

(提案8)

(案)

提 言

# 100万人ゲノムコホート研究の実施に向けて



平成25年(2013年) ○月 ○日

日 本 学 術 会 議

第二部

ゲノムコホート研究体制検討分科会

この提言は、日本学術会議第二部ゲノムコホート研究体制検討分科会ヒト生命情報統合研究推進小委員会での審議結果を、第二部ゲノムコホート研究体制検討分科会において取りまとめ公表するものである。

### 日本学術会議第二部ゲノムコホート研究体制検討分科会

委員長	浅島 誠	(連携会員)	独立行政法人日本学術振興会理事
副委員長	本庶 佑	(連携会員)	京都大学大学院医学研究科特任教授
幹事	小原 雄治	(第二部会員)	情報・システム研究機構国立遺伝学研究所特任教授
幹事	菅野 純夫	(連携会員)	東京大学大学院新領域創成科学研究科教授
	斎藤 成也	(第二部会員)	情報・システム研究機構国立遺伝学研究所集団遺伝研究部門教授
	山本 雅之	(第二部会員)	東北大学大学院医学系研究科教授
	山本 正幸	(第二部会員)	公益財団法人かずさ DNA 研究所所長
	巖佐 庸	(連携会員)	九州大学大学院理学研究院教授
	春日 雅人	(連携会員)	独立行政法人国立国際医療研究センター総長
	高木 利久	(連携会員)	東京大学大学院新領域創成科学研究科教授
	辻 省次	(連携会員)	東京大学大学院医学系研究科脳神経医学専攻神経内科学教授
	大江 和彦	(特任連携会員)	東京大学大学院医学系研究科社会医学専攻医療情報経済学分野教授

### ヒト生命情報統合研究推進小委員会

委員長	浅島 誠	(連携会員)	独立行政法人日本学術振興会理事
副委員長	本庶 佑	(連携会員)	京都大学大学院医学研究科特任教授
幹事	小原 雄治	(第二部会員)	情報・システム研究機構国立遺伝学研究所特任教授
幹事	菅野 純夫	(連携会員)	東京大学大学院新領域創成科学研究科教授
	春日 雅人	(第二部会員)	独立行政法人国立国際医療研究センター 総長
	山本 雅之	(第二部会員)	東北大学大学院医学系研究科教授
	喜連川 優	(第三部会員)	情報・システム研究機構国立情報学研究所所長、東京大学生産技術研究所教授
	高木 利久	(連携会員)	東京大学大学院新領域創成科学研究科教授
	田中 耕一	(連携会員)	株式会社島津製作所田中耕一記念質量分析研究所所長
	辻 省次	(連携会員)	東京大学大学院医学系研究科脳神経医学専攻神経内科学教授

橋本 信夫	(連携会員)	独立行政法人国立循環器病研究センター理事長・総長
福嶋 義光	(連携会員)	信州大学副学長・医学部長、医学部遺伝医学・予防医学講座教授
一圓 剛		ヒュービットジェノミクス株式会社社長
大江 和彦		東京大学大学院医学系研究科社会医学専攻医療情報経済学分野教授
川本 俊弘		産業医科大学医学部教授
清原 裕		九州大学大学院医学研究院基礎医学部門社会環境医学講座環境医学分野教授
田中 英夫		愛知県がんセンター研究所疫学・予防部部長
野木森 雅郁		アステラス製薬株式会社社長
松田 文彦		京都大学大学院医学研究科附属ゲノムセンターセンター長

報告書および参考資料の作成にあたり、以下の方々に御協力いただきました。

佐藤 孝明	株式会社島津製作所フェロー
松尾 恵太郎	九州大学大学院医学研究院基礎医学部門社会環境医学講座予防医学分野教授
箕輪 真理	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准教授
岡本 忍	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准教授
川本 祥子	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准教授

本件の作成にあたっては、以下の職員が事務を担当した。

事務	中澤 貴生	参事官(審議第一担当)
	伊澤 誠資	参事官(審議第一担当)付参事官補佐
	藤本紀代美	参事官(審議第一担当)付審議専門職

## 要 旨

### 1 作成の背景

2012年8月のゲノムコホート研究体制検討分科会提言「ヒト生命情報統合研究の拠点構築」において、生物としてのヒトを総合的に理解し、単に病気だけでなく、人間の持つ健常形質の多様性や老化にも目をむけた研究を推進するために、100万人規模のゲノムコホートを構築し、多様な生命情報を蓄積し、多次元かつ膨大な情報を最新の情報科学を用いて統合解析する「ヒト生命情報統合研究」推進の必要性を提言した。ゲノムコホート研究体制検討分科会ではその提言をさらに発展させるため、分科会のもとに小委員会を設置し、研究推進の具体的な方策を検討してきた。本提言はそれを取りまとめたものである。

我が国における「ヒト生命情報統合研究」として、40歳以上の健常者100万人規模のコホートを構築し、疾患罹患情報などの医療情報や環境・生活習慣情報、ゲノム情報に加え、生体試料バンクの構築と生体試料を用いたオミックス解析（生体分子の網羅的解析）、膨大な情報を効率的に扱う計算科学、多様な情報や異種のデータを統合し表現型との関連を統計学的に解析する新たな生命情報解析理論の構築などを行うことを提案する。そのためには医学、生物学の研究者のみならず、理学、薬学、工学、情報科学の研究者を動員し、さまざまな技術革新を行い、予防医学の基盤を構築し国民の健康に資するとともに、生命科学の発展にも寄与する研究と技術開発を行う体制をつくる必要がある。

### 2 提言の内容

#### (1) 実施体制

「ヒト生命情報統合研究」の実施体制は、中核拠点、地域拠点、分析・解析拠点、生体試料バンクからなる。

中核拠点は、統一プロトコルの作成、組織構築、制度設計を行い、既存のコホートを用いたパイロット研究を実施する。その上で、地域拠点の数と場所や規模を地域バランスと実績を考慮して決定し、統一プロトコルに則った事業への参画が可能なことを条件に全国で公募する。そして、事業の3年目をめどに10万人規模のコホート研究を進める地域拠点を全国で10カ所程度構築する。事業開始後は、中核拠点に設置される事業運営委員会で事業全体の情報の集約と方針決定を含む企画運営を行う。地域拠点は、中核拠点と緊密な連携をとり本事業を実施する。そのために、住民コホート・職域コホートからなる10万人程度のコホートを構築し、地域での健康づくりの土壌を醸成しつつ事業に協力する地域住民（事業参加者）の健診を実施し、生体試料・追跡データの取得などを行う。大規模研究への参画に加えて、各拠点での独自の研究を実施し成果を公表するとともに、健診データを利用した住民の健康づくり啓発に努める。そのため、得られた情報・生体試料を分析・解析拠点、生体試料バンクに集積すると同時に、地域拠点でも保管し、地域拠点における独自

の研究開発に活用する。また、それぞれに運営委員会、倫理審査委員会、広報委員会を置き、事業の運営、倫理審査、人材の育成、広報についても、中核拠点と緊密に連携しつつ、地域特性や事業参加者の意識を尊重した活動を実施する。

分析・解析拠点は、統合オミックス解析センターとインフォマティクスセンターから構成され、得られた生体試料の分析・解析を実施するとともに、情報統合データベースの構築により、臨床情報や疾患罹患情報の蓄積を行う。また、それらの多様な情報と疾患との関連を解析する。

生体試料バンクは、生体試料の集約と一元管理および、収集された生体試料のデータベース登録と品質管理を主要な業務とする。また、生体試料の分散管理と事業継続計画の策定や、生体試料の再配分とそれに関わる情報の管理も行う。

## (2) 産学官連携

製薬企業のみならず、診断薬企業、医療デバイス企業、食品関連産業、IT企業などのさまざまな企業が本事業に積極的に参画し、企業の視点からデータ解析や生体試料解析を行い、その成果の産業化を目指すことは現在の社会に成果を還元する上では重要であると考えられる。そのためには、産業界との連携を促進するためのインフォームド・コンセント（説明と同意）などの工夫が必須であり、個人情報保護法、薬事法の面からも医療情報並びに生体試料の利活用に関する倫理問題などの課題が解決される必要もある。産業界の意識調査を十分に行った上で産学官民にわたるコンソーシアムを結成し、必要な運営規則などを定め、運営体制を整える。また、事業の成功、成果の産業化にあたっては各省庁間の連携が重要である。

## (3) 倫理的配慮・社会との接点

ゲノムコホート研究は、試料を提供する事業参加者の利益のためというよりも、国民全体のよりよい医療の実現に必要なものであり、提供を受けた生体試料やデータの具体的な利用方法や予想される研究成果を試料採取時に明確に示すことは困難であることが多い。したがって、提供試料の使用目的を限定せず、また将来の健康情報の提供を了解する「包括同意」の取得が不可欠である。そのためには、社会の合意形成、倫理審査の在り方、個人情報の保護などの倫理的課題について十分に検討し、明確な倫理指針を打ち出す必要がある。

## (4) 人材育成

本事業には、さまざまなレベルの人材が多数必要であるが、中でも、遺伝医学・ゲノム学教育者、ゲノム医学コミュニケーター、ゲノム医学リサーチコーディネーター、ゲノム医学カウンセラー、匿名化処理従事者、電子カルテ管理者、生命情報学者（バイオインフォマティシャン）、生物統計学者、統計遺伝学者などが決定的に重要である。これらの人材は、広く、医学・生物学研究、医療の現場やさまざまな

企業でも重要な役割を果たすと考えられている。

## 目 次

1	はじめに	1
2	ゲノムコホートの戦略的ミッション	2
(1)	グランドデザイン	2
(2)	実施体制	2
(3)	ゲノムコホートにおける生体試料バンクの役割	3
(4)	データベース構築とデータの公開・共有	3
(5)	既存コホートとの連携可能性と対策	4
(6)	人材育成	4
(7)	生命科学への展開	5
3	運営組織	6
(1)	組織構成	6
(2)	中核拠点	6
①	事業運営委員会	7
②	事業実施連絡会議	9
③	中央倫理審査委員会	9
④	データ公開・共有管理委員会	10
⑤	広報委員会	10
⑥	外部評価委員会（事業外独立組織）	10
(3)	地域拠点	10
①	事業参加者のリクルート	10
②	健診データの収集と登録	11
③	生体試料の蓄積	11
④	地域拠点における解析研究	11
(4)	解析・分析拠点	11
①	統合オミックス解析センター	12
②	インフォマティクスセンター	12
(5)	生体試料バンク	12
4	事業参加者	14
(1)	リクルートの方法	14
(2)	追跡方法	14
5	分析・解析方法の開発と標準化	15
(1)	ゲノム解析	15
(2)	オミックス解析	15
(3)	測定・分析方法の標準化	15
(4)	データ取得・変換方法の開発	16
6	情報統合データベース	17
(1)	今後開発が必要なデータベース	17

(2)	データベースとその統合化に必要な技術	17
(3)	情報解析技術	18
(4)	計算機環境	19
7	産学官連携	20
(1)	ヒト生命情報統合研究に関する可能性と期待	20
(2)	製薬産業と『ヒト生命情報統合研究』	20
(3)	バイオマーカーの発見と診断薬企業	20
(4)	医療デバイス企業	21
(5)	食品関連産業（食品、健康食品企業）	21
(6)	IT企業	21
(7)	産学連携体制の必要性和各省の役割分担	21
8	倫理的配慮・社会との接点	23
(1)	倫理的課題とその対策	23
(2)	倫理審査	23
(3)	インフォームド・コンセントと個人情報保護	24
(4)	社会的な合意形成	24
9	人材育成	25
(1)	遺伝医学	25
①	遺伝医学・ゲノム学教育者	25
②	ゲノム医学コミュニケーター	25
③	ゲノム医学リサーチコーディネーター	25
④	ゲノム医学カウンセラー	25
(2)	医療情報整備	26
(3)	情報処理・解析	26
①	バイオインフォマティシャン	26
②	データマネージャー	26
<付録1>		27
<付録2>		28
<付録3>		33
<参考資料1>	審議経過	35



