

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 8 日現在

機関番号：17102

研究種目：新学術領域研究（研究領域提案型）

研究期間：2010～2014

課題番号：22132005

研究課題名（和文）ゲノム情報解析にもとづく性差構築機構の解明

研究課題名（英文）Studies on molecular mechanisms of sexual differentiation based on genome informatics approach

研究代表者

須山 幹太（Suyama, Mikita）

九州大学・生体防御医学研究所・教授

研究者番号：70452365

交付決定額（研究期間全体）：（直接経費） 52,800,000円

研究成果の概要（和文）：次世代シーケンサーの性能向上などの技術革新により、比較ゲノム解析やエピゲノム解析が急速に進展した。このようなハイスループットなゲノム解析をもとに、基本的な生命現象である「性差」を対象とし、特にそこに見られる遺伝子発現制御をゲノムワイドに包括的に解明することを目的として研究を行った。その結果、性差に関連する様々な事象、たとえば、体細胞に見られる男女差や、妊娠時の子宮内膜に見られる大きな遺伝子発現の変化を明らかにした。また、この過程で、性差だけでなく、広く分子医科学研究に有用な方法論の開発を行った。

研究成果の概要（英文）：Comparative genomics and epigenomic analyses have rapidly been developed with the technological progress of next-generation sequencing platforms. These high-throughput genome analysis tools enabled us to conduct studies focusing on some of the fundamental biological processes such as sexual differentiation in molecular levels. The major objective of our research was to comprehensively reveal gene regulation associated with the process in genome-wide fashion. We reveals that there is sexually different expression of genes even in somatic cells. We also elucidated a molecular mechanisms behind decidualization of endometrial stromal cells by a genome-wide gene expression analysis combined with the data from epigenetic changes. During this project, we have also developed some computational methodologies useful for general analyses in the field of molecular biology and medicine.

研究分野：比較ゲノム

キーワード：比較ゲノム バイオインフォマティクス 遺伝子発現制御 シス因子

1. 研究開始当初の背景

様々な生物種のゲノム配列決定が急速に進んでおり、研究開始当初でヒトやマウスをはじめとする約50種の脊椎動物ゲノムの比較解析が可能であった。また、2007年頃から次世代シーケンサーが急速に普及し、生命医科学研究の完了、さらにそれに続くエピゲノム解析の急速な進展から、ゲノムワイドな遺伝子発現制御の分子基盤の解釈が可能になりつつあった。これらの大規模データ解析には、バイオインフォマティクス手法が必須である。しかし、このような大規模配列データを新しい分野であるため、データの解析手法が十分に確立していなかった。「性差」の確立や維持において、ゲノムワイドな遺伝子発現の違いや、エピゲネティックな制御の研究がようやく可能になってきた時代的背景であったといえることができる。

2. 研究の目的

次世代シーケンサーの普及により、比較ゲノム解析やエピゲノム解析が急速に進展した。このようなハイスループットなゲノム解析をもとに、基本的な生命現象である「性差」を対象とし、特にそこに見られる遺伝子発現制御機構の解明を目的として研究を行った。とくに RNA-seq にもとづく遺伝子発現の性差や、それにヒストン修飾や DNA メチル化などのエピジェネティックな情報も複合的に解析することで、その分子基盤の解明を目指した。

3. 研究の方法

遺伝子発現を網羅的に調べるには、以前はマイクロアレイが主流であったが、次世代シーケンサーの普及にともない RNA-seq 法が広く用いられるようになってきた。これはマイクロアレイに比べ、RNA-seq 法はダイナミックレンジが広いことと、配列レベルでの解析のためスプライスバリエーションなどの検出も可能であるからである。そこで、まず、RNA-seq のデータから様々なスプライスバリエーションを検出する方法を開発し、さらに比較ゲノム解析を用いることで選択的スプライシングに関与すると考えられるシス因子の検出を行った。

次世代シーケンサーから得られるデータを処理し解釈できるかたちにするには、コンピュータを用いた一連の解析が必要である。解析のための要素プログラムはすでに多くの研究者により開発・公開がされており、それらを使ってデータ解析を進めるが、これを効率的に行うには、一連の操作をパイプライン化する必要がある。また、これらの要素プログラムは常に改良が加えられている。そのためその動向にも注意しながら解析パイプラインを維持することが必要である。また、データを深く解釈するには、独自の的方法論を開発することが重要となるため、それらの開発も進めてきた(研究成果(4)を参照)。

また、遺伝子発現制御を解析するには、RNA-seq のような実際の発現レベルのデータだけでなく、ENCODE などにより多くの知見が急速に蓄積しているヒストン修飾などのエピゲノムデータも複合的に解析することが重要である。そこでヒストン修飾や DNA メチル化のデータも合わせて解析した。

