

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年3月31日現在

機関番号：32612

研究種目：若手研究（B）

研究期間：2010～2011

課題番号：22700305

研究課題名（和文） RNA二次構造を考慮したSNP解析

研究課題名（英文） Analysis of SNPs that induce structural change of RNA secondary structures

研究代表者

佐藤 健吾（SATO KENGO）

慶應義塾大学・理工学部・講師

研究者番号：20365472

研究成果の概要（和文）：RNA二次構造の変異を検出するために、各アレル間の距離をそれらが形成する二次構造分布の間の距離と定義し、これを効率よく計算するアルゴリズムを開発した。本手法をSNPデータベースに適用した解析結果から、同義置換SNPやタンパク質非コード領域内のSNPによる二次構造の変異が疾患関連遺伝子において統計的に有意に現れることが示唆された。

研究成果の概要（英文）： We developed an efficient algorithm for detecting structural changes of RNA secondary structures induced by SNPs on RNA sequences. Comprehensive analysis on the SNP database using our method revealed statistical significance of structural changes induced by synonymous SNPs on coding regions and SNPs on non-coding regions of mRNAs from clinically associated genes.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,500,000	450,000	1,950,000
2011年度	600,000	180,000	780,000
年度			
年度			
年度			
総計	2,100,000	630,000	2,730,000

研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：RNA二次構造、一塩基多型、バイオインフォマティクス

1. 研究開始当初の背景

これまでのSNP解析は、アミノ酸の変異が起こる非同義置換SNPに関するものが中心であった。アミノ酸配列の変異によってタンパク質が変化し、最終的にその機能や表現型に変化が引き起こされるというストーリーは直感的に理解しやすく、実際そのようなSNPの例は非常に多く見つかっている。しかし一方で、アミノ酸の変異を起こさない同義置換SNPの中には、mRNAの二次構造の変化による翻訳効率の著しい低下を引き起こ

し、その結果、表現型に大きな変化をもたらす事例が報告されている。

2. 研究の目的

本研究の目的は、アミノ酸の変異を起こさない同義置換一塩基多型(SNP)やイントロン、UTRなどのタンパク質非コード領域におけるSNPがmRNAの二次構造に与える影響を計算する手法を開発することである。さらに、本手法を大規模SNPデータベースに適用し、疾患と二次構造の変異をもたらすSNP

の関連を網羅的に調査することによって、mRNA の二次構造に起因する様々な生物学的な知見を得ることを目指す。

3. 研究の方法

RNA 二次構造の変異を検出するために、与えられた RNA 配列が取りうるすべての二次構造の確率分布を考慮し、各アレルの距離をそれらが形成しうる二次構造分布の間の距離と定義し、これを効率よく計算するアルゴリズムを開発する。具体的には、配列長が同じ二つの二次構造の間の距離を異なる塩基対の個数（二次構造のハミング距離）とし、形成しうる二次構造に関する期待値を計算する。

4. 研究成果

大規模データベースに登録されている SNP データで解析を行った。代表的な SNP データベースである NCBI dbSNP には 1800 万弱のヒト SNP が登録されており、疾患に関連することがわかっている SNP にはアノテーションが付けられている。その結果、スコア上位 SNP には難治性の遺伝病が多く含まれていることがわかった。これまでの非同義置換 SNP のみ注目した手法ではこのような SNP を発見することは難しい。また、大規模な解析結果から、同義置換 SNP やタンパク質非コード領域内の SNP による二次構造の変異が疾患関連遺伝子において統計的に有意に現れることが示唆された。以上により、本手法の有効性を示すことに成功したと言える。

	All			Clinical associated		
	N	MFE	BP	N	MFE	BP
coding synonymous	29265	1.25	8.69e-3	205	1.27	9.15e-3
coding non-synonymous	41970	1.30	9.56e-3	2676	1.36	9.59e-3
intron	4036291	1.19	6.19e-3	1493	1.27	8.08e-3
near-gene-5	87625	1.27	6.67e-3	74	1.30	8.39e-3
near-gene-3	86873	1.28	6.70e-3	90	1.38	1.01e-2
5'-UTR	16482	1.43	8.99e-3	49	1.43	1.22e-2
3'-UTR	97716	1.19	8.15e-3	62	1.40	7.75e-3

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 13 件)

1. Kato, Y., Sato, K., Asai, K., Akutsu, T.: Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming, *Nucleic Acids Res.*, in press (2012) (査読有)
2. Sakakibara, Y., Hachiya, T., Uchida, M., Nagamine, N., Sugawara, Y., Yokota, M., Nakamura, M., Poperdorf, K., Komori, T., Sato, K.: COPICAT: A software system for predicting interactions between proteins and chemical compounds, *Bioinformatics*,

vol. 28, no. 5, pp. 745-756 (2012) (査読有)

