

ゲノム情報と電子化医療情報等の統合によるゲノムコホート研究の推進
大規模分子疫学コホート研究の推進と統合

わが国におけるこれからの大規模分子疫学コホート研究の構築と推進に向けた研究のあり
方に関する提言

実施期間:平成 23 年度～平成 25 年度
代表機関名:(独)国立がん研究センター
総括責任者:堀田 知光
研究代表者:津金昌一郎

目次

はじめに

I. 大規模分子疫学コホートの構築

1. 標準プロトコールによる対象者のリクルート
 - 1-1) 対象者の定義と登録方法
 - 1-2) 実施体制
 - 1-3) 調査に必要な費用
 - 1-4) 共通プロトコールを用いて集団健診会場で調査を行った場合の課題
2. 調査票
3. 生体試料の収集
4. 生体試料の保管方法
5. 災害時バックアップ体制
6. 追跡調査方法について
 - 6-1) 異動・生死
 - 6-2) 死因
 - 6-3) がん罹患
 - 6-4) 循環器疾患罹患
 - 6-5) 糖尿病
 - 6-6) 精神疾患(うつ病、認知症)
 - 6-7) 医療情報の利活用
7. 倫理的・法律的・社会的問題について
 - 7-1) 標準プロトコールの利用と課題
 - 7-2) 保存試料・情報の将来の利用目的等を幅広く設定した場合の同意のあり方について
 - 7-3) 新たな情報ソース活用の可能性と医療等 ID
 - 7-4) ゲノム解析結果の利活用について
 - 7-5) 偶発的所見について
8. 個人情報およびデータ管理のあり方
 - 8-1) 疫学指針・ゲノム指針に適応した ID 管理について
 - 8-2) 個人情報保護・安全管理について

II. ゲノム解析

1. 血液試料からの核酸抽出と保管システムについて
2. ゲノム網羅的SNP解析
 - 2-1) サブコホート 2900 人による日本人多型データの収集
3. リシークエンシング等による次世代分子疫学的コホート解析
 - 3-1) 稀な多型検出のためのエクソームデータ解析パイプラインの検討
 - 3-2) 末梢血を用いたメチロームデータと生活習慣等の関連解析
 - 3-3) インフォマティシヤンの育成
4. 生活習慣・環境要因と相関する候補遺伝子多型解析
 - 4-1) 遺伝環境相互作用について
 - 4-2) Mendelian randomization 法について
 - 4-3) リスク予測モデルについて
5. 大規模分子疫学コホート研究におけるゲノム解析のデザイン・サンプルサイズの検討

III. 他研究との統合

1. 調査票情報の統合に関する提言
2. 生体試料情報統合に関する提言

おわりに

はじめに

ライフ・イノベーションによる健康大国の実現のため、総合科学技術会議では平成 23 年度科学・技術重要施策アクション・プランを策定した。このプランでは、疾患予防の観点から「予防医学の推進による罹患率の低下」を重点課題とし、「ゲノムコホート研究と医療情報の統合による予防法の開発」を推進することとしている。そこで、基本的に健常な一般住民よりなる十万人規模のコホート研究を推進するため、①対象者の生活習慣・環境要因等に関する情報収集のための追跡調査(疫学・コホート研究)の実施、②血液等の生体試料を採取、保存し、ゲノム、バイオマーカー等の解析を実施するための基盤の整備、③対象者の医療情報を、ICT 及びネットワークを効果的・効率的に活用して、収集し、対象者一人一人の疫学・コホート研究の上記データとの統合を実現し、疾患リスクにおけるゲノムと環境因子との相関や疾患メカニズムの解明、薬物反応(作用、副作用)の発見・評価の実現等と、最終的には人類の健康・医療の増進に貢献することが期待されている。

本プロジェクトは、上記アクション・プラン実施の準備段階研究として、数十万人規模の分子疫学コホート研究を実施にあたっての研究目的の設定と研究推進体制(対象者の登録方法、追跡体制、生体試料の採取・保管体制、個人情報取扱い等)の検討、標準プロトコル案の策定及び実際の地域住民集団を用いた適用可能性の検証による標準プロトコルの改善と最適化を実施し、最終的に、大規模分子疫学コホートを構築・推進するために必要な方法論を確立し、その最適プロトコルを提示するものである。ここでは、近年の ICT 戦略に基づいて進展している電子化医療情報を効果的・効率的に活用し、将来的に、コホート研究のデータと医療情報の統合させる方策についても検討する。さらに、今後 10-20 年間に本格化する分子疫学コホート研究の解析において予想される必要技術・試料・情報の予備的検討を通して、特にゲノム解析に必要な各種研究資源・体制・データベース構造・データ解析・保護技術等に関する基礎的情報も提示する。

本書では、上記の内容について、本プロジェクトにおいて検討した内容を、「わが国におけるこれからの大規模分子疫学コホート研究の構築と推進に向けた研究のあり方に関する提言」として取りまとめた。

II. 大規模分子疫学コホートの構築

1. 標準プロトコールによる対象者のリクルート

1-1) 対象者の定義と登録方法

分子疫学コホートの構築のためには、生体試料の収集が必須であるので、特定健診会場など会場招致方式の住民健診会場で、健診時に余分に試料の提供の同意を依頼するのが最も現実的で効率的である。

- 対象者は、我が国における特定健診対象年齢と同じである市区町村に住居登録されている40-74歳の日本人男女とするのが現実的・効率的である。同意を得て、生活習慣情報(調査票)、健(検)診結果、血液試料の収集を行うことが最低限必要である。そのほか、目的に応じて尿試料などの他の生体試料収集を行うことを推奨する。また、生活習慣の変化を把握するため、数年毎の繰り返し調査が推奨される。
- 住民健診会場などで調査を実施するため、市町村、および、健診実施団体、医師会などの理解と協力が必須である。健診受診率をあげることが住民の健康づくりの一つの大きな目標であることから、健診会場で研究を行うことによる混雑などで、住民から苦情が発生し、次年度の健診受診率が低下することはあってはならないことであり、最大限の工夫・努力をするとともに、市町村等健診実施主体側に研究の意義を説明し、調査で得られた情報を還元することで、市町村の健康施策づくりに役立つ可能性があること、住民の健康向上につながることを目的としていることを理解していただく必要がある。また、研究参加による住民への利点があれば(研究用に提供される血液の一部を使い、抗ヘリコバクターピロリ抗体とペプシノゲンの測定による胃の健康度チェックなど)、研究を行うことが受診勧奨につながる場合もある。
- 健診会場で待ち時間が増えることは、研究への同意率が下がることのみならず、次年度に受診者が健診を受けない理由となり、健診受診率の低下につながり得ることから、協力自治体も研究への協力を避ける一因となる重要な要素である。待ち時間を最小限にするために、説明は健診受付の待ち時間にDVDを用いる、集団で効果的に行う、など工夫はしているが、さらに健診会場で誘導する人員や説明同意を行う人員を増員するなど、さらなる改良が必要である。
- 本研究におけるコホート調査でのインセンティブとして、食事調査の結果と、健康についてこれまでの研究からわかる範囲での結果を返却し、調査票への回答などにかかる時間への対価として1,000円相当の品物を渡している。その他、将来の(次世代への)研究成果の還元が間接的なインセンティブになるかもしれない。地域の実情に応じて工夫して行われ、参加率の向上が図られるべきであるので、標準プロトコールでは、調査票の配布・回収の方法を特に定めない。ただし、血液試料等の収集を必須とする標準プロトコールでは、主なリクルート方法が住民健診会場に限定される。住民健診受診率の低い地域で研究への同意率を維持するには、地域住民の参加を促す工夫が必要になり、研究を行うことで受診勧奨につながる場合もある。ある地域において研究の特典付与により特定健診の受診率が高まったという報告がある。一方、ヒトゲノム・遺伝子解析研究に関する倫理指針(ゲノム指針)では「試料・情報の提供は無償であること」としており(第3の7の(11)の細則)、参加者と自治体の双方が適切なメリットを見いだせるような研究計画の立案・遂行を、地域研究事務局が担うべきである。

