

平成 26 年度 研究開発実施報告書

ライフサイエンスデータベース統合推進事業「統合化推進プログラム」研究代表者

[菅野 純夫]

[東京大学 新領域創成科学研究科・教授]

[疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オミクスデータの
統合]

§1. 研究実施体制

(1)「東大」グループ(研究機関別)

① 研究代表者:菅野 純夫 (東京大学新領域創成科学研究科、教授)

② 研究項目

- ・ヒトオミクスデータの統合
- ・培養細胞系データの統合
- ・フォーマット、オントロジーの標準化

(2)「がんセンター」グループ(研究機関別)

① 主たる共同研究者:土原 一哉 (国立がん研究センタートランスレーショナルリサーチ分野、分野長)

② 研究項目

- ・ヒトオミクスデータの統合
- ・培養細胞系データの統合

(1)「DBCLS」グループ(研究機関別)

① 主たる共同研究者:河野 信 (大学共同利用機関法人 情報・システム研究機構ライフサイエンス統合データベースセンター、特任助教)

② 研究項目

- ・ブラウザの構築
- ・フォーマット、オントロジーの標準化

§ 2. 研究実施内容

本研究開発課題では、DataBase of Transcriptional Start Sites (DBTSS) を拡張し、ヒト疾患ゲノム統合データベース KERO (Kashiwa Encyclopedia of Regulatory Omics) を構築・公開する。KERO では、ヒトゲノム多型・変異に生物学的機能注釈を与えるべく、ゲノム変異位置、近傍のエピゲノム (ヒストン修飾、DNA メチル化パターン)、トランスクリプトーム情報 (発現量、スプライスパターン) をヒトゲノム情報に統合する。データの統合は当初ががんゲノム解析を志向したものを中心に行うが、最終的には疾患の別を超えたデータの統合を目指す。特に本研究では、培養細胞系あるいは生物種を超えてマウスをはじめとする動物モデル系から得られたオミクスデータに焦点を当てる。これにより臨床検体で集積が乏しいエピゲノム情報を充実させ、同時に生物学的機能解析の実践の場としてのモデル系におけるオミクス情報を整備する。

平成 26 年度において、研究代表者らは、がん細胞培養細胞株をモデル系にしたデータベース DBTSS (Release 9.0) (<http://dbtss.hgc.jp/>) の構築と公開を行った。26 種類の細胞株について、それぞれ全ゲノムシーケンス、遺伝子発現情報、エピゲノム情報 (8 種類のヒストン修飾と DNA メチル化情報) を統合したものである。DBTSS では外部参照データとして、東北メディカルメガバンク、長浜コホート、日本ファーマコゲノミクスコンソーシアムにより産出された日本人多型データ約 5000 人分が無償、自由に閲覧可能とした。それぞれのデータベースのダウンロードサイトより関連ファイルを取得、それぞれのサイト担当者との協議の上、ファイル加工を行い DBTSS に格納した。また内部的に 500 症例の臨床がん体細胞突然変異データを格納し、制限アクセスを可能としている。また公開可能な部分についての情報を記載した論文を NAR DB issue 2015 に発表した。

また一方で、研究代表者らはこれまでに、マウス免疫担当細胞を種々の刺激源で刺激した詳細な経時計測系において、多階層オミクスデータを収集、データベースとして収載している (転写開始点データ、ヒストン修飾等について計 400 計測点)。これらをモデルデータに、ヒト細胞系へのデータの写像を行ったデータベースの構築を行うべくデータの成形を進めており、現在までに 2 つの実験系をモデルケースに、その基礎的なデータ加工を完了している。これらは文部科学省新学術領域「ゲノム支援」の一環として産出されたデータであるが、同研究課題は今年度に最終年度を迎え、データ公開が可能となる。これらについてもヒトの場合と同様の枠組みで、可及的速やかに公開の準備を整えている。

また、上記のモデルデータについての RDF 化によるデータ統合プラットフォームの整備を開始した。現在までに格納されたデータについて、統合データベースの観点からの検索可能性を考慮した RDF 化を行っている。特に転写開始点データを始めとする遺伝子発現データから開始し、エピゲノムデータ、ゲノム多型データへと進めるものとする。エピゲノムデータについては、IHEC (国際エピゲノムコンソーシアム)、ゲノム多型データについては東大コホート、東北メディカルメガバンクとの連携を密にし、フォーマットの統一を図っている。

§3. 成果発表等

(3-1) 原著論文発表

① 発行済論文数 (国内 (和文) 0 件、国際 (欧文) 3 件):

② 未発行論文数(“accepted”、“in press”等)(国内(和文) 0件、国際(欧文)0件)

③ 論文詳細情報

- * 1. Ayako Suzuki, Hiroyuki Wakaguri, Riu Yamashita, Shin Kawano, Katsuya Tsuchihara, Sumio Sugano, Yutaka Suzuki, Kenta Nakai, “DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data” *Nucl. Acids Res.* 43 (Database issue):D87-91, 2015 (doi: 10.1093/nar/gku1080)
2. Ayako Suzuki, Hideki Makinoshima, Hiroyuki Wakaguri, Hiroyasu Esumi, Sumio Sugano, Takashi Kohno, Katsuya Tsuchihara, Yutaka Suzuki. Aberrant transcriptional regulations in cancers: genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines. *Nucleic Acids Research* 42 (22): 13557-13572 2014 (doi: 10.1093/nar/gku885).
3. Kyoko Matsumoto, Ayako Suzuki, Hiroyuki Wakaguri, Sumio Sugano, Yutaka Suzuki. Construction of mate pair full-length cDNAs libraries and characterization of transcriptional start sites and termination sites. *Nucleic Acids Research* 42(16): e125. 2014 (doi: 10.1093/nar/gku600)

