による gplot パッケージ中の heatmap2 関数により作成した。

## C. 研究結果

## ハウスキーピング遺伝子群：

前述したとおり、ハウスキーピング遺伝子群は一般的に発現量の変化しにくいものとされている。しかしながら、ヒト ES/iPS 細胞においてはハウスキーピング遺伝子群の発現の挙動が一般細胞と異なり、必ずしも一定でないということが我々の解析によっても明らかとなった。そこで、このバイオインフォマティクス解析においては、CT 値による解析をより正確なものにするために、ATCB、RAF1、および GAPD の 3 遺伝子のみをコントロール遺伝子として取り扱うことと決定した。

## プレ解析：

NIBIO ヒト幹細胞応用開発室において、ヒト ES/iPS 細胞の未分化維持培養過程におけるヒト幹細胞の未分化マーカーや早期分化マーカーなどの遺伝子群の発現の安定性を確認するために、Human Stem Cell Pluripotency Array を用いて、対象株について異なる継代数で複数回、測定された。その結果について、我々はまず、上記のコントロール遺伝子を用いて CT 値の算出を行った。

## 遺伝子発現プロフィール解析

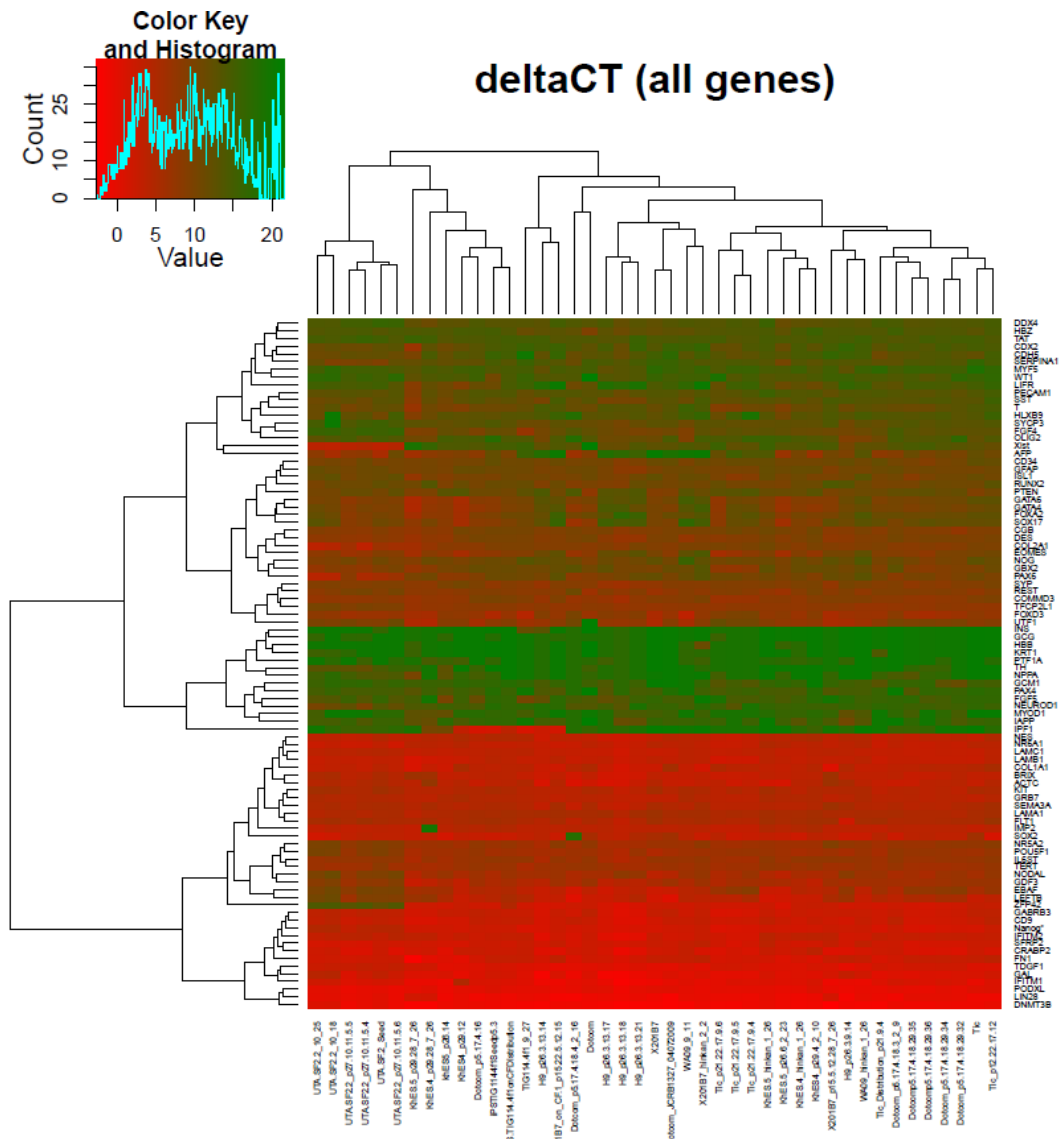
これまでに、このサンプルごとの CT 値比較解析により、同一の細胞株、同一の培養方法によってえられた細胞サンプルであっても、継代数が異なる場合、発現量が一定している遺伝子がある一方、ある程度発現量の異なる遺伝子が存在することが明らかとなった。つまり、未分化維持培養を続けていたとしても、継代ごとに遺伝子発現プロフィールが異なるということである。我々は、このことに注目をして、より詳細にサンプル間の遺伝子発現プロフィール解析を行った。

