



バイオメディカル情報解析分野

Division of Biomedical Information Analysis

この分野に向いていそうな方。

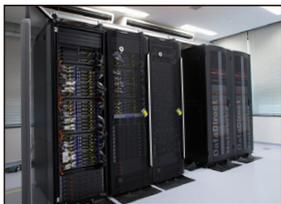
- 情報解析に興味のある方
- プログラミングに興味のある方
- 大規模データを俯瞰的にみて規則を発見したい方
- 次世代シーケンサによる全ゲノムデータ解析に興味のある方
- 大規模ゲノムコホートプロジェクトのゲノム解析に関わりたい方
- 実験系の先生と連携して実験系サイドと情報系サイド双方を学びたい方

バイオメディカル情報解析分野
Division of Biomedical Information Analysis

生命医療に関係するビッグデータを
自在に解析できる人材を育てる。



本研究室は、ゲノム情報などのライフサイエンス分野における超大規模ヘテロデータに対し、情報科学・数理統計学の知識を基にスーパーコンピュータを活用して解析を行い、新たな知識を発見するための研究を行っています。10年前の国際ヒトゲノムプロジェクトでは、各国の多数の研究者が数千億円以上かけ、ヒト一個体分に相当するリファレンスゲノムを解読しました。現在、個人をターゲットとし、少量の血液から簡単に短時間でパーソナルゲノムを決定できる時代を迎えつつあります。それに伴い、ゲノムの他、オミクス情報やそれに付随するバイオメディカル情報（医療情報、アンケート情報、生理学検査情報）、すなわち超ヘテロビックデータを解析するための技術革新が求められています。



今日、医学生物学の様々な分野においてはこれらの大量情報解析の需要が高まっており、データ駆動型サイエンスへのパラダイムシフトを迎えています。このデータに切り込むにあたり、本研究室は [情報科学] [バイオインフォマティクス] [数理統計学] [集団遺伝学] [分子進化学]などを専門とするスタッフ陣が連携し研究を推進しています。さらに研究活動として、ヒトゲノム解析、エビゲノム解析、メタゲノム解析、転写・翻訳制御解析などを駆使し、大規模データからの仮説を提示、実験的検証も視野に入れた活動を行っています。大学院教育としては各学生の興味に応じた [大規模計算技術] [データ解析] [統計モデリング] [効率的なアルゴリズム開発] [ソフトウェア開発] [分子生物学実験]などを含めた研究指導を行い、情報科学の立場からライフサイエンスへの貢献を目的とした、ビッグデータを自在に解析できるデータサイエンティストを育成します。



長崎 正朗 | Masao Nagasaki
教授 (兼 東北メディカル・メガバンク機構 インシリコ解析室室長)



Riu Yamashita | 山下 理宇
准教授

Faculty Members

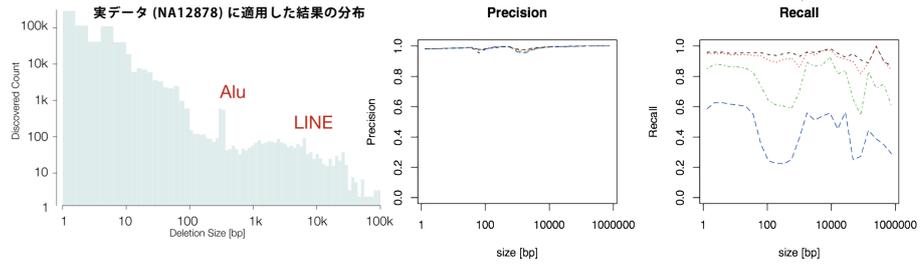


助教 成相直樹 小島要 山口由美 山岸潤也 佐藤行人 河合洋介

iSVP: 複数のSV検出アルゴリズムを組み合わせたdeletionコールパイプラインの構築



ヒトゲノムに多様性を与えている変異は、SNP以外にも deletion, insertion, duplication, inversionなどの様々な構造変異(SV)が存在し、それらのジェノタイプを決定することは重要な課題である。本研究では、NGSデータから deletionを予測する異なるアルゴリズム(Read Pair, Split Read, Assemblyタイプ)の適合率と再現率を様々なサイズのdeletionに対してシミュレーションデータをベースに評価した。さらに、結果を適切にマージすることで、約30×のシーケンズ深度があれば1bp-1Mbp程度までのdeletionを精度よくかつ包括的に検出できることを明らかにした。



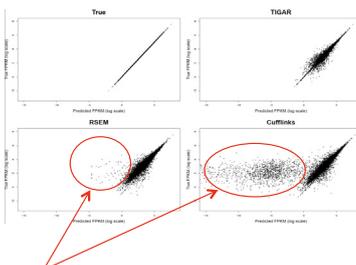
Mimori et al., BMC Systems Biology, 2013.

TIGAR: 網羅的かつ高精度なRNA-Seqデータからの転写産物定量アルゴリズムの開発



近年、次世代シーケンズによるRNA-Seq技術により、遺伝子発現を網羅的かつ塩基レベルの高解像度で解析することが可能となった。我々は既存手法よりも高精度な遺伝子発現レベル推定を実現するため、モデルの複雑さ(転写産物の候補の数)、隠れ変数(マッピングの曖昧さ)、パラメータ(転写量)を、変分ベイズ推定により同時に最適化する統計的手法TIGARを開発した。

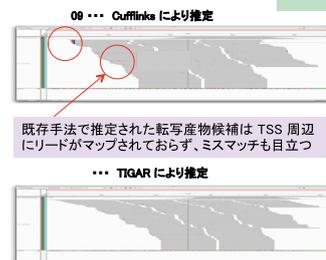
• 実際の転写量と推定量の相関



RefSeqから10,000 transcript isoformをランダムに選んで、そこからRNA-Seqデータを生成した。(シミュレーション実験による評価)

既存手法では発現量が小さい転写物の遺伝子発現レベル推定に失敗するが、TIGARでは既存手法に比べて高精度に推定できる

マウスのES細胞から取得したRNA-SeqデータからTIGARと既存手法(Cufflinks)で転写量を推定した。



RefSeq ID	gene	locus	length	cuff_FPKM	TIGAR_FPKM
NM_001110009	Apoc1	chr7:2027482-9-20278007	430	6267.79	4.77
NM_007469	Apoc1	chr7:2027482-9-20277945	490	197.331	4147.92

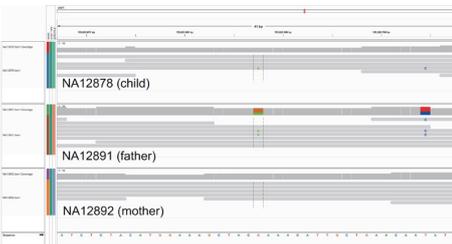
Nariai et al., Bioinformatics, 2013.

PedigreeCaller: 家系情報を用いた高精度な変異検出手法

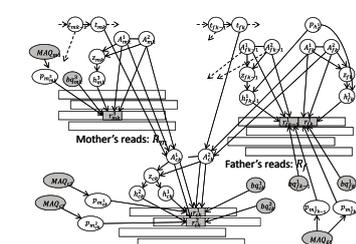


個人におけるゲノムにの単塩基置換などの変異の検出では、次世代シーケンズからのデータをヒトファレンスゲノムにマッピングを行い、シーケンシングにおける誤りやマッピングにおける誤りを考慮した統計手法が用いられる。PedigreeCallerでは、各個人間の家系情報が得られる場合にグラフィカルモデルと呼ばれる統計手法により、各個人のシーケンズからの情報を統合し、最も確率の高い遺伝子型を推定することで、変異の検出が行われる。シミュレーションデータや実際ヒトの家系データより、各サンプルを独立に解析する場合に比べ、より高精度な変異コールが得られることが検証されている。

IGVによる、トリオ家系サンプルにおけるシーケンズデータ



グラフィカルモデルによる家系データからの各個人の情報の統合



Kojima et al., Bioinformatics, 2013.

Latest Publications

最近の出版論文

2014

- Kaname Kojima, Naoki Nariai, Takahiro Mimori, Yumi Yamaguchi-Kabata, Yukuto Sato, Yosuke Kawai and Masao Nagasaki, HapMonster: a Statistically Unified Approach for Variant Calling and Haplotyping Based on Phase-informative Reads, Lecture Notes in Bioinformatics, accepted.
- Tomohiko Ohtsuki, Naoki Nariai, Kaname Kojima, Takahiro Mimori, Yukuto Sato, Yosuke Kawai, Yumi Yamaguchi-Kabata, Tetsuo Shibuya and Masao Nagasaki, SVEM: a Structural Variant Estimation Method using Multi-Mapped Reads on Breakpoints, Lecture Notes in Bioinformatics, accepted.

2013

- Takahiro Mimori, Naoki Nariai, Kaname Kojima, Mamoru Takahashi, Akira Ono, Yukuto Sato, Yumi Yamaguchi-Kabata and Masao Nagasaki, iSVP: an integrated structural variant calling pipeline from high-throughput sequencing data, BMC Systems Biology, 7(Suppl 6):S8, 2013.
- Kaname Kojima, Naoki Nariai, Takahiro Mimori, Mamoru Takahashi, Yumi Yamaguchi-Kabata, Yukuto Sato, and Masao Nagasaki, A statistical variant calling approach from pedigree information and local haplotyping with phase informative reads, Bioinformatics, 29(22):2835-2843, 2013.

- Naoki Nariai, Osamu Hirose, Kaname Kojima and Masao Nagasaki, TIGAR: transcript isoform abundance estimation method with gapped alignment of RNA-Seq data by variational Bayesian inference, Bioinformatics, 29(18):2292-2299, 2013.

Photo Gallery [フォトギャラリー]