## 1 はじめに

危機的な少子高齢化時代を迎える我が国にとって、病気の超早期発見と発症前の治療的介入による予防法の確立は、活力のある健康長寿社会を構築するために不可欠である。このためには、大きな集団の長期観察によって個人の健康に関わる多様な情報と生体試料を蓄積し、最先端の測定・分析技術と統計学、計算科学を駆使したヒト生命情報の統合解析を行う、大規模ゲノムコホート研究が必要である。対象集団の規模は、現在の我が国の疾患発症率をもとに多くの重要な疾患の病因に迫ることが可能な100万人に設定した。その成果は、疾患の原因解明と予防・治療法の開発を通じた、世界に一步先じた高齢化社会の健康長寿モデルの構築につながる。加えて、人間の持つ健常形質の多様性の解明や、生物が共有する基本的生命現象を発見する大きな可能性を秘めた究極の「ヒト生物学」として、生命科学分野に多大な貢献が期待される。また、研究に必要なさまざまな先端解析技術の開発と実用化・汎用化は、我が国の科学技術全般や産業界に大きな技術革新をもたらす。超高齢化社会の我が国にとって、予防に関する情報を用いた新たな健康産業の創出や、保健医療情報の電子化による新時代の保健医療システムの構築も極めて重要である。以上のように、大規模ゲノムコホート研究は、今後の我が国の科学技術政策や保健・医療政策の方向性の決定に大きな影響を与えると断言しても過言ではない。したがって、その実施において国のとるべき姿勢を明確にし、組織構築、制度設計および解析技術開発など事業に関わる課題を明らかにした上で、国家事業として強力に推進することが強く望まれる。

## 2 ゲノムコホートの戦略的ミッション

### (1) グランドデザイン

ヒトはその遺伝的背景、環境、生活習慣において極めて多様な集団であり、同一の遺伝的背景や環境で行われるモデル動物による研究の成果をそのままヒトにあてはめることは極めて難しい。生物としてのヒトを総合的に理解し、単に病気だけでなく、人間の持つ健常形質の多様性や老化にも目をむけた研究を推進するためには、地域ベースで大規模な健常者集団を長期にわたって追跡するゲノムコホートを構築し、多様な生命情報を蓄積することが不可欠である。そのような研究では、新しい測定・分析技術による生体分析情報、すなわち個人のゲノム情報に加え血液や尿中のタンパク質、代謝物などの生化学的データ、細胞生物学的データ、生理学的データ、MRI や X 線 CT などによる画像診断データなどのバイオマーカーの系統的収集が不可欠である。そういった多次元かつ膨大な情報を環境・生活習慣情報や疾患罹患情報などの医療情報と統合し、最新の情報科学を用いて解析する「ヒト生命情報統合研究」を実施することによってはじめて、生物学的事実にもとづいたヒトの健康や病気の総合的理解が可能となる。また得られた結果をさまざまな疾患の患者集団を対象とする疾患コホートと比較解析することで、病気の発症前に治療介入する「先制医療」のための情報基盤が構築され、健康長寿社会の実現に向けた新たな予防医学の提唱が可能となる。さらには、生物が共有する基本的生命現象を発見する大きな可能性も秘めている。

このような研究を実施するためには医学、生物学の研究者のみならず、理学、薬学、工学、情報科学の研究者を総動員し、新しい計測機器による微量分析の技術開発、膨大な情報を効率的に扱う計算科学、多様な情報や異種のデータを統合し表現型との関連を統計学的に解析する新たな生命情報解析理論の構築など、多くの課題を克服する必要があるが、逆に本事業の推進によって、こういった領域におけるさまざまな技術革新が飛躍的に進むことが強く期待される。

### (2) 実施体制

100 万人規模の生体試料バンクの構築と生体試料、臨床情報、疾患罹患情報を統合したヒト疾患解析による多因子疾患の原因解明と予防・治療法の開発を目標に、中核拠点を軸とした組織構築、制度設計を行い、既存のコホートを用いたパイロット研究を経て、コホート事業の開始に向けた解析基盤、情報基盤を整備する。そして、事業の3年目をめどに、全国で10カ所程度の実施拠点を構築し、全国体制下の事業組織による統一基準でのヒト生命情報統合解析の実施基盤を構築する。その基盤は、中核拠点、分析・解析拠点、地域拠点、生体試料バンクからなる。また、事業の成果を産業へ結びつけるための仕組みとして、産学連携コンソーシアムを形成する必要がある。さらに、このような網羅的解析に基盤を置く研究では、新しい解析方法を積極的に取り込んでいく必要があるため、研究から得られる成果とその公共性を鑑みて、一定条件を満たす研究については事業に協力する地域住民（事業参加者）の包括同意に基づく研究遂行を可能とするさまざまな制度面の整備を行う必要がある。また、医療情報

を遺漏無く定期的に把握していくためには、電子化された医療情報の活用が欠かせない。既存の電子化された医療情報を最大限活用するとともに、医療マイナンバーの導入や医療情報の相互利用を、国策として早期に実現するための働きかけが必要であろう。

本事業実施の所要経費は、準備期間（3年間）、ゲノムコホート実施期間（13年間）を合わせて約1000億円と見込まれる。そのうちゲノムコホート本体事業費（付録1参照）以外の産学連携や新規技術開発に必要な約300億円の経費は、民間企業より集める等の工夫が必要である。

### (3) ゲノムコホートにおける生体試料バンクの役割

「生体試料バンク」とは、「人体に由来する試料およびそれに関する情報を、医学と科学の研究に利用するために、体系的に収集・保管・分配するシステム」であると定義される。生体試料バンクの構築は、国民の健康・福祉の向上、生命科学研究の発展、経済における国際競争力の維持などにおいて必要不可欠である。長期間の追跡による大規模ゲノムコホートと生体試料バンクを組み合わせ、大規模な試料と情報の集積を実施することが、疾患発生メカニズムの解明や、効果的な予防法や治療法の開発に重要であることは、国際的に広く認識されている。

大規模生体試料バンクの整備と公正な運営は以下の4つの意義を持つ。

- 1) 研究者が個別に生体試料を収集し、個別に研究を実施すると、労力やコストの限界から研究規模が限定され、また種々のバイアスを回避できないために十分な精度をもって試料や臨床情報を収集できないなどの理由から、信頼性の高い結果を導き出し難くなる。結果的に貴重な研究費が浪費されることになるが、統一した基準で収集した大規模生体試料と臨床情報は研究の質を担保する。
- 2) 試料確保に割かれている研究者の多大な労力を真の研究活動に充てることができ、研究活動の効率化が図れる。
- 3) 一定の基準の下、生体試料バンクで収集した生体試料や臨床情報を他の研究者に適正に配分することで、我が国全体の研究水準が向上する。
- 4) 集積された生体試料を厳重な品質管理の下に長期保管することで、将来の生命科学研究における分析・解析にとって貴重な研究資源を供給できる。

### (4) データベース構築とデータの公開・共有

情報処理とデータベースの開発・運用においては、中核拠点、地域拠点のそれぞれの役割分担と連携が重要課題である。中核拠点は、試料採取、データ取得や解析方法などに関する統一プロトコルを整備し示すとともに、各地域拠点とのデータ利用の相互連携を図ることにより、100万人規模の生体試料バンクやデータベースを構築し本研究事業に参加する研究者にそれらを提供する。一方、地域拠点は、統一プロトコルに沿って情報および生体試料を収集するとともに、それぞれの地域の特色を生かし、自立的に独自の創意工夫や研究開発を実施する。

情報処理とデータベースの構築においては、中核拠点と地域拠点の連携によるデータの蓄積や集約を行うための技術面、データ共有の面からの支援を念頭に技術開発を進めることが求められる。本研究により得られた生体試料の分析・解析結果はインフォマテイクスセンターのデータベースで一元管理し、地域拠点からの柔軟性の高い検索とアクセスを可能にするための技術開発と運用を実施する。地域拠点では、分析・解析拠点により開発された標準化のための統制語彙やプロトコル、情報統合データベースを利用して臨床情報や疾患罹患情報を蓄積することにより、分析・解析拠点へのデータ統合の円滑化を図る。

大規模な国の予算的支援を得て実施されるゲノムコホート研究で得られる研究資源は国民の共有財産であるため、本研究から得られる臨床情報、生体分析情報、環境・生活習慣情報などを速やかに研究者コミュニティに提供し、最大限の利活用を促すことで我が国の科学技術の発展に寄与しなければならない。また、本研究で得られるデータの提供については、集団の解析結果のデータのように公開による提供が可能なデータ（データ公開）と、制限付きアクセスにより外部研究者に提供するデータ（データ共有）に分類して対応することが適切である。また、研究成果は、我が国固有のヒト生命情報として科学技術振興機構（JST）バイオサイエンスデータベースセンター（NBDC）を通して公開し、内外の研究者の利活用を促す。

#### (5) 既存コホートとの連携可能性と対策

本事業は、実施期間中に集積される情報量、測定・分析手法、データ解析手法などにおいて、従来の疫学コホート事業を遥かに凌駕する全く新しい挑戦的な試みである。その一方で、この事業の成功には地域住民の積極的参加や自治体の協力が不可欠であることから、全国ですでに実施されている疫学コホートのノウハウや自治体単位で整備された病院間ネットワークなどの基盤を最大限に活用することが重要である。そこで、既存のコホートとは以下のように連携し、ネットワークを構築する。

- 1) 本事業は、中核拠点において実施計画を十分に議論した後、先行している国際チームの経験を参考にしつつ事業参加者登録システム、追跡方式、生体試料保存方法、データの整理保存方式などを含めて統一プロトコルを策定し、そのプロトコルに沿った形で実施する。
- 2) 統一プロトコル策定と並行して、全国に配置する地域拠点の数と場所や規模を地域バランスと実績を考慮して決定し、統一プロトコルに則った事業への参画が可能なことを条件に既存のコホート事業も含めて全国で公募する。

#### (6) 人材育成

遺伝子解析技術の進歩がヘルスケアの改善に資することが期待される中、そのような「新規」の健康・医療情報を効率的かつ安全に事業参加者に還元できるよう、必要な人材をいかに普及・配置させるかが、喫緊の課題となっている。

一般住民を対象としたゲノムコホート研究においては、大きく、ゲノム教育・ア

ウトリーチ、健康・医療情報の整備、大規模かつ継続的な情報処理・解析などの活動に関わる遺伝医学・ゲノム学教育者（臨床遺伝専門医）、ゲノム医学コミュニケーター、ゲノム医学リサーチコーディネーター、ゲノム医学カウンセラー、生命倫理学者、生命情報学（バイオインフォマティクス）専門家、ゲノム解析専門家などの人材の育成が必要である。

## (7) 生命科学への展開

これまでの生命科学の革新的な技術の発展により、我々の生命科学の情報は飛躍的に増大した。とりわけ、ヒトゲノム解読により生物種間の差異などが明確となり、ヒトとしての生命機能の解明が極めて重要となってきた。従来の生命科学においては、遺伝学的な改変などにより、マウスをはじめとするモデル生物における病態解明および生理機能の解明が飛躍的に進んだ。一方でヒトにおける生命科学情報の統合的な集積は、臨床家による病態関連情報の散発的な集積に頼ってきたため、極めて不十分である。したがって、本研究を通じて遺伝的背景と環境因子との相互関係がもたらす疾患原因並びに病態を解明することで、ヒトの生命現象の解明に対する大きな貢献が期待される。

