様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月 １０ 日現在

機関番号：１２６０８
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：２２１２５００８

研究課題名（和文）染色体動態解析およびゲノム比較進化解析のための情報処理技術の確立

研究課題名（英文）Development of bioinformatics platform for analyzing chromosome dynamics and comparative genomics.

研究代表者
伊藤　武彦（ITOH, TAKEHIKO）

東京工業大学・生命理工学研究科・教授

研究者番号：９０５０１１０６

交付決定額（研究期間全体）：（直接経費）　 123,200,000円

研究成果の概要（和文）：主にIlluminaシークエンサのデータを入力としたChIP-seq解析アルゴリズム、ゲノムアセンブル解析アルゴリズム、点/構造変異解析アルゴリズムを新規に開発、改良を実施し、個別のプログラムとしてまとめ、論文などを通じて公開を図った。
またこれらのアルゴリズムを用いて、新学術他班員などとの共同研究により、出芽酵母、分裂酵母、ヒト、マウス、病原性バクテリアなど種々の生物種のデータについて各種タンパク質局在プロファイルの解析、新規ゲノム配列決定、SNP検出、ゲノム構造変異解析を実施した。

研究成果の概要（英文）：We developed novel ChIP-seq analysis, genome assembling and detecting point or structural variation algorithms from NGS (illumina) sequence data. These algorithms were built up as individual programs and opened to the public via paper or our HP.
And we also performed analysis of the distribution of protein localization, denovo assembling of genomes, and SNPs or structural variation detection against S.cerevisiae, S.pombe, Human. Mouse and some bacterias with other many researchers.

研究分野：ゲノム情報

キーワード：染色体動態　バイオインフォマティクス

１．研究開始当初の背景

本研究が開始された当時は、illumina 社製 GAIIx や ABI 社製 SOLiD などのいわゆる第二世代型次世代シークエンサ（以後次世代シークエンサと略す）が普及し始め、ゲノム情報に基づいた様々な研究が広く一般化し始めた頃であった。例えば、新規ゲノム配列の決定においても微生物等では、sanger 法からの転換が図られた頃であり、遺伝子発現解析やエピゲノム解析でも、マイクロアレイの利用から次世代シークエンサの利用への転換が図られていた頃である。

次世代シークエンサの最大の特徴は、従来の sanger 法と比して、塩基単価が極めて安い事である。そのため、個々の read は sanger 法と比べて短い(36-100bp 程度)という欠点はあったものの幅広い分野への適用が試みられており、今まではラボ単位ではできなかったような研究もコストの大幅な削減により可能となり、数多くの研究者が次世代シークエンサを用いた解析を目指すようになっていた。

コストの大幅な削減による影響は大きく、例えば表現型の異なる変異体が得られた場合、その原因遺伝子の特定に従来は順遺伝学的手法が用いられていたが、次世代シークエンサ登場以降は、全ゲノムをシークエンスし変異を同定する手法が普及する等、研究の方法自体へのブレークスルーをもたらすことにも繋がる様な状況であった。

このように次世代シークエンサの普及は、分子生物学分野に大きな変革をもたらしつつあったが、一方同時に問題となっていたのが、得られる膨大なデータの効果的な情報処理であった。Sanger 法と異なり、一度に得られるデータ量は数 Gb から数百 Gb とヒトゲノムの何倍にも相当し、それらのデータから意味ある結果を引き出すためには大規模情報解析が必要不可欠なことは明らかであった。

もちろん次世代シークエンサ登場に伴い、個々の目的に応じた解析ソフトウェアは世界中で開発されていたが、それらの使用においても計算機に関する知識が必要であったり、適切な前処理が必要であったりと使いこなしたり、使い分けるには計算機の専門家の知見が必要であった。また、得られた情報解析結果の評価を適切に行わない限り、何か結果は出るがそれが果たして生物学的に意義のあるものか否かの適切な判断をするのも困難であった。