3. Okada, Y., Saito, Y., Sato, K., Sakakibara, Y.: Improved measurements of RNA structure conservation with generalized centroid estimators, *Front. Gene.*, 2:54 (Aug. 2011) (査読有)
4. Poolsap, U., Kato, Y., Sato, K., Akutsu, T.: Using binding profiles to predict binding sites of target RNAs, *J. Bioinform. Comput. Biol.*, vol. 9, no. 6, pp. 697-713 (Dec. 2011) (査読有)
5. Hamada, M., Yamada, K., Sato, K., Frith, MC., Asai, K.: CentroidHomfold-LAST: Accurate prediction of RNA secondary structure using automatically collected homologous sequences, *Nucleic Acids Res.*, vol. 39, pp. W100-106 (Jun. 2011) (査読有)
6. Sato, K.*, Kato, Y.*, Hamada, M., Akutsu, T., Asai, K.: IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming, *Bioinformatics*, vol. 27, no. 13, pp. i85-i93 (Jul. 2011) (*Joint First Authors) (査読有)
7. Kielbasa, SM., Wan, R., Sato, K., Horton, P., Frith, MC.: Adaptive seeds tame genomic sequence comparison, *Genome Res.*, vol. 21, no. 3., pp. 487-493 (Mar. 2011) (査読有)
8. Saito, Y., Sato, K., Sakakibara, Y.: Fast and accurate clustering of noncoding RNAs using ensembles of sequence alignments and secondary structures, *BMC Bioinformatics*, 12(Suppl 1):S48 (Feb. 2011) (査読有)
9. Hamada, M., Sato, K., Asai, K.: Improving the accuracy of predicting secondary structure for aligned RNA sequences, *Nucleic Acids Res.*, vol. 39, no. 2, pp. 393-402 (Jan. 2011) (査読有)
10. Hamada, M., Sato, K., Asai, K.: Prediction of RNA secondary structure by maximizing pseudo-expected accuracy, *BMC Bioinformatics*, 11:586 (Nov. 2010) (査読有)
11. Saito, Y., Sato, K., Sakakibara, Y.: Robust and accurate prediction of noncoding RNAs from aligned sequences, *BMC Bioinformatics*, 11(Suppl 7):S3 (Oct. 2010) (査読有)

12. Kato, Y.*, Sato, K.*, Hamada, M., Watanabe, Y., Asai, K., Akutsu, T.: RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming, *Bioinformatics*, vol. 26, no. 18, pp. i460-i466 (Sep. 2010) (*Joint First Authors) (査読有)
13. Sato, K., Hamada, M., Mituyama, T., Asai, K., Sakakibara, Y.: A non-parametric Bayesian approach for predicting RNA secondary structures, *J. Bioinform. Comput. Biol.*, vol. 8, no. 4, pp. 727-742 (Aug. 2010) (査読有)

[学会発表] (計 11 件)

1. 佐藤健吾, 加藤有己, 浅井潔, 阿久津達也, 榎原康文: 双対分解による RNA 構造アラインメント, 第 28 回情報処理学会バイオ情報学研究会 (Mar. 28, 2012, 仙台) (口頭発表)
2. 佐藤健吾, 加藤有己, 浅井潔, 阿久津達也: Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming, 第 34 回分子生物学学会年会 (Dec. 14, 2011, 横浜) (口頭発表、ポスター発表)
3. Sato, K., Kato, Y., Akutsu, T., Asai, K., Sakakibara, Y.: Simultaneous aligning and folding RNA sequences by dual decomposition, CBI/JSBi 2011 合同大会 (Nov. 10, 2011, 神戸) (口頭発表、ポスター発表)
4. Sato, K.*, Kato, Y.*, Hamada, M., Akutsu, T., Asai, K.: IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming, ISMB/ECCB 2011 (Jul. 17, 2011, ウィーン) (*Joint First Authors) (口頭発表)
5. Abe, M., Hase, S., Ogawa, M., Okada, Y., Sato, K., Saito, Y., Sakakibara, Y.: Comprehensive analysis of small non-coding RNAs in medaka transcriptome by deep RNA-seq approach, the 16th Annual Meeting of the RNA society (Jun. 15, 2011, 京都) (口頭発表)
6. Saito, Y., Sato, K., Sakakibara, Y.: Fast and accurate clustering of noncoding RNAs using ensembles of sequence alignments and secondary structures, the 9th Asia Pacific Bioinformatics Conference (APBC 2011) (Jan. 14, 2011, インチョン) (口頭発表)
7. Saito, Y., Sato, K., Sakakibara, Y.:

Robust and accurate prediction of noncoding RNAs from aligned sequences, the International Conference on Bioinformatics (InCoB 2010) (Sep. 26, 2010, 東京) (口頭発表)

8. Kato, Y.*, Sato, K.*, Hamada, M., Watanabe, Y., Asai, K., Akutsu, T.: RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming, the 9th European Conference on Computational Biology (ECCB 2010) (Sep. 27, 2010, ゲント) (*Joint First Authors) (口頭発表)
9. 佐藤健吾, 加藤有己, 阿久津達也, 浅井潔: 期待精度最大化に基づく RNA シュードノット予測, 第 22 回情報処理学会バイオ情報学研究会 (Jul. 29, 2010, 東京) (口頭発表)
10. 加藤有己, 佐藤健吾, 浜田道昭, 浅井潔, 阿久津達也: RNA-RNA interaction prediction using integer programming with threshold cut, 第 21 回情報処理学会バイオ情報学研究会 (Jun. 19, 2010, 那覇) (口頭発表)
11. Sato, K., Whittington, T., Bailey, T., Horton, P.: Improved prediction of transcription binding sites from chromatin modification data, the 2010 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2010) (May 5, 2010, モントリオール) (口頭発表)

[その他]

ホームページ等

Rtips:

<http://rna.naist.jp/>

DAFS:

<http://www.ncrna.org/software/dafs/>

RactIP:

<http://www.ncrna.org/software/ractip/>

IPknot:

<http://www.ncrna.org/software/ipknot/>

CentroidFold:

<http://www.ncrna.org/software/centroidfold/>

BPLA kernels:

<http://bpla-kernel.dna.bio.keio.ac.jp/>

6. 研究組織

(1) 研究代表者

佐藤 健吾 (SATO KENGO)

慶應義塾大学・理工学部・講師

研究者番号: 20365472

(2) 研究分担者

該当なし

(3) 連携研究者
該当なし