ゲノムコホート研究は、バイオインフォマティクス、数理科学、構造生物学、分子生物学、進化学、発生生物学、系統分類学、計算機学、統計学、人文社会科学、病理学、社会医療学、薬学、環境学などの多くの学問分野と連携し、基礎生物学からの知識に基づき、ヒトの生命現象を解析することとなる。そしてその成果は、遺伝病やさまざまな遺伝形質の解明にとどまらず、生物の高次機能や環境への適応と進化といった生命科学において極めて重要な課題への展開が期待される。

### 3 運営組織

#### (1) 組織構成

多種多様で膨大なデータの品質管理や情報解析の効率性の観点からは、1カ所の拠点で100万人規模のデータを一元的、統一的に取得し、それを一括して解析することが理想である。しかしながら、現在我が国にはその受け皿となれるような組織や人材がないこと、1ヶ所で100万人の事業参加者のリクルートは困難であることなどを考えると、すでにいくつかの地域で実施されているコホート研究を生かしつつ、それらも含め複数の地域拠点を設置し、事業参加者のリクルートを全国規模で行うことが必要であろう。また、各拠点の規模と数は、おおよそ10万人の事業参加者からなる地域拠点が地域性も考慮して全国に10カ所程度あることが、効率性やデータの品質管理の点から望ましい。その際、各地域拠点が全く独立してコホート事業を実施するのであれば、単に10万人規模のコホートが乱立するだけになってしまう。そこで、中核拠点を設けて試料の採取やデータの取得や解析方法などに関する統一プロトコルを策定し、また、データの共有の仕組みを設けて事業全体を統括する。中核拠点、地域拠点、分析・解析拠点、生体試料バンクの連携を図ることにより、100万人規模のゲノムコホート研究が実現できる。

しかしながら、分析・解析拠点にすべてのデータや試料の解析機能を過度に集中させると、地域拠点は事業参加者のリクルートと一次データの産出のみとなり、インセンティブや活力が失われてしまう。地域拠点の活力や特色を生かしつつ既存のコホート事業もうまくこの枠組みに取り入れていくためには、地域拠点が自立的にイニシアティブをとって事業を進めることを可能としなければならない。分析・解析拠点と地域拠点のバランスの良い事業計画が求められる所以である。

中核拠点は、データ取得のためのガイドラインやデータの共有の在り方を定め、分析・解析拠点では、標準化された方法で生体試料の測定・分析とデータ解析を実施するが、各地域拠点はそれに従いつつ、一方でそこで定められていること以外の具体的な決まり事を決め、独自の創意工夫や独自の研究開発を行う。このような「連邦型」の拠点形成を実現するためには中核拠点、地域拠点、分析・解析拠点、生体試料バンクのそれぞれが以下のような機能を持つことが求められる。

#### (2) 中核拠点

事業全体の司令塔として、事業の統括運営を行う。計測項目や基準の統一と倫理規約の統一は最も重要な役割である。主な業務として、事業組織の構築、制度設計と事業計画の立案を行う。環境省が実施した子供の健康と環境に関する全国調査（エコチル調査）におけるコアセンターは、研究実施の中心機関として調査の総括的な管理・運営を行っており、各地域のユニットセンターの管理・支援、リスク管理、広報・コミュニケーション活動などを行う調査全体の中央事務局として必要な役割を担っている。中核拠点の設立においては、こういった前例を参考にしつつ、中央倫理審査委員会の運営や社会の合意形成を目的とした活動を行うのが適切であると考えられる。

事業の全体計画に基づき、本事業の意思決定機関である事業運営委員会を組織し、事業計画の立案、統一プロトコル策定および地域拠点の選定と事業実施規模の確定を行う。運営委員会の定期的開催で各地域拠点での事業の進捗の把握に努め、情報の集約と方針決定を行う。事業の中央倫理審査委員会を立ち上げ、事業および付随する研究の倫理審査を実施する。また、実施組織外に外部評価委員会を設置し、第三者による評価を行う。ゲノムコホート研究は長期の研究活動になることから、本研究で得られたデータ集積や生体試料の保管については、中核拠点において継続性を保証できる責任体制を構築して運営・管理することが重要である。同時に、国の財政支援としても継続性のある活動を維持する方策を明確にすることが重要である。

なお、データベースの構築や情報解析に必要な技術開発については、中核拠点の直下に置かれる分析・解析拠点において、実施されるものとする。

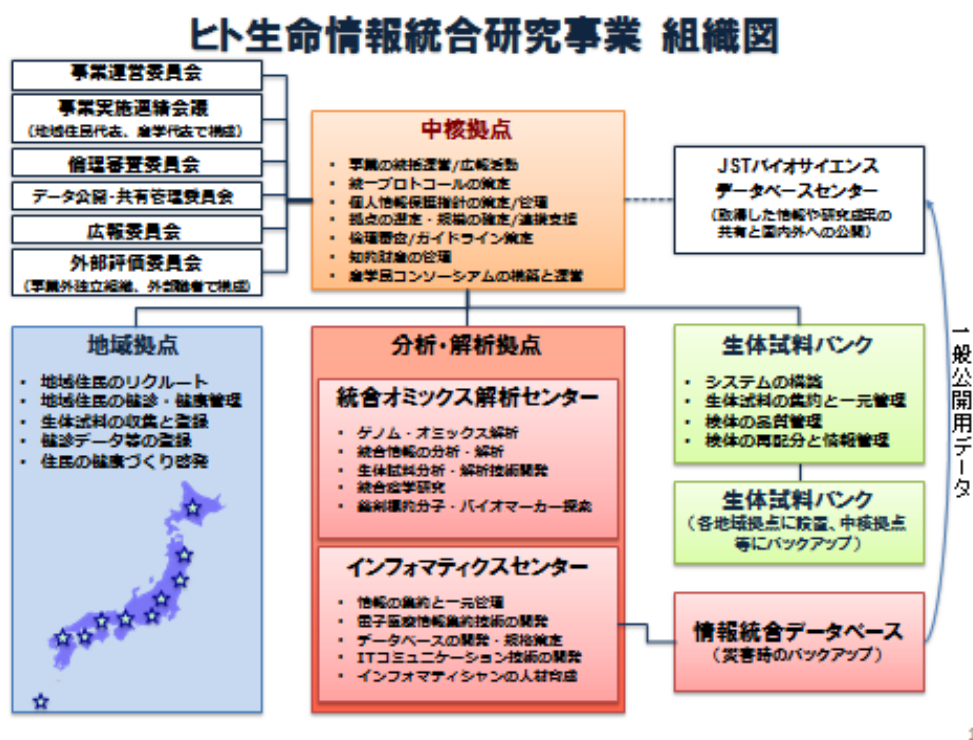


図1 ヒト生命情報統合研究事業組織図

以下に、中核拠点の業務を列挙する。

#### ① 事業運営委員会

事業計画の立案、統一プロトコル策定および地域拠点の選定と事業実施規模の確定を行う。生体試料バンクの運営と試料の管理、分析・解析拠点との連携によるデータ分析・解析と情報統合データベースの構築、国内外の関連研究機関（NBDCや海外のバイオバンク事業など）との間の連携を統括する。また、本事業全体に必要な費用について予算を立案し、地域拠点の運営費、分析・解析拠点、生体試料バンクなどの費用も含め、中核拠点から全体の予算を申請するものとする。この委員会の主たる業務を以下に列挙する。

## ア 統一プロトコルの策定

地域拠点で事業参加者の登録、生体試料採取、データの取得を円滑に実施するために統一プロトコルを策定し、生体試料の採取と処理、データの取得方法、データの標準化について統一規格を提示する

プロトコルの中には、どのような生体試料やデータをどのようなスケジュールで採取するのかといった研究計画、生体試料の扱いやデータの品質管理や標準的な解析の標準業務手順の策定を含む。具体的には、ベースライン調査方法や疾患罹患調査方法に加え、ゲノム解析に関して標準化を図る。加えて新しい物理化学的計測技術による解析データ、すなわち血液の代謝物といった生化学的データ、細胞生物学的データ、生理学的データ、MRI や X 線 CT などによる画像診断データ、さらには心理学や行動学的データに関しても標準化が必須である。また、採取する生体試料の量に関しては、新技術の開発にともない将来的に測定すべき新たな項目も多く存在すると考えられるため、十分量の生体試料（血液、尿など）を確保する必要がある。したがって、国内外の既存の事業での実績を参考としつつも、将来試料の枯渇が生じないように採取量をゼロベースで検討する。

## イ 個人情報保護指針の策定・個人情報の管理

事業の円滑な実施には、事業参加者の個人情報の厳格な保護に加え、研究活動の透明性、説明責任などの倫理的妥当性を担保した包括同意が必須である。そこで新ゲノム指針をふまえつつ、本事業により適合する形の倫理規定を策定し、事業実施主体と事業参加者個人との契約に基づく登録制度を設計する。また、個人情報保護のための匿名化の方法も検討し、統一基準を策定する。

種々のデータの共有のための条件となるインフォームドコンセント（IC）の内容を地域拠点の間で統一するために、雛形を作成する。ゲノムコホートは前向き研究であるため、特定の疾患などとの関連は予測できない上、将来全く新しい分析技術が開発される可能性も高いため、個人情報保護を最優先しつつあらゆる研究に生体試料や情報が利用できる「包括同意」が必要である。

長期にわたる本事業の期間中の任意の時期に、事業参加者が自己の意思で受診する医療機関や健診機関での健康医療情報を電子的にかつ標準化された形式で網羅的に収集する必要がある。そのためには各個人に固有の ID が必要となるが、現在検討中の医療マイナンバーが導入される場合はこれを利用するなどの検討も必要である。また、同一個人ごとに統合できる情報基盤を整備する必要がある。一方、匿名化処理との関係を整理し、試料・情報の管理方法を検討する必要もある。

## ウ 地域拠点の選定・規模の確定



既存のゲノムコホート事業の例などを参考に、地理的立地や人の移動などの条件を検討し、周辺の住民も併せて40代以上の人口が十分に確保でき、かつ高い割合で追跡が可能と考えられる地域に地域拠点を設定する。その際、事業の規模から考えて、その地域の中核となる研究機関や医療施設を地域拠点の中心施設とする。このような地域拠点を全国に10か所程度配置し、100万人の事業参加者を募る。また地域拠点への支援として、測定・分析に関する技術支援や人材の供給を行う。

## エ 知的財産の管理

本事業から産生される知的財産を一元的に管理し、外部への許諾などの手続きを実施する。許諾に関するルールについては、一定の基準のもとで手続きが実施されるよう検討する。

## オ 産学官民コンソーシアムの構築と運営

本事業の成果として、疾患感受性遺伝子や疾患関連バイオマーカーなどの発見が期待されるが、これらは新しい創薬のシーズを生み出すものであるため、産業化を視野に入れて産業界との早い時期からの連携が必須である。またコホートデータと統合することが有用と考えられるライフログの取得などを手掛け、いわゆるビッグデータをすでに保持している、あるいはそのための技術を持つ情報系企業などとの共同研究なども推進すべきである。そのために、産業界の意識調査を十分に行った上で産学官民にわたるコンソーシアムを結成し、必要な運営規則などを定め、運営体制を整える。