1-2) 実施体制

住民健診会場で行うべき、主な研究側の役割は、

1. 説明同意取得
 2. 研究で必要な項目の問診
 3. 尿試料取得時の補助
 4. 血液試料取得時の補助(場合によっては分注作業)
 5. 会場誘導
 6. 可能であれば調査票確認
- である。

1-3) 調査に必要な費用

分子疫学コホートとして国際的に通用する規模が数十万人以上であること、日本の各地の多様性から、一地域で実施することは現実的ではない。そこで、数か所の地域を地域研究事務局として、それをまとめるコホート中央研究事務局が設置されることが想定される。標準プロトコールに基づく研究において、中央研究事務局は研究の統括的な運営を担い、各地域研究事務局が地域の実情に応じた研究の遂行を分担するモデルを提案する(表-1)。

表-1. 標準プロトコールによる研究体制のコホート中央研究事務局と地域研究事務局の役割

中央研究事務局 (データシステム管理者 データマネージャー 解析担当者 サイトとの窓口担当者 研究倫理支援担当者 広報担当者)	研究の計画と遂行	標準プロトコールの立案、SOPによる共有化と企画運営 サイトとの連絡、訪問、応援、助言等 委員会、会議や研修会等の企画・開催
	社会的側面への対応	倫理・法・社会面への対応 リスク管理・危機管理 広報
	研究の運営と管理	研究全体の体制維持・運営 研究費の獲得、サイトにおける研究費の獲得支援 サイトの体制維持・運営補佐
地域研究事務局	研究の計画と遂行	地域人口、対象年齢(40-74歳)人口から参加者(目標)の立案 参加者(目標)年間スケジュールと全体スケジュールの立案 サイトにおける研究計画書の準備、IRB審査
	研究実施準備	サイトにおける調査手順書、安全管理マニュアルの作成 人材の確保と研修教育 研究資材の調達管理 外部機関との委託契約
	調査票の配布	研究対象候補者全員に配布(郵送の場合)
	対象者リクルート	人員の配置と交通の手配、資材の運搬
	試料・情報収集	採血、遠心分注、一時保管(-80℃) 問題事例の記録と報告
	結果返却	
	試料・情報管理	一時保存 中央事務局へ送付
	対象者追跡	
	地域住民への成果還元	地域研修会等の実施

コホート中央研究事務局には、データシステム管理者、データマネージャー、解析担当者、サイトとの窓口担当者、研究倫理支援担当者および広報担当者を配置するのが望ましい。今回、中央研究事務局の費用は、追跡をどの程度行うかにより様々であるので、試算対象外とする。また、あくまでもコホート調査の立ち上げと維持の部分の費用であり、調査の結果得られた試料・情報の解析・研究の費用は含めていない。

地域研究事務局のモデル地域として、人口 10 万人程度、対象年齢(40-74 歳)人口 4 万人、参加者(目標)1 万人を 4 年間でリクルートする設定で試算を行った。対象地区を 4 つに分け、各年度に研究対象となる地区で調査票と説明書を対象候補者 1 万人に配布、特定健診会場にて 100 人×25 日間=2,500 人のリクルートを行う計画とする。地域に応じた研究計画書の検討、実施準備、調査票の配布、試料・情報収集、結果返却、試料・情報管理、対象者追跡等を実施する。また地域住民へのアウトリーチ活動も必要である。この流れに添って、必要な研究費用の試算を行った。

地域研究事務局で具体的に行うことは、以下の項目である。費用は、表-2 を参照。

- 1) 調査開始前
 - ① 研究計画書、説明・同意文書等の作成、倫理審査対応・承認手続き
 - ② 自治体・健診実施団体・医師会等との打合せ
 - ③ 調査手順書の作成
 - ④ 人材の確保と研修教育
 - ⑤ 研究資材の調達管理
 - ⑥ 外部機関との委託契約
- 2) 調査実施中
 - ① 調査票を 1 万人に配布(今回は郵送を想定)、特定健診会場にて同意取得
 - ② 人員の配置と交通・宿泊の手配

- ③ 資材の運搬
- ④ 採尿・採血・遠心分離・分注
- 3) 調査実施後
 - ① 生体試料の一時保管(−80℃での試料凍結保管)
 - ② データ整理
 - ③ 研究参加者への結果返却
 - ④ 問題事例の対応、記録と報告
 - ⑤ 中央研究事務局へ資料送付
 - ⑥ 地域住民への成果還元として地域講演会等の実施
 - ⑦ 対象者の異動・死亡追跡情報収集
 - ⑧ 自治体への研究結果の還元

仮想モデル地域での、初年度の費用は 47,690(千円)であった。この中には継続的な人件費(研究員 1 名、事務員 1 名)の他、設備費としてディープフリーザー(DF)2 台分 6,000(千円)が含まれる。2 年度目から 4 年度目までの費用は DF を除く 41,690(千円)である。4 年間で総計 172,763(千円)であった。なお、以上の試算には、いずれも消費税は入っていない。本試算の結果をもとに実情に合わせて練り直すことで、特定の地域への応用が可能である。

表-2. 調査実施費用試算

	単価	数	初年度 年額	2-4年度 年額	4年合計
人件費					
研究員(非常勤)		1	3,900,000	3,900,000	15,600,000
事務員(非常勤)		1	3,200,000	3,200,000	12,800,000
IC、アンケート確認25名×30日	8,000	750	6,000,000	6,000,000	24,000,000
クオカード(アンケート記入への謝礼)	1,040	2,500	2,600,000	2,600,000	10,400,000
研修費用(25名)	22,000	25	550,000	550,000	2,200,000
国内旅費(25名×30回)	2,000	750	1,500,000	1,500,000	6,000,000
国内旅費(会議等4名×4回)	50,000	16	800,000	800,000	3,200,000
調査票	193	13,000	2,510,300	2,510,300	10,041,200
説明文書	40	10,500	420,000	420,000	1,680,000
チラシ	11	10,500	115,500	115,500	462,000
送信用封筒	39	10,500	409,500	409,500	1,638,000
返信用封筒	28	10,000	280,000	280,000	1,120,000
封入発送作業	94	10,000	940,000	940,000	3,760,000
送料	144	10,000	1,440,000	1,440,000	5,760,000
同意書(複写式)	46	3,000	138,000	138,000	552,000
採血調査票(複写式)	45	3,000	135,000	135,000	540,000
ペプシゲン ヘルコバクターピロリ抗体IgG測定	1,250				
チューブ類、遠心分注費用等概算	750				
生体試料処理費用概算	2,000	2,500	5,000,000	5,000,000	20,000,000
データ加工			2,000,000	2,000,000	8,000,000
読み込み (読取、画像データ化、パンチデータ化、栄養計算結果出力、封入)	795	2,500	1,987,500	1,987,500	7,950,000
送料	82	2,500	205,000	205,000	820,000
その他消耗品(月額10万)	100,000	12	1,200,000	1,200,000	4,800,000
その他通信運搬費(月額3万)	30,000	12	360,000	360,000	1,440,000
その他雑役務費(月額10万)	100,000	12	1,200,000	1,200,000	4,800,000
設備費(DF)(初年度のみ)	3,000,000	2	6,000,000	0	6,000,000
追跡調査費用(人件費5名 30日)	8,000	150	1,200,000	1,200,000	4,800,000
データ入力費用(人件費5名 60日)	8,000	300	2,400,000	2,400,000	9,600,000
アウトリーチ活動費(委託費、HP等、ポスター等含む)			1,200,000	1,200,000	4,800,000
合計			47,690,800	41,690,800	172,763,200