UTA-SF2-2 は、東京大学浅島研究室にて無フィーダー・無血清にて樹立され寄託されたヒト iPS 細胞である。NIBIO ヒト幹細胞応用開発室では、UTA-SF2-2 の資源化にあたり、ウシファイブロネクチンならびに hESF9a 培地を用いた無フィーダー・無血清の培養工程における品質管理法を策定し、実際に資源化をおこなっている。その過程で、品質管理の一環として、RNA サンプルを 3 回、異なる継代数で採取し、遺伝子発現プロフィール解析が行われた。遺伝子発現プロフィールは 3 サンプル間で非常によく一致し、このサンプルにおいて

遺伝子発現が安定していることが明らかとなった。

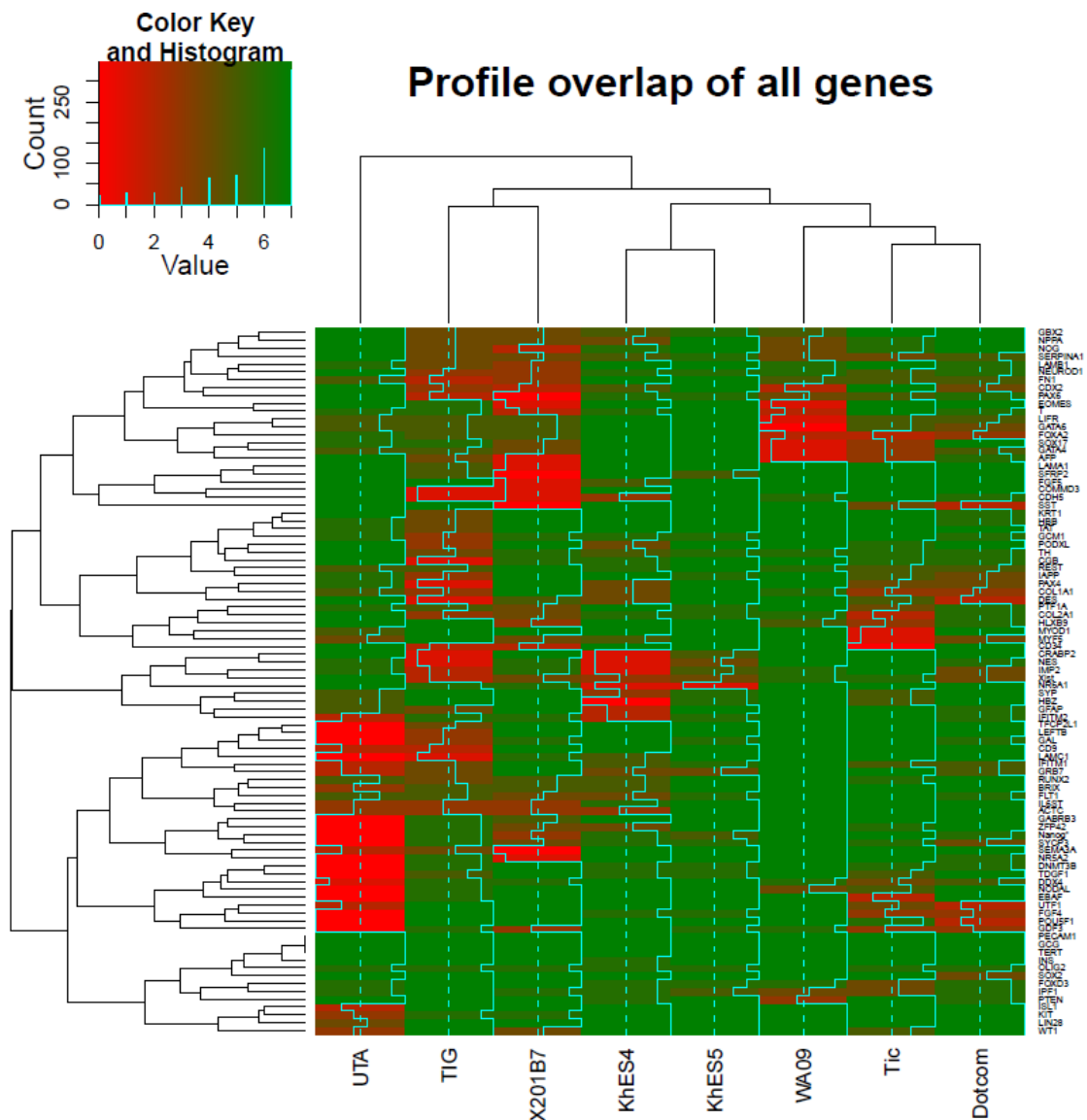
一方、Dotcom は分化誘導に適しているため資源化細胞の需要も高いが、一般には安定して未分化に維持して培養することが難しいとされている。この Dotcom は未分化維持培養の過程においても、たびたび分化する傾向があるため、資源化や研究で品質管理を行った場合に、免疫染色や FCM 解析の結果、未分化細胞の割合が通常より低い（全体の 80% 以下）と判定されることがある。今回、このバイオインフォマティクス解析をおこなった