## カ 人材育成

本事業の発展には、膨大な臨床情報とゲノム情報の統合解析技術を駆使できる人材育成が急務であるため、中核拠点での基礎的な研修期間ののち、産学官民コンソーシアムの関連機関や地域拠点へ派遣し、実地体験による訓練を中心とした人材育成プランを作成・実施する。

## ② 事業実施連絡会議

本事業にはさまざまな利害関係者が関与するので、その間をつなぎ、連絡調整をする部門が必要となる。地域住民や産業界、学術団体など、各団体の代表を集めた協議会を定期的で開催する。特に、社会の合意形成の前提として、科学者コミュニティからゲノムコホート研究についての支持を得ておく必要があるため、日本学術会議と緊密な連携をとる。

## ③ 中央倫理審査委員会

複数の外部識者を含む委員により、事業および付随する研究の倫理審査を実施

する。また各拠点における倫理審査が均質になるように、審査基準などのガイドラインを策定し、地域拠点の倫理審査委員会に通知する。

#### ④ データ公開・共有管理委員会

研究者コミュニティがデータを広く利用できるように、データの公開や共有の方針、データ共有により行われた研究成果を論文として発表する際の方針などを策定する。公開条件・安全対策などを検討した上で、随時データを公開する。事業外部へのデータ公開については、NBDC で実施している NBDC ヒトデータベースの枠組みを利用することを検討する。また、情報の公開・共有を進める研究倫理については、データ共有の申請者の所属機関での研究倫理審査委員会で審査され、その承認が得られていることを前提とするが、個々の申請の審査と承認は本委員会で担当する。

#### ⑤ 広報委員会

本事業の活動について、一般向け（リクルート活動として）および外部研究者向けに広報活動を行い、マスメディアとの協同による社会への積極的な情報提供やアウトリーチ活動など、社会の合意形成に必要な活動も担当する。また、本事業の実施期間中に、成果報告を兼ねたシンポジウムなどの研究会を開催する。

#### ⑥ 外部評価委員会（事業外独立組織）

外部委員を主とする事業の評価委員会を組織し、第三者による事業評価を実施する。加えて、本研究におけるデータ共有、生体試料の提供、研究倫理および研究コンプライアンスに対し適切な助言を行う。

### (3) 地域拠点

地域拠点に選定された各地域では、地域での健康づくりの土壌を醸成しつつ、定期的な健診による地域住民の健康管理を通じた予防医学研究の実施が望まれる。具体的実施項目として、統一プロトコルに則った生体試料（血液、尿ほか）の収集と登録、バイオマーカー測定、生理学・身体機能検査などの健診データの登録があげられる。また、大規模研究への参画に加えて、各拠点での独自の研究を実施し成果を公表するとともに、健診データを利用した住民の健康づくり啓発に努める。英国の UK Biobank における Ethics & Governance Framework (EGF)などを参考にしつつ<sup>1</sup>、地域特性や事業参加者の意識を尊重した教育活動および広報活動を実施する。

#### ① 事業参加者のリクルート

本事業は「ゲノムコホート」であるため、基本的に健常者をリクルートする。

---

1 <http://www.ukbiobank.ac.uk/ethics/>

まずは地域住民、自治体などとの密接な関係の構築に努め、健康づくりの土壌を育てる。そして、定期的な健診を実施し、地域住民の健康管理を通じた予防医学研究を推進する。

## ② 健診データの収集と登録

リクルートされた住民に対して試料採取と健診のスケジュールを作成し、包括同意の取得ののち健診を実施する。そして中核拠点から求められた規格のデータを取得し、データベースに蓄積する。その際、連結可能匿名化を実施するが、対応表をどこで管理するかについては、各地域拠点で管理することも含め十分な検討を要する。データの安全な保管のために生体認証などを含めたセキュリティ対策を施したサーバ施設を確保し、管理体制を作る。蓄積されたデータのうち事業全体で共有すべきものは中核拠点が管理するデータベースに格納する。この体制において、中核拠点に蓄積されるデータについては、取得された包括同意に基づき、各地域拠点がデータベース全体を利用することはもちろん、外部公開できる部分については審査を行った上で第三者に提供することを可能とする。

## ③ 生体試料の蓄積

各拠点で収集された試料については、今後の解析技術の進展などによりどのようなデータが取得できるか予想がつかないこともあり、また量に限りがあることも考えると、一定量を生体試料バンクで一元管理し、分析・解析拠点での測定・解析に供することが前提となる。一方、試料の一部は各地域拠点で保管し独立したバンクとして運営するように計画を立て、各地域拠点の裁量で地域拠点としての特色を持たせた独自の解析に用いることを可能とする。

## ④ 地域拠点における解析研究

各地域拠点では、本研究の統一プロトコルに沿った研究開発に加えて、地域特性や専門性を生かした独自の研究開発を実施することが奨励される。例えば、2カ所の地域拠点が連携して進める研究に関しては、地域拠点間で共同研究契約を締結することにより、互いに保持している試料を分譲しデータを共有することで、20万人規模の新たな解析が可能になると思われる。中核拠点はこれらの連携の在り方の基本方針を定める。

なお、分析・解析作業においては、各地域拠点がすべての機能を持つことは効率的ではないので、各地域拠点の特色を生かし、分析・解析機能ごとに役割分担することも考える。中核拠点はそのために全体の調整を担当する。

## (4) 解析・分析拠点

得られた生体試料の分析・解析を実施するとともに、情報統合データベースの構築により、臨床情報や、疾患罹患情報の蓄積を行う。また、それらの多様な情報と疾

患との関連を解析する。このためには以下の2つのセンターが必要とされるが、これらは必ずしも中核拠点の中に配置する必要はなく、分析・解析拠点として独立させることも視野に検討する。

#### ① 統合オミックス解析センター

得られた検体を用いたオミックス解析（生体分子の網羅的解析）のための分析・測定のみならず、そのために必要な技術の開発と標準化、計算科学・情報学などを駆使したデータ解析法を開発をする上で極めて重要な役割を担う。したがって技術開発に重点を置いた機能を持たせ、既存の施設を含め全国で3～4ヵ所設置し、各拠点が特定の分析・測定に特化したものとするのが適当である。解析拠点には関連技術を持つ企業の参加を促し、産学連携の開発体制を整備することも重要である。また、それらの設置場所については、生体試料バンクからの距離なども考慮に入れて議論すべきである。

#### ② インフォマティクスセンター

健診や生体試料の分析・解析で得られる情報の集約と一元管理を行い、こういった多様な情報、あるいは診療・健康情報とオミックスデータといった異種データを統合するデータ標準化形式の策定とデータベースの構築を行う。また、電子医療情報を用いた疾患罹患情報の集約技術を開発し、情報を集積して分析・解析データと統合する。加えて、情報通信技術（ICT）の開発と、それを用いた事業参加者との双方向コミュニケーションシステムの構築を試みる。解析によって得られた情報や研究成果の概要は公開用のデータとしても整備し、我が国固有のヒト生命情報としてNBDCを通して公開する。また、産学官民コンソーシアムなどからの要請により、個別のデータ解析を受託の形で実施し、また共同研究契約のもとで研究開発を進め、疾患関連因子の同定、薬剤標的分子・診断マーカーなどの探索研究が実施できる仕組みを検討する。

### (5) 生体試料バンク

生体試料バンキングシステムの構築と、LIMS（Laboratory Information Management System）の導入による、生体試料の集約と一元管理および、収集された生体試料のデータベース登録と品質管理を主要な業務とする。また、生体試料の分散管理と事業継続計画（BCP）の策定や、生体試料の再配分とそれに関わる情報の管理も行う。活動は、個人情報保護法などの関連法規制に従って実施されるため、厳格なセキュリティ管理と組織の継続性が必要である。

100万人の事業参加者の生体試料を全て一元管理する施設を中央に設置し、加えて災害などの非常時に備えた分散管理のためのバックアップ施設を別の地域に設置する。それぞれのコホート実施地域で収集された生体試料の一部は各地域拠点の独自の研究活動に使用するため、それぞれの地域拠点に管理・保管するバンクを設置すると

いう二重構造とする。

長期にわたる生体試料の保管では、試料の劣化を最小限に抑える保管条件の検討をあらゆる角度から検討すべきである。保管方法と温度、定期的な品質管理の方法、非常時の生体試料のレスキュー法など、国内外の既存の生体試料バンクの経験を参考としながらも、新たな保管方法や保管設備の開発も視野に入れて事業実施計画を策定することが極めて重要である。

バンキング事業は、多額の設備投資と維持・管理費が必要とされるため、必ずしも施設を新設するのではなく、民間への業務委託も可能性として検討する。

## 4 事業参加者

### (1) リクルートの方法

10万人規模のコホートを全国に10か所程度構築するために10ヶ所の地域拠点を設置し、地域住民、地元企業などの職員や社員を対象とした住民コホートと職域コホートからなる100万人の追跡集団を設ける。各地域拠点で10万人の事業参加者を得る場合、組織づくり、労力、費用、追跡データ（診療記録）を入手する方法などを考慮に入れると、現在我が国で広く行われている特定健診に相乗りすることが最も現実的である。また、住民コホートを構築する上で、既存のゲノムコホートとの連携を重視し、本事業の実施に関する条件を提示して、条件が合えば本事業への一部参画を可能とする。職域コホートを実施する地域拠点は、ある程度の規模の企業にゲノムコホートへの参画を提案し、賛同する複数の企業を集めたコンソーシアムをつくり、事業参加者のリクルートと追跡調査を行う。

### (2) 追跡方法

100万人のコホート構築とその長期にわたる追跡調査は、大学など既存の研究施設だけでは対応できないと考えられる。したがって、地域拠点には本事業を担当する専属の施設を設けて、十分な人員を配置し予算を投下する必要がある。住民コホートおよび職域コホートの場合は、担当者は経験のある疫学研究者であることが想定されることから、地域拠点の中心施設はこの疫学研究者が所属する既存の研究施設に併設するか、あるいは新たに設置する。配置する人員は、コホート研究の専門的知識を有する医師、リサーチナース、データマネージャ、生物統計学者、情報処置技術者などである。それぞれの人員や予算の規模は、担当するコホートの大きさによって決定する。

住民コホートは、3～5年ごとに健診を行って健康状態を把握し、臨床情報および生体試料を収集する。その他、原則として年1回アンケート調査や電話調査によって事業参加者の安否を確認するとともに、健康状態を調査する。事業参加者に何らかの疾患が発症した場合、あるいは発症が疑われる場合は、原則的に医療機関より臨床情報を入手し、診断を確認する。職域コホートについても同様の追跡調査を行う。

登録された事業参加者の医療情報を追跡する1つの方法として、重要な役割を果たし得るのは保険者である。国内のどこの保険医療機関を受診しても診療報酬請求情報（レセプト情報）が医療機関から保険者に電子的に提出されているので、保険者は被保険者のすべての受療イベント情報を保有している。したがって、事業参加者が加入している保険者を参加登録時に把握しておくとともに、事前に同意文書を取得して必要な受療イベント情報を収集できる体制を構築する。また、「行政手続における特定の個人を識別するための番号の利用などに関する法律案」（いわゆるマイナンバー法案）で導入が準備されているマイナンバー制度については、将来的にその活用も考慮するが、そのためには関連の法令整備が必要である。

## 5 分析・解析方法の開発と標準化

本研究では、膨大な数の生体試料を複数回収集し、多様な生体分子の分析・測定を実施するため、どこでどのような分析・測定をいかなる解析法で行うかを十分に検討し、プロトコルを策定することが極めて重要である。また、先端的な技術や機器を用いたオミックス解析などについては、各地域拠点にすべての解析を実施する解析センターを設置することは効率的ではないので、それぞれの拠点において解析の内容を整理し、役割分担ができるようなネットワークを構築する。さらに、現在の技術に固執することなく新技術も積極的に取り入れ、新たな手法での分析・解析に柔軟に取り組むことが重要である。