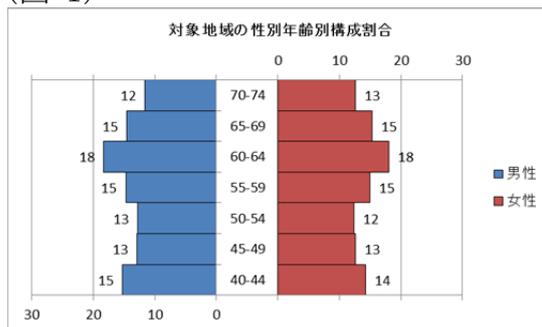
- 4年間で1万人の同意を得られるような分子疫学コホートの構築のためには、追跡実施のためのコホート中央研究事務局の人件費など諸費用を除いて、毎年5,000万円が少なくとも必要である。継続的な研究費の獲得が必要である。
- 標準プロトコールのような公益性の高い研究計画には利益相反があってはならない。公的資金や結果の客観性が保証される資金で賄われるべき研究である。しかしながら、用途について研究費ごとに制限が異なるのは問題である。実際に本研究では研究用採血での保険にかかる費用がまかなえないなど、他の研究費では認められるものが使えないことがあった。また、説明書や同意書などの大量に使用する印刷物に研究名やロゴを入れることを義務付けている研究費で、途中で名称が変わり、年度が変わると印刷物などが使えなくなったりするのも無駄である。封筒や調査票などの印刷物はロットにより単価が大きく異なるが、その都度使用する数だけ注文しなくてはならないなどの方法も、無駄が多い。
- 研究者のみならず、事務局業務にも十分なスキルと経験が求められるが、そのような人材を単年度雇用や非常勤雇用で賄うことは無理がある。研究体制を明確にし、研究倫理に基づく適正な研究費の使用をモニターするシステムを構築し、年度を超えたプロジェクトの実施にも柔軟に対応できるような資金の運用方法を検討すべきである。
- 大規模分子疫学コホート研究は長期縦断研究で長期間に及び、繰り返し調査があること、集団健診会場以外でも同意を得るためには、本研究で費やした以上の作業と人員を増やす必要があり、長期間の一定した、さらなる予算の確保が必要となる。
- この研究体制モデルにおいて、コホート中央研究事務局の担う役割は、各サイトにおける新規コホート研究を実施する際の共通課題を検討し、コホートデータ統合を実現するものである。そのため、専任の ELSI 担当者など、標準プロトコールの運営に必要な担当者を配置し、諸課題に関する研究の適切な運営方法について継続的に検討することも大切である。

1-4) 共通プロトコールを用いて集団健診会場で調査を行った場合の課題

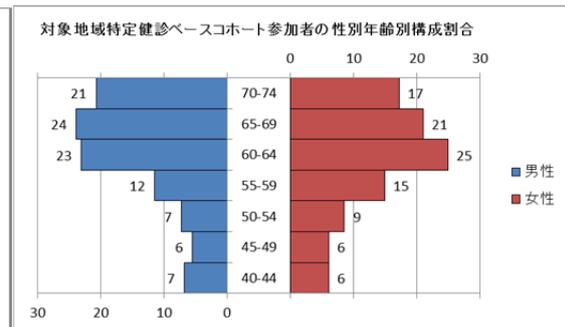
二つの地域で、共通プロトコールを用いて、40-74歳の健診受診者に、集団健診会場で同意を得て、生体試料の収集を行ったところ、70~80%以上の高い同意率が得られた。

しかし、地域一般集団の性別年齢別構成割合(図-1)と比較すると、研究参加者は60歳以上が多くなるという偏りが生じた(図-2)。

(図-1)



(図-2)



- 同意率がある程度高い共通プロトコールの使用により、各新規コホートが統合可能となり、国際的評価にも耐えうる、日本人の代表性と公益性の高い、国民の健康の維持・増進、がんなどの疾病予防にとって質の高い大規模分子疫学コホート研究の構築が実現できる。しかし、異なるプロトコールを用いている既存のコホート研究を統合するには、本研究で検討しているように、統合手法の開発が必要である。今後は、特に年齢構成の点で、より地域全体を代表する一般集団から同意を得て生体試料の収集を行うための検討が必要である。
- 特定健診会場での生体試料収集は、年齢構成に偏りが生じることから、より一般集団に近づけるために、職場健診・人間ドックでの調査の協力、および、健診実施団体の協力や研究者による実

施により研究用採血の機会を設けるなど、対象地域の住民から高い同意を得るための工夫が必要である。

- 集団健診会場以外で簡易に生体試料を収集できるよう、血液以外から(たとえば、口腔内粘膜など)から DNA サンプルをとる方法も今後は検討する余地があるかもしれない。

2. 調査票

調査票は、食習慣などの生活習慣要因や、既往歴、家族歴、社会経済状況、心理要因等を把握するためのものである。将来の他研究集団との統合解析の実現可能性を考慮し、各質問項目については、その質問方法や選択肢(カテゴリー分け)を先行研究と合わせ、また、原則として、国内外の先行研究において妥当性の確認されている質問方法を採用する。妥当性の確認されていない質問方法の場合は、妥当性研究が必要である。

3. 生体試料

生体試料としては、血液試料(血漿、血清、白血球 DNA、赤血球)、尿試料を基本として収集する。生体試料の収集拠点が国内広域にまたがるため、生体試料の収集・処理の方法を調査地域間で標準化しておく必要がある。また、生体試料を中央研究事務局へ輸送する方法についても、標準化しておく必要がある。生体試料の収集・処理から中央研究事務局への輸送までの方法は、詳細を手順書に定めて、地域における研究実施担当者に周知することが重要である。

4. 生体試料の保管方法

生体試料の保管は、調査地域毎に行うのではなく、人的・物的リソースを集中させ中央研究事務局で行うことが効率的であると考えられる。ただし、全ての生体試料を一カ所で集中保管すると、大規模災害等で生体試料を破損するリスクが高まるため、分散保管を併用するのが望ましい(詳しくは、「5. 災害時バックアップ体制」を参照)。生体試料の保管方法としては、 -80°C の超低温槽を用いた保管システムが十分機能することが確認されており、現時点で適当な方法と考えられる。近年の技術革新により、液体窒素を用いて、ロボットが生体試料の管理を行うような大規模保管システムが開発されているが、長期運用の実績はまだ乏しい。また、初期費用が高額になるため、本提言書で試算した初年度調査実施費用(約 50,000 千円)では導入が困難である。