２．研究の目的

本研究では、次世代シークエンサの普及という状況を踏まえた上で、申請者らが独自に開拓してきた ChIP-on-Chip 解析に基づいた染色体情報研究をさらに発展させ、出芽酵母、分裂酵母の 2 種類の酵母をモデルとし、有糸分裂、および配偶子形成に伴う染色体構造と動態を次世代型シークエンサにより解析し、情報学的に染色体動態を再構築することを目指すとともに、他の班員と共同でハエ、マウス、ヒト、イネも含め染色体の動態および、精度の高いゲノム配列の新規決定、さらには染色体構造のあらゆる種類の変化（点変異、転座、重複、欠失、逆位等）を迅速かつ詳細に明らかにするための情報処理技術の開発を行うことを第一の目的とした。

本技術は新学術領域における他班員に供与するとともに、いくつかの生物種について得られた知見からゲノム比較進化情報解析を行い、染色体構築原理の普遍性と多様性の分子基盤を明らかにする。これら一連の研究により体系的に染色体の構造と機能を理解するための情報プラットフォームを構築し、我が国独自の染色体システム生物学を発展させる契機とすることを目指す。

次世代シークエンサの利用においては扱う情報量が膨大になるため、情報処理にも目的に合致した工夫が必要となるが、今回、本研究領域に参加し、他班員との共同研究を積極的に推進することで、他研究者空からのニーズを積極的に取り入れ、汎用性の高いアプリケーション開発を実施する。タイリングアレイ解析のパイオニアとして培った情報解析技術を存分に活かし、次世代シークエンサを用いた情報解析においても、この分野でのリーダーとなることを目指す。

３．研究の方法

本研究では、研究分担者である白髭や新学術における他の班員(岩崎、篠原、岡田、石井など)が取得したサンプルに対して、シークエンス拠点となっている東大分生研にて SOLiD, Illumina を用いた各種次世代シークエンサによる実験解析をまず行う。そこで得られたデータを、東工大にて情報解析し、その結果を元に各研究者との解釈に関する議論や q-PCR など他実験手法に基づいた検証を通じたフィードバックを行うことでよりよいアルゴリズムの構築を実施し、研究を継続的に行った。解析に用いたデータは、各種タンパク質に関する ChIP-seq データ、酵母からヒトまで幅広く含んだ点変異、構造変異解析用データ、新規ゲノム配列決定用データ等多岐に渡る。

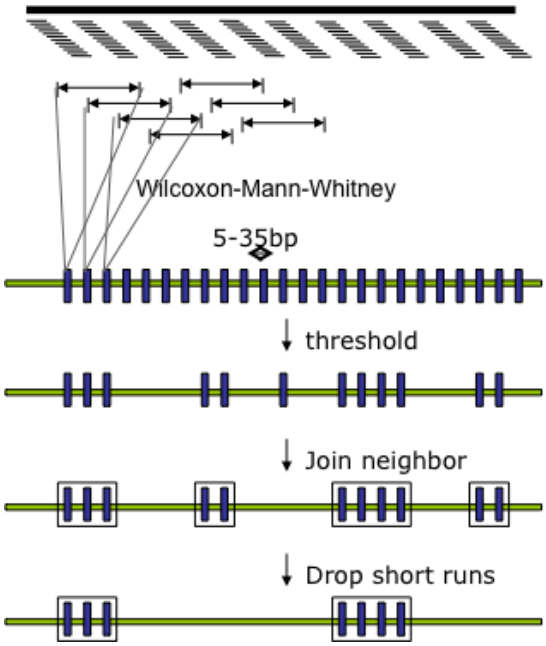
また、各種アルゴリズムの中で新規性および汎用性の高いものに関しては、研究者一般が利用可能な形でのパイプライン化、スタンドアローンで利用可能なプログラム化をはかり、幅広く公開を実施した。

４．研究成果

研究期間を通して、主に Illumina シークエンサのデータを入力とした ChIP-seq 解

析アルゴリズム、ゲノムアセンブル解析アルゴリズム、点変異解析アルゴリズムを開発、改良を実施し、新学術他班員との共同研究により、様々な生物種について各種タンパク質局在プロファイルの解析、SNP 検出、ゲノム構造変異解析を実施した。以下、特筆すべき研究事例をいくつか紹介する。

研究期間の前半を中心に注力した研究分野として、ChIP-seq 解析アルゴリズムの確立が挙げられる。ChIP-on-chip 解析で培ったノウハウやコントロールデータに対しての有意検定アルゴリズムを、ゲノムをあるウィンドウに分割し、各ウィンドウにマップされた補正後の read 数を用いての Wilcoxon 検定により実現した。下図参照。