### (1) ゲノム解析

技術的にも価格的にも、迅速かつ信頼度が高く汎用性の高い解析手法はいまだ開発されていないのが現状である。ゲノム情報は不変かつ明確な文字列表現があるため、高品質のゲノム DNA を保管し、技術が成熟した時に解析するほうがより効率的かつ安上がりである。したがって 100 万人の解析を実施するための設備投資を事業当初から行うことは不適切である。加えて全検体の解析を解析拠点で実施する必要は必ずしもなく、技術的要件やデータの質の担保があれば既存の公的研究機関や企業に外注することも考慮する。一方、中小規模な解析は、技術開発やインフォマティクス分野の開発および教育には欠かせないものであるため、必要最小限の設備は解析拠点に必要である。

### (2) オミックス解析

産出されるデータが定性的なゲノム解析とは異なり、オミックス情報（転写物、代謝物、タンパク質、ペプチド、脂質、糖など）やエピゲノムの網羅的解析は、分析・測定値が使用機器、検体の処理、試薬の調製、測定環境など多様な因子に影響を受ける。また、質量分析計の検出波からの定量には、確率モデルの適用など、高度な数理的手法を用いる必要がある。したがって、定量性を担保するためには、可能であれば一ヶ所・同一条件で全検体を測定するべきである。

### (3) 測定・分析方法の標準化

測定・分析が外部のプラットフォームを含む複数の研究機関で実施された場合、得られるデータに互換性があり、一定の質を担保できるような統一ルールが必要である。また、測定結果は使用する機器や試薬、検体の処理方法に加えて測定者の技量にも依存するため、試薬や消耗品の一括購入と品質管理、測定技術者の技術講習などが必要となる。そこで、統合オミックス解析センターでは、定期的な進捗報告会の開催、必要に応じた技術者のトレーニングなどを実施し、測定・分析結果の質の担保に努める。

#### (4) データ取得・変換方法の開発

ヒト生命分子の網羅的測定・分析は、上述のようにいまだ確立された方法論がないが、将来革新的な技術・方法が開発される可能性が高い。そういった技術を基盤として産出される測定値や分析結果を研究に利用可能なデータとして蓄積するためには、検出器から出力された生データに対して適切な数学的・情報学的変換処理を施す必要がある。したがって、統合オミックス解析センターに計算科学者、数学者、物理化学者で構成される解析方法開発グループを置くことが望まれる。

表 1 代表的な分析・解析項目

分析・解析項目	方 法
ゲノム	次世代シーケンサー・マイクロアレイ
中間形質	
水溶性低分子代謝物	質量分析など
脂質	質量分析など
ペプチド	質量分析など
タンパク質	質量分析など
糖	質量分析など
転写物	次世代シーケンサー・マイクロアレイ
生化学検査	通常臨床検査
血液学検査	通常臨床検査
細胞免疫学的検査	フローサイトメトリー

出典 本小委員会作成



## 6 情報統合データベース

### (1) 今後開発が必要なデータベース

#### 1) 事業内部情報共有データベース

各地域拠点で収集された生体試料のメタデータや解析データなどを横断的・縦断的に取得できるデータベース。

#### 2) オミックス情報データベース

生体試料から得られるゲノム配列やオミックス解析の生データと実験メタデータ、解析情報などを格納するデータベース。

#### 3) 健康・臨床情報データベース

電子医療情報 による疾患罹患情報やバイオマーカーに関する高精度の臨床情報を格納するデータベース。匿名化処理が必要。

#### 4) 環境・生活習慣データベース (医薬品データベースも含む)

生活環境や食生活などの生活習慣に関する情報を長期間追跡し記録するためのデータベース。匿名化処理が必要。

#### 5) 一般公開データベース (上記の中で公開可能なもの)

個人情報やプライバシーに配慮した上で、一般に公開できる集団の解析データを分類し公開するデータベース。

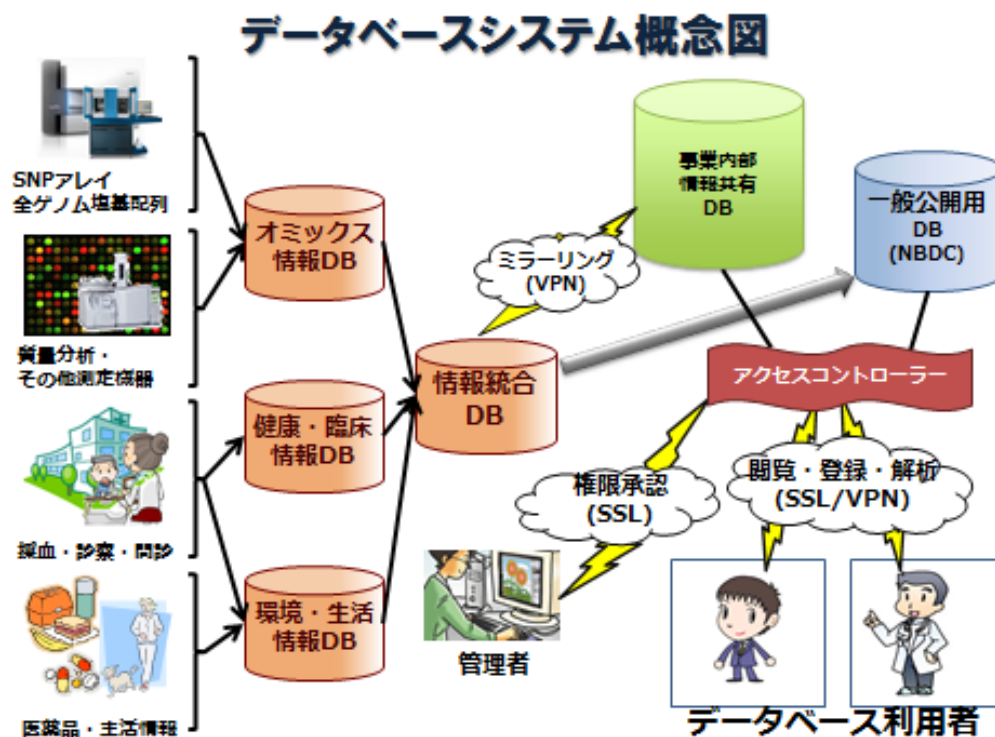


図2 データベースシステム概念図

### (2) データベースとその統合化に必要な技術

#### 1) 標準化対応の医療情報システムの開発と導入整備

現在規格が統一されていない電子カルテシステムや検査情報管理システムの

出力データを標準化するための機能開発。

2) データの品質管理 (QC) 技術

各地域拠点で生産されたデータを集約するための標準化や、データ監視、バージョン管理に関する技術開発。

3) 画像データ管理、メタデータ付与、検索技術

MRI や X 線 CT などの医療診断画像データから特徴量を抽出する、あるいはメタデータのタグ付けをする技術の開発が必要。

4) 匿名化、暗号化技術

個人情報保護のための匿名化、暗号化などのステップが必須。また、疾患発症追跡のための匿名化情報と臨床情報や時系列実験データを参照可能な情報集約システムの開発と導入が必要。

5) 秘匿検索技術

データベースに格納された暗号化データを検索するための技術。

6) 非構造化データ管理、検索、情報抽出技術

臨床情報や、疾患罹患情報に関連する非構造化データ・半構造化データの取り扱いを実現し、必要な情報を抽出し処理する情報基盤技術。

7) 大規模データ分散管理技術

各地域拠点、分析・解析拠点に分散したデータを保持・処理するための技術と基盤ソフトウェア。故障や天災などによるハードウェアの障害にも強いことが必要。

8) データ統合化技術

多様で質の異なる大量のデータから、利用者が知識発見を行うためにデータを統合し俯瞰するための技術開発が必要。

9) 高度質問応答システム

情報検索の専門家ではない現場の担当者、または一般の利用者が必要な情報にたどり着けるように、高度な言語横断情報検索と質問応答システムの開発が必要。

10) セキュアネットワーク構築

本事業では大量の個人情報を扱うため、情報ネットワークに対する安全性を確保しつつ、利用者の利便性を損なわないための技術開発が重要。

(3) 情報解析技術

1) ゲノム・オミックスデータの統合解析、メタ解析

全ゲノム情報と、各種オミックス情報の関連解析を実施するための技術開発。

2) 疾患関連因子の同定

新規疾患関連因子の発見につながるような、各種オミックス統合解析手法やソフトウェアを開発。

3) 健康・疾患とオミックス、環境・生活習慣の関連解析

環境因子と疾患との関連解析やリスク評価方法の確立と有意性の検証などの疫学解析を実施。患者検体と健常者検体の比較解析により疾患感受性遺伝子の同定を実施。

4) 医薬品の副作用とゲノム・オミックスとの関連解析

レセプト情報などのデータベースを活用し、投薬情報と疾病（副作用など）発生情報や安全対策措置の効果を評価するための情報基盤の整備。

5) 薬剤標的分子・診断バイオマーカー探索

薬剤の標的分子を明確にし、その機能を特異的に制御する薬剤（分子標的薬剤）の探索を実施。マイクロアレイ解析やオミックス解析を組み合わせ、診断バイオマーカー解析手法の開発を実施。

**(4) 計算機環境**

インフォマティクスセンターにおいては、事業参加者の個人ゲノムデータの蓄積の他に、ゲノム解析用並びにオミックス解析用、さらにデータベース運用のための計算機が必要である。既存の国内外のヒトゲノムを扱っているデータベースでの実績も考慮に入れると、ゲノム情報だけでも圧縮時 100GB/一人分が最低必要であり、中間ファイルや関連データなども含めると数十倍を確保する必要がある。そのため、最終的に 100 万人のデータを格納するためには 2,500PB のディスク、100 万コアを持つ計算機が必要になると考えられる。

**表 2 データ解析センターの計算機スペック（予想）**

	2015 年（1 万人）	2020 年（10 万人）	2030 年（100 万人）
コア数	20,000	100,000	1,000,000
ディスク	25PB	250PB	2,500PB

（参考）理化学研究所・京コンピューターのコア数は 2012 年度、705,024 である。

出典 本小委員会作成

## 7 産学官連携

### (1) ヒト生命情報統合研究に関する可能性と期待

ゲノムコホートを用いたヒト生命情報統合研究からは、新しい創薬の戦略、疾病の予防などに関わる新たな健康産業の創出、さらに臨床開発に不可欠なバイオマーカーの同定など、多くの成果が期待できる。個人情報保護法、薬事法の面からも医療情報並びに生体試料の利活用に関する倫理問題などの課題が解決され、企業が積極的に事業に参画できる環境が整えば、世界的な視野での研究開発力の高揚になると期待されている。

### (2) 製薬産業と『ヒト生命情報統合研究』

日本の製薬企業から創出された同じ作用機序の中で最初に発明された医薬品（New Class の医薬品）の割合は、低分子化合物の発見、開発が主流を占めていた 1980 年代には世界の 30%を占めていたが、ここ 10 年では 10%を切るようになってきた。日本の製薬業界は、独自性のある医薬品開発の新たな創薬戦略を立て、早急に実現する必要性に迫られている。

世界の医薬品開発は、従来のモデル動物主体の研究からヒトゲノム情報や臨床情報を活用した研究へと大きくシフトしており、特定の分子や遺伝子異常を狙った分子標的薬など、患者一人ひとりに最適な治療のための新薬開発が進んでいる。また実際の治療にあたっては、治療薬の効果や副作用を投薬前に予測するコンパニオン診断薬を使った患者の選択あるいは有効性の評価が行われるようになってきた。このような流れを背景に、本研究における多くの健常者の長期追跡から得られる膨大な健康関連情報は、新たな創薬のシーズを数多く含む貴重な研究開発基盤として、製薬企業が期待するところは極めて大きい。

### (3) バイオマーカーの発見と診断薬企業

日本の市場における診断薬市場は医薬品に比べると遥かに小さいが、欧米においては、FDA（アメリカ食品医薬品局）の報告にもあるように<sup>2</sup>、薬効評価や有効例、無効例の鑑別に役立つバイオマーカー探索の激しい競争が始まっている。診断薬市場は個人に最適な医療が広まるにつれ世界的に大きな市場に成長すると考えられており、生体試料バンクと経時的な医療情報を利用した疾患の発症から進展に至る過程の詳細な解析は、こういったバイオマーカーの発見に直結する研究として診断薬企業からの期待が高い。

また、いままで医療産業への参入を躊躇していた企業の中にも、今後世界の潮流となるゲノム情報やバイオマーカー情報をもとにした個人に最適な医療のための研究開発への本格参入を検討するところもある。自社の持つ技術の発展的応用と複数の企業の技術連携による新産業創出に期待するところは大きい。

---

2 Aliza Thompson 2012 America 糖尿病学会講演 (ADA)

#### (4) 医療デバイス企業

我が国の医療市場における医療デバイスの総売上は 1.5 兆円を超える。しかしながら市場の大部分を外国製品に奪われ、海外への流出額は 1 兆円を超えており、特にステント、ペースメーカーなどの埋込型の医療機器で顕著である。日本の精密機械・医療素材産業には素材や製品を開発する能力があるが、特殊な販売チャネルの存在や単独の企業での開発力が弱いために、国産品の開発は遅れている。デバイス開発における最重要課題は品質、有効性、安全性などの評価方法の確立である。大規模ゲノムコホートで得られるさまざまな情報を利用することで、日本独自の医療デバイス開発への道が拓けるといふ期待は大きいものがある。

#### (5) 食品関連産業（食品、健康食品企業）

特定保健用食品（特保）に代表される「食」による疾患予防は、食品関連産業にとって将来の成長分野と考えられている。保険制度の違いはあるものの、米国における健康食品の市場は 7.5 兆円（日本は 7,300 億円）と極めて大きく、今後の我が国での成長が期待される。一方、小規模企業にとって特保取得のための投資は、現状の特保制度下では負担が大きい。また、一般住民コホートにおける介入試験や、健常者を対象とした臨床試験は、現在の社会システムでは構築が難しい。このような観点から、特定の食品（例えば乳製品）などの効果を大規模データから解析することは、新商品の開発戦略に大きなメリットをもたらすと期待されている。

#### (6) IT 企業

ゲノムコホートにおける医療情報の統合化には、膨大な情報を集約・管理するための技術革新が必要である。医療情報を扱う IT 企業では、従来の電子カルテなどの医療現場での情報収集システムから、日常生活における健康管理を含めた健康増進を目的とするコミュニケーションシステムまで、幅広い分野への参画が検討されている。いずれの企業も、大規模ゲノムコホート研究の情報の集積・処理のみならず、医療情報の利活用による新たなビジネスも今後の大きな収益の柱と考えている。また一部の企業は、「健康でいきいきと生きる町」というインフラそのものをビジネス化するために、海外にフィールドを設定して研究をすでに開始している。今後、ゲノムコホートの多様な情報を利用した新たなビジネスを見据えた IT 企業での研究開発が、積極的に展開されるであろう。

#### (7) 産学連携体制の必要性と各省の役割分担

本事業の推進においては、長期的な事業計画や実施基盤の構築と並んで、事業に関連する省庁が所掌する分野の産業や地方自治体などの協力が求められる。このために、下図に示すように、内閣府が本研究の統括の役割を担う。文部科学省は中核拠点の運営、分析・解析拠点の構築と運営、人材育成、国際連携を担い、厚労省は、医療

情報の収集と分析のための組織構築、生体試料バンクの構築と運営、地域拠点における医師会との協力関係の構築などを担当する。さらに情報ネットワークの構築、分析技術の開発などについては経済産業省がその役割を担う。総務省は、マイナンバー法を活用しながら、地域における医療情報収集技術の開発と IT 関連の成果収集に大きな貢献が期待される。

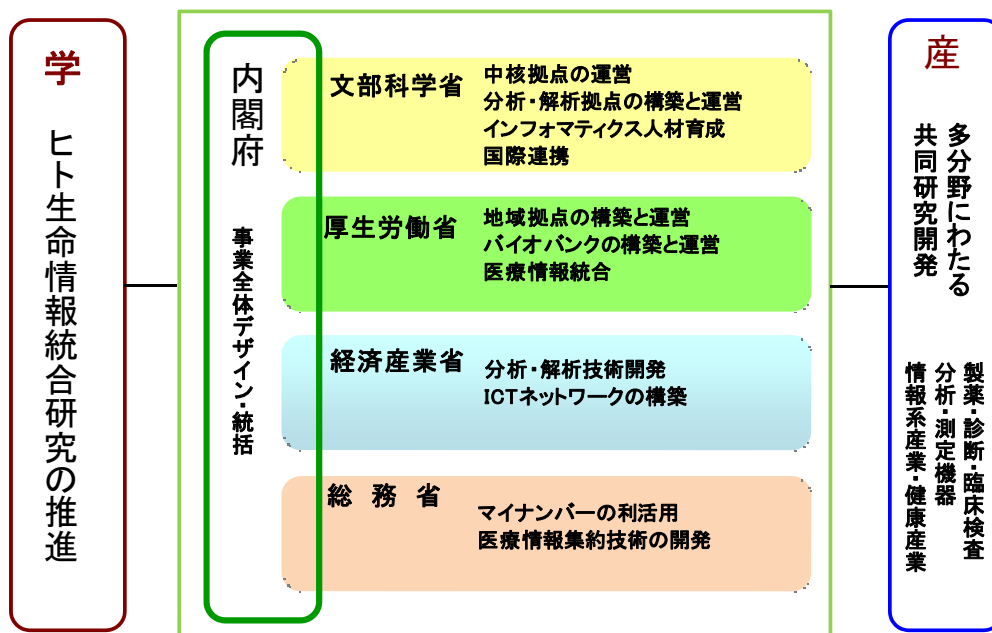


図3 産学官連携体制と各省に期待する役割

## 8 倫理的配慮・社会との接点

### (1) 倫理的課題とその対策

人を対象としたすべての医学研究は、研究の内容と目的を事業参加者に説明し、十分な理解の下に、同意を得た後に開始される。従来の単一遺伝子疾患を対象とした遺伝医学研究では、「患者・被験者-医師・研究者」の枠組みの中で十分なインフォームドコンセントのプロセスが実施され、患者・被験者にとっても研究目的や研究成果は比較的理解しやすいものであった。

一方、ゲノムコホート研究では、提供を受けた生体試料やデータの具体的な利用方法や予想される研究成果を生体試料採取時に明確に示すことは困難であることが多い。ゲノムコホート研究は、生体試料を提供する事業参加者の利益のためというよりも集団全体のよりよい医療を実現するために必要なものであり、そのためにはより多くの人々の参加が求められていることについて広く社会の理解を得ておく必要がある。

本研究においては、事業参加者から提供試料の使用目的を限定しない包括同意や将来の健康情報の提供についての了解を得ておくことが不可欠である。本研究の活動には、厳格なセキュリティ管理と組織の継続性が必要であり、かつ個人情報保護法などの関連法規制に従って実施されるため、実施に際してこれらの倫理的課題について十分に検討し、適切なプロセスを構築しておく必要がある。

また、事業参加者一人ひとりを追跡するための医療マイナンバーを医療や研究でのみ限定的かつ安全に使用できるようにするには、目的外使用の禁止やその罰則規定の適切な整備が必要である。さらに、今後開始される大規模疫学研究においても、同意した人にものみ使用できることを明確にするよう、個人情報保護関連法令と関連倫理指針の整備と改訂も必要不可欠である。

### (2) 倫理審査

ゲノムコホート研究では、提供された事業参加者の試料や診療・健康情報などは多数の研究機関において多様な研究目的に利用される。事業参加者が健常者の場合は、自らの試料や診療・健康情報などを提供する意義は、専ら医科学研究の推進や公衆衛生の向上といった公共的利益に貢献するということになる。事業参加者は、個別的な利益を得ることは原則としていないにも関わらず情報漏洩などさまざまなリスクにさらされることになるため、慎重に審査する必要がある。

ゲノムコホート研究の倫理審査のレファレンスとしては、現状では「ヒトゲノム・遺伝子解析研究に関する倫理指針」（2013年4月改正）が用いられることになるが、準備期間の間にこの倫理指針に基づいてゲノムコホート研究を円滑に進めることができるかどうかを十分に精査しておく必要がある。また、諸外国での取り組み（UK Biobank、EUのゲノム解析に関する Legal Regime など）および我が国で先行して実施されているコホート調査（ながはまコホート、東北メディカルメガバンク、エコチル調査など）で行われた倫理審査にむけた取り組みを参考にすべきである。特に、な

がはまコホートでは、現行の倫理指針を踏まえながら、事業に携わるすべての者が遵守すべきことと事業の基本的な仕組みを明確にした独自のルール「ながはまルール」<sup>3</sup>を策定し、市の条例として法的拘束力を持たせた上で、事業参加者から包括同意を取得して研究を進めている。

倫理審査委員会の設置に関しては、基本的には、中央倫理審査委員会に加えて、地域ごとに倫理審査委員会を設置する。中央倫理審査委員会においては、全体の研究計画の倫理的問題について審議する。一方、地域の倫理審査委員会においては、各地域固有の倫理的課題および各地域で実施される独自の研究計画について審議する。

### (3) インフォームド・コンセントと個人情報保護

ゲノムコホート研究におけるインフォームド・コンセントのプロセスにおいては、事業参加者の利益の保護や安全確保の視点から、プライバシーの権利保護と被験者情報に関する秘匿性・守秘義務の確保が極めて重要である。不正使用、あるいは漏洩などにより遺伝情報に基づく被害・差別・偏見が一例でも起きた場合には、ゲノム研究・遺伝医学研究全体に不信感をもたれることになり、ゲノムコホート研究全体の実施計画に影響を与えることになる。インフォームド・コンセントは手続き論的なものになりがちであるが、プライバシー保護と厳重な秘匿性を可能とする研究計画を立て、それを着実に実行する組織構築および人材育成を行うことが求められる。

本研究への協力を得るにあたり、研究の目的、研究内容に加えて、研究者コミュニティに対して連結可能匿名化の上でデータや生体試料が提供されることを十分に説明の上、包括同意を得るようにする。研究者コミュニティとのデータ共有や生体試料の提供についても、事業参加者のプライバシー保護に最大限の配慮をし、連結可能匿名化の上で適切に提供するようにする。

### (4) 社会的な合意形成

ゲノムコホート研究の実施に際しては、事業参加者の同意だけではなく、広く一般社会からも本研究の推進についての支持を得ておく必要がある。しかし、国民の間にある遺伝学・ゲノム学についての理解と、先端技術としての遺伝学・ゲノム学の成果・可能性の間には隔たりがあるのが現状である。この隔たりを埋めるためには先行事例で行われている方法を参考にすべきである。例えば、ながはまコホートでは、事業参加者の市民が中心となって結成されたNPO法人が大学、行政と一体となって事業を支えており、事業参加者の勧誘や研究者を招いての講演会や健康づくりイベントなどを通じた市民の啓発活動のみならず、共同研究者として事業へ参画するなどさまざまな活動を行っている<sup>4</sup>。また、新たなゲノム科学研究の進展の状況と起こりうる社会的課題を調査するとともに、それを基盤として多分野の専門家および一般市民が協

---

3 ながはまルール <http://www.city.nagahama.shiga.jp/index.cfm/9,3185,19,158,html>.