5. 災害時バックアップ体制

コホート研究は、長期にわたる研究なので、研究期間中に大規模災害等が発生し、生体試料を破損するリスクを想定しておく必要がある。効率の面では、中央研究事務局で生体試料の保管を集中的に行うのが望ましいが、集めている生体試料を調査地域や関係機関で多重的に管理するなどして、災害時のバックアップ体制を構築しておく必要がある。ただし、調査地域や関係機関にバックアップ体制を構築する場合、中央研究事務局ほど人的・物的リソースを割くことは通常困難なため、停電や超低温槽の故障等の事故により生体試料を破損するリスクを高める可能性について十分注意が必要である。

6. 追跡調査方法について

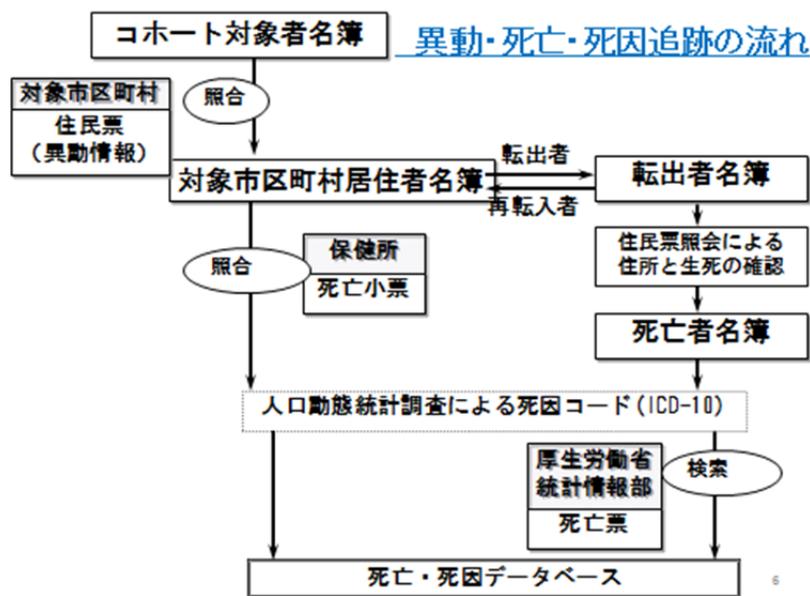
高齢化社会が急速に進行する我が国において、死因及び、①がん②脳卒中③急性心筋梗塞といった循環器疾患、④糖尿病、⑤精神疾患(うつ病、認知症等)の5大疾病の予防法の確立が求められ、それら疾病の追跡が必要である。以下にこれらの追跡方法についての概要と検討事項を示す。また、追跡調査における既存医療情報の利活用について整理した。なお、以下の方法は、研究に対する本人同意の得られていることが前提である。

6-1) 異動・生死(図-3)

死因は、厚生労働省へ統計調査の調査票情報の提供を申請し、承認後に提供される人口動態調査情報で把握することとなるが、その際に必要な情報は、死亡時の住所と死亡年月日である。これらの情報を把握するためには、研究期間中、住民票照会による対象者の住所異動と生死の確認の継続が必須になる。

- 住民票の写し(紙)が発行された後は、追跡情報を整理するために、電子化処理に要する日数と人件費が必要となる。統合の際には、各施設で電子化したフォーマットやソフトが異なる場合は、書式を整える必要があるが、これは不可能ではない。
- 住民票の写しの交付等の事務処理の実際は、個人情報保護法関連の条例で定められているため、役所によって異なっている。大規模分子疫学コホート研究では、各自治体への照会にかかる費用が膨大になりかねない。よって、データベース整備の実現を目指し、それまでは照会にかかる費用を免除するような制度が望ましい。
- もし、米国で行われている National Death Index(NDI)のような電子化システムが存在し、Social Security Number などの照合キーを用いて個人の同定をしたうえで研究利用可能ならば、現在、死因を把握する前提として行っている、住民異動・生死の確認のための住民票照会・請求は不要となる。照合キーを用いて個人を同定したうえで死因を把握できる電子化システム整備が望まれる。

図-3 異動・死亡・死因追跡の流れ



6-2) 死因(図-3)

上述の方法による住民票照会・請求にて対象者の死亡を確認した場合には、所属機関や統計法(平成19年法律第53号)第33条の規定に基づき、人口動態調査の死亡票、死亡小票、転写 CD-R の利用申請を行い、事務局で保持している対象者の氏名、性別、生年月日その他、死亡年月日、死亡時の住所情報から、死因の把握が可能である。

- 人口動態調査の死亡票、死亡小票、転写 CD-R の利用を行うならば、申請作業・照合作業に関わる人員が必要であるが、情報源は共通なので統合は容易である。
- 現在は、個人の同定を研究対象者の氏名、性別、住所地、生年月日、死亡住所、死亡年月日、で行っているが、NDIのように Social Security Number といった照合キーが存在すれば、個人の同定がより確実に行われ、照合したレコードに該当する死亡票・死亡小票の原本を直接閲覧することも必要なくなり、より効率的に行えることは追記しておく。個人の同定の効率化・正確性をあげるためには、共通番号制度におけるマイナンバーなど照合キーを、公益性が大きい医学研究で利用できるように、法制度を整備する必要がある。

6-3) がん罹患

我が国でがんの罹患を把握するためには、(1)地域がん登録、を活用することが最も効率的と考えられる。しかし、地域がん登録は2012年に全国に展開されたが、まだ開始されたばかりの地域や精度

があまり高くない地域では、(2) 協力医療機関からの情報の系統的収集、(3) 自己申告による医療機関からの情報収集、(4) 死亡小票から、といった、そのほかの方法で情報を補充する方法を考えざるを得ない。

- 2013年12月にがん登録推進法が成立した。現在行われている、地域がん登録の限界として、がん診療拠点病院以外の施設からの地域がん登録への届出が少なく、また、対象者が他県の医療機関でがんと診断された場合、本来、対象者が在住住所の地域がん登録に登録されるはずであるが、登録されずに漏れが生じているのが現状である。全国がん登録が開始されると、すべての病院に、がん患者の届け出が義務づけられ、今後は全国がん登録として、国立がん研究センターに全国のがん罹患データが集約し、定められた研究利用申請後、データが利用可能になる予定である。がん罹患データが全国がん登録から把握されれば、情報源が同じであるので統合は容易である。今後の全国がん登録が精度よく管理されること、公益性の高い疫学研究では、全国がん登録を用いてがん罹患把握を可能にすることを提言とする。
- 全国がん登録が整備されるまでの間は、地域がん登録を活用することが有効であるので、対象地域の地域がん登録システムの精度を確認しながら、精度向上の必要がある場合には、対象地域の病院へ地域がん登録への疾病登録を促すことが、また研究の追跡精度の向上においても重要である。
- 現在は、院内がん登録・地域がん登録を利用するために、対象者名簿との照合が必要であるが、マイナンバーなどの照合キーがあれば作業の効率性・正確性が増すため、全国がん登録も含めて、公益性の高い疫学研究に利用可能とすることを検討することが重要である。

6-4) 循環器罹患

わが国には、がんと異なり、循環器疾患の登録システムが存在していない。しかし、がんと同様、循環器疾患は早期死亡の主要な疾患であり、登録される循環器疾患は、わが国の公衆衛生の向上のために重要な知見が得られるだけでなく、研究の国際的競争力を高める上でも、これらの予防を対象とした研究は有用である。

- 共通の循環器登録票を用いれば、データの統合は容易であるかもしれない。しかし、我が国に登録システムがない現状では、協力医療機関や研究者の熱意と能力に依存しているのが実情であり、データの精度や質を評価できる指標がない。がん登録と同様、循環器疾患も全国で登録されるようなシステム導入の検討が必要である。