サンプルは、少なくとも 80% 以上未分化細胞が含まれるという細胞集団のサンプルから RNA を抽出し、遺伝子発現を測定したものであるが、個々のサンプル間の遺伝子発現プロフィールは大きく異なる場合があることが判明した。つまり、免疫染色や FCM 解析の未分化/分化マーカー発現解析のみでは判明しなかった遺伝子発現の変化が、PCR アレイによる網羅的遺伝子発現測定とバイオインフォマティクス解析を組み合わせることによって詳細に解析できたということを示している。



**図1 遺伝子発現プロファイル解析例（ヒートマップ）**

32 サンプル（8 細胞株）について 90 遺伝子の発現量を  $\Delta CT$  値で示したもの。各サンプル（横軸）について、遺伝子発現プロファイルの類似度をユークリッド距離によって定義し、階層的クラスタリングを行なうと共に、各遺伝子（縦軸）についてサンプル群による発現パターンで階層的クラスタリングを行なった。



**図2 各細胞株を特徴付ける遺伝子発現パターンの比較**

まず 8 つの各細胞株毎に、各遺伝子の発現量の分布を計算した。次に、各細胞株と各遺伝子に対して、発現量分布の 25%-75%の範囲が別の細胞株の対応する範囲と重なる場合には 1 を重ならない場合には 0 を与えた。最後に、他の全ての細胞株との比較でこの数を足し合わせたものを元の細胞株と遺伝子に割り当てた（最小が 0 で最大が 7。この数を「細胞株非特異的発現量」と呼ぶ）。ヒートマップは、細胞株非特異的発現量を示したもので、遺伝子と細胞株のそれぞれについて、階層的クラスタリングを実行している。

## D. 考察

図 2 から、限られた少数の遺伝子のみが、細胞株に特異的な発現分布を示すことがわかった。特に、UTA は他の細胞株に比べて、より多くの特異的な遺伝子を持っている。また、Dotcom と Tic は、この解析による遺伝子発現パターンの観点からは極めて類似していることが示された。(但し、分化マーカーとして知られている GCM1、LAMB1、NES の発現量分布は、Tic に特異的である。) さらに、KhES4 と KhES5 は同じクラスタに入ることが示された。