この解析アルゴリズムは、酵母を中心として種々のクロマチンタンパクの局在解析に用いられ、特筆すべき業績としてコヒーシンローダーである Scc2 の局在がセントロメアおよび転写活性の高い遺伝子に普遍的に見られ、このセントロメアに置ける局在がキネトコア構成タンパクに依存していることを示したこと（Current Biology, 2011）、および、染色体の長さによって本質的な複製のメカニズムが異なることを示したこと（Nature, 2011）などがあげられる。最終的にこのアルゴリズムは改良後 DROMPA（Gene to Cells, 2013）としてまとめられ、世界中の研究者に広く用いられている。

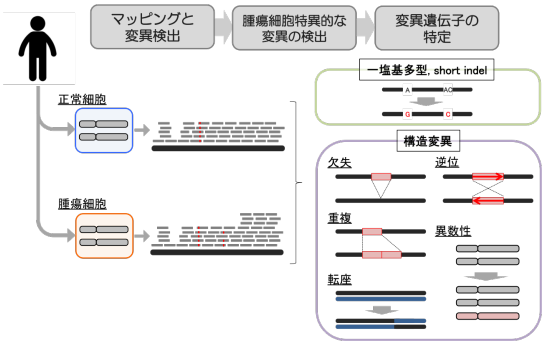
その後も、酵母、マウス、ヒトなど様々な生物種について、Mcm4, BrdU など各種タンパク質局在プロファイルの解析を実施し、これらの結果は既知遺伝子情報との相関、さらには RNA-seq と合わせることで多くの論文にまとめられている。

研究期間の中間から後半ではゲノム点変異、構造変異を精度高く効率的に抽出する手法および新規ゲノム配列決定手法の研究開発に注力した。

ゲノム変異箇所の抽出は、次世代シークエンサ利用における最も基本的な使用法の一つであるが、一般に考えられているよりも、情報解析手法は確立されていない。点変異の検出でさえも、参照配列とターゲット配列とがどの程度近縁かにより、マッピングベースの変異解析が有効な場合もあれば、アセンブル後アライメントにより比較解析する手法が有用な場合も存在する。これら様々なケースに対応できるように、多面的な角度から変異検出アルゴリズムの開発を実施した。

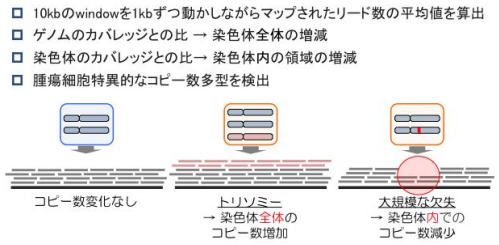
特徴的な例としては、大量の近縁なバクテリアゲノム解析を実施するケースに対応したアルゴリズムの開発が挙げられる。通常、複数株間の比較による系統解析では、系統推定の段階においてマルチプルアライメントが必要になる。しかし、数多くの株をシークエンスし、比較する場合にはこのマルチプルアライメント時の計算コストが極めて大きい。そこで新たに開発したアルゴリズムでは、参照配列にマッピング、変異箇所を検出しマルチプルアライメントの回避に成功した。また、マッピング時の read 数が少ない等配列をどれかの株で確定できないサイトも考慮できるように工夫する等のアルゴリズムも導入し、院内感染等のデータ解析を実施した。

また、がんなどの疾患サンプルからのゲノム構造変異検出を念頭においたアルゴリズムの開発も実施した。正常細胞、腫瘍細胞双方由来のシークエンスデータを元に、CNV，大規模な挿入・欠失、転座を検出する手法の確立に成功した。

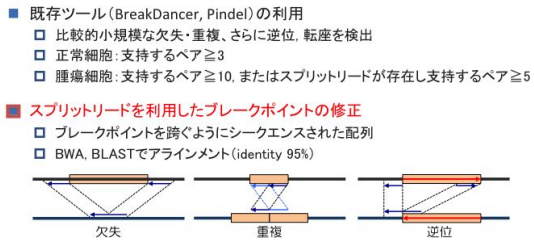


同様の研究は先行事例が多く存在するが、トリソミーやモノソミー、あるいは数 Mb にもわたるような大規模な重複、欠失の検出が不十分であったり、構造変異の検出においてブレークポイントの位置が不正確であったりするなどの問題点が存在した。これらの問題点を解決することで、既存手法よりも高精度な変異検出手法の確立を実施した。CNV の検出では、腫瘍細胞由来および正常細胞由来のシークエンスデータをヒト