4. 研究成果

(1) トランスクリプトームに見られる性差遺伝子発現の網羅的な解析にはマイクロアレイや RNA-seq が主に用いられるが、RNA-seq の利点としては、配列レベルの解析であるため選択的スプライシングの差異も検出できることが挙げられる。そこで、RNA-seq データから選択的スプライシングを検出し、それらを制御するシスエレメントの同定(Bioinformatics 2013)や、それらシスエレメント間に存在するネットワークを明らかにした(Nucleic Acids Res. 2010)。ここで用いた方法を性差に着目した解析に適用することで性特異的な選択的スプライシングの検出も可能になると考えられる。そこでまず、ヒト B 細胞の RNA-seq のデータを公共データベースから取得し、体細胞である B 細胞においても遺伝子発現強度に性差が見られることを明らかにした。さらに同サンプルを用い、選択的スプライシングにおける性差の検出についても検討を進めている。

(2) 子宮内膜の脱落膜化に伴うエピゲノム変化

妊娠時に子宮内膜は脱落膜化という形態的および機能的な分化が起こる。この分化時の遺伝子発現変化やクロマチンのヒストン修飾の変化を調べるため、山口大学・医学部・産科婦人科学の杉野先生のグループと共同で研究を行った(論文)。一般に遺伝子発現の変化はヒストン修飾などのクロマチン構造の変化と関連していると考えられる。そこで、次世代シーケンサーを使い、子宮内膜間質細胞における RNA-seq 解析による網羅的な遺伝子発現と、遺伝子発現制御に直接関わっていると考えられる H3K27ac と H3K4me3 の2種のヒストン修飾を ChIP-seq 法によりゲノムワイドに決定し、脱落膜化を誘導した細胞のものと比較した。その結果、脱落膜化に伴い、H3K27ac と H3K4me3 は多くの遺伝子座において増加することを見出した。また、RNA-seq の結果から、脱落膜化に伴い 881 遺伝子の発現が亢進していた。その内、223 の遺伝子のプロモーターにおいて、H3K27ac あるいは H3K4me3 が増加が認められた。パスウェイ解析から、これらの遺伝子の中にはインスリンシグナリングと関連しているものが多く見られた。

(3) DNA メチル化状態の不均質性の解析
DNA のメチル化は遺伝子発現制御において重要な役割を果たしている。特に発生初期や配

偶子形成時に劇的な DNA メチル化の変化が見られることが知られている。次世代シーケンサーを応用したバイサルファイト法により、DNA のメチル化状態をゲノムワイドに配列レベルで明らかにすることが可能になった。次世代シーケンサーを用いることで、一度に数千万から数億本におよぶ 100 塩基程度の配列断片のメチル化情報が得られる。そこで得られた同一配列断片に存在する複数のメチル化サイトを解析することで、メチル化が配列上で相関しているのか、あるいは独立に起きるのかを知ることができる。そこでマウスの様々な段階にある生殖細胞や ES 細胞のゲノムワイドなメチル化状態を調べた次世代シーケンサーのデータを用い、その結果、多くの座位で隣接するメチル化サイトに連鎖が観測された。これは、メチル化が変動しているような状態、あるいは中間的なメチル化状態において、特にインプリント遺伝子の様に、父方あるいは母方由来の一方のアリルがメチル化による不活化を受けるような座位において、隣接するメチル化サイト間で強い連鎖が見られた。

(4) 比較ゲノミクス解析のための方法論の開発

転写因子の ChIP-seq のピーク中にシスエレメントである転写因子結合部位を見つけ出すには、生物種間での配列保存性を評価することが有効である。そこでシスエレメント探索を効率的に行うため、任意の部位でのゲノムアライメントを簡便に取得し、配列の進化的保存性を評価するためのプログラムである GenomeCons (<http://bioinfo.sls.kyushu-u.ac.jp>) を開発した(論文)

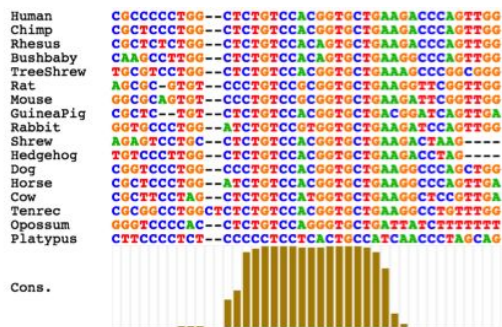


図1 GenomeCons による進化的に保存したシス因子の例。アライメント中で4種類の塩基をそれぞれ対応する色で塗り分けている。アライメントの下部の棒グラフが各座位における保存の強さを表している。転写因子結合部位に相当する中央の15残基が局所的に強く保存しているのがわかる。

GenomeCons を用いることで、コンピュータを用いたバイオインフォマティクス解析に不慣れた実験系の研究者も容易に比較ゲノム情報にアクセスすることが可能で、例えば ChIP-seq の結果得られたピーク中に進化的

に保存したシスエレメントがあるかどうかを簡単に調べることができる。

また、シスエレメントが同定できても、それが遠位エンハンサーとして働くことも考えられるため、その標的遺伝子を見つけるのが困難な場合が多い。その際、一つの有力な指標となるのは、染色体の空間配置である。もし、二つの座位が空間的に近接していれば、配列上は離れていても作用を及ぼすことが可能であると考えられる。そのような空間的近接性を実験的に網羅的に調べる方法として Hi-C 法があり、既にそのデータが公共データベースにて公開されているが、データ解析上、高度な技術を要するため、そのデータの活用は十分には進んでいない。そこで、既存の Hi-C データを誰もが活用できるようにするためのプログラムとして ChromContact の開発を進めてきた。このプログラムにより、既知の遠位エンハンサーが検出できることを確認した。さらに、GenomeCons 同様、一般の研究者に活用してもらうため、ウェブサーバーとしての公開準備を進めている。