(3-2) データベースおよびウェブツール等の構築と公開

- ① 公開中のデータベース・ウェブツール等
別紙1参照。

(3-3) その他の著作物(総説、書籍など)

1. 関真秀, 鈴木絢子, 鈴木穰「1細胞ゲノム・トランスクリプトーム解析」次世代生物学の扉を開く1細胞解析法, 秀潤社, pp. 244-250, Vol. 34, No.3, 2015. (2015年2月)
2. 鈴木絢子, 鈴木穰, 菅野純夫「急速に普及するRNA-Seqで遺伝子発現をみる」次世代シーケンズ解析スタンダード NGSのポテンシャルを活かしきる WET & DRY, 羊土社, pp.204-215, 2014. (2014年9月)
3. 鈴木絢子, 鈴木穰「がん細胞の多様性の解明に向けたシングルセル解析」個別化医療を拓くがんゲノム研究, 羊土社, pp.110-116, 2014. (2014年8月)
4. 鈴木絢子, 鈴木穰「次世代シーケンサー」医学のあゆみ 医学・医療のいまがわかるキーワード 2014, 医歯薬出版株式会社, Vol. 249, No.5, pp.380-381, 2014. (2014年5月)

(3-4) 国際学会および国内学会発表 (公開)

- ① 招待講演 (国内 3件、国際 0件)
1. 鈴木穰、遺伝子発現制御に影響を及ぼすヒトゲノム多型/変異の同定と解析に向けて、アジレントゲノミクスフォーラム、東京、2014年6月5日
2. 鈴木穰、進化するRNA-Seq:臨床検体からシングルセル解析まで — ウェット・ドライ解析の実験ノート、東京、2014年6月24日
3. 鈴木穰、がん細胞のマルチオミクス解析、164委員会、第46回研究会 地方分科会、仙台、2014年10月24日

② 口頭講演 (国内 2 件、国際 0 件)

〈国内〉

1. 鈴木 穰、Integration of Transcriptome Database, DBTSS, with Multi-Omics Data and Disease-associated Human Variations, 日本分子生物学会、横浜、2014 年 11 月 25 日

2. 鈴木 穰、統合データベース解析、第 3 回生命医薬情報学連合大会、仙台、2014 年 10 月 3 日
〈国際〉

1. 発表者、タイトル、学会名、場所、月日
なし。

③ ポスター発表 (国内 2 件、国際 3 件)

〈国内〉

1. Ayako Suzuki, Sumio Sugano, Takashi Kohno, Katsuya Tsuchihara, Yutaka Suzuki. “Multi-layered analysis of transcriptional aberrations in lung adenocarcinoma cell lines”. 日本分子生物学会. Pacifico Yokohama, 横浜、2014 年 11 月 25 日.

2. Ayako Suzuki, Takashi Kohno, Katsuya Tsuchihara, Yutaka Suzuki. “Integrative analysis of genome, transcriptome and epigenome in 26 lung adenocarcinoma cell lines”. 日本癌学会, Pacifico Yokohama、横浜、2014 年 9 月 26 日.

〈国際〉

*1. Ayako Suzuki, Takashi Kohno, Katsuya Tsuchihara, Yutaka Suzuki. “Integrative multi-omics analysis in lung adenocarcinoma cells”. The 16th annual Advances in Genome Biology and Technology (AGBT) meeting. Marco Island, Florida, USA, 25-28 Feb. 2015.

2. Ayako Suzuki, Takashi Kohno, Katsuya Tsuchihara, Yutaka Suzuki. “Integrative analysis of regulatory aberrations in lung adenocarcinoma cell lines”. The 64th Annual Meeting of the American Society of Human Genetics (ASHG). San Diego Convention Center, California, USA, 18-22 Oct. 2014.

3. Yutaka Suzuki “Aberration of Transcriptional Regulations in Cancer Cells”, Cold Spring Harbor Meeting, NY, USA, 6-10 May 2014.

(3-5) 知的財産権の出願

なし

(3-6) 受賞・報道等

なし

§4. 研究開発期間中に主催した活動(主催したワークショップ等)

年月日	名称	場所	参加人数	概要
2015 年 3 月 23 日	チーム内ミーティング	東京大学	10 人	研究進捗報告のための

	グ(非公開)			ミーティング
2014年11月22日	分子生物学会展示	パシフィコ 横浜	100人	データベースの紹介

以上

別紙1 既公開のデータベース・ウェブツール等

No.	研究開発課題名	データベース、ウェブツール等の名称	概要	URL	公開開始日	関連論文 (論文リストに記載があれば、その番号でも可)	備考
01	疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オミクスデータの統合	KERO/DBTSS	本データベースではヒトゲノム多型・変異に生物学的機能注釈を与えるべく、ゲノム変異位置、近傍のエピゲノム(ヒストン修飾、DNAメチル化パターン)、トランスクリプトーム情報(発現量、スプライスパターン)をヒトゲノム情報に統合する。	http://dbtss.hgc.jp	2014/9/15	1	