本報告には示していないが、同様の解析を特定のマーカー遺伝子のみについて行なったところ、各細胞株(特に UTA) を特徴付ける遺伝子のほとんどは、分化マーカーであることが分かった。

本研究で新たに定義した「細胞株非特異的発現量」は、各遺伝子の発現量分布がどの程度、細胞株に固有かを定量化した点がユニークだと考えられる。これまでに報告された類似の解析のほとんどが、細胞株ペアの比較にとどまるのに対して、この量を用いることで、3 つ以上の細胞株についての特徴を解析することが可能になった。

但し、ここで定義した「細胞株非特

異的発現量」は、その遺伝子の発現量の大小を示しているわけではない点に注意する必要がある。特異的な遺伝子として同定されたものが、各細胞株でどのような発現量を示すかは元のデータに戻って調べねばいけない。

本研究では、遺伝子の数に比べてサンプル数は比較的少ない(各細胞株に 3-5 サンプル)ため、発見された遺伝子発現パターンが細胞株を十分に代表できているかどうかについては、今後の実験で検証していく必要があるだろう。そのためにも、本研究で得られた網羅的な遺伝子発現情報などのデータを蓄積し、将来の解析と組み合わせられる形で整理しておくことが重要である。

## E. 結論

網羅的遺伝子発現測定とバイオインフォマティクス解析を組み合わせ、ヒト ES/iPS 細胞の株間の差異や細胞の培養の状態の微妙な差異をより正確に的確に検出できることが確認された。ヒト iPS 細胞としての幹細胞特性検査の結果を効率的に解析していくためには、バイオインフォマティクス解析が不可欠である。その際に、遺伝子発現プロフィールだけでなく、



培養工程、培養に使用したマテリアル、染色体数や免疫染色・FCM 解析結果等の品質評価の情報をトレースできるようにセットで情報を管理していくことが、より発展した品質管理につながり、より詳細な細胞特性解析を可能にするだろう。今後は、このような膨大な情報を適切に管理し、解析に役立てていくことがより一層重要になってくる

## F . 健康危険情報

特筆すべき事項はなし。

### 参考論文

[1] Tang C. K., Aoshi T., Jounai N., Ito J., Ohata K., Kobiyama K., Aoshi T., Jounai N., Ito J., Ohata K., Kobiyama K., Dessailly B.H., Kuroda E., Akira S., Mizuguchi K., Coban C., Ishii K., The chemotherapeutic agent DMXAA as a unique IRF3-dependent type-2 vaccine adjuvant, *PLoS One*, 8(3):e60038. 2013

[2] Dessailly BH, Dawson NL, Mizuguchi K, Orengo CA., Functional site plasticity in domain superfamilies, *Biochim Biophys Acta*, 1834(5):874-89, 2013

[3] Tripathi L., Kambara, H., Chen Y. A. , Nishimura Y., Moriishi, K., Okamoto T., Morita, E., Abe, T., Mori, Y., Matsuura, Y., Mizuguchi K., Understanding the Biological Context of NS5A-Host Interactions in HCV Infection: A Network-Based Approach, *Journal of Proteome Research*, 12(6):2537-51, 2013

[4] Fujita J., Miyazaki Y., Hirose M., Nagao C., Mizohata E., Matsumoto Y., Mizuguchi K., Inoue T., Matsumura H., Expression, purification, crystallization and preliminary crystallographic study of FtsA from methicillin-resistant *Staphylococcus aureus*, *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 69(Pt 8):895-8, 2013

[5] Nystrom J., Igarashi Y., Ito M., Morita M., Nakatsu N., Yamada H., Mizuguchi K., Toxygates: interactive toxicity analysis on a hybrid microarray and linked data platform, *Bioinformatics*, 1;29(23):3080-6, 2013

[6] Yoshimaru T., Komatsu M., Matsuo T., Chen Y. A., Murakami Y., Mizuguchi K., Mizohata E., Inoue T., Akiyama M., Yamaguchi R., Imoto S., Miyano S., Miyoshi Y., Sasa M., Nakamura Y., Katagiri T.,

Targeting BIG3-PHB2 interaction to overcome tamoxifen resistance in breast cancer cells, *Nat Commun*, 4:2443, 2013

[7] Andrabi M., Mizuguchi K., Ahmad S., Conformational changes in DNA-binding proteins: Relationships with precomplex features and contributions to specificity and stability., *Proteins*, in press