ゲノムにマッピングし、染色体ごとに 10kb の window を 1kb ずつ移動させながら window 内のカバレッジの平均値とその染色体のカバレッジとの比をとることでコピー数を算出することで実現するアルゴリズムを構築した。下図参照。



また構造変異の検出においては、既存の BreakDancer, Pindel の利用時における条件検討を実施し、さらにスプリットリードを利用し、それらのリードをゲノムに対して再アライメントした結果を用いてブレークポイントの修正を行うようにアルゴリズム開発を実施した。



有効性検証のため、実際のがん細胞由来サンプルと同一個人正常細胞由来のサンプル比較を実施した所、既存ツールをかけたのみでは、ブレークポイントが一致したケースは見られなかったが、修正後では 1,400 カ所以上の一致を見た。これにより、がん細胞特有の変異同定が容易になり、より高い精度での変異解析が可能となった。

ゲノムアセンブル解析においては、既存の Velvet, SOAPdenovo や ALLPATHS-LG などと同様に次世代シークエンサのデータ解析に適した De bruijn アルゴリズムを用いた新規アセンブラを開発した。最大の特徴は、野生種のゲノムなど高いヘテロ接合性をもつゲノムの再構築時に生じるグラフ構造中のバブル構造を、read の coverage 情報等を活用しつつ、うまく解決する事で対処している事にある。本プログラムの使用により、バクテリアから高等真核生物に至るまで次世代シークエンサのデータから新規ゲノム配列の決定が可能となった。本手法は、新規ゲノム配列決定のみならず、前述の点変異、構造変異検出においてもアセンブルした配列同士の比較による決定を可能としている。特に、比較対象間で変異が大きい場合等に有効である。

以上見て来たように研究期間を通して、ChIP-seq 解析アルゴリズム、ゲノムアセンブル解析アルゴリズム、変異解析アルゴリズムを新規に開発し、改良を重ね、最終的なパイプラインとしてまとめあげた。さらにこれらを利用して分担者、および他班員との共同研究により、出芽酵母、分裂酵母、ヒト、マウス、病原性バクテリアなど種々の生物種のデータについてタンパク質局在プロファイルの解析、SNP 検出、ゲノム構造変異解析を実施した。

5．主な発表論文等

〔雑誌論文〕（計 16 件）

① Kajitani R, Toshimoto K, 他 11 名, Itoh T, Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads., Genome Res. 24:1384-95 (2014) 査読有 doi: 10.1101/gr.170720.113.

② Jeppsson K, Kanno T, Shirahige K, Sjögren C, The maintenance of chromosome structure: positioning and functioning of SMC complexes., Nat Rev Mol Cell Biol. 15:601-14 (2014) 査読有 doi: 10.1038/nrm3857.

③ Suda N, Itoh T, 他 5 名, Shirahige K, Tickle C, Tanaka M, Dimeric combinations of MafB, cFos and cJun control the apoptosis-survival balance in limb morphogenesis., Development.141:2885-94 (2014) 査読有 doi: 10.1242/dev.099150.

④ Hattori Y, Usui T, Satoh D, Moriyama S, Shimono K, Itoh T, Shirahige K, Uemura T, Sensory-neuron subtype-specific transcriptional programs controlling dendrite morphogenesis: genome-wide analysis of Abrupt and Knot/Collier., Dev Cell. 27:530-44 (2013) 査読有 doi: 10.1016/j.devcel.2013.10.024.

⑤ Kon A, 他 37 名, Shirahige K, Miyano S, Ogawa S, Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms., Nat Genet. 45:1232-7 (2013) 査読有 doi: 10.1038/ng.2731.

⑥ Nakato R, Itoh T, Shirahige K, DROMPA: easy-to-handle peak calling and visualization software for the computational analysis and validation of ChIP-seq data., Genes Cells. 18:589-601 (2013) 査読有 doi: 10.1111/gtc.12058.

