

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)

「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」
研究開発期間：平成26年4月～平成29年3月

研究開発終了報告書

研究代表者：菅野純夫
(東京大学 大学院新領域創成科学研究科 教授)

§1. 研究開発実施の概要

本研究開発では、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ベースの構築を試みた。データの統合は本研究開発機関においては、がんゲノム解析を志向したものを中心に行い、最終的には疾患の別を超えたデータの統合を目指すべく、その予備的検討を並行させた。

本研究では、特に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ーミクスデータに焦点を当てた。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ーミクス情報の整備を進めた。個々の臨床ゲノム解析からは頻度の低いゲノム多型・変異情報について相当な生物学的検証を行う機会を拡大することを期待したものであり、また逆に多くのモデル系研究に対しては、積極的にヒト疾患の診断あるいは治療を志向した応用展開を促進することを企図したものである。統合に用いるデータは、代表者らがそれまでにデータ産生に参画し1次解析を行ったもの、将来的に同様の作業を行うことが予定されたものを核として行った。

研究開発初年度、当初の計画書に記載した通り、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 **KERO**(Kashiwa Encyclopedia of Regulatory Omics; <http://kero.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クス、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。**KERO** 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人分を閲覧可能とした。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**KERO** に格納し、2次利用用途で提供したものである。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。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。

初年度以降、我々は上記の関連データセットの拡充と情報の更新を継続する一方で、マウスモデル実験系のデータの収載を行った。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、多階層オーミクスデータを収載、本データベースから公開している(転写開始点データ、ヒストン修飾等について計400計測点)。マウスデータの収載を可能にするために関連するブラウザ、データ加工/格納の枠組みの整備を行った。上記のマウスデータをモデルケースに、ゲノム-ゲノムアラインメントを通じて、マウス多層オーミクスデータをヒトゲノム情報へと写像した。さらに最終年度には、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種々のマウスデータを開発された枠組みで実装、公開した。これらは平成27年度に最終年度を迎えており、本データベースの構築なしには散逸の危機があったものと考えている。

また、上記のモデルデータについて、**RDF**化によるデータ統合プラットフォームの整備を行った。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**RDF**化を進めている。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た。

§2. 研究開発実施体制

1. 研究グループ

(1)「東大」グループ(研究代表者グループ)

人員構成

氏名	所属機関	役職	研究開発項目	参加時期
菅野 純夫	東京大学 新領域創成科学研究科	教授	統括	H26.4～ H29.3
鈴木 穰	同上	教授	統括、データ収集及び一次加工、データベース運用	H26.4～ H29.3
入江 拓磨	同上	研究員	データ解析	H26.4～ H29.3
横田 真澄	同上	学術支援職員	データ入力補助	H26.4～ H29.3
横山 裕子	同上	学術支援職員	データ入力補助	H26.4～ H29.3

担当項目

研究の統括
データの収集と1次加工
データベースへの投入と公開用データベースの運用

(2)「がんセンター」グループ(主たる共同研究者グループ(1))

人員構成

氏名	所属機関	役職	研究開発項目	参加時期
土原 一哉	国立がん研究センター トラंसレーショナルリサーチ分野	分野長	統括、データ収集、オントロジーに留意したメタデータの整備	H26.4～ H29.3

担当項目

データ収集とヒト疾患応用研究を志向した知識システムの構築

(3)「DBCLS」グループ(主たる共同研究者グループ(2))

人員構成

氏名	所属機関	役職	研究開発項目	参加時期
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	統括、データベース構築	H26.4～ H29.3

担当項目

データベース構築

2. 有識者会議等

外部有識者会議は実施していない。ただし、IHEC との連携を実現するため、平成 27 年 6 月

DBCLS が国際版バイオハッカソンを開催した際に、ヨーロッパ、アメリカの IHEC のデータエコシステム実務者を招聘し、緊密な協議を行った。その情報を下記に示す。

(1) 会議概要

名称	国際版バイオハッカソン
目的	エピゲノム、トランスクリプトームについて RDF 化スキーマについてのコンセンサスを形成する。
委員数	2人

(2) 委員一覧

氏名	所属機関	役職	備考
Lillian Rosa Ashmore	Baylor College of Medicine, USA	博士研究員	
Alessandro Vullo	The European Bioinformatics Institute	博士研究員	

(3) 開催歴

年月日	場所	参加人数	主な議題
平成 27 年 6 月 12 日	慶應義塾大学 先端生命科学研究所 鶴岡メタボロームキャンパス	5 人	IHEC との連携を実現するためヨーロッパ、アメリカの IHEC のデータエコシステム実務者を招聘し、緊密な協議を行った。

§3. 研究開発の目的、実施内容及び成果

1. 研究開発の背景

1. 本研究開発開始時でのねらい

本研究では、ヒト疾患ゲノム解析より見出された塩基多型あるいは突然変異が、発現制御の最終産物としてのトランスクリプトームに対していかなる影響を及ぼすのか、近傍エピゲノム、トランスクリプトーム情報(多階層オーミクス情報)を網羅的に記載するデータベースの作出を試みた。最終的に幅広い疾患ゲノムについてのデータ統合を目指す、当面のデータ統合は、がんゲノム解析から見出された体細胞突然変異の機能的解釈を志向した。様々な臨床検体、モデル実験システムより得られたオーミクスデータの統合を端緒とする。代表者らが個別のプロジェクトとして参画する国内のがんゲノム解析に国際エピゲノムコンソシアム(IHEC)より産出される標準エピゲノム、トランスクリプトームデータを核とするものとした。

研究計画開始当時、標準、疾患ゲノムデータベースは国内外で急速に整備が進められはじめており、その流れは現在もさらに加速している。当時、正常細胞での多階層オーミクスデータも早晩に整備されると見込まれており、実際、平成28年にIHECから第一期のエピゲノムカタログが公開された。本研究では、それらのデータ統合に加えて、多くのモデル実験系から産出されるデータの収載する点に特色を求めた。モデル系については、慎重に選択された実験系から産出されたデータを収集した。当時から現在に至るまで、がんゲノム解析に限っても、臨床検体に付随する多くの技術的、倫理的問題が存在する。これを回避するために、マウス等の動物モデルあるいは培養細胞モデルを用いて様々な機能解析を行うのが多くの局面において、常法である。多くの利点を有する一方で、これらのモデル系ではどの程度 *in vivo* の系を反映するのか、それぞれのモデル系にもどってその意義を検証する必要がある。臨床データ体系に照らしたデータの統合、比較が必要となるが、現在にいたるまで、その目的に適したデータベースは存在しない。

本研究課題で構築を試みたデータベースは、ヒト疾患関連遺伝子に直接、従事するゲノム研究者に資するだけでなく、モデル実験系から見出された知見をヒト疾患の診断、治療応用へ展開する機会を拡大することを企図した。また、データ取得背景が明確に規定されたデータ系列は、種々のシステム生物学的なアプローチを志向した研究にも共通基盤情報資源を与えることを期待した。さらに、セルベースアッセイの場としての培養細胞系のオーミクス基盤情報の整備は、ファーマコゲノミクスの効率化、薬剤スクリーニングの国際競争力を確保する上でも喫緊の課題である。

上記の付帯状況、その要求性は現在でも変わらない。実際、本研究で提案したデータベースの基盤整備は、広く現在の基礎的研究に端を発し、診断・治療応用への社会還元サイクルを加速しはじめている。がん代表される多くの疾患は、実体として多様なゲノム変異の集合である。これらのうちに適当な分子標的薬が入手可能なものについては、検出された変異パターンに最も適合した投薬を行う「最適化医療」が喧伝されるようになった。またゲノム変異に限らず、エピゲノム、トランスクリプトームについても、疾患発症には同様の多様性が関与している可能性も示唆されており、エピゲノムトランスクリプトーム状態を直接の標的とする抗がん剤の開発も開始されている。いずれにオーミクス階層についても、症例あるいは疾患ごとに類似した変異パターンを有する検体、実験系を求めればその解析、データの確保は困難である。本研究課題によるデータ統合は、個々の出現頻度はまれであっても、他の疾患、あるいはモデル実験系を俯瞰することで有効なデータの活用を行うことができるのではないかと考えた。症例あるいは時には生物種を超えての横断的検索を可能とし、全体あるいは少なくとも部分的に類似を示す“パターン”を発見できる機会を拡大するべくデータ統合を試みたものである。疾患あるいはモデル横断的な治験、臨床、基礎研究のいずれの目的に対しても、その援用知識空間を拡大することなくして、多くの複雑な疾患の治療・診断法を、現実的な規模感で効率的に開発することは困難であるとの考えに基づいている。

2. 本研究開発開始時での背景

イルミナ HiSeq を主軸とする近年の次世代シーケンズ関連技術の進展は著しい。ヒトゲノム配列を大量に決定し疾患関連多型・変異を同定、診断・治療に応用しようという試みは、現在に至るも数多い。特に研究開発当時、国際がんゲノムコンソシアム(ICGC)およびがんゲノムアトラス

(TCGA)に代表されるいわゆるがんゲノム解析において、累計1万を超える検体の全ゲノムあるいは全エクソーム解析の完了が報告された直後であった。その結果また代表者ら自身関わった国内のがんゲノム解析結果から、がんゲノムでの突然変異はがん種あるいは症例ごとに大きな多様性を示し、例外的な遺伝子を除いて個々の遺伝子が集団中に10%を超える頻度で出現することは非常にまれであることが明らかとなりつつあった。これはふたつの問題を提起すると考えた。第一に、疾患間で同一遺伝子の変異が低頻度にしか共起しないために、統計的手法でいわゆるがん原性の駆動力となるドライバー変異を同定するためには多数の検体の解析が必要である。第二に、同定された変異に関する生物学的機能解析に際し、臨床検体に比して変異パターンあるいはエピゲノム、トランスクリプトームパターンが相応である検証モデル系の選択が困難である。また、これまでの解析は主としてエクソン部分のアミノ酸置換変異について行われてきたが、遺伝子発現制御に機能異常を及ぼすプロモーター、エンハンサーあるいはスプライスサイトでの変異の重要性が指摘され、その解析も始められつつあった。実際、ICGCの第二期では発現制御に変動を及ぼす変異の全ゲノム解析からの同定が主要解析対象のひとつと位置づけられている。発現制御変異については、さらにその機能的影響の評価が困難である。情報源のがんゲノム変異に遺伝子発現制御の情報が含まれない以上、いかに情報学的手法を駆使してもゲノム変異と発現プロファイルの因果関係を類推するには限界がある。急速に蓄積するこれらの臨床がんゲノムデータから、その生物学的意味の抽出を試みるに際し、可能な限りのエピゲノム、トランスクリプトームデータの統合は必須であると考えた。

他方、エピゲノム、トランスクリプトームについても、国際ヒトエピゲノムコンソシアム(IHEC)および米国ロードマップ計画で活性化、抑制性のヒストン修飾、RNA polymerase IIをはじめとした基本転写因子の結合状態、DNAのメチル化状態を記載した、いわゆるクロマチンマップの作製が急速に進められていた。数年以内に、ヒト正常組織について1000人を超える検体からのエピゲノム情報が収集、標準カタログとして公開されることが想定されていた(平成28年に実際に公開された)。同時に、がんゲノムデータを始めとして多くの疾患ゲノム、これらの大量に産生される多階層オミクスデータを正常状態として関連づけるデータの統合は、早晚、国内の多くのサイトで進められることが想定された。(想定は国外に関しては、正しく、国内に関しては必ずしも著しい進展は見られなかった。)いずれにせよ、統合されるデータについて多くの場合、その由来する検体は同一個体でなく、ゲノタイプ、エピゲノムタイプ、トランスクリプトームタイプを直接関連付けることは不可能であると考えられた。

ヒト臨床検体を直接対象とした解析に並行して、試料の由来、実験環境の統一あるいは実験精度の高いオミクスデータの集積を目指して、培養細胞系あるいはマウスモデルを用いたモデル系でのオミクスデータの収集も進められていた。当時、ENCODE計画から、100種類のヒト培養細胞株について体系的な多階層オミクスデータが公開されたところであった。また、TALEN/CRISPR等の遺伝子編集技術が普及したことで、ノックアウト細胞、マウスの作出が加速、ゲノタイプと対応した形でのオミクスデータが蓄積することが想定されていた。大規模計画として顕在化しないものの実際、代表者らの参画する新学術領域「ゲノム支援」においても独自に作出された遺伝子改変細胞、モデル生物について収集されたゲノム解析および関連するオミクス解析は数多かった。ただし、多くの研究課題が最終的にヒト細胞のモデルを志向しているにも関わらず、モデル系の援用がヒト疾患機序の解明と治療法の開発等、ヒトの応用研究に直結している局面は少なかった。一方で培養細胞系、その他のモデル実験系は、いわゆるファーマコゲノミクスにおいて種々の薬剤スクリーニング系にも用いられる。恣意的に選択された細胞系において薬効を示した化合物に対して、その作用機序を明らかにするために改めてオミクス解析を行うことも多い。しかし、あらかじめ細胞系のオミクス状態の記載が完了していれば、詳細かつ体系的な構造-活性相関解析が可能になる等、飛躍的に効率的なスクリーニングを構築することが可能となると考えた。また、薬剤が当初期待されたものと異なる薬効を示すことがあるが、横断的なオミクス情報の検索が可能であればそのパターン比較から薬剤の新たな応用の発見につながる可能性について期待していた。米英でも同様の目的で情報の整理統合が進められていたが、掲載されているオミクス情報が少数の遺伝子のゲノムでの変異情報に限定されること、また対象となる細胞系が欧米人

由来のものに集中、民族固有性を有する疾患の解析、薬剤感受性の解析には不十分であることが予想された。以上の研究計画開始時の展望は3年後の現在に至るまでの期間に、実データの蓄積の進行という形で実証されており、内外で近年のうちに、その成果として結実する成果報告が相次ぐものと予想している。

3. 本研究開発開始時での目的

本研究では、ヒトがんゲノム多型・変異情報への生物学的機能情報注釈の付加を目的に、ヒト臨床検体に限らず培養細胞あるいはマウスをはじめとする動物モデル系において急速に収集の進む多階層オミクスデータの統合を試みた。我が国においても潜在的に利用可能であるデータは数多く、配列データはNBDC/DDBJに集約されることが期待されていた。しかしそれらを直接検索、応用可能なオミクスデータと統合する試みはなかった。またモデル系オミクスデータについてはデータ自体が散逸する可能性が懸念されていた。本研究では代表者らの保有するオミクスデータの統合を契機に、ヒト疾患応用を集約点に疾患あるいは種を超えて我が国に潜在するオミクスデータを整理統合することを試みた。日本人ゲノムに生じる疾患関連変異解析を志向して、我が国に特徴的な個々に精度が高く独自性、多様性に優れた幅広いデータを集約することで、国際的にも高い発信力を備えたデータベースの構築を行うものであった。当初、収載を計画したデータセットを下記の図に示す。現在までに、その大部分について、当初の計画を上回る規模でのデータの公開を完了している。また統合化推進機構の仲立ちを通じて、東北メディカルメガバンクデータベース、東京大学ヒトゲノム多型データベースとの連携を図ることができた。エピゲノム情報に関してはIHECとの緊密な連携を図ることができた。これらについては、一般に無償で実際にデータベース公開している。新規診断薬の開発等、その成果の産業的波及効果の真の結実には至らないものの、実務的な基盤整備には一定の道筋をつけることができたのではないかと考えている。

公開を予定していたデータの一覧

	名称	件数 (保有しているもの)	概要
1	ヒトがんゲノムデータ	300	肺腺・小細胞がん、食道がん、胃がん、大腸がんの全ゲノム/エキソームデータ
2	ヒト正常ゲノムデータ	500	正常日本人全ゲノム/エキソームデータの頻度情報
3	ヒトエピゲノム、トランスクリプトームデータ	300	肺腺がん、大腸がん(未完成):正常肝臓、正常大腸
4	ヒト培養細胞オミクスデータ	30 (300計測点)	肺腺がん、大腸がん、胃がん(未完成)、初代培養正常細胞
5	マウスオミクスデータ	10 (400計測点)	初代培養マクロファージ、B細胞等の免疫担当細胞
6	その他のモデル動物データ	100 (500計測点)	「ゲノム支援」で収集、公開予定のデータ

2. 研究開発対象のデータベース・ツール

(1) データベース

主要なもの

正式名称	略称	概要
ヒト臨床ゲノム多層オミクス統合データベース	KERO	日本人ゲノムの多型・変異情報を核に、培養細胞系、マウスモデル系での実験系からえられたエピゲノム・トランスクリプトームデータを統合したデータベースである。

上記以外のもの

正式名称	略称	概要
ヒト臨床ゲノム多層オーミクス統合管理システム	KERO-UT	上記が公開サーバー上に構築したものであるのに対して、多層オーミクスデータの実装を試験的に運用するものである。
臨床情報管理システム	KERO-NCC	臨床情報について試験的に統一するがんセンター内部に設置した試験データベースである。
KERO RDF ビューワー	KERO-DBCLS	RDF 化を実装、SPARQL 検索を試行する試用データベースである。最終的には RDF 対応サーバーとして NBDC に移管する予定である。

(2) ツール等

正式名称	略称	概要
BEACON 検索ツール	なし	外部の BEACON 検索を実装するデータベースと呼応して、本データベース中に格納するゲノム多型・変異あるいはその他のオーミクス情報の検索を可能とする枠組み。
外部データセット更新ツール	なし	dbSNP、ExAc、TCGA、ICGC、COSMIC 等、外部参照データの更新に関する作業を低減するツール。
RDF 化スキーマ構築ツール	なし	ゲノム、エピゲノム、トランスクリプトームに関する RDF 化について該当するスキーマに基づいてデータを成形し、検索を試行する一連のツール群。

※データベース、ツールの詳細は別紙参照。

3. 達成目標及び実施計画

(1) 当初の実実施計画・達成目標

研究開発計画とその進め方

研究開発開始当初、以下のような研究計画を立案した。実際に公開することのできたデータコンテンツ、ウェブブラウザ、および関連ツールの整備について、概ね当初の実実施計画に記載した通りの、あるいはそれを上回る成果を上げることでできたと考えている。

平成 26 年度

・ヒトオーミクスデータの統合

代表者らがこれまでに参画、収集した臨床検体でのゲノム、エピゲノム、トランスクリプトーム情報を整理統合する。代表者らが DDBJ/JGA を通じて公開している 100 症例の肺腺がんゲノム情報、近日中に公開予定である 100 症例の食道がんゲノム情報、大腸がんゲノム情報、また数年以内に IHEC より公開を予定している 20 種類の標準エピゲノム情報、230 症例の肺腺がんトランスクリプトーム情報をモデルにデータの成型と統合を行う。これらについては、代表者らがそのデータの産出、1次加工を行ってきたものであり、個々に参画する研究グループ内で構築されたデータベースへの格納が完了している。標準的なパラメータによるデータ集計(BWA を用いたマッピング、GATK での塩基変異の検出、MACS での ChIP Seq ピークの検出、rpkm による遺伝子発現量の記載)も終了している。またデータ加工に比較的ルーズな閾値を用いているために柔軟に事後的なパラメータ設定が可能である。

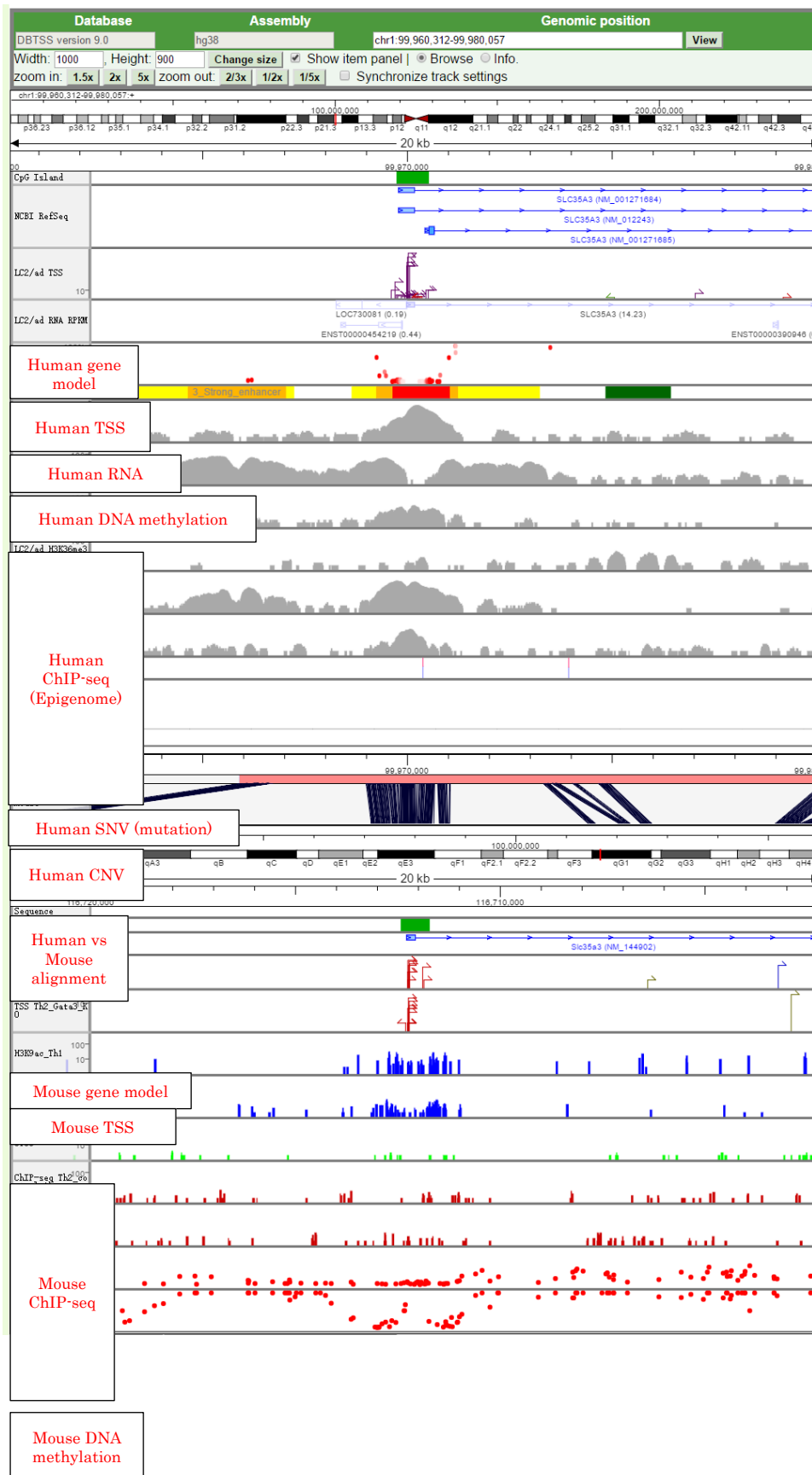
・培養細胞系データの統合

多階層オーミクスデータの統合の観点から、臨床検体に由来するデータを補完する意味で、代表者らは 26 種類の日本人、欧米人由来の肺腺がん細胞株について全ゲノムシーケンス(平均 depth x30)、一連の ChIP Seq(8種類のヒストン修飾)、RNA Seq(各 5000 万リード)、

DNA メチル化(バイサルファイトシーケンス:平均 depth x100)解析を終了している。また、今後、胃がん、食道がんについて総計 100 種類の細胞株を目処に同様のデータ産生を行う予定であるが、これらについてもデータベース化を行う。

・ブラウザの構築

ブラウザは現在代表者らが運用している転写開始点データベース(<http://dbtss.hgc.jp/>)を発展的に援用することにより行う。現在のトランスクリプトームデータにゲノム変異、エピゲノム情報を統合する。個々の要素については、それぞれの研究計画の目的に沿って構築したものであるが、データテーブルの統合、本研究に特化したいくつかの検索システムの実装により、横断的な検索とブラウズを比較的簡便に実現することが可能である。特に代表者らは代表的なパスウェイについて、文献情報にもとづいて描画した独自のパスウェイマップを用いたデータ検索エンジンを実装している(実際にその成果として公開されたデータを開発したブラウザ上に表示した例を下図に示す)。多くの個別遺伝子研究者にむけて、遺伝子を入口としたデータの閲覧を可能とすることで統一的なインターフェースを利用しつつも幅広く利用されるデータベースの開発を目指す。



MAPK/Erk in Growth and Differentiation

Cancer type: Lung adenocarcinoma 26 cell lines (each) | Cell: LC2/ad | Unit: RPKM log2 fold against normal cell

Tint control: light ——— deep

Coloring: RPKM log2 fold: <-6.2 (blue) 0 (white) 6.2< (red)

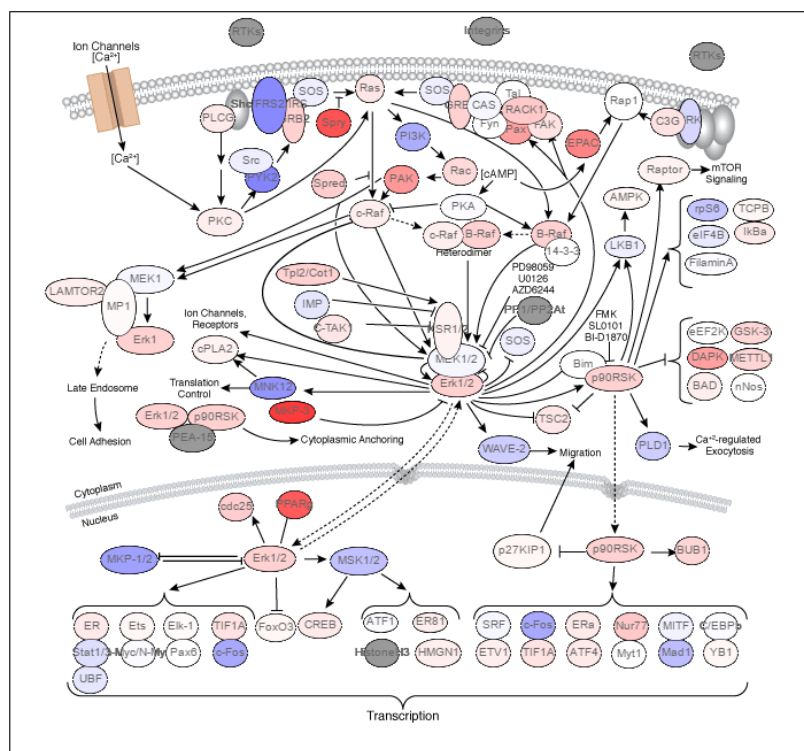


Illustration reproduced courtesy of Cell Signaling Technology, Inc.

・フォーマット、オントロジーの標準化

データベースの構築、統合化に関してはNBDCから提供されるガイドライン、オントロジーに従う。サンプルや実験条件などのメタデータについても、他分野との横断した利用や検索を可能とするために、他の統合化推進プログラムで既に利用されているオントロジーや BioPortal などのオントロジーのデータベースから世界的によく利用されているオントロジーを選択して利用する。上記オントロジーでカバーできない部分に関しては、DBCLS 等の協力を得ながら独自のオントロジーの開発も行う。また次世代シーケンスデータに関しては、標準的なデータフォーマットである BAM/WIG ファイルの形で提供する。

以上の項目については、少なくとも既に DDBJ/JGA に登録を完了しているデータについては、初年度において速やかに、実際に利用可能な形でデータベースの一般公開を目指す。(実際に初年度においてそのほとんどの公開を完了することができた。)

平成 27 年度

・モデル生物データの統合

代表者らは、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、多階層オー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。UCSC Genome Browser でのゲノム座標変換系を用いて、データをヒトに写像し、擬似的にヒトと同様の検索を可能とする。また代表者らは新学術領域「ゲノム支援」に中核データ産生機関として参画する過程で、マウス以外の多くのモデル生物(サル、ラット、メダカ、ゼブ

ラフィッシュ等)についても、多階層的オーミクスデータの収集を完了している。本研究課題は今年度に最終年度を迎え、そのほとんどのデータは来年度までに公開が可能となる。これらについても同様の枠組みで収載する。

平成 28 年度

・知識発見を志向した検索エンジンの開発と実装

各臨床検体、培養細胞あるいは動物モデル系についてゲノム、エピゲノム、トランスクリプトームの各階層でのパターン分類を行うことでデータベースの横断的検索を積極的に促進すべく、知識発見を志向したより複雑な検索システムの開発を行う。具体的には、がんドライバー遺伝子に着目して、その各階層でのパターンをモデル系と臨床検体で対応付ける。遺伝子全体からみると類似性を示さなくとも、特定のパスウェイ、遺伝子群に限って類似性を示すモデル系-臨床系データの相互に網羅的な対応付けを行う。これにより、それぞれのモデル系がモデルとして機能する臨床データ系をひも付ける。この部分についても現在、データベースに実装、公開されている(下図)。

以上、全開発期間を通じて、収載され、現在までに新規開発ブラウザ、検索システムを用いて公開されているデータセットを下表に示す。

・ヒト(がん細胞、正常細胞)

Data contents	Data accession	Data type (Datasets#)	Publication
SAEC (Normal Human Small Airway Epithelial Cell): RNA-Seq and ChIP-Seq	DRA002311	RNA-seq (1), ChIP-seq (8)	A. Suzuki et al., NAR, 2014
Lung cancer 26 cell-lines: Bisulfite sequencing	DRA001841	BS-seq (26)	A. Suzuki et al., NAR, 2014
Lung cancer 26 cell-lines: RNA-seq	DRA001846	RNA-seq (26)	
Lung cancer 26 cell-lines: Whole-genome sequencing	DRA001859	gDNA (26)	
Lung cancer 26 cell-lines: ChIP-seq	DRA001860	ChIP-seq (208)	
DLD1 (Hypoxia with non-tagged RNAi)	SRA003625	TSS-seq (1), RNA-seq (4), ChIP-seq (4)	Tanimoto et al., HUGO J, 2011
DLD1 (Hypoxia with HIF1A RNAi)	SRA003625	TSS-seq (1), RNA-seq (4), ChIP-seq (4)	
DLD1 (Normoxia with HIF1A RNAi)	SRA003625	TSS-seq (1)	
DLD1 (Hypoxia with HIF2A RNAi)	SRA003625	TSS-seq (1)	
DLD1 (Normoxia with non-targeted RNAi)	SRA003625	TSS-seq (1)	
DLD1 (Normoxia with HIF2A RNAi)	SRA003625	TSS-seq (1)	

Beas2B overexpress STAT6 IL4+	SRA008161	TSS-seq (1), RNA-seq (1), ChIP-seq (4)	
Beas2B overexpress STAT6 IL4-	SRA008162	TSS-seq (1), RNA-seq (1), ChIP-seq (4)	Kanai et al., DNA Res., 2011
Beas2B parent IL4+	DRA000017	TSS-seq (1)	
Beas2B parent IL4-	DRP000018	TSS-seq (1)	
Beas2B STAT6 siRNA- IL4+	DRA000021	TSS-seq (1)	
Beas2B STAT6 siRNA-	DRA000022	TSS-seq (1)	
Beas2B STAT6 siRNA+ IL4+	DRA000019	TSS-seq (1)	
Beas2B STAT6 siRNA+	DRA000020	TSS-seq (1)	
Ramos IL4+	SRA008163	TSS-seq (1), RNA-seq (1), ChIP-seq (4)	
Ramos IL4-	SRA008164	TSS-seq (1), RNA-seq (1), ChIP-seq (4)	
MCF7 O2 21%	SRA003625	TSS-seq (1), ChIP-seq (2)	Tsuchihara et al., NAR, 2009
MCF7 O2 1%	SRA003625	TSS-seq (1), ChIP-seq (2)	
TIG O2 21%	SRA003625	TSS-seq (1), RNA-seq (1), ChIP-seq (5)	Tsuchihara et al., NAR, 2009
TIG O2 1%	SRA003625	TSS-seq (1), RNA-seq (1), ChIP-seq (5)	
HEK293 O2 21%	SRA003625	TSS-seq (1), RNA-seq (1), ChIP-seq (1)	Tsuchihara et al., NAR, 2009
HEK293 O2 1%	SRA003625	TSS-seq (1)	
Hela		TSS-seq (1)	
Fetal Heart	DRA000024	TSS-seq (1)	Yamashita et al., Genome Res., 2011
Fetal Kidney	DRA000025	TSS-seq (1)	
Fetal Liver		TSS-seq (1)	
Fetal Thymus		TSS-seq (1)	
Fetal Brain	DRA000023	TSS-seq (1)	
Brain	DRA000026	TSS-seq (1)	
Heart	DRA000027	TSS-seq (1)	
Kidney	DRA000028	TSS-seq (1)	
Adipose		TSS-seq (1)	
Adrenal		TSS-seq (1)	
Brain		TSS-seq (1)	

Brain2	TSS-seq (1)
Breast	TSS-seq (1)
Colon	TSS-seq (1)
Heart	TSS-seq (1)
Kidney	TSS-seq (1)
Liver	TSS-seq (1)
Lung	TSS-seq (1)
Lymph	TSS-seq (1)
Muscle	TSS-seq (1)
Ovary	TSS-seq (1)
Prostate	TSS-seq (1)
Testis	TSS-seq (1)
Thyroid	TSS-seq (1)
UPF1 siRNA1 control	RNA-seq (6)
UPF1 siRNA1 control	RNA-seq (6)
UPF1 siRNA1	RNA-seq (6)
UPF1 siRNA2	RNA-seq (6)
UPF1 no siRNA1	RNA-seq (1)

・マウス

Data contents	Data accession	Data type (Datasets#)	Publication
NIH3T3		TSS-seq (1)	
Mouse embryo 7d Mouse embryo 11d Mouse embryo 15d Mouse embryo 17d		TSS-seq (4)	Wijaya et al., Genome Inform., 2009
10T1/2 0h		TSS-seq (1)	
ATDC5 0h		TSS-seq (1)	
ChIP-seq Th1,TH2 anti-Chd4	DRA000928	ChIP-seq,(8) TSS-seq (3)	Hosokawa H. et al., Proc Natl Acad Sci USA., 2013
TSS-seq Th1,Th2	DRA000928		
ChIP-seq Th1,TH2 Ab	GSE51079,SRP007 894	ChIP-seq (2)	
ChIP-seq Th1,TH2 GATA3	GSE51079,SRP007 894	ChIP-seq (3)	
ChIP-seq Th1,TH2 H3K27me3,H3K9ac,H3K4 me3	GSE51079,SRP007 894	ChIP-seq (5)	Hosokawa H. et al., Proc Natl Acad Sci USA., 2013
ChIP-seq 3xFlag-GATA3	DRA001102	ChIP-seq (2)	

RNA-seq Th1,Th2	DRA001102, DRA000928	RNA-seq (3)	
ChIP-seq Mouse Liver 1D4	DRA001050	ChIP-seq (20)	Yoshitane H. et al., Mol Cell Biol., 2014
ChIP-seq Mouse Liver CLSP4	DRA001050	ChIP-seq (10)	
RNA-seq Mouse Liver	DRA001050	RNA-seq (8)	
micro RNA-seq Mouse Liver	DRA001050	miRNA (8)	
RNA-seq Bmap Ko Mouse Liver	DRA001278	RNA-seq (4)	
RNA-seq WT Mouse Liver	DRA001278	RNA-seq (4)	
ChIP-seq Th1,Th2 Ab	GSE51079	ChIP-seq (3)	Tumes DJ. et al., Immunity., 2013
ChIP-seq Th1,Th2 Ezh2	GSE51079	ChIP-seq (3)	
Reg1[-/-] LPS stimulated	DRA003234	ChIP-seq (4)	Mino T. et al., Cell., 2015
Roquin [San/San] LPS stimulated	DRA003234	ChIP-seq (4)	
TP3_wild_H3K27me3	DRA000485	ChIP-seq (2)	Tanaka S. et al., Blood., 2012
TP3_KO_H3K27me3	DRA000486		
MLL_AF9_AML_Ezh2_WT_input	DRA000487	ChIP-seq (1)	
TP3_4_input	DRA000488	ChIP-seq (1)	
Tet2 KD	DRA000858	ChIP-seq (2)	
Ezh2 KD	DRA000858	ChIP-seq (2)	
Tet2 Ezh2 KD	DRA000858	ChIP-seq (4)	Muto T. et al., J Exp Med., 2013
Blastocyst	DRA000484	ChIP-seq (1)	Kobayashi H. et al., PLoS Genet., 2012
ESC	DRA000484	ChIP-seq (1)	
GV_oocyte_Dnmt3L	DRA000484	ChIP-seq (1)	
GV_oocyte_wild	DRA000484	ChIP-seq (1)	
Sperm	DRA000484		
10.5 female PGC	DRA000607	ChIP-seq (1) ChIP-seq (1)	Kobayashi H. et al., Genome Res., 2013
10.5 male PGC	DRA000607	ChIP-seq (1)	
13.5 female PGC	DRA000607	ChIP-seq (1)	
13.5 male PGC	DRA000607	ChIP-seq (1)	

16.5 female PGC	DRA000607	ChIP-seq (1)
16.5 male PGC	DRA000607	ChIP-seq (1)

・ヒト mutation (TCGA/ICGC その他の外部研究室データ・正常日本人等)

Data provider	Disease type	Number of samples	Reference
National Cancer Center Hospital East and University of Tokyo	Lung adenocarcinoma	97	PLoS One. 2013 Sep 12;8(9):e73484.
National Cancer Center Hospital East	Small cell lung cancers	57	J Thorac Oncol. 2014 Sep;9(9):1324-31.
ICGC	43 of ICGC DCC Project Codes	6,590	https://dcc.icgc.org/
Dr. Meyerson's Lab.	Lung adenocarcinoma	183	Cell. 2012 Sep 14;150(6):1107-20.
Dr. Ogawa's Lab.	Myelodysplasia	29	Nature. 2011 Sep 11;478(7367):64-9
	Clear-cell renal cell carcinoma	106	Nat Genet. 2013 Aug;45(8):860-7
TCGA	Gastric adenocarcinoma	295	Nature. Published online 23 July 2014
	Urothelial bladder carcinoma	131	Nature. 507 (7492):315-22.
	Glioblastoma	291	Cell. 155 (2):462-477.
	Clear cell renal cell carcinoma	446	Nature. 499 (7456):43-49.
	Endometrial carcinoma	373	Nature. 497 (7447):67-73.
	Acute myeloid leukemia	200	NEJM. 368:2059-2074.
	Breast tumors	507	Nature. 490 (7418):61-70.
	Squamous cell lung cancers	178	Nature. 489 (7417):519-525.
Colon and rectal cancer	224	Nature. 487 (7407):330-337.	

	Ovarian carcinoma	316	Nature. 474 (7353):609-615.
	Glioblastoma	91	Nature. 455 (7216):1061-1068.
HGVD	Normal (Japanese)	1,208	URL:http://www.genome.med.kyoto-u.ac.jp/SnpDB
ToMMo	Normal (Japanese) *Not open data	1,070	URL:http://humandbs.biosciencedbc.jp/en/hum0015-v1
JPDSC (Japan PGx Data Science Consortium)	Normal (Japanese)	2,994	URL:http://humandbs.biosciencedbc.jp/hum0013-v1
ExAC	Various disease	60,706	URL:http://exac.broadinstitute.org/

・IHEC (一般未公開)

Data contents	Data accession	Data type (Datasets#)
Cytotrophobalst, term placenta (XY)	JKU001	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Cytotrophobalst, term placenta (XX)	JKU002	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Cytotrophobalst, 1st trimester placenta (XY)	JKU003	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Cytotrophobalst, 1st trimester placenta (XX)	JKU004	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Cytotrophobalst, 1st trimester placenta (XX)	JKU015	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Syncytiotrophoblast, 2nd trimester placenta	JKU005	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Syncytiotrophoblast, 2nd trimester placenta	JKU006	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Syncytiotrophoblast, 1st trimester placenta	JKU007	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Syncytiotrophoblast, 1st trimester placenta	JKU008	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Extravillous cytotrophobalst, 1st trimester placenta	JKU016	RNA-seq (1), ChIP-seq (7)
Extravillous cytotrophobalst, 1st trimester placenta	JKU017	RNA-seq (1), ChIP-seq (7)
Endometrial epitherium, follicular phase	JKU009	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Endometrial epitherium, follicular phase	JKU010	ChIP-seq (7)
Endometrial epitherium, secretory phase	JKU011	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Endometrial epitherium, secretory phase	JKU012	ChIP-seq (7)
Endometrial stroma cell	JKU013	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Endometrial stroma cell	JKU014	BS-seq (1), RNA-seq (1), ChIP-seq (7)

Aorta EC	JTK001	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Aorta EC	JTK002	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Pulmonary Artery EC	JTK003	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Coronary artery EC	JTK004	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Coronary artery EC	JTK005	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Umbilical vein EC	JTK006	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Umbilical vein EC	JTK007	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Aorta EC	JTK008	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Coronary artery EC	JTK009	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Pulmonary Artery EC	JTK010	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Cardiac Microvascular EC	JTK011	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Endocardiac cells (ENDC)	JTK012	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Aorta EC	JTK013	RNA-seq (1), ChIP-seq (3)
Pulmonary Artery EC	JTK014	RNA-seq (1), ChIP-seq (2)
Coronary artery EC	JTK015	RNA-seq (1), ChIP-seq (3)
Umbilical artery EC	JTK016	RNA-seq (1), ChIP-seq (2)
Umbilical vein EC	JTK017	RNA-seq (1), ChIP-seq (2)
Umbilical vein EC	JTK018	RNA-seq (1), ChIP-seq (2)
Aorta EC	JTK019	RNA-seq (1), ChIP-seq (3)
Brachiocephalic artery EC	JTK020	RNA-seq (1), ChIP-seq (2)
Aorta EC	JTK021	RNA-seq (1), ChIP-seq (2)
Coronary artery EC	JTK022	RNA-seq (1), ChIP-seq (2)
Umbilical artery EC	JTK023	RNA-seq (1), ChIP-seq (2)
Pulmonary Artery EC	JTK024	RNA-seq (1), ChIP-seq (2)
Brachiocephalic artery EC (BCAEC)	JTK025	RNA-seq (1), ChIP-seq (2)
Common Carotid Artery (CCA)	JTK026	BS-seq (1), RNA-seq (1), ChIP-seq (4)
Endocardiac EC	JTK027	ChIP-seq (3)
Common Carotid Artery (CCA)	JTK028	ChIP-seq (3)
Common Carotid Artery (CCA)	JTK029	ChIP-seq (3)
Common Carotid Artery (CCA)	JTK030	ChIP-seq (3)
Common Carotid Artery (CCA)	JTK031	ChIP-seq (3)
Common Carotid Artery (CCA)	JTK032	ChIP-seq (3)
Common Carotid Artery (CCA)	JTK033	ChIP-seq (3)
Common Carotid Artery (CCA)	JTK034	ChIP-seq (3)
Carotid Artery EC	JTK035	ChIP-seq (3)
Subclavian Artery EC	JTK036	ChIP-seq (3)
Common Carotid Artery (CCA)	JTK037	ChIP-seq (3)
Umbilical artery EC	JTK038	ChIP-seq (3)
Endocardiac EC	JTK039	ChIP-seq (3)
Aorta EC	JTK040	ChIP-seq (3)
Coronary artery EC	JTK041	ChIP-seq (3)
Umbilical vein EC	JTK043	ChIP-seq (3)
Umbilical vein EC	JTK044	ChIP-seq (3)
Liver, hepatocytes	JNC001	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Liver, hepatocytes	JNC002	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Liver, hepatocytes	JNC003	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Liver, hepatocytes	JNC002	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Liver, hepatocytes	JNC003	BS-seq (1), RNA-seq (1), ChIP-seq (7)

Liver, hepatocytes	JNC004	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Liver, hepatocytes, hepatitis B virus-positive chronic hepatitis	JNC005	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Liver, hepatocytes, hepatitis C virus-positive chronic hepatitis	JNC006	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Ascending colon, surface epithelial cells	JNC0015	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Ascending colon, surface epithelial cells	JNC0016	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Ascending colon, surface epithelial cells	JNC0017	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Ascending colon, surface epithelial cells	JNC0018	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Descending colon, surface epithelial cells	JNC0019	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Descending colon, surface epithelial cells	JNC0020	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Descending colon, surface epithelial cells	JNC0021	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Descending colon, surface epithelial cells	JNC0022	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Rectum, surface epithelial cells	JNC0023	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Rectum, surface epithelial cells	JNC0024	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Rectum, surface epithelial cells	JNC0025	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Rectum, surface epithelial cells	JNC0026	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0027	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0028	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0029	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0030	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0027	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0028	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0029	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0030	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0027	BS-seq (1), RNA-seq (1)
Stomach, foveolar epithelial cells	JNC0028	BS-seq (1), RNA-seq (1), ChIP-seq (7)
Stomach, foveolar epithelial cells	JNC0029	BS-seq (1)
Stomach, foveolar epithelial cells	JNC0027	BS-seq (1)
Stomach, foveolar epithelial cells	JNC0028	BS-seq (1)
Stomach, foveolar epithelial cells	JNC0029	BS-seq (1)

(2) 期間中に追加・削除・変更した実施計画・達成目標

本研究課題で提案するデータ統合は、代表者らがこれまでにデータ収集と1次解析に関わってきたデータについて行うものであって、研究開発当初から、データの取得、公開に障害はないと考えていた。実際、その面での公開範囲の縮小を強いられることはなかった。むしろデータ量の蓄積は予定よりも順調に推移し、現在までに公開が完了しているデータセットは当初予定の規模を超える。データベースサーバーについては代表者らが他のデータベースを開発、運用している東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピューターシステムを用いた。同センターではこれまでに多くのヒトに由来するデータ等の機微情報を扱

ってきており、ヒト臨床検体試料由来のデータに関する倫理的な配慮についても問題はなかった。また当初の統合に用いたデータは適切な倫理審査の後に、DDBJ/JGA よりすでに公開されているものであった。以上、本研究開発には、不確定性の強い要素はほとんど含まれず、忠実な計画の履行を行うことで大きく計画変更を強いられる可能性は低いと考えていた。この点についても研究計画は順調に推移した。

また、知的財産権等に関する考え方について、本研究で構築される情報基盤は広く我が国の公益に供するべきであるという考えに基づいて、知的財産権の取得については考慮しないこととした。実際に、本研究で作成されるデータベースについては、産業界からも学界からも制約なく利用を可能としている。収載される一切の関連データは無償、あらゆる再利用権を付与して公開されている。データベース自体が NBDC のアーカイブサイトから無償で提供可能にする予定である。逆にその用途に合致しないデータは当面の収載範囲に含めなかった。

全体として、達成成果は十分なものであったと考えているが、研究計画最終年度に企図していた臨床データの統合を本研究開発期間中には完了することができなかった。これについては、現在、主たる共同研究者の河野 (DBCLS) は、がんセンターとの共同研究において、ゲノム情報と動的に変化する臨床情報を高いデータ機密性を持って対応付け、検索を可能とする枠組みを構築している。データベースのこの部分は、正常、病変部あるいは再発前後でのゲノムデータ、および薬剤投与等の治療履歴が詳細に記載されている。これについてはヒトゲノム情報の機微性から一般に公開することは不可能であるが、本データベースで構築される生物学的な知見を臨床データベースに接続することは可能である。当初、がんセンターをモデル施設に、実際の診断、治療を志向した応用研究を可能とする枠組みの構築をする。ヒトに関するデータ等の機微情報に関しても、制限付きアクセスで利用可能となる部分については明示し、そのシーケンスデータ (FASTQ データ等) について、データ供給者からの入手の便宜を図る。得られる枠組みは同様のデータを蓄積する各医療機関に無償で供与するべく引き続き検討を進める予定である。

4. 実施内容

(1) 実施内容

研究開発の実施は、研究分担者が密接な連絡を保ちつつ遂行することでできた。これは申請時に記載したように、参画する分担者が千葉県柏地区に地理的に集中しており、頻繁に会合を行うことが可能であったことによると考えている。本データベースの構築趣旨上、定量的にその貢献を評価することは難しいが、各研究分担班の役割は以下の通りであった。

菅野、鈴木は研究を統括した。東京大学柏シーケンス拠点で産生されたデータ収集と 1 次加工を行った。ヒト臨床ゲノムデータを中心に、同時に可能な限り広い範囲で関連研究機関と連携し、産生されるデータとの接続を模索、相互にさらなるデータベースの拡充を図った。取得、成形されたデータは河野らが国立がんセンター東病院および DBCLS 柏オフィスで開発するデータベースに投入、データベースの作成に資した。また、菅野、鈴木は試作されたデータベースを一般公開するに際し、医科学研究所スーパーコンピューターシステムへの移行と運営を行った。

土原は、データの収集、オントロジーに留意したメタデータの整備を行った。土原がこれまでに産生したあるいは現在産生を行っているゲノム変異、エピゲノムデータは、本データベースで収集される臨床ゲノムデータの中で最大のものである。さらに土原は、模擬的に最終ユーザーとしてデータベースのヒト臨床がん研究への応用的活用を試行した。土原自身が試験的に本データベースを活用することで、種々の閾値の設定を含めたパラメータの最適化を行った。また、臨床医としての利用者の視点から見た検索項目の設定を行った。本データベースにおけるシグナル伝達パスウェイごとのデータ統合は土原の立案、実装による。

河野は、データベースの実務を担当した。データベース本体および初年度のウェブシステムは、菅野、鈴木らがこれまでに公開した転写開始点データベースを発展的に移行する形で省力的に開発し、特に外部データベースデータと連携を進めるツールの開発を行った。

BEACON 検索の実装等、その他の外部のデータベースとの連結を可能とする一連の手法群は河野の開発による。また土原、鈴木と情報交換を密にし、NBDC/DDBJ に準拠したオントロジーの整備、メタデータの付加、公開データのフォーマットを行った。特に格納された各階層でのオーミクスデータの RDF 化について、そのスキーマの作成に主導的役割を果たした。現在も DBCLS の関連研究者、さらには IHEC 担当者と協議して、統一的なスキーマの枠組みの構築を検討している。

(2) データベースの利便性に関する利用者ニーズと具体的な対応

本研究開発期間において、データ利用者の拡大を目指して、毎年、分子生物学会年会においてブースを出展した。また、分子生物学会、がん学会、人類遺伝学会、その他の国内外の学会で口頭発表により、また羊土社、秀潤社をはじめとする国内総説出版により、データベースの周知を行った。いずれについても、現行での本データベースに収載されたデータの潜在的な有効性については一定の評価が得られた。ただし、そのユーザーインターフェイスについての整備不備を指摘する意見も数多く収集された。Help を中心とした Documentation の充実、遺伝子名の先読み補完機能の実装、利用者からのフィードバックを容易にする Wiki/SNS の整備、ダウンロードサイトの負荷軽減等、数多く浮き彫りになったこれらの課題について、今後、ひとつずつ検討を加えていく予定である。

(3) 持続的なデータベース運用体制の構築に向けた取り組み

本研究開発においては、データ統合、公開の枠組みの構築等、実際のデータの収載、データベースとしての公開の多くは、研究期間の前半に達成した。研究期間後半では、データ収集自体でなく、データの効率的な利用と将来的な発展性をにらんで、知識発見の枠組みの構築と RDF 化を主軸とするデータ統合化の枠組みの開発に注力してきた。実際、期間後半では、データ投入、加工、検索機能の更新については、前期に開発したツール群を活用することにより、半自動的な処理で行われ、省力化、省コスト化を実現している。今後も、現在、開発が完了している部分については最小限のコストでデータベースの維持が可能であると考えている。生じた人的な余力は、外部データベースとの統合作業に割くことが可能であり、上記のようにいくつかの主要データベースとの統合を完了、あるいはそれに向けた実務レベルでの作業を完了している。

データ産生者との接点として、代表者らは、今後も科研費新学術領域「先進ゲノム支援」シークエンス拠点のひとつとして、継続的にデータ産生を行っていく予定である。データ産生者側の視点からも、多階層オーミクスデータとその活用は取得データの生物学的意味付けの基幹をなす解析体系であり、逆に本データベースの成果をフィードバックする形で、効率的にこれらの課題についての生物学的な知識発見を支援していこうと考えている。

(4) 統合化推進プログラムの他のチームや DBCLS との連携

JST の仲介により、本 NBDC 統合化推進事業で関連するデータベース構築を行っている徳永チーム(開発データベース;ヒト多型データベース HGVD)と緊密な連携を実現すべく、実務者会議を年に数度、定期的に行った。ヒト多型データに関して、その位置、頻度に関して相互に密接なリンクで結ぶことができた。またヒトゲノムデータの RDF 化について協議を行い、そのベンチマークテストを実施することができた。

(5) データ産出を行う研究組織や研究室、プロジェクトとの連携

本データベースの作成は、東京大学、国立がん研究センターの研究者に DBCLS の研究者を加えて行っている。菅野、鈴木らは新学術領域「先進ゲノム支援」において中核支援機関のひとつとして、多くの生物種、実験系においてオーミクスデータを産生している。該当データの1次加工、2次以降の情報解析について、情報系研究者との連携も密であり、多くのデータを今期研究機関に本データベースに収載することが可能であった。また鈴木は

CREST-IHEC において、データ収集、加工、公開について、実務を担当している。鈴木を窓口として IHC データとの連携、一部データの登録、あるいは国際的な RDF 化スキーマの設定に向けて密接な協議を行うことができた。開発期間中に土原は大規模臨床ゲノムシーケンスコンソーシアム SCRUM JAPAN を立ち上げた。産出されつつあるデータは AMED で開発される臨床ゲノムデータベースに収録される予定であるが、将来的には同データベースと本データベースとの連携を図ることも可能であると考えている。

(6) 人材の育成

本研究計画当初においては、菅野、鈴木の下、博士課程学生に、特にデータベースのコンテンツに関する活用の実務を担当させた。該当者はデータ解析者としてがんセンターに常任研究員として雇用され、現在、若手実験手法、情報解析手法をつなぐ貴重な人材として、活躍している。また、同博士が開発したデータベースの枠組み(がん細胞の多層オーミクス情報の統合)、実践したデータベース活用について、さらには自身の経験談も交えて、NBDC の人材育成セミナーで 2 度にわたって講演を行った (<http://events.biosciencedbc.jp/training>; 2015.08.04 AJACS 鳥取大学; 2016.09.12 AJACS 東京女子医科大学)。また、同氏の転出後、研究期間後半においては、別の博士研究員を雇用して、その任に充てた。こちらの人材についても将来の実験系情報系ゲノム科学を担う人材として必要な経験を習得していると考えている。

(7) その他

研究実施期間中に行われたヒトゲノム情報倫理関連指針の改定により、多くのプロジェクトがデータの扱い、データベース登録要件とその審査、登録の実践について、実務上の混乱が発生していた。本データベースの構成員がその実務を実行あるいは適切な助言を与えることにより回避された混乱も多いのではないかと自負している。実際、本データベースが窓口となつて、NBDC ヒト DB に配列登録が実現したヒト臨床検体由来オーミクスデータも相当数存在する。

§4. 主要なデータベースの利活用状況

1. アクセス数

(1) 実績

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
KERO	訪問者数	15,693	16,406	17,322	17,784
	訪問数	30,679	30,012	31,324	32,985
	ページ数	1,385,881	1,360,024	1,260,056	1,307,171

表 1-1 研究開発対象の主要なデータベースの利用状況(年度別)

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
KERO	訪問者数	1,308	1,367	1,444	1,976
	訪問数	2,557	2,501	2,610	3,665
	ページ数	115,490	113,335	105,005	145,241

表 2-2 研究開発対象の主要なデータベースの利用状況(月間平均)

(2) 分析

本データベースは、訪問者数、ページ数ともに暫時的に増加傾向にあり、本データベースに関する関心の増大を反映していると考えている。関連学会への出展、利用者講習会の開催等でアピールに努めた際にも利用者数は増加し、普及活動が有効に機能していると考えている。ただし、新規のデータを更新した際のアクセス数は期待したほどの増加を示さないことから、毎回のアップデートについて、継時的に告知する枠組みが必要であると思われた。メール配信あるいは Wiki、SNS の活用等、情報発信の観点からの開発も必要であると考えている。

2. データベースを利用して得られた研究成果事例

代表者らは、本データベースの構築を報告した (Suzuki et al. NAR DB Issue 2015)。さらに、他の外部機関の研究者との共同研究により、がん細胞の多層オーミクス解析を行い、がん細胞における遺伝子発現制御異常のパターンの網羅的解析を報告した (Suzuki et al. NAR2014)。また、代表者がヒト正常細胞における転写開始点、転写開始終結点についてのデータを収集し、本データベースに格納したことを報告した (Matsumoto et al NAR 2014)。

3. その他

本計画により開発されたデータベースは、国内外からの多くのウェブアクセスとデータのダウンロードあるいは関連した問い合わせを受けている。また、積極的に講演会等での紹介を行いデータベースの有効活用の推進を試み、また日本語総説により、データベースの周知に努めた結果、利用者は漸次、増加傾向にある。

§5. 研究開発期間中に得られた科学・技術や産業に対する波及効果

蓄積されつつあるゲノム多型・変異情報を活用すべく、その生物学的情報を多層オーミクス解析に求める試みが各所で始められている。国内では、その試みはオーミクス計測技術導入の観点からその基盤整備が遅れる大学等の公的研究機関より、民間企業においてより鋭敏な反応が見られるように思う。本データベースが、全体として低調なデータ共有化の現状に対して効率的な利用を考えるうえで一石を投じる、モデルケースとして機能することを期待している。

実際、代表者自身、本活動が契機となって、データベースに掲載されている個別の事象についてあるいはその技術的な改良、さらには生物学的な検証を目的に、代表者は、日立製作所、富士

フィルム、協和発酵キリン、ゲノメディア社、シスメックス社等、バイオテック関連企業、製薬企業、診断機器企業、バイオインフォマティクス企業の幅広い業種の企業との共同研究開発計画を開始することができた。また関連オーミクスデータを基盤情報として、AMED の主導する GapFree2 を核として、製薬企業(武田、第一三共、アステラス)のコンソシアムが形成され臨床検体を用いた同種データの収集と解析を開始している。いずれも本データベースに収載されたデータの潜在的な有効性に期待が集まっていることの証左であると考えている。しかし、これらの期待を実際の成果として結実させるためにはより一層の開発が必要とされることも同時に明らかとなった。ゲノム多型・変異データについては、現在、本データベースがカバーする日本人ゲノム多型・変異の情報は依然として限られている。これは、現在までに集積されたゲノム配列の 1 次データ量が依然として限られていたことに主因があるが、最近になって本邦でも、大規模なゲノム配列シーケンス計画が開始された。ただ、ヒトゲノム情報に不可分な個人情報の扱いの問題もあり、大量に産出されるデータを特に民間に対して広く有効活用する明確な用途はまだ立っていない。実際、ヒトゲノム情報の取扱いについては、今春の個人情報保護法の改正をにらんで、多くの混乱があり、鈴木はいくつかの倫理関連学会、研究会で口演、意見交換に参加した。新法制下でのデータの有効利用の形を早急に示す必要があることが求められている。

多層オーミクスデータについて、本データベース構築の過程で試行したゲノムデータとの統合は、見出されたゲノム多型・変異の生物学的解析を進める有効な手段であったが、その結果、さらに大規模でのデータの収載とその統合的解析の手法開発が必要であることが明らかとなった。例えば、抗がん剤開発には少なくともいくつかのモデルデータについては、実際に薬剤を作用時に惹起されるオーミクス状態の摂動を記載、解析することが必要である。また薬剤耐性細胞の同定と解析にはシングルセル解析が必須である。これらのいずれについても最近、データ取得法についての進展が著しいが、そのデータの集積、解析手法については、依然として、具体的な取り組みがなされていない。

以上、多層オーミクスデータのゲノム変異・多型への統合について、本研究開発がその将来展望に向けて一定の波及効果を及ぼすことができたと考えているが、その結実までにはより一層の集約的な開発が求められていると感じている。

§6. 今後の展開

国内外でゲノム多型・変異情報の蓄積はますます加速している。我が国でも東北メガバンクをはじめ日本人のゲノム多型を収集する試みが一定の成果をあげている。各国でもそれぞれの民族的背景について、ゲノム情報、遺伝子多型の頻度情報を格納したデータベースが急速に整備されている。一方で、ENCODE、IHEC 計画でもデータコーディネーションセンターの整備が急速に進められている。しかし、現在のこれらのデータベースは、疾患特異的にゲノム多型を扱うもの、あるいはオーミクスの各階層に特化したものであって、本質的にその横断的な検索には不向きである。反面、ゲノム、エピゲノム、トランスクリプトーム解析においては、シーケンスプラットフォーム、鋳型調製、品質管理を含めた実験手法、そのデータの一次解析について、驚くほど画一化が進められており、必ずしも実験系に特化した処理を必要としない。データ集約点として、我が国では DDBJ あるいは NBDC がオーミクスデータを集約しているが、主として格納、公開されているのは FASTQ ファイルを中心とした生データである。それに対し本データベースは、将来的にもこれらのデータベース群の下流に位置し、全体としてヒトゲノムデータの生物学的解釈を主眼に、DDBJ/JGA に格納される配列データに多階層オーミクスデータから見て遺伝子機能情報をメタデータとして補完する体制の確立を目指した。今後もゲノムおよび多階層オーミクスデータについての集積は増大すると思われる。潜在的に大きな価値を有するこれらのデータを DDBJ/JGA に格納されたまま死蔵する危機を回避すべく、本データベースを最終的には NBDC で直接管理運営する体制について協議したい。

またデータ共有に関しては、ヒトゲノム情報の活用が最も深刻な問題である。現在のところ個人情報の扱いの制約からその 1 次データについては、狭義の意味での共用は許容されない。また、同データは付帯する患者臨床情報データなしにはその価値を大きく変じるが、この部分についてはさらに厳密な機微性が必要となる可能性がある。しかし、代表者はこのような制約条件下にあっても、依然としてその共通利用化に向けての研究開発の試みを放棄すべきでないと考えている。例えば、多くの臨床研究においては、個々の個所のゲノム変異情報よりもその集団中の変異情報を利用することも多い。現行法制下でも、ゲノム配列については個人の全体配列情報は秘匿されるものであっても、個別の変異箇所の集団中の変異頻度については一般公開が可能である。ただし、個々に異なる臨床情報に対して、事前に想定されない検索条件に合致する症例について、部分集団中のある変異頻度情報を計算するには、結局、個々のデータについてのアクセスを行う必要がある。ただし、この部分を計算機上でのあるいは臨床研究コーディネーターの人的資源を投入することで秘匿検索することができれば、厳密には個人情報保護に抵触することなく、データの活用が可能かもしれない。この点についての研究開発は喫緊の課題であると考えている。

知的財産権の取得については、代表者は、上記の将来的な開発項目も加えて、本研究計画で構築される情報基盤は、広く我が国の公益に供すべきであって、その産官学での有効活用を妨げる知財権は行使すべきでないと考えている。代表者らはこれまでも一貫したポリシーとして、ゲノム研究、特に、その基盤研究段階においては、過剰な知的財産権の主張は、新規の研究シーズ発見への障害となる局面が多いと位置付けてきた。作成されるデータベースについては、産業界からも学界からも制約なく利用を可能できるものとしたい。収載される一切の関連データは無償、あらゆる再利用権を付与して公開する。特に DDBJ/NBDC のアーカイブサイトから無償で提供可能とする。逆にその用途に合致しないデータ、ツールは収載範囲に含めるのは、適当でないと考えている。

§7. 自己評価

本研究課題は、革新的な技術開発を提案するものでも、以前では想像もできなかったデータベース構築の枠組みを提案するものではない。それにも関わらず、代表者らは、創出されるデータベースは、基礎生物学的研究からヒト臨床応用研究にいたるまで幅広い研究分野において、基盤的なデータベースとして大きな役割を果たすものができるかと自負している。次世代シーケンス技術をはじめとする大規模ゲノム解析手法の普及によって、ゲノム規模のデータが取得されているにも関わらず、それらが個別遺伝子研究にのみ活用されたのちに、十分にデータ抽出されないまま死蔵される機会は今後さらに増すと考えられる。本データベースの構築を端緒に、それらについてヒト疾患についての応用を焦点に新たな付加価値を付加することを可能とする枠組みの整備を期待している。本データベースは、その個々の活用事例の集積としての成果だけでなく、その問題提起の意味でも有意義な開発を行うことができたのではないかと考えている。

§8. 外部発表等

1. 原著論文発表

(1) 論文数概要

種別	国内外	件数
発行済論文	国内(和文)	0件
	国際(欧文)	3件
未発行論文 (accepted, in press 等)	国内(和文)	0件
	国際(欧文)	0件

(2) 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)

2. 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)
5. 鈴木穰 編「実験医学別冊 NGS アプリケーション RNA-Seq 実験ハンドブック」, 羊土社, 2016. (2016年3月)
6. 鹿島幸恵, 鈴木絢子, 鈴木穰 「一細胞オーミクス解析 ～がん微小環境の解明にむけて～」, 週刊医学のあゆみ, 258巻1号 (2016年7月2日号)
7. 鹿島幸恵, 鈴木絢子, 鈴木穰 「一細胞解析技術」, 週刊医学のあゆみ, 258巻4号 (2016年7月23日号)
8. 鹿島幸恵, 鈴木絢子, 関真秀, 鈴木穰 「一細胞解析(仮)」, 実験医学別冊『エピジェネティクス実験スタンダード(仮)』, (2017年5月発行予定)

3. 国際学会発表及び主要な国内学会発表

(1) 概要

種別	国内外	件数
招待講演	国内	7件
	国際	4件

口頭発表	国内	4 件
	国際	1 件
ポスター発表	国内	3 件
	国際	5 件

(2) 招待講演

〈国内〉

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
4. 鈴木穰、1 細胞解析、平成 27 年度文部科学省新学術領域研究生命科学系 3 分野 がん・ゲノム・脳 支援活動 合同シンポジウム、東京、2015 年 8 月 19 日
5. 鈴木穰、肺腺がん細胞株のシングルセル解析: 遺伝子発現の多様性の解明に向けて、第 74 回日本癌学会学術総会、名古屋、2015 年 10 月 10 日
6. 鈴木穰、がん細胞のマルチオミクス計測、CREST シンポジウム「トランスオミクスによる生命システムの解明」、東京、2015 年 3 月 4 日
7. 鈴木穰、シングルセル多層オミクス解析によるがん細胞の多様性の分析、第 20 回日本がん分子標的治療学会学術集会、別府、2016 年 5 月 31 日

〈国際〉

1. Yutaka Suzuki, “Nanopore sequencing for genotyping pathogens of tropical diseases”. London Calling 2015. London, UK, 15-16 May 2015.
2. Yutaka Suzuki, “MinION sequencing of malaria parasites”. London Calling 2016. London, UK, 26-27 May 2016.
3. Yutaka Suzuki, “Nanopore sequencing for genotyping pathogens of tropical diseases”. the 8th Asian Network of Research Resource Centers (ANRRC) International Meeting. Kyoto, Japan, 22 September 2016.
4. Yutaka Suzuki, “Nanopore sequencing for genotyping pathogens of tropical diseases”. Joint International Tropical Medicine Meeting 2016. Bangkok, Thailand, 7-9 December 2016.

(3) 口頭講演

〈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
3. 鈴木穰、シングルセル多層オミクス解析によるがん細胞の多様性の分析、日本がん分子標的治療学会学術集会、別府、2016 年 5 月 31 日
4. 鈴木穰、進展するゲノム解析技術の育種への応用、日本育種学会、鳥取、2016 年 9 月 24 日

〈国際〉

1. Yutaka Suzuki. “Multi-omics analysis of cancer cells”. The 11th International Workshop on Advanced Genomics. Tokyo, Japan, 20-22 May 2015.

(4) ポスター発表

〈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.
3. 鈴木絢子、鈴木穰、土原一哉、MinION を用いた肺腺癌細胞における変異及び融合遺伝子の検出と解析、日本分子生物学会・日本生化学会(BMB2015)、横浜、2015年12月3日

〈国際〉

1. Yutaka Suzuki “Aberration of Transcriptional Regulations in Cancer Cells”, Cold Spring Harbor Meeting, NY, USA, 6-10 May 2014.
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. 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.
4. Yutaka Suzuki, Arthur E Mongan, Josef Tuda, Junya Yamagishi, “Nanopore sequencing for genotyping pathogens of tropical diseases”, 5-9 May 2016.
5. Yutaka Suzuki, “Nanopore sequencing for genotyping pathogens of tropical diseases”. Cold Spring Harbor Meeting, NY, USA, 10-14 May 2016.

4. 知財出願

なし

5. 受賞・報道等

(1) 受賞

なし

(2) メディア報道

なし

(3) その他

なし

§9. 研究開発期間中の活動

1. 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
平成 26 年 2 月 14 日	チーム内打ち合わせ(非公開)	国立がんセンター東病院	5 人	キックオフミーティング準備のための打ち合わせ
平成 26 年 4 月 30 日	チーム内打ち合わせ(非公開)	東京大学フューチャーセンター	5 人	同上
平成 26 年 5 月 1 日	チーム内打ち合わせ(非公開)	東北大学	5 人	同上
平成 26 年 5 月 15 日	チーム内打ち合わせ(非公開)	国立がんセンター東病院	5 人	同上
平成 26 年 6 月 2 日	キックオフミーティング(非公開)	JST 東京本部	15 人	本研究開発費で行う研究の役割分担の調整
平成 26 年 9 月 29 日	チーム内打ち合わせ(非公開)	JST 東京本部	5 人	進捗状況確認のための打ち合わせ
平成 27 年 3 月 23 日	チーム内打ち合わせ・菅野班ミーティング(非公開)	国立がんセンター東病院	8 人	同上
平成 27 年 7 月 10 日	医療情報セキュリティープライバシーセミナー(非公開)	東大フューチャーセンター	10 人	医療情報を匿名化して扱う方法をいくつかの企業に提案説明してもらうためのセミナー
平成 27 年 8 月 3 日	医療情報セキュリティープライバシーセミナー(非公開)	JST 東京本部	30 人	同上
平成 28 年 1 月 8 日	チーム内作業打ち合わせ(非公開)	東京大学フューチャーセンター	3 人	RDF 化、Beacon 化作業の打ち合わせ
平成 28 年 3 月 10 日	実務者連絡会(非公開)	JST 東京本部	15 人	進捗状況確認と今後の方針決定のための打ち合わせ
平成 28 年 6 月 7 日	チーム内作業打ち合わせ(非公開)	東京大学フューチャーセンター	3 人	RDF スキーマ仕様等の打ち合わせ
平成 28 年 8 月 1 日	チーム内打ち合わせ(非公開)	東京大学フューチャーセンター	5 人	進捗状況確認のための打ち合わせ
平成 28 年 9 月 12 日	情報交換会(非公開)	JST 東京本部	15 人	進捗状況確認と今後の方針決定のための打ち合わせ
平成 28 年 9 月 15 日	チーム内打ち合わせ(非公開)	東京大学フューチャーセンター	3 人	進捗状況確認のための打ち合わせ
平成 28 年 10 月 28 日	チーム内打ち合わせ(非公開)	東京大学フューチャーセンター	2 人	同上
平成 28 年	チーム内打ち合わせ(非公開)	東京大学フ	4 人	同上

11月7日	公開)	ユーチャー センター		
-------	-----	---------------	--	--

2. 主催したワークショップ、シンポジウム、アウトリーチ活動等

年月日	名称	場所	参加人数	目的・概要
平成 26 年 11 月 25 日 -27 日	日本分子生物学会年会 特別企画「使ってみよう バイオデータベースーつ ながるデータ、広がる世 界 (BioDB)」	パシフィコ 横浜	5 人	目的:データベースの一般へ の利用と普及の促進 概要:DB の紹介と、主に今後 の開発展望の発表
平成 27 年 12 月 1 日- 4 日	日本分子生物学会・日 本生化学会合同大会 「使ってみようバイオデー タベースーつながるデー タ、広がる世界 (BioDB)」	神戸ポート アイランド	5 人	目的:データベースの一般へ の利用と普及の促進 概要:DB の紹介、主にブラウ ザ機能・パスウェイ検索機能
平成 28 年 11 月 30 日- 12 月 2 日	日本分子生物学会年会 特別企画「使ってみよう バイオデータベースーつ ながるデータ、広がる世 界 (BioDB)」	パシフィコ 横浜	5 人	目的:データベースの一般へ の利用と普及の促進 概要:DB の紹介、主に前年 の内容に加え、RDF 化の進 捗

以上

別紙1 研究開発対象のデータベース等

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
1	DBTSS		ヒト、マウス、その他の真核生物の正確な転写開始点を収録したデータベースです。mRNA 5'端に存在するキャップ構造を特異的に選別するオリゴキャッピング法を用いて取得した最新の完全長cDNA配列と、その配列から明らかになった転写開始点およびプロモーター領域の情報を公開しています。ゲノムビューワ上に、オリゴキャッピング法で同定されたcDNA領域、RefSeqやEnsemblで公開されている転写開始点、CpGアイランド、SNP位置が表示されます。	http://dbtss.hgc.jp/	2014年9月14日	継続・発展	データベース等	対象外	調整中	調整中	論文リスト1,2,3
2	KERO		本データベースはヒトゲノム多型・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したものである。特に、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ている。	http://kero.hgc.jp/	2015年7月9日	新規	データベース等	対象外	調整中	調整中	論文リスト1,2,3