4 健康づくり0次クラブ <http://zeroji-club.com/>



同して議論を深める機会を設け、具体的な対応策を提案、実施することも必要であろう。

## 9 人材育成

### (1) 遺伝医学

遺伝医学に関連しては、以下の4種の人材が必要であり、現在でも不足している。早急に教育システムを立ちあげ研修コース、認定制度を設けることが肝要である。

#### ① 遺伝医学・ゲノム学教育者

当該地域の大学などに「ゲノム教育学講座」を設置し、そこに集団遺伝学、臨床遺伝学、生命倫理学などの実践的知識を有する教員を着任させ、後進の遺伝医学・ゲノム学教育者の育成に従事する。同時に教育講演会を適宜開催し、住民の日常的な医療ケアに直接従事する医師や、その他の医療従事者にゲノム情報の取り扱い・意義・限界を教育する。また、以下のゲノム医学に関する職種の教育にも研修コースの講師などとして関与する<sup>5</sup>。その養成には、臨床遺伝専門医の到達目標の中に集団遺伝学、遺伝疫学およびゲノム科学を加えることなどが望ましい。

#### ② ゲノム医学コミュニケーター

遺伝子差別のリスクに留意しつつ、ヘルスケアにおけるゲノム医学的取り組みに対する市民の意識と受容を促すべく、パンフレットやホームページなどの作成、サイエンスカフェの設営などを通して、ゲノム解析技術やその研究成果に係る情報を一般市民にわかりやすく伝達する。地域（コホート研究）ごとに、その事務局内に同職種の人材を配置する。国全体として研修コース・認定制度を設け、当該職種の技能の標準化を図ることが望ましい<sup>6</sup>。

#### ③ ゲノム医学リサーチコーディネーター

ゲノムコホート研究において、事業参加者のリクルートに係る業務（同意取得、検体の採取、基盤的臨床情報の取得など）を支援する。国全体として研修コース・認定制度を設け、当該職種の技能の標準化を図ることが望ましい<sup>7</sup>。

#### ④ ゲノム医学カウンセラー

希少で家族性の遺伝疾患に関わる問題だけでなく、ゲノムコホート研究に関連

---

5 日本人類遺伝学会と日本遺伝カウンセリング学会が認定している「臨床遺伝専門医制度」<<http://www.jbmg.jp>> は、到達目標を定め3年間の臨床遺伝に関する研修を行った医師を対象に筆記試験・面接試験を行っており、966名が認定されている。

6 NPO法人「遺伝カウンセリング・ジャパン」<<http://www.npo-gc.jp>>は認定遺伝カウンセラーを支援する目的で設立され、活動の一つとしてサイエンスカフェを積極的に実施しているので、ゲノム医学コミュニケーター（genomic medicine communicator）の活動を考える際に参考となる。

7 ゲノム医学リサーチコーディネーターの養成には、日本人類遺伝学会と日本疫学会が協同して運営しているゲノム医学リサーチコーディネーター制度<<http://jshg.jp/qualifications/gmrc.html>>が参考になる。

して、遺伝や遺伝性疾患に関する説明・支援が必要な事業参加者にゲノム医学リサーチコーディネーターと共同で対応する。遺伝カウンセラーと同様、国全体として研修コース・認定制度を設け、当該職種の技能の標準化を図る<sup>8</sup>。その養成には、認定遺伝カウンセラーの到達目標、教育課程の中に、集団遺伝学、遺伝疫学およびゲノム科学を加えることが望ましい。

## (2) 医療情報整備

ICT を活用して被験者の健康・医療情報の抽出、整備を支援する人材である、医療情報技術者が必要である。当該地域のゲノムコホート研究における、ICT を用いた情報システムの運用実務を担う技術職であり、医療情報技術士の資格を要する。

## (3) 情報処理・解析

100 万人規模の大量かつ多様な情報を適切に扱うには、遺伝カウンセラー、匿名化処理従事者、電子カルテ管理者、などさまざまなレベルの人材が多数必要であるが、中でも、生命情報学者（バイオインフォマティシャン）、生物統計学者、統計遺伝学者などの専門家が決定的に重要である。これらの人材は、いまでも大幅に不足しており、また、促成での教育が困難であるから、事業の立案時から、人材の育成を、大学などの教育機関と連携して取り組むべきである。専門性が高く、法律上の繊細な取り扱いが要求される大量の情報を、大規模な集団で、広範囲かつ永続的に維持管理し解析するために、上述の専門家を育成しつつ、適切な場所に配置することが、喫緊の課題である。

ゲノム情報という、これまでにないビッグデータを、大規模な集団で、広範囲かつ継続的に追跡調査し解析するための人材である。

### ① バイオインフォマティシャン

ゲノム解析、ゲノム疫学研究に関わる知識と経験を有し、研究のデザイン、大規模情報の処理、独自性の高い方法論の開発や医学生物学的モデルの構築、医学的な意味付けなど、の高度な研究を行う。医工連携を軸とした学部横断的分野での修士ないし博士号の取得を通じて当該人材を育成する。

### ② データマネージャー

ゲノムコホート研究における生体試料・情報の実務的管理（匿名化含む）を行う。より厳密な倫理的配慮が必要となるという点以外は、通常の臨床研究のデータマネージャーとほぼ同様の職種である。

---

<sup>8</sup> 日本人類遺伝学会と日本遺伝カウンセリング学会では「認定遺伝カウンセラー制度」<<http://plaza.umin.ac.jp/~GC/>>を運営している。現在、全国に認定遺伝カウンセラーを養成する10の大学院修士課程があり、その卒業生のみが受験資格を得ることができる。2013年現在138名が認定されており、すでにさまざまな遺伝子医療の現場で臨床遺伝専門医と連携し、高度な遺伝カウンセリングを担当している。

## <付録1>

### 本事業実施の所要経費の内訳

所要経費（16年間681億円）

（1）準備段階（3年間15億円）

- ・人件費 3億円
- ・外注費（調査、設計） 3億円
- ・人材育成 3億円
- ・計算機 3億円
- ・会合費、海外調査等 3億円

（2）ゲノムコホート実施段階（13年間500億円）

- ・生命分子統合解析（ゲノム、オミックス解析等）100億円
- ・新規ゲノムコホート実施（一次データの解析含む）400億円

（3）情報集積・共有・統合段階（11年間120億円）

- ・データ集積のためのネットワーク運営 30億円
- ・データの標準化、統合化とソフトウェア開発 20億円
- ・データベースの構築維持 20億円
- ・データベースの構築、活用のための人材育成 20億円
- ・データ格納サーバ 30億円

（4）データベースの公開、活用段階（当面5年間46億円）

- ・データベース公開 15億円
- ・データ活用技術開発 15億円
- ・解析用スパコン 15億円
- ・データベースや技術の普及 1億円

民間企業などからの貢献（16年間300億円）

- ・人件費（研究員、技術員など）160億円（60人／年、16年）
- ・設備費（ITインフラ整備含む）100億円
- ・消耗品費 40億円

出典 本小委員会作成

## ＜付録 2＞

### 本文 17 ページ「6 情報統合データベース」の補足

#### (1) 今後開発が必要なデータベース

##### 1) 事業内部情報共有データベース

各地域拠点で収集された生体試料のメタデータを集約し、一元管理を行う LIMS システムの開発。また、適切なアクセス制御をしながらすべてオミックス解析データを横断的・縦断的に取得できるデータウェアハウス開発が必要である。

##### 2) オミックス情報データベース

生体試料から得られるゲノム配列、発現、エピゲノム、プロテオミクス、メタボロミクスなどのオミックス解析の生データと実験メタデータ、解析情報などを格納するデータベースの開発が必要である。実験手法や検体のメタデータの記述に関しては、統制語彙などを用い標準化されたデータ形式で格納するか、適宜、標準形式に変換する機能が求められる。

##### 3) 健康・臨床情報データベース

電子医療情報による疾患罹患情報やバイオマーカーに関する高精度の臨床情報を格納するデータベースの開発。疾患罹患の記述に関しては、統制語彙などを用い標準化されたデータ形式で格納するか、適宜、標準形式に変換する機能が求められる。中央拠点を介して事業参加者で共有する部分に関しては、個人情報保護法に配慮し連結不可能な匿名処理など、匿名化の処理を行うことが必要である。

##### 4) 環境・生活習慣データベース（医薬品データベースも含む）

健康な人の生活環境や食生活などの生活習慣に関する情報を長期間追跡し記録するためのデータベースの開発。環境・生活習慣の記述に関しては、統制語彙を利用し標準化されたデータ形式で格納するか、標準形式に変換する機能が求められる。中央拠点を介して事業参加者で共有する部分に関しては、個人情報保護法に配慮し連結不可能な匿名処理など、匿名化の処理を行う事が必要である。

##### 5) 一般公開データベース（上記の中で公開可能なもの）

本事業で蓄積される多種多様で大規模なデータの中から、個人情報やプライバシーに配慮した上で、一般に公開できる集団の解析データを分類し公開するデータベース。必要であれば、アクセス制御を行い閲覧できるデータを利用者により制御する。

#### (2) データベースとその統合化に必要な技術

##### 1) 標準化対応の医療情報システムの開発と導入整備

現在多くの企業により販売されている電子カルテシステムや検査情報管理システムは、データ出力規格が統一されておらず、多施設データ収集や統合は容易でない。各アプリケーションには、標準化されたデータ形式（SS-MIX2 形式など）でのデータ入出力ができるような機能拡張が必要である。また、分野ごと専門用語の異表記を統一してデータを統合するために、統制語彙や標準オントロジーの

開発が必要である。

## 2) データの品質管理 (QC) 技術

各地域拠点で生産されたデータを分散したデータベースに登録する際には、データの正確性や相互矛盾などを監視し修正したり、表記揺れを自動訂正するための技術が重要になる。そのためには、標準化されたデータモデルの構築やデータウェアハウスの利用、データプロファイリング、データ監視ツールやバージョン管理に関する技術開発が必要である。

## 3) 画像データ管理、メタデータ付与、検索技術

MRI や X 線 CT などの医療診断画像は、容量も大きく、そのままでは計算機で取り扱うことが困難なデータである。これらのデータから画像解析技術によって特徴量を抽出する技術を開発する。また、画像、動画からテキストを抽出してテキストマイニングによりメタデータのタグ付けをする技術の開発が必要である。

## 4) 匿名化、暗号化技術

疾患発症追跡のための医療情報ネットワーク構築には、個人情報保護のために個人情報を非個人情報化 (匿名化) とデータ自体の暗号化などのステップが必須である。また、疾患発症追跡のための匿名化、暗号化技術の他に、連結可能匿名化、連結不可能匿名化情報などさまざま匿名化処理を施した情報と臨床情報を利用して時系列実験データを参照可能な情報集約システムの開発と導入が必要である。

## 5) 秘匿検索技術

データベースに格納された暗号化データを、利用者が、問い合わせ先のデータセンターに情報を知らせることなく検索するための技術開発。秘匿検索では、データ問い合わせを暗号化する必要があるが、その際、検索速度と安全性を両立した検索システムと運用法の構築が必要である。