⑦ Deardorff MA, Bando M, Nakato R, Watrin E, Itoh T, 他 35 名, Shirahige K. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin

acetylation cycle., Nature. 489: 313-7 (2012) 査読有 doi: 10.1038/nature11316.
⑧ De Piccoli G, Katou Y, Itoh T, Nakato R, Shirahige K, Labib K, Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases., Mol Cell. 45:696-704 (2012) 査読有 doi: 10.1016/j.molcel.2012.01.007.
⑨ Takami H, Noguchi H, Takaki Y, Uchiyama I, Toyoda A, Nishi S, Chee GJ, Arai W, Nunoura T, Itoh T, Hattori M, Takai K, A deeply branching thermophilic bacterium with an ancient acetyl-CoA pathway dominates a subsurface ecosystem., PLoS One. 7:e30559 (2012) 査読有 doi: 10.1371/journal.pone.0030559.
⑩ Kuwahara T, 他5名, Itoh T, Nakayama-Imaohji H, Ichimura M, Itoh K, Ishifune C, Maekawa Y, Yasutomo K, Hattori M, Hayashi T, The lifestyle of the segmented filamentous bacterium: a non-culturable gut-associated immunostimulating microbe inferred by whole-genome sequencing., DNA Res. 18:291-303 (2011) 査読有 doi: 10.1093/dnares/dsr022.
⑪ Kegel A, Betts-Lindroos H, Kanno T, Jeppsson K, Ström L, Katou Y, Itoh T, Shirahige K, Sjögren C, Chromosome length influences replication-induced topological stress., Nature. 471:392-6 (2011) 査読有 doi: 10.1038/nature09791.
⑫ Maruyama H, Shin M, Oda T, Matsumi R, Ohniwa RL, Itoh T, Shirahige K, Imanaka T, Atomi H, Yoshimura SH, Takeyasu K, Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus kodakarensis. Mol Biol Cell. 22:386-98 (2011) 査読有 doi: 10.1091/mbc.E10-08-0668.
⑬ Kurze A, Michie KA, Dixon SE, Mishra A, Itoh T, Khalid S, Strmecki L, Shirahige K, Haering CH, Löwe J, Nasmyth K, A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion., EMBO J. 30:364-78 (2011) 査読有 doi: 10.1038/emboj.2010.315.
⑭ Hu B, Itoh T, Mishra A, Katoh Y, Chan KL, Upcher W, Godlee C, Roig MB, Shirahige K, Nasmyth K., ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex., Curr Biol. 21:12-24. (2011) 査読有 doi: 10.1016/j.cub.2010.12.004.
⑮ Pauli A, van Bemmel JG, Oliveira RA, Itoh T, Shirahige K, van Steensel B, Nasmyth K., A direct role for cohesin in gene regulation and ecdysone response in Drosophila salivary glands., Curr Biol. 20:1787-98 (2010) 査読有 doi: 10.1016/j.cub.2010.09.006.
⑯ Liu J, Zhang Z, Bando M, Itoh T,他10名, Shirahige K, Krantz ID, Genome-wide DNA methylation analysis in cohesin mutant human cell lines., Nucleic Acids Res. 38:5657-71 (2010) 査読有 doi: 10.1093/nar/gkq346.

〔学会発表〕(計2件)

① 吉村大, 後藤恭宏, 小椋義俊, 林哲也, 伊藤武彦, Whole Genome Shotgun を用いた病原菌に関する疫学研究のための解析手法の開発, ゲノム微生物学会, 2014/3/7, 東京農業大学, 東京都・世田谷区
② Miki Okuno, Yukiko Kodama, Takehiko Itoh, A whole genome comparison between lager brewing yeast Weihenstephan 34/70 and its ancestral strains., Yeast 2013, 2013/8/29, Westend of Goethe University, フランクフルト(ドイツ)

6. 研究組織

(1)研究代表者
伊藤　武彦 (ITOH TAKEHIKO)
東京工業大学・大学院生命理工学研究科・教授
研究者番号：90501106

(2)研究分担者
白髭　克彦 (SHIRAHIGE KATSUHIKO)
東京大学・分子細胞学研究所・教授
研究者番号：90273854