

平成 29 年 6 月 5 日現在

機関番号：82626

研究種目：若手研究(B)

研究期間：2015～2016

課題番号：15K16089

研究課題名（和文）ゲノム3次元構造データに基づく共局在遺伝子の網羅的探索

研究課題名（英文）Finding 3D co-localization of genomic elements from HiC data

研究代表者

齋藤 裕 (Saito, Yutaka)

国立研究開発法人産業技術総合研究所・人工知能研究センター・研究員

研究者番号：60721496

交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：ハイスループット実験技術HiCによって得られたゲノム3次元構造データから、遺伝子の共局在を網羅的に探索する情報解析手法Coseargeを開発した。Coseargeによって胚性幹細胞のHiCデータを解析した結果、先行研究において提案されたtopologically associated domainとは異なる新しい共局在遺伝子を多数発見できた。特に、胚性幹細胞に特異的な遺伝子であるSOX2とDNAダメージ修復関連の遺伝子が共局在するという新しい知見が得られた。

研究成果の概要（英文）：We developed Cosearge, a bioinformatics method for finding 3D co-localization of genomic elements from HiC data. We applied Cosearge to HiC data from embryonic stem cells, and detected a number of novel co-localized genes distinct from topologically associated domains proposed in existing studies. In particular, we found that SOX2 and DNA damage response genes are co-localized in embryonic stem cells.

研究分野：エピゲノム情報解析の新規手法開発

キーワード：ゲノム3D構造 エピゲノム HiC/3C

1. 研究開始当初の背景

ゲノムの形成する染色体規模の3次元構造において、複数の遺伝子領域が空間的に近接(共局在)することで協調的な発現制御を受ける例が報告されており、遺伝子制御の新たなパラダイムとして注目を集めていた。HiCなどの実験技術の登場によりゲノム3次元構造の大量データが得られるようになったが、遺伝子の共局在について網羅的な解析を行うための情報基盤は整備されていなかった。そこで本研究は、HiCデータから共局在遺伝子を網羅的に探索する情報解析手法の開発を目的とした。

2. 研究の目的

ゲノムはヒストンやDNA結合タンパク質との相互作用によって折り畳まれ、染色体規模の巨大な3次元構造を形成している[Cavalli et al., Nat Struct Mol Biol, 20(3):290-299, 2013]. ゲノム3次元構造の有する様々な機能の中で特に大きな注目を集めているのが、遺伝子領域の空間的な近接、すなわち「共局在」である。図1のように、配列上で離れた位置にある遺伝子や異なる染色体上の遺伝子が、ゲノム3次元構造を介して共局在することが報告されている。共局在遺伝子は共通の転写因子の標的になりやすく、多くの場合に協調的な発現制御を受ける。また、ゲノム3次元構造は細胞分化や疾患によって変化するため、それぞれの組織には特異的な共局在遺伝子が存在すると考えられている。例えば、胚性幹細胞(ESC)において多能性に関与する転写因子の標的遺伝子が特異的に共局在していることが、ごく最近の研究によって明らかになった[de Wit et al., Nature, 501(7466):227-231, 2013.]. こうしたゲノム3次元構造における遺伝子の共局在は非常に新しい研究分野であり、様々な細胞種の共局在遺伝子の全容とその組織特異性について、大部分は謎に包まれている。

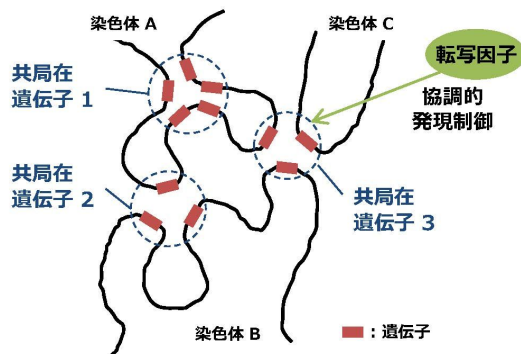


図1 ゲノム3次元構造と遺伝子の共局在

ゲノム3次元構造を測定するための実験技術として、近年、次世代シーケンシングを応用したハイスループットな方法が登場している[Dekker et al., Nat Rev Genet, 14(6):390-403, 2013.]. これらは実験プロトコルの種類によりHiC, 3C, 4C, 5Cなど色々

な名称で呼ばれるが、本報告書では統一してHiCと表記する。図2のように、HiCは3次元構造上の距離を配列リード数によって測定する実験技術であり、出力はゲノムの任意の2点間の距離を表す行列となる。HiC行列から得られるゲノムの全遺伝子ペアの距離情報は、共局在遺伝子を同定するために極めて有用であり、先行研究においても幾つかの情報解析手法が提案されてきた。しかし、既存手法は全て仮説検証型、すなわち、あらかじめ共局在遺伝子の候補として特定の遺伝子セットだけに注目して、それらが共局在しているか否か判定することを目的としている。こうした手法では、どの遺伝子セットに注目すべきか手掛かりのない状態から、共局在遺伝子を新規に発見することは不可能である。上述したESCの研究も、多能性に関与することが既に分かっている転写因子の標的遺伝子だけに注目して共局在を検証したものであり、その他の遺伝子の共局在についてはHiC行列が得られているにも関わらず全く発見できなかった。HiC行列のゲノムワイドな3次元構造データを最大限に活用して、事前知識に依存せず共局在遺伝子を発見できる手法は、未だ存在しなかった。

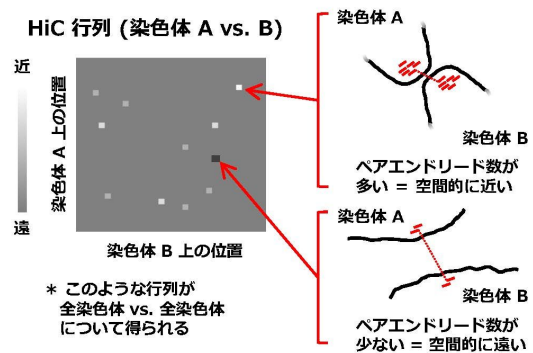


図2 HiC とゲノム3次元構造データ

以上のような背景から、HiCによって得られるゲノム3次元構造データに基づいて、共局在遺伝子を網羅的に探索する情報解析手法の開発を本研究の目的とした。

3. 研究の方法

(1) ハイスループット実験技術HiCによって得られるゲノム3次元構造データに基づいて、遺伝子セットの共局在の強さを評価するためのスコアを設計した。先行研究において提案された共局在スコア[Paulsen et al., Nucleic Acids Res, 41(10):5164-5174, 2013.]を改良して、探索アルゴリズムに適した新しい共局在スコアを開発した。先行研究では遺伝子セットの共局在を判定する統計検定が提案されており、セット内の全遺伝子ペアに関するHiC行列の各要素の平均値を計算して、この値をランダムサンプリングした遺伝子セットと比較することで、経験的なP-valueを求めている。しかし、この方法はP-valueが非常に小さい(共局在が強い)場

合に、必要なサンプリング回数が増加して計算コストが膨大になってしまう。そこで本研究では、P-value の代わりに少数のサンプルから計算可能な Z-score を求めて、これを共同在スコアとした。

(2) 高い共同在スコアを示す遺伝子セットを発見するために、網羅的な探索アルゴリズムを開発した。ここでは貪欲探索に基づくシンプルなアルゴリズムを採用した。次の方法で共同在遺伝子の網羅的な順位付きリストを作成した。

全遺伝子から探索を行い、第1位の共同在遺伝子とする。

これらの遺伝子を取り除いて再度探索を行い、第2位の共同在遺伝子とする。以上を逐次的に繰り返していき、各順位の共同在遺伝子の P-value を既存手法によって計算し、 $P=0.01$ などの閾値を下回った時点で終了する。

(3) 様々な細胞の HiC データに探索アルゴリズムを適用した。共同在遺伝子に機能アノテーションを行うため、ENCODE などの公開オミクスデータを活用したエピゲノムデータ解析を行った。ChIP-Seq データを利用して共同在遺伝子に共通の転写因子が結合しているか解析し、RNA-Seq データを利用してこれらが全て高発現もしくは低発現となっているか調べることで、転写因子による協調的な発現制御を検証した。

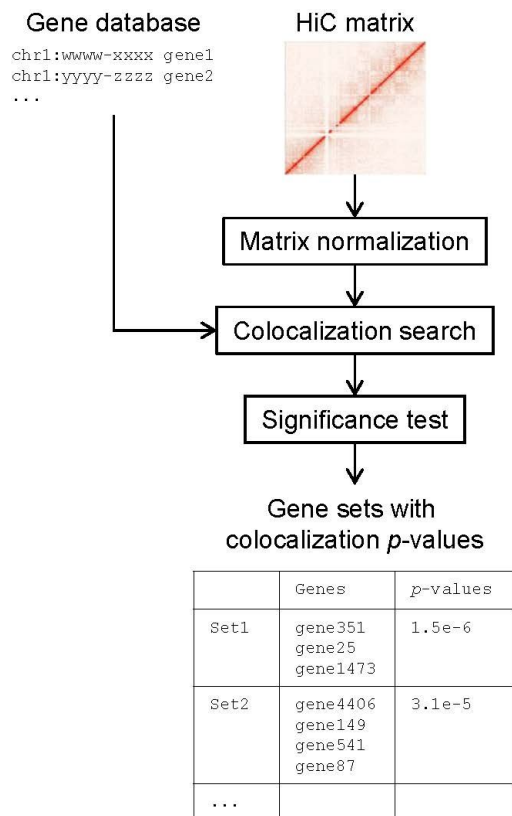


図3 Cosearge アルゴリズム

4. 研究成果

研究の方法(1),(2)のアルゴリズムをソフトウェア Cosearge として実装した。これにより HiC データに基づく共同在遺伝子の網羅的探索が初めて可能になった(図3)。

まず、ヒト ESC と IMR-90 細胞株の HiC データに Cosearge を適用して、共同在遺伝子の網羅的探索を行った。ゲノム 3 次元構造の先行研究では、topologically associated domain (TAD) という概念が広く受け入れられている。これはゲノム 3 次元構造が配列上で連続した領域からなる複数のドメインに分割できるという考え方である。Cosearge によって発見された共同在遺伝子を TAD と比較した。その結果、共同在遺伝子の大部分は複数の TAD にまたがって存在していることがわかった。これにより、TAD の概念に依存した先行研究のアプローチでは、ゲノム上に散在する共同在遺伝子を網羅的に検出できないことが示された(図4)。

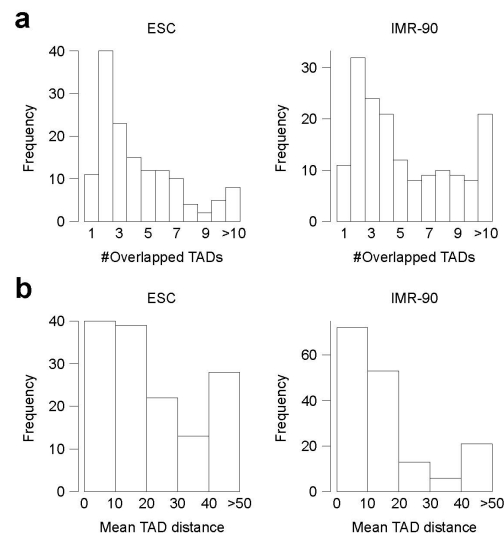


図4 共同在遺伝子と TAD の比較

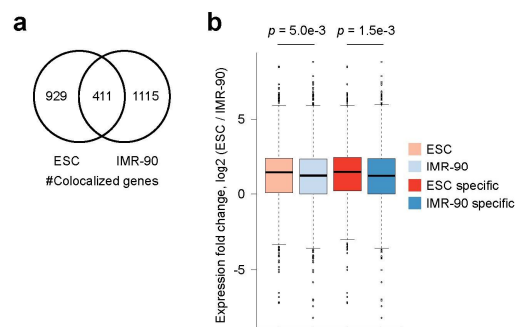


図5 共同在遺伝子の組織特異性

次に、ESC と IMR-90 の共同在遺伝子を比較することで、共同在遺伝子の組織特異性を解析した。その結果、多くの共同在遺伝子は ESC と IMR-90 のどちらか一方にしか現れない組織特異的な共同在遺伝子であることがわかった。また、これらの組織特異的な共同在遺

伝子は ESC と IMR-90 の間で遺伝子発現が変化していることがわかった。これにより、ゲノム 3 次元構造における遺伝子共局在が発現制御に関与していることが示された(図 5)。

さらに、ESC の共局在遺伝子に対して ENCODE の公開エピゲノムデータを利用した機能アノテーションを行った。(図 6)

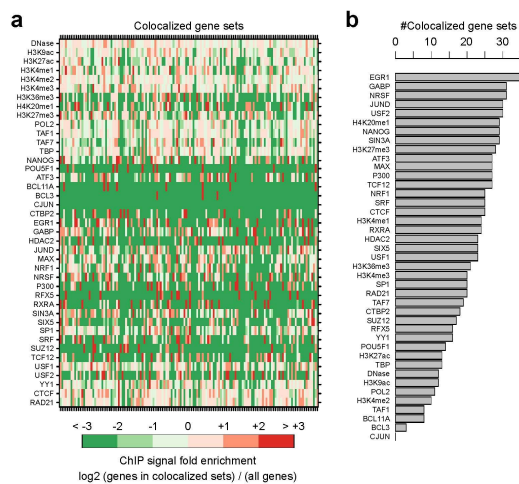


図 6 共同在遺伝子のエピゲノム状態

その結果、いくつかの共同在遺伝子について ESC 特異的な転写因子である NANOG や POU5F1 の ChIP-seq シグナルが検出された。これらの共同在遺伝子は NANOG や POU5F1 によって協調的な発現制御を受けていると考えられる。また、これらの共同在遺伝子に含まれる遺伝子を調べたところ、ESC 特異的遺伝子である SOX2 と DNA ダメージ修復に関与する遺伝子が共同在している例が発見された。DNA ダメージ修復は ESC において重要な機能をもつことが知られているが、その遺伝子が SOX2 と共同在していることは本研究で初めて得られた知見である。この結果は、Cosearge による事前知識に依存しない網羅的探索が、先行研究では得られなかったまったく新しい共同在遺伝子を発見できることを示している。

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 3 件)

Yutaka Saito, Chie Sugimoto, Toutai Mituyama, *Hiroshi Wakao. Epigenetic silencing of V(D)J recombination is a major determinant for selective differentiation of mucosal-associated invariant T cells from induced pluripotent stem cells. *PLOS ONE*, 12(3):e0174699, 2017. doi: 10.1371/journal.pone.0174699. (査読有)

Yuzo Takayama, Tamami Wakabayashi, Hiroko Kushige, Yutaka Saito, Yoichiro Shibuya, Shinsuke Shibata,

Wado Akamatsu, Hideyuki Okano, *Yasuyuki S. Kida. Brief exposure to small molecules allows induction of mouse embryonic fibroblasts into neural crest-like precursors. *FEBS Letters*, 591(4):590-602, 2017. doi: 10.1002/1873-3468.12572. (査読有)
Yutaka Saito, *Toutai Mituyama. Detection of differentially methylated regions from bisulfite-seq data by hidden Markov models incorporating genome-wide methylation level distributions. *BMC Genomics*, 16(Suppl 12):S3, 2015. doi: 10.1186/1471-2164-16-S12-S3. (査読有)

[学会発表](計 7 件)*: 発表者

若尾 宏*, 杉本 智恵, 光山 統泰, 齋藤 裕. iPS 細胞からの自然免疫型 T 細胞分化誘導の分子機序の統合的解析. 第 39 回 日本分子生物学会年会 (MBSJ 2016). パシフィコ横浜 (神奈川県), 11/30 - 12/2, 2016.

若林 玲実*, 高山 祐三, 櫛笥 博子, 渋谷 陽一郎, 齋藤 裕, 赤松 和土, 芝田 晋介, 岡野 栄之, 木田 泰之. 薬剤によるマウス胎児線維芽細胞 (MEFs) からの神経堤細胞の分化誘導法の開発. 第 39 回 日本分子生物学会年会 (MBSJ 2016). パシフィコ横浜 (神奈川県), 11/30 - 12/2, 2016.

齋藤 裕*. ワークフローの紹介 (BS-seq). Galaxy Workshop Tokyo 2016. 東京大学先端科学技術研究センター (東京都), 4/28, 2016.

齋藤 裕*, 光山 統泰. ゲノム 3 次元構造データに基づく共同在遺伝子の網羅的探索. 第 38 回 日本分子生物学会年会, 第 88 回 日本生化学会大会, 合同大会 (BMB 2015). 神戸ポートアイランド (兵庫県), 12/1 - 12/4, 2015.

Yutaka Saito*, Toutai Mituyama. Detection of differentially methylated regions from bisulfite-seq data by hidden Markov models incorporating genome-wide methylation level distributions. The 26th International Conference on Genome Informatics (GIW / InCoB 2015). Tokyo (Japan), Sep 9 - Sep 15, 2015.
齋藤 裕*, 光山 統泰. ゲノム 3 次元構造データに基づく共同在遺伝子の網羅的探索. NGS 現場の会 第四回研究会. つくば国際会議場 (茨城県), 7/1 - 7/3, 2015.

齋藤 裕*, ワークフローの紹介 (BS-seq). Galaxy Workshop Tokyo 2015. 東京大学先端科学技術研究センター (東京都), 4/28, 2015.

〔その他〕

Cosearge ソフトウェア

<https://github.com/yutaka-saito/cosearge>

6. 研究組織

(1) 研究代表者

齋藤 裕 (SAITO, Yutaka)

国立研究開発法人産業技術総合研究所・人

工知能研究センター・研究員

研究者番号：60721496