## 6) 非構造化データ管理、検索、情報抽出技術

臨床情報や、疾患罹患情報に関連する構造化されない文章や数値データ、また測定機器から産出される画像データなどの非構造化データ・半構造化データは、そのままの形では従来のデータベースで取り扱うことが困難である。このような非構造化データの取り扱いを実現するために、テキストマイニング、自然言語処理、データマイニング、機械学習、パターン認識、人工知能の分野の技術を利用し、コホートに関係する情報を抽出し処理する情報基盤を開発する。

## 7) 大規模データ分散管理技術

大規模なデータを分散した解析拠点に置かれたデータベースで管理するために、大量のマシンでデータを保持・処理するための技術と基盤ソフトウェアが必要である。これらのソフトウェアは、大量かつ大容量のデータ処理に対応可能であるとともに、故障や天災などによるハードウェアの障害にも強いことが求められる。

## 8) データ統合化技術

多様で質の異なる大量のデータから、利用者が知識発見を行うためには、統計的手法などのデータマイニングや最新の可視化技術を応用して、大量で質の異なるデータを統合し俯瞰するための技術開発が必要である。

#### 9) 高度質問応答システム

必ずしも情報検索の技法に熟達していない現場の臨床技師や医師、または一般公開されたデータベースの利用者が効率よく必要な情報にたどり着けるようにするため、情報検索の技術に高度な言語横断情報検索と質問応答システムの開発が必要である。そのためには、柔軟なデータ構造、メタデータの付与などデータベース側の高度化とともに、自然言語からデータベース問い合わせ語を自動生成するような技術開発が必要である。

#### 10) セキュアネットワーク構築

本事業では大量の個人情報扱うため、外部端末の接続によるウィルスの拡散や、未許可利用者による情報の漏えいやネットワークの不正利用などの問題は致命的であり、情報ネットワークに対する安全性の確保が重要である。そこで、内部ネットワークに接続する端末を、いったん、LAN とは隔離されて存在する検査専用のネットワーク（検疫ネットワーク）に接続し、コンプライアンス検査を行うシステムが必要になる。そのために、利用者の利便性を損なわないかたちで高度なネットワーク認証や検疫ネットワークへの接続可能な技術的の開発が重要である。さらに、ネットワーク認証・検疫の状況把握や分析が可能なログ取得やサーバの運用法の開発が必要である。

### (3) 情報解析技術

#### 1) ゲノム・オミックスデータの統合解析、メタ解析

全ゲノム配列の決定、DNA マイクロアレイを用いた網羅的一塩基多型 (SNP) ・コピー数変異 (CNV) 解析、マイクロ RNA プロファイリング、エピジェネティクス解析、発現解析を行う。血液や尿中の代謝物やタンパク質の網羅的な測定とプロファイリングを行い、プロファイルデータとゲノム情報との関連解析を行う。

#### 2) 疾患関連因子の同定

メタボロミクスやプロテオミクス、マイクロ RNA の超高感度検出装置、定量計測装置から得られた大量の情報から多種因子同時分析技術、疾患関連因子のプロファイリングなどを行い、新規疾患関連因子の発見につながるような、各種オミックス統合解析手法やソフトウェアを開発する。

#### 3) 健康・疾患とオミックス、環境・生活習慣の関連解析

コホートデータからの罹患率や有病率の算定、環境因子と疾患との関連解析やリスク因子間の交絡の明確化、リスク評価方法の確立と有意性の検証などの疫学解析を行う。また、多因子疾患や中間表現型の感受性遺伝子の探索、および環境因子を含めたリスク評価の手法の開発を行う。特異的患者検体と健常者検体の比較解析により疾患感受性遺伝子の同定を行う。

#### 4) 医薬品の副作用とゲノム・オミックスとの関連解析

個人情報の保護などに配慮しながら、レセプトなどのデータベースを活用し、副作用などの発生に関しての医薬品使用者母数の把握や投薬情報と疾病（副作用など）発生情報の双方を含む頻度情報や安全対策措置の効果の評価のための情報基盤の整備。膨大で多様な安全性情報を医学・薬学・薬剤疫学・生物統計学・情報工学などの専門家が効率的・効果的に活用できるよう、組織・体制の強化を図るとともに、電子レセプトなどのデータベースから得られた情報を活用し、薬剤疫学的な評価基盤を整備することが必要である。

#### 5) 薬剤標的分子・診断バイオマーカー探索

薬剤の標的分子を明確にし、その機能を特異的に制御する薬剤（分子標的薬剤）の探索を実施する。そのために、癌関連タンパク質や細胞膜上の膜タンパク質など一連の薬剤標的タンパク質群における立体構造を予測し、特異的に結合する医薬品候補化合物をシミュレーションにより高精度に探索する技術の開発と洗練が必要である。さらに診断バイオマーカーの探索として、マイクロアレイ解析やオミックス解析を組み合わせ、疾患特異的に変化し、血液などの体液中に分泌される疾患関連蛋白質を網羅的かつ効果的に探索するための解析手法の開発を行う。

### (4) 計算機環境

各地域拠点における個人ゲノムデータの蓄積の他にデータ解析センターにおいては、ゲノム解析用並びにオミックス解析用、さらにデータベース運用のための計算機が必要である。現在この参考となるのは、国内では国際塩基配列データベース運用機関である国立遺伝学研究所のスーパーコンピュータ並びに、オミックス解析拠点である東京大学医科学研究所ゲノムセンターのスーパーコンピュータ、海外ではNGS解析センターであるBGI（中国）、Sanger Center（UK）などである。本コホート研究開始から数年はデータ量も少ないと考えられることから、スタート時は既存解析センタークラスの計算機環境でスタートし、その後見直すことが必用である。2013年時点で必要なディスク容量については、国際塩基配列データベース（INSD）のShort Read Archive（SRA）のエントリー調査並びに、1000 Genome Projectの情報などがある。これら既存の国内外のヒトゲノムを扱っているデータベースでの実績も考慮に入れると、ゲノム情報だけでもfastq形式のデータとして圧縮時100GB/一人分が最低必要であり、中間ファイルや関連データなども含めると数十倍を確保する必要がある。そのため、最終的に100万人のデータを格納するためには2,500PBのディスク、100万コアを持つ計算機が必要になると考えられる。（参考までにコホート関係者の中では概算として1人1テラのディスクが必要とされているらしい）

表3 2012年の既存NGS解析拠点の計算機スペック（参考）

	コア数	フロップ数(T)	ディスク	メモリ
国立遺伝学研究所 (SC性能国内28位)	5,616	165.1	5PB	最大10TB shared memory system, >4GB/core
東京大学医科学研究所 ゲノムセンター (SC性能国内23位)	16,128	148.4	2PB	
BGI（北京）	20,000	220	20PB	
SangerCenter（UK）	20,000	-	20PB	

（注）国立遺伝学研究所において2014年に実施される中間増強により、フロップ数、ディスク共に倍加の予定

出典 本小委員会作成



## <付録3>

### 用語集

#### 一塩基多型

ゲノムの塩基配列上で特定の場所の一つの塩基が個人によって異なる部分。英語では Single Nucleotide Polymorphism と呼び、その頭文字をとって SNP と略される。

#### インフォームド・コンセント

研究の内容について説明を受け、十分に理解した上で自らの自由意志に基づいて研究への参加に同意すること。

#### 疫学

特定の集団における健康に関連する状況あるいは事象の、分布あるいは規定因子に関する研究を行う医学の一分野。

#### オミックス解析

生体内のゲノム、転写物、タンパク質、ペプチド、脂質、糖、代謝物などの網羅的解析。

#### ゲノムコホート研究

コホート研究に、ゲノム情報や経時的に変化する生体分子の情報を融合した、新世代の生命科学研究。

#### コホート研究

ヒトの集団を長期間観察し、疾患の発症や増悪に関係する因子を解明する研究手法。観察対象となる集団をコホートという。

#### 生体試料

ヒトから採取する血液・尿、組織および血液や組織から抽出した DNA や RNA など。

#### 多因子疾患

複数の遺伝因子に加え、環境・生活習慣や老化が関わって発症する病気。生活習慣病やがんなどが代表例として挙げられる。

#### 追跡

対象者を長期間観察し、臨床検査値の変化、疾患の新規発症や再発、死亡（死因を含む）等を把握すること。

### ながはま0次予防コホート事業（ながはまコホート）

京都大学が、滋賀県長浜市、一万人の長浜市民と連携して実施している、日本初の本格的な大規模ゲノムコホート研究。

### バイオインフォマティクス

計測された生命情報の情報解析法を開発し、生物学的な発見を導き出す学問領域。当該領域に関わる研究者や技術者をバイオインフォマティシャンという。

### バイオマーカー

ゲノム、生体試料中の特定の物質、臨床検査値などの中で、疾患の発症や増悪と関連し病態把握や予測に貢献する因子。

### ベースライン調査

コホート研究で事業参加者を登録する際に行う検査・調査。

### ライフログ

ヒトの生活や動きなどを映像や音声、位置情報などを経時的にデータとして記録すること。

### 臨床情報

血圧や心電図などの検査結果、罹患・治療・投薬に関する情報、問診調査結果など個人の保健・医療に関する全ての情報。

### 連結可能匿名化

生体試料や情報を用いて解析を行う際に、研究者に個人が特定できないように個人識別情報（氏名、住所など）を記号や番号で置き換える操作を匿名化と呼び、記号・番号のことを匿名化記号・番号と呼ぶ。個人識別情報と匿名化記号・番号の対応表を残す場合を連結可能匿名化、対応表を残さない（二度と対応が付けられない）場合を連結不可能匿名化と呼ぶ。

## <参考資料 1> 審議経過

平成 24 年 (2012)

- 1 月 27 日 日本学術会議幹事会 (第 144 回)  
ゲノムコホート研究体制検討分科会設置
- 2 月 20 日 日本学術会議幹事会 (第 146 回) 委員の決定
- 4 月 4 日 ゲノムコホート研究体制検討分科会 (第 1 回)  
役員を選出、今後の進め方について、活動予定について
- 5 月 7 日 分科会 (第 2 回)  
ヒアリング、委員からの報告、今後の進め方について
- 5 月 28 日 分科会 (第 3 回)  
ヒアリング、委員からの報告、提言の構成について
- 6 月 28 日 分科会 (第 4 回)  
ヒアリング、提言の構成について
- 7 月 9 日 分科会 (第 5 回)  
提言案について
- 7 月 24 日 分科会 (第 6 回)  
提言案について
- 11 月 22 日 分科会 (第 7 回)  
ヒト生命情報統合研究推進協議会設立に向けて
- 12 月 21 日 日本学術会議幹事会 (第 167 回)  
ヒト生命情報統合研究推進小委員会設置

平成 25 年 (2013)

- 1 月 23 日 分科会 (第 8 回)  
シンポジウムの総括、今後の在り方について
- 2 月 22 日 日本学術会議幹事会 (第 169 回) 委員の決定
- 4 月 9 日 ヒト生命情報統合研究推進小委員会 (第 1 回)  
役員を選出、サブ委員会の設立、今後のスケジュール
- 5 月 28 日 小委員会 (第 2 回)  
サブ委員会からの報告、今後の活動、小委員会の対応
- 6 月 25 日 小委員会 (第 3 回)  
サブ委員会からの説明、小委員会のまとめ、今後のスケジュール
- 7 月 3 日 分科会 (第 9 回)  
提言案について
- 月○日 日本学術会議幹事会 (第○回)  
日本学術会議第二部ゲノムコホート研究体制検討分科会 提言  
「100 万人ゲノムコホート研究の実施に向けて」について承認