さらに、次世代シーケンサーを用いたエピゲノム解析において注意すべき点として、そのデータに含まれるノイズへの対処が挙げられる。そこで、ChIP-seq のピーク検出プログラムがノイズに対してどの程度ロバストであるかを評価した。人工的に生成したノイズを様々な割合で混ぜたデータを作成し、既存のピーク検出プログラムを適用することで、ノイズに対するロバスト性を評価した。その結果、一般にピーク検出が難しいとされる転写活性領域に特徴的なヒストン修飾である H3K36me3 のような広範囲にわたるピークであっても、50%のノイズ含量までは、どのプログラムもノイズなしのものと同程度のピーク検出が行われることがわかった。

(3) 領域内グループとの共同研究

この領域では、次世代シーケンサーを積極的に活用してきた。そこで得られたデータを解釈する上で、バイオインフォマティクスを駆使したデータ解析が必須である。私たちのグループは、これまでの経験を活かし、シーケンサーデータ解析のためのパイプラインを整備や、新規プログラムの開発を行うことで、領域内のデータ解析をサポートしてきた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計11件)

T. Sato, M. Suyama, "GenomeCons: a web server for manipulating multiple genome sequence alignments and their consensus sequences", *Bioinformatics* 31, 1293-1295, 2015 (査読あり), DOI: 10.1093/bioinformatics/btu803.

H. Okae, H. Chiba, H. Hiura, H. Hamada, A. Sato, T. Utsunomiya, H. Kikuchi, H. Yoshida, A. Tanaka, M. Suyama, T. Arima, "Genome-wide analysis of DNA methylation dynamics during early human development", PLoS Genet. 10, e1004868, 2014 (査読あり), DOI: 10.1371/journal.pgen.1004868.

M. Yoshihara, L. Jiang, S. Akatsuka, M. Suyama, S. Toyokuni, "Genome-wide profiling of 8-oxoguanine reveals its association with spatial positioning in nucleus", DNA Res. 21, 603-612, 2014 (査読あり), DOI: 10.1093/dnares/dsu023.

I. Tamura, Y. Ohkara, T. Sato, M. Suyama, K. Jozaki, M. Okada, L. Lee, R. Maekawa, H. Asada, S. Sato, Y. Yamagata, H. Tamura, N. Sugino, "Genome-wide analysis of histone modification in human endometrial stromal cells", Mol. Endocrinol. 28, 1656-1669, 2014 (査読あり), DOI: 10.1210/me.2014-1117.

T. Baba, H. Otake, T. Sato, K. Miyabayashi, Y. Shishido, C.Y. Wang, Y. Shima, H. Kimura, M. Yagi, Y. Ishihara, S. Hino, H. Ogawa, M. Nakao, T. Yamazaki, D. Kang, Y. Ohkawa, M. Suyama, B.C. Chung, K. Morohashi, "Glycolytic genes are targets of the nuclear receptor Ad4BP/SF-1", Nat. Commun. 5, 3634, 2014(査読あり), DOI: 10.1038/ncomms4634.

[学会発表](計41件)

佐藤哲也, 須山幹太
ピーク検出プログラムのロバスト性を評価する方法.
第8回日本エピジェネティクス研究会年会、2014年5月26日(東京大学、東京).

M. Yoshihara, L. Jiang, S. Akatsuka, S. Toyokuni, M. Suyama
Genome-wide profiling of 8-oxoguanine reveals its association with spatial positioning in nucleus.
The 4D Nucleome 2014, 2014年12月18日 (Aki GROUND HOTEL, Hiroshima).

佐藤哲也、大川恭行、須山幹太
RNA-seq データを利用したヒストン修飾領域同定法.
第36回日本分子生物学会年会、2013年12月5日(神戸ポートアイランド、神戸).

M. Suyama

Mechanistic analysis of transcription by using RNA-seq data.
Frontiers in Bioinformatics, 2012年10月20日 (Girona, Spain).

佐藤哲也、須山幹太
ChIP-seq データを用いた核内受容体共役因子の探索.
第34回日本分子生物学会年会、2011年12月14日 (パシフィコ横浜、横浜).

[図書](計0件)

[産業財産権]
出願状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

取得状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
取得年月日:
国内外の別:

[その他]
ホームページ等
<http://bioinfo.sls.kyushu-u.ac.jp/genomecons/>

6. 研究組織

(1) 研究代表者

須山 幹太 (SUYAMA, Mikita)
九州大学・生体防御医学研究所・教授
研究者番号: 70452365

(3) 連携研究者

内山 郁夫 (UCHIYAMA, Ikuo)
基礎生物学研究所・理論生物学領域・助教
研究者番号: 90243089