[8] Hobro A. J., Standley D.M., Ahmad S., Smith N.I., Deconstructing RNA: optical measurement of composition and structure., *Phys Chem Chem Phys*, 15(31):13199-208, 2013

[9] Takemura N., Kawasaki T., Kunisawa J., Sato S., Aayam L., Kobiyama K., Aoshi T., Ito J., Mizuguchi K., Karuppuchamy T., Matsunaga K., Miyatake S., Mori N., Tsujimura T., Satoh T., Kumagai Y., Kawai T., Standley D., Ishii K., Kiyono H., Akira S., Uematsu S., Blockade of TLR3 protects mice from lethal radiation-induced gastrointestinal syndrome, *Nature Communications*, in press

[10] 水口賢司, 創薬支援のデータベースとバイオインフォマティクスに

よるデータ統合, *SAR News*, 24:2-6, 2013

## G. 研究発表

### 1. 学術論文発表

1 . Nagao C., Nagano N., Mizuguchi K., Prediction of detailed enzyme functions and identification of specificity determining residues by random forests, *PLoS One*, 9(1):e84623. 2014

### 2. 学会発表

#### 【国際学会：一般講演】

1 . Ahmad S., Mizuguchi K., Global gene co-expression patterns improve consistency between experimentally detected host factors crucial for influenza virus life cycle, *Keystone Symposium on Advancing Vaccines in the Genomics Era*, Rio de Janeiro, Brazil, 2013.11.1

2 . Ito J., Ahmad S., Ishii K., Mizuguchi K., A Comprehensive Analysis of miRNA Expression Profile in Human Serum Collected from Type-A Influenza Vaccine Clinical Trial, *Keystone Symposium on Advancing*

Vaccines in the Genomics Era,  
Rio de Janeiro, Brazil, 2013.11.2

3. Shirai H., Ikeda K., Yamashita K., Tsuchiya Y., Sarmiento J., Liang S., Mizuguchi K., Morokata T., Higo J., Standley D.M., Nakamura H., High-resolution modeling of antibody structures by a combination of bioinformatics, expert knowledge, and molecular simulations., Antibody Engineering and Therapeutics Conference, Huntington Beach, CA, USA, 2013.12.8

【国内学会：一般講演】

4. 岡村美菜子, 柳原佳奈, Ahmad S., 劉有容, 平田みつひ, 二川浩樹, 水口賢司, 古江 - 楠田美保, ヒト多能性幹細胞から肝細胞に優れた細胞株を予測するための評価方法の開発, 第 86 回日本組織培養学会年会, 大阪, 2013.5.30
5. 木田博, 濱野芳匡, 井上義一, 水口賢司, Tripathi L., 広瀬雅樹, 矢野幸洋, 多田康子, 西川博嘉, 坂口志文, 熊ノ郷淳, 特発性非特異的間質性肺炎における疾患特異的自己抗体の検索, 第 16 回間質性肺炎細胞分子病態研究会, 東京, 2013.8.24

6. 五十嵐芳暢, Nystrom J., 森田瑞樹, 伊藤真和吏, 中津則之, 山田弘, 水口賢司, Toxygates, トキシコゲノミクスデータと linked data の統合解析プラットフォーム, トーゴの日シンポジウム 2013, 東京, 2013.10.4

7. 伊藤真和吏, 森田瑞樹, 五十嵐芳暢, 陳怡安, 長尾知生子, 坂手龍一, 増井徹, 水口賢司, 創薬・疾患研究のための生命科学分野のデータベース横断検索 システム Sagace, トーゴの日シンポジウム 2013, 東京, 2013.10.5

8. 陳怡安, Tripathi L., 水口賢司, TargetMine, a data warehouse system for target discovery, トーゴの日シンポジウム 2013, 東京, 2013.10.5

9. 藤田純三, 宮崎祐満, 廣瀬未果, 長尾知生子, 溝端栄一, 松本佳巳, 水口賢司, 井上 豪, 松村浩由, メチシリン耐性黄色ブドウ球菌(MRSA)由来 FtsA の結晶化, 平成 25 年度日本結晶学会年会及び総会, 熊本, 2013.10.12

10. Andrabi M., Mizuguchi K., Ahmad S., DNA-binding-induced conformational changes in protein, 第 51 回日本生物物理学会年会, 京都, 2013.10.28

11. 池田和由, 伊東純一, 水口賢司,

富井健太郎, PoSSuM Updates and Integration With ChEMBL For Application of Drug Reuse, CBI学会 2013大会, 東京, 2013.10.28

12. Ahmad S., Mizuguchi K., Sequence-based prediction of interacting residue-pairs in proteins to integrate prediction of partners and binding sites, 日本バイオインフォマティクス学会 2013年年会 (JSBi 2013), 東京, 2013.10.29

13. 土屋裕子, 水口賢司, Analysis of antibody-antigen interactions and prediction of their complex structures, CBI学会 2013大会, 東京, 2013.10.29-30

14. 長尾知生子, 長野希美, 水口賢司, A random-forest based method that can predict detailed enzyme functions and also identify specificity determining residues, CBI学会 2013大会, 東京, 2013.10.29

15. 陳怡安, Tripathi L., 水口賢司, Applications of an integrated data warehouse system in to investigate complex biological systems, CBI学会 2013大会, 東京, 2013.10.29

16. 田端桂介, 有本大, Tripathi L., 水口賢司, 森田英嗣, フラビウイ

スタンパク質と宿主因子の相互作用ネットワーク解析, 第20回トガ・フラビ・ペスチウイルス研究会, 神戸, 2013.11.9

17. 岡村美菜子, 柳原佳奈, Ahmad S., 劉有容, 平田みつひ, 水口賢司, 二川浩樹, 古江-楠田美保, ヒト多能性幹細胞から肝細胞内胚葉分化誘導効率に優れた細胞株, 日本口腔組織培養学会, 日本歯科大学 東京, 2013.11.23

18. 伊藤真和吏, 五十嵐芳暢, 陳怡安, 長尾知生子, 坂手龍一, 水口賢司, 生命科学分野の横断検索サービスとセマンティック・ウェブ, 第36回日本分子生物学会年会, 神戸, 2013.12.6

#### 【国際学会：招待講演】

19. 水口賢司, Data integration and protein network analysis for early stage drug discovery, Structural Life Science 7<sup>th</sup> International Conference on Structural Genomics (ICSG2013-SLS), 札幌, 2013.7.31

#### 【学会以外のセミナー、講演会等】

20. 水口賢司, 医薬基盤研における創薬支援データベースの開発, シリーズ研究講演会「薬づくりの新しいR

- & Dモデルを探る」第1回, 東京, 2013.6.20
21. 水口賢司, 創薬支援のためのデータ統合とデータベース開発, 東北大学大学院情報科学研究科、仙台, 2013.7.10
22. 水口賢司, Computational and systems approaches to early stage drug discovery, 九州大学 生体防御医学研究所附属生体多階層システム研究センター, 福岡, 2013.9.25
23. 水口賢司, データ統合とネットワーク解析による創薬初期研究の支援, 第3回シスメックスプロテインカンファレンス、東京, 2013.10.18
24. 水口賢司, データ統合とネットワーク解析による創薬支援, 第9回 霊長類医科学フォーラム, つくば, 2013.11.14
25. 水口賢司, 創薬の初期研究におけるデータ統合:ターゲットと安全性の評価, 第345回 CBI学会研究講演会, 東京, 2014.1.9
26. 水口賢司, ‘アジュバントゲノミクス’に向けた統合データベースの現状, 第7回次世代アジュバント研究会, 大阪, 2014.1.21
27. Ahmad S., Ito J., Mizuguchi K., An integrated map of influenza-virus life-cycle host factors and their predicted micro-RNA regulators for bottom-up bio-marker discovery, 第7回次世代アジュバント研究会, 大阪, 2014.1.21
28. 伊東純一, Ahmad S., Tripathi L., 石井健, 水口賢司, インフルエンザワクチンが誘発する発熱」の予測へ向けた血清中マイクロ RNA マーカーの探索, 第7回次世代アジュバント研究会, 大阪, 2014.1.21
29. 水口賢司, データベースは、創薬初期でのターゲット評価と安全性の予測に役立つか?, 「創薬研究におけるバイオデータベース講習会」(第三回データベース講習会@大阪(池田)), 大阪, 2014.1.24
30. 水口賢司, 計算生物学によるシステムの理解から創薬へ, 京都大学理学部, 京都, 2014.2.18