6-5) 糖尿病罹患

循環器疾患と同様、わが国には、糖尿病登録は存在しない。我が国の糖尿病の有病率は増加しており、糖尿病の予防要因を明らかにしていくことは重要な我が国の課題である。系統的レビュー・メタ解析から得られたエビデンスに基づき(PLoS ONE 8(9): e74699)、以下のように糖尿病に関する追跡情報の収集方法と標準化する方法が提案される。

- 1～5年毎に糖尿病に関する情報(HbA1c、自己申告の糖尿病診断や治療状況など)を収集することにより、糖尿病自己申告と検査値を組み合わせることで糖尿病罹患を定義することが望ましい。ただし、大規模研究では全サンプルでHbA1c測定が困難な場合もあり、自己申告の情報のみを用いて糖尿病を定義することも考えられる。
- 標準化された基準で定義された糖尿病罹患は、メタ解析手法で統合することが可能である。異なる基準で定義された糖尿病罹患は、罹患率の違いなどによる異質性に十分に留意して、変量効果モデルで統合することが望ましい。

6-6) 精神疾患(認知症・うつ病)

わが国では、急激な高齢化とともに、認知症患者数は増加の一途をたどっている。また、認知症になることで周囲の精神的・身体的負担、社会の経済的負担が大きくなることを考慮すると、認知症に罹患することなく天寿を全うすることは高齢化社会の重要な課題である。一方、海外における先行研究から、うつ病の初回エピソードがみられる平均年齢は20歳代であり、40代以上のコホート研究では発

症を把握するのが困難であるが、近年、うつ病が認知症のリスク因子であるという報告が相次ぎ、高齢者のうつ病も予防すべき非常に重要な課題となっている。現在、定められた認知症登録やうつ病登録は存在しないので、本研究で行った登録法の可能性・実現性の検討をもとに提言とする。

- 認知症登録には、(1)協力医療機関からの情報の収集、(2)介護保険情報からの収集の可能性が考えられたが、協力医療機関の負担、介護保険情報の妥当性など、登録の実現性については今後も検討が必要である。
- うつ病登録には、(1)協力医療機関からの情報の収集、(2)レセプトからの情報の収集、(3)自己申告による情報の収集の可能性が考えられたが、協力医療機関の負担、レセプト情報から正確に対象者を把握する方法の検討、自己申告による発症把握の妥当性検討など、登録の実現性については今後も検討が必要である。
- 医療機関に登録票の作成を依頼する方法では、認知症と正常のグレーゾーンである軽度認知機能障害 MCI (Mild Cognitive Impairment) の把握は不可能である。近年の疫学研究では、MCI の状態で、運動習慣や食習慣の改善を行うことにより、認知機能の低下を予防する可能性が示唆されている。MCI を把握するためには、見かけ上、地域で健康に暮らしている方を対象に、メンタルヘルス検診を実施し、MCI を把握する以外にない。今後は、認知症のみならず、MCI の発症登録もすすめるべきであると考えらる。

6-7) 医療情報の利活用

(1) 病理組織

今後のがんの原因究明・予防法開発研究に必須であると考えられる、腫瘍組織を分子病理学的特徴(体細胞ゲノム・遺伝子変異、DNAメチル化プロファイル、mRNA・タンパク質発現プロファイルなど)をもとにサブタイプに分けリスク要因との関連を検討する際の腫瘍組織検体の利用においては、ホルマリン固定パラフィン包埋(formalin-fixed paraffin-embedded; FFPE)検体の利用が現実的である。ただし、RNAでの発現解析は、全トランスクリプトーム解析ではなく、特定の候補遺伝子によるクラスター分類までなら一部可能なレベルである。標本のブロック番号は、個人情報に準じた扱いをするので、FFPE検体の提供には各施設の了解が必要と考えられる。一方、FFPE検体から未染色スライドを作製する場合は、スライドに匿名化された番号を書けば良いので、より収集の手続きがしやすい。必要に応じ業務委託を利用し、匿名化された未染色スライドを、症例が十分蓄積されてから収集するのが現実的である。

(2) レセプト情報・健診情報

国保など保険者単位でデータが得られる場合も一部あるが、費用・時間がかかること、保険の種類が変わった場合に追跡できなくなるなど、非効率的である。ナショナルデータベース(NDB)がコホート独自のデータとリンケージ可能になれば、統合されたレセプト情報・健診情報の効率的な利用が可能となる。データリンケージについては、共通番号制度におけるマイナンバーが利用できることも必要である。

(3) 電子カルテ情報

大規模コホート研究の追跡調査に既存の電子化医療情報(電子カルテ等)を利用することは、登録者からの同意があれば個人情報保護法上、特に問題はない。しかし、かな氏名、生年月日等、個人情報そのものを用いた既存の電子化医療情報からのデータ抽出方法の精度は不十分で、効率が悪く、また、突合キーとして、氏名など明らかな個人情報を各施設に提供しなければならず、個人情報管理の観点からもリスクを伴う。

データ抽出の精度をあげるには、大規模コホート研究の登録者情報と既存の電子化医療情報との照合キーとしてマイナンバーが活用できることが必要である。データ抽出の効率を上げるには、全国規模の悉皆性のある電子化医療情報のデータベースが整備され、コホート独自の情報とマイナンバーによりリンケージできるようになることが必要である。具体的には、まずは、NDB、DPC データがマイナンバーでリンケージ可能となるよう整備すべきである。その他の電子カルテ情報については、データ形式の標準化をさらに進め、悉皆性のある大規模データベースを構築していく必要がある。

7. 倫理的・法律的・社会的問題について

7-1) 標準プロトコールの利用と課題

- 倫理的・法律的・社会的問題(ELSI)に配慮した標準プロトコール作成の経緯:

平成 24 年にパイロット調査として開始された本プロジェクトでは、研究推進に伴う ELSI について、特に、ゲノム試料収集を含む研究資料の収集と利用についての説明と同意、同意の撤回のあり方に重点を置き、外部委員を含む ELSI 検討委員会を設置し、議論した。この委員会には、推進委員会の外部メンバーである生命倫理や法律の専門家が含まれ、ELSI に配慮し、かつ今後の社会状況の変化にも対応できるようなインフォームド・コンセントや結果の公開等について検討を重ね、その結果を踏まえ標準プロトコールを完成させた。

- 標準プロトコールの課題:

- 研究施設における倫理審査委員会の役割の増大:

平成 25 年度に全部改正したゲノム指針が施行され、その指針の前文に「ヒトゲノム・遺伝子解析研究に多様な形態があることに配慮して、本指針においては基本的な原則を示すこととし、研究者等が研究計画を立案し、その適否について倫理審査委員会が判断するに当たっては、この原則を踏まえつつ、個々の研究計画の内容等に応じて適切に判断することが求められる。」という一文が追加され、多様な計画の申請が可能となった。標準プロトコールにおいても、当初の包括的同意に基づく将来の研究利用のためには、再同意取得の必要性も含め、倫理審査委員会による審査を前提として計画されている。倫理審査委員会の果たすべき役割は今後ますます大きくなると考えられる。

- 社会状況に応じた更新の必要性:

ゲノム指針の全部改正後、疫学研究に係わる指針と臨床研究に係わる指針の統合が進められている。このような流動的な状況下で最善のプロトコールの作成に努めたが、標準プロトコールを最新に保つには、研究方法と解析技術、法律と指針など ELSI を取り巻く状況を見渡して常にアップデートすること、古いバージョンで既に収集された研究試料をうまく利活用できるような仕組みづくりが必要である。

7-2) 保存試料・情報の将来の利用目的等を幅広く設定した場合の同意のあり方について

- 分子疫学コホート研究における包括的同意:

ゲノム指針において、「個人情報を取り扱うに当たっては、その利用の目的(利用目的)をできる限り特定しなければならない。」「また、利用目的を変更する場合には、変更前の利用目的と相当の関連性を有すると合理的に認められる範囲を超えて行ってはならない。」と記されている。しかしながら、コホート研究とは研究開始時に収集した保存資料に将来の追跡情報をリンクしながらデータベースを構築し、その時点で研究目的に最適の解析を行うという研究デザインであり、当初の試料・情報の収集時に利用目的を限定することは困難である。よって、利用目的等を具体的にはっきりと示すことが難しく、幅広い同意の範囲を取らざるをえない。

さらに、将来の研究利用時の再同意取得を前提とするものではなく、具体的研究課題毎の倫理審査委員会の承認により利用されるという方針を明らかにする必要がある。本提言書では、以上のような同意を分子疫学コホート研究における「包括的同意」ととらえた。標準プロトコールの説明文書は、最初の試料・情報収集時に将来にわたる包括的同意を得るためのものである。

- 包括的同意によらない研究計画は実現困難:

包括的同意によらない研究計画にするには、収集した試料・情報を利用する研究課題毎に再同意を得るという方針をとる必要がある。しかしながら、標準プロトコールの実施に際し、事情の異なる各地域において 10 万人以上という対象者数から利用のたびにインフォームド・コンセントの再取得を課せられると、膨大なコストとデータのロスが生じ、研究の存続自体が危うくなる。一方で、コホート研究の当初の目的であり、試料・情報収集時に同意を得ている研究で想定される範囲内の解析であれば、再同意がなくても参加者のリスクは最小限であり、研究結果によって得られる確

かなエビデンスという公衆衛生上のメリットをリスクが上回るとは考えにくい。

- 包括的同意における ELSI 補うために求められる研究体制:

- 人員とコスト:

分子疫学コホート研究では、研究者と参加者が双方向性の緊密なコンタクトを取ってその都度再同意を得るように促す意見もあり、倫理的には理想の姿である。しかしながら、大規模長期観察型の研究から期待される科学的成果を最大限にするためには、包括的同意のある研究を行うべきであり、また包括的同意で懸念される ELSI を補うためには、研究体制を整備する上で、研究倫理支援や広報活動などのための手厚い人員やコストを割く必要がある。

- 研究者に求められる誠実な対応:

最初のリクルートにおけるインフォームド・コンセントの段階で、分析計画をできるだけ詳細に記して丁寧に説明し、研究参加者への適切な研究内容の開示と同意撤回の手順を明確に提示しておくことである。あわせて、個人情報のための安全措置(匿名化と対応表の管理など)と、他の研究やバンクへの試料・情報提供にあたっての適切な処置(対応表を渡さないなど)等の具体的な手順を明記した安全管理マニュアルの遵守が必須である。情報が集約されるコホート中央研究事務局には、専任のデータ管理者およびデータマネージャーの配属が望ましい。

- 説明文書におけるゲノム解析の明快な説明の必要性:

全ゲノム解析の可能性の記載:特に対象年齢層である中高年において、社会のリテラシーとしてゲノム解析という概念が十分に浸透していない現状で、どんなに言葉を尽くしたとしても、一般住民に対しその意義を十分に説明できているかどうかについては、実際、非常に困難であるといわざるを得ない。そこで、特に現段階で考えられる最大限の解析である全ゲノム解析について、本研究の標準プロトコールにおける説明文書で明確に記述した。

個人情報との関連の記載:あわせて、その解析結果が公表される可能性、施設の倫理審査委員会による審査など一定のルールに基づき、利用目的や利用者が拡大された上で公的なデータとして共有される可能性、さらには個人情報の漏えいなどの条件が揃った場合には、ゲノムDNAの塩基配列情報から個人が特定できないわけではないことについて、説明文書において明らかにした。

- 研究計画・成果などの情報公開の必要性:

さまざまな考えを持つ幅広い参加者の存在:ELSI 上の課題として、このような幅広い同意における自己情報コントロールの権利が取り上げられている。いかなる場合も個人の自由を尊重すべきであるという理念に基づく。しかしながら、標準プロトコールにおいて一般住民を対象に参加を呼びかけていると、研究参加への動機は利他的な、社会のために役立つなら協力したいという善意に基づくことがほとんどであることに気づかされる。このような対象者に対しては、むしろ保存・収集された試料を「自分から離れたものであり、公衆衛生のための利用ならば抵抗がないので、細かな連絡でなるべくわずらわされたくない」という考えを尊重する必要がある。他方、何らかのインセンティブを期待したうえで参加であれば、「あくまで自分の試料・情報であり、将来にわたり利用目的を細かく知ったうえで制限したい」という考えもある。大規模長期追跡型の研究において、後者からの要求に応じることは、現状では難しいといわざるを得ない。

情報公開の必要性:さまざまな参加者への研究者の誠実な対応として、研究計画の概要や研究成果の情報公開を遅滞なく、綿密に行うことで補うべきである。少なくとも、中央研究事務局でそのような業務を行う広報担当者の配置などにより体制を整え、研究関連情報を系統的にわかりやすく公開するとともに、地域研究事務局の担当者と連絡をとりながら対象者に向けたアウトリーチを行う必要がある。

7-3) 新たな情報ソース活用の可能性と医療等 ID

- 医療情報等に関わる共通番号制活用の提言:

- 新たな情報ソース活用の可能性:

標準プロトコールでは、本人の同意を得たうえで第三者から直接情報収集が行われる。例えば、健診実施機関から健診情報を得てデータベースに蓄積するなどである。本人からの情報提供にはさまざまなバイアスがかかる場合があり、第三者からの客観的な情報の収集が不可欠である。そこで、その過程についてできるだけ丁寧に説明するように標準プロトコールを作成した。しかしながら、今後ビッグデータの活用が盛んになると、新たな情報ソースから健康情報を得て研究に利用することが望ましいような状況が生まれる可能性もある。

- 医療情報等に関わる共通番号制の活用:

新たな情報ソースからの健康情報の入手には、同意の問題と正確な突合という実務的な問題がある。同意については、倫理審査委員会の審査を受けることになる。突合の問題では、氏名や生年月日などの個人情報を突合項目として追跡情報等を追加すると、個人情報を取扱わなくてはならず、手間がかかり、欠損データが生じるリスクがある。そこで、個人情報を扱わずに情報を正確に突合できる仕組み作りに工夫が必要となる。研究開始前に医療情報等に関わる共通番号制が存在し、その使用についての同意を提供者からあらかじめ取得するのが理想的である。この点に関しても、現時点では国の方針が明らかになっておらず、標準プロトコールの説明文書では実現可能性を示すにとどまっている。今後、疫学研究と関連のある制度の整備状況をモニターし、利用可能な情報を適切に運用できる体制を作るために、積極的に働きかけることが必要である。

7-4) ゲノム解析結果の利活用について

- ゲノム解析結果の利活用をめぐる問題:

ゲノム情報を用いた個別化医療推進の実現可能性や必要性は広く認識されているが、生活習慣病に関する研究成果としてのゲノム情報の利活用のルールについては、国内はもとより国外においても一致した結論に至っていない。また、今後、遺伝情報による差別を禁止し、悪用を防ぐためにどのような法律や規制が定められるべきか、産業利用も睨みつつ議論が始まったところである。また、研究を取り巻く倫理観も時代に応じて変化している。

- 前向きな合意形成の必要性:

大規模分子疫学コホート研究の成果を次の世代に引き継ぐためには、その恩恵を誰もが受けられるような未来というしっかりとしたビジョンを国民の間で共有することが大切である。具体的には、研究者の責務として、予防医療の枠組み作りと普及に向けた取り組みや、遺伝情報を自分の健康や幸福のために上手に利用できる社会を築くという合意を形成するようなコミュニケーションの実施が求められる。

- 社会的合意形成のための提言:

ゲノム解析結果の幅広い利活用を前提とする研究の実施にあたっては、情報公開、広報(啓発)等の様々なコミュニケーションにより社会の不安を解消し、遺伝情報に基づく個人の多様性に基づく差別の問題を含めて、国民の正しい理解と判断を得るための努力を積極的に実施するとともに、法規制としても遺伝情報の社会にとって有益な活用の促進と不適切な扱いを禁ずる制度や遺伝情報による差別を禁ずる制度構築の可能性についても検討すべきである。

7-5) 偶発的所見について

- 偶発的所見への対応をめぐる状況:

標準プロトコールでは、遺伝子解析結果は原則非開示であるが、個人の健康に重大な影響を与えることが判明し、かつ、有効な対処法があるとき(偶発的所見)については、承認された研究計画に従い対応することとしている。偶発的所見への対応については、行わなくてはならないという立場と、むしろ行うべきではないとする立場があり、研究の多様性等に対応して、世界でも様々な議論があり、定まっていない。標準プロトコールでは一般住民が対象であり、全ゲノム解析を含む詳細な解析を予定していることから、偶発的所見の開示対象となりえる疾患が広範囲に及ぶと予

想される。しかしながら、その全ての分野に対する判断を1つの研究施設でカバーすることは難しい。

- 国としての対応方針の必要性:

すでに遺伝診療としての導入が普及している遺伝病については、比較的すぐに偶発的所見の報告対象とすべき遺伝子及び遺伝子変異がリスト化できそうであるが、現在のところまだ出来ていないため、関連学会等の動きを注目していく。また、現状では研究機関の倫理審査委員会ごとに個別の判断が行われることになるが、非常にセンシティブな判断であるため、少なくとも国内で一定のコンセンサスをもつ必要がある。よって、国としての遺伝情報開示検討のための委員会等の設置が望ましい。

8. 個人情報およびデータ管理のあり方

8-1) 疫学指針・ゲノム指針に適応した ID 管理について

分子疫学コホート研究には、ゲノム指針が適用されることとなるが、コホート研究として追跡調査をおこなううえで現状のゲノム指針の原則に従うことは困難である。ゲノム指針においては、研究に用いられる血液、組織など人体の一部そのもののみならず、その他の研究に用いられる情報も含めて試料等としている。その上で、「試料等の提供が行われる機関の長は、試料等を外部の機関に提供する際には、原則として試料等を匿名化しなければならない」こととなっている。そのためゲノム指針の原則にしたがうと、試料等の提供が行われる機関から、外部の機関に個人情報の提供を行うことができない。また、「試料等」に研究に用いられる情報がすべて含まれるために、追跡調査を行う場合には、たとえば住民異動情報を提供する市町村も試料等の提供が行われる機関として共同研究機関にならなければならない。

現状においては、分子疫学コホート研究の実施には一つの研究においてゲノム指針と疫学指針の両者を適用しなければ運用できない。すなわち、前述のように追跡調査において住民異動情報を提供するのみの機関の位置づけは疫学指針に従った運用とし、同意取得時や遺伝子解析を行う際の匿名化などはゲノム指針に従った運用とするなどである。現状では、このような運用について倫理審査委員会の承認を得る必要がある。指針の改訂が望まれる。

なお、改正ゲノム指針では前文に「また、ヒトゲノム・遺伝子解析研究に多様な形態があることに配慮して、本指針においては基本的な原則を示すこととし、研究者等が研究計画を立案し、その適否について倫理審査委員会が判断するに当たっては、この原則を踏まえつつ、個々の研究計画の内容等に応じて適切に判断することが求められる」の文章が追加された。この点を踏まえ、現在検討中のいわゆる「統合指針」において、ヒトゲノム・遺伝子解析研究に関する適切な方針を盛り込むことも必要であろう。

8-2) 個人情報保護・安全管理について

コホート研究の実施・運営においては、疾病、生活習慣、健(検)診、死亡、異動にかかわる個人情報を扱うことが必須となるため、中央研究事務局及び各調査実施地域において個人情報保護・安全管理マニュアル(ゲノム指針に準拠)を定め、研究対象者に危険・不利益が及ばないように管理を徹底する。以下、個人情報を扱う際の原則について、簡単に記述する。

- 個人情報を扱う際の原則

- i) 研究班内の個人識別情報としては、コホート ID 番号のみ(必要に応じてエラー確認のため、性、生年月日を付記)を使用し、そのまま個人識別が可能な個人識別情報(氏名、住所(地番まで)など)については、可能な限り使用を制限する。
- ii) コホート ID 番号とそのまま個人識別が可能な個人識別情報(氏名、住所(地番まで)など)の両方を含む対象者名簿およびマスターファイルについては、外部ネットワークから独立したコンピューターに保管する、取り扱う者を制限するなど、特に厳重に取り扱う。
- iii) 遺伝情報はコホート ID 番号以外の匿名化番号で扱う。個人情報と匿名化番号との対応表は、個人情報管理室で管理され、施設内の研究部門や(共同研究機関を含む)施設以外の機関に提供されないようにする。

そのほか、研究者間でのデータ受取の手段に関する基本原則(原則として郵便を用い、郵便

を用いる際には追跡機能がある手段、電子媒体にはパスワードをかけるなど)、情報処理室の確保(施錠可能、入室者制限など)、使用するパソコン環境整備(使用者限定、パスワード管理、外部ネットと独立など)、個人情報入りメディア・紙媒体の管理の管理(施錠可能な保管庫、消去方法など)、廃棄方法、作業の外注(個人情報取り扱いに関する業務委託契約書など締結、守秘義務など)などについても記述されるのが望ましい。

II. ゲノム解析

本プロジェクトにおけるパイロット研究等をもとに、研究資源・体制・データベース構築・データ解析・情報保護技術などの観点から基礎的情報の整理を行い、今後 10-20 年後に本格化する次世代大規模分子疫学コホート研究におけるゲノム解析のあり方に関する提言とする。

1. 血液試料からの核酸抽出と保管システムについて

- 現行技術による基本的な DNA 試料の抽出・保管:

JPHC-NEXT 等の新規分子疫学コホートでは、将来実施されるであろう多角的・多目的なゲノム・エピゲノム解析を想定し、核酸抽出の元となる血液試料を長期に渡って安定的に保管しておく必要がある。JPHC Study において約 20 年前に収集され、 -80°C の超低温槽で保管されてきた Buffy coat から現時点において標準的な方法で核酸 DNA を抽出したところ、質・量とも概ね良好な DNA サンプルが得られた。この DNA サンプルを用いて、本プロジェクトで計画されたゲノム網羅的 SNP 解析、ゲノム網羅的メチル化解析、エクソームシーケンセス等のゲノム・エピゲノム解析は特に問題なく実施できたので、JPHC-NEXT 等の新規分子疫学コホートで収集される血液試料を Buffy coat まで分注して -80°C の超低温槽で保管しておけば、将来行うことが想定される多角的・多目的なゲノム・エピゲノム解析を実施できる可能性は十分に高いものと思われる。ただし新規分子疫学コホートにおいても、血液試料の収集・処理が複数の調査地域で行われる事が想定されるので、血液保管の前段階として、血液試料の収集・処理の過程を出来るだけ標準化しておく事が望ましい。

- 多種、複数回の生体試料収集に対応した ID 管理:

各種バイオマーカー情報等の継時的な変化及び追跡可能性を担保する上で複数年において生体試料を収集することが望まれる。匿名化後の解析を容易にする上で、複数種類及び複数時点に対応した匿名化 ID システムが必須。

2. ゲノム網羅的 SNP 解析

2-1) パイロット研究課題「サブコホート 2900 人による日本人多型データの収集」に基づく提言

- ゲノム解析による品質検査を組み込んだ生体試料管理の高度化:

末梢血収集に使ったボックス内で性別不一致が顕著に多いボックスが存在した。複数地域・時点において、様々な熟練度の担当者に関わる大規模な試料・情報の収集においては一定の確率で、このような取り違えなどのエラーが起きる。それを最小限にするための、明解かつ簡潔な標準作業手順書の作成とそれに基づく研修、ダブルチェックなどの作業フロー構築、DNA 試料の SNP アレイ解析等による品質検査と ID 管理 (ID トラッキング)、元試料の分注保管などを組み合わせた、試料管理の高度化が必要である。

- 日本人とその生活習慣・環境要因の地域多様性に対応した全国からのサンプリングの必要性和解析のあり方:

日本人集団内におけるゲノム多型による集団の構造化が確認された。バイアスとしての構造化への対応については、コホート研究内であれば地域を考慮した matched case-control study が有効である。

3. リシーケンシング等による次世代分子疫学的コホート解析

3-1) パイロット研究課題「稀な多型検出のためのエクソームデータ解析パイプラインの検討」に基づく提案

- 日本人多型・変異解析用リファレンスゲノムデータの整備:

HapMap や 1000 人ゲノムプロジェクト等で豊富なデータが蓄積されている Caucasian のゲノムに比べて、日本人のゲノムを次世代シーケンサーを用いて解析すると、ミスマッピング由来も含めて多くの indel や CNV 等の構造多型が call される。日本人リファレンスゲノムの構築が必要である。その際、

上記 2-1)の知見に基づき、日本の各地域からの十分な数のサンプリングに基づくデータ構築が必要である。

- 次世代シーケンサーデータに基づく多型・変異 call システムの調査・開発・検証の継続:

シーケンサーのハードウェアの技術革新に加えて、それと連携した多型・変異コールのツールの進化・更新も早い。世界の最新技術の情報を的確に捕捉し、その時点で日本人に最適のパイプラインを構築・運用するためのバイオインフォマティクス研究者・技術者チームを維持する必要がある。

特に insertion/deletion や、コピー数等構造多型の解析ツールの開発と、それを検証するための日本人「正解」データセットの確保、及びデータ解析ツールによる予測結果に対して、系統的に検証実験をする体制が必要である。

第3世代(PCRを用いない、一分子シーケンシング)のシーケンサーを含め、異なる chemistry のシーケンサーを併用するなど、異なる read 長・insert 長のデータを組み合わせた多型・変異解析が望ましい。

一定の頻度を持つ構造多型については、日本人リファレンスゲノムデータベースの中で、それと強い連鎖不平衡にある tagging SNP や SNP haplotype の情報を整備し、カスタム SNP アレイによる解析を可能にすることも検討すべきである。

3-2) パイロット研究課題「末梢血を用いたメチロームデータと生活習慣等の関連解析」に基づく提言

- 次世代分子疫学コホートにおける位置付け:

既に報告されている、年齢・喫煙と末梢血全血のメチル化プロファイルとの明らかな相関が、複数地域から構成される日本人集団においても再現された。曝露の影響をこの様な客観的な指標で評価することが出来れば、調査票の情報を補完することが可能になるかもしれない。

このように DNA メチル化は加齢や、様々な既知・未知の生活習慣や環境曝露等の影響を受けて変化するが、比較的安定な形質と考えられ、その変化が、既存の指標では説明しきれない疾患リスクの個人差を捕捉するための新たな指標となる可能性が期待される。次世代分子疫学コホートで解析を計画すべき項目であると考えられる。

- 白血球の分画成分の影響への対応の必要性:

しかし血液中の白血球は、多種の分化した細胞が、様々な割合で混在しており、その混合物である全白血球のメチル化と特定の形質や生活習慣・環境要因等との相関が得られても、分画成分比の変動によるものか、実際に局所的あるいはゲノム網羅的なメチル化の変化によるものかの判別が出来ない。そこで理想的には、特定の分化段階にある白血球を均一な集団として分離し、そのメチローム情報を取得・解析する必要がある。

- 日本人末梢血白血球の分画別標準エピゲノムリファレンスデータ構築の必要性:

大規模分子疫学コホート研究においては、少なくとも現時点においては上記の、特定の白血球分画分取は現実的でないと考えられるため、データ解析において、臨床検査としての分画比のデータ(但し、T 細胞・B 細胞の割合も必要)を用いた調整などの方法が考えられる。そのためには、十分なサイズの、日本人の白血球分画別メチローム情報のリファレンスデータベースの構築が必要である。このような日本人エピゲノムリファレンスデータは IHEC(国際エピゲノムコンソーシアム)等でもまだ整備されておらず、学術的意義も高い。

具体的には、いわて東北メディカル・メガバンク機構の計画等における検討が望まれる。

- 実験バイアスを考慮したエピゲノム解析:

正常白血球のメチル化レベルの変動幅は一般に小さく、解析には高い精度の定量性が求められる。従って、現在のアレイ技術によるメチローム解析では、解析はできるだけまとめて行い、チップ上にサンプルをランダムに配置するなどの基本操作により、測定バイアスの抑制に十分配慮する必要がある。今後、シーケンシング等によるメチル化解析に移行した時は再度、検討が必要である。

- エピゲノム生物学・病理学研究者を含めた学際的解析チームの必要性:

エピゲノム修飾の個人差、その成因と変動のメカニズム、機能的影響などについて、特に病変部位以外のいわゆる正常組織に関しては、未だ未解明の部分が大きい。しかしながら、細胞の分化と老化、機能発現の中核をなす分子機構であり、明確に定義され、追跡される大規模人間集団である住民コホート研究の試料の解析データが医学・生物学にもたらす波及効果は測り知れない。そのためにも、疫学者・バイオインフォマティクスのみならず、エピゲノム生物学・病理学の研究実績を持つ研究者を含めた学際的なチームの構築が必須である。

3-3) 採択条件「バイオインフォマティクスの人材育成の努力」に基づく提言

- 疾患研究現場におけるバイオインフォマティクスのキャリアパス整備:

臨床試験等を扱う生物統計学や疫学の研究者らは、研究デザインから参画することで単なる解析担当者に留まらない地位を得ている。分子疫学コホート分野においては、従来のアルゴリズム開発を目的としたバイオインフォマティクスとは異なる、解析そのもの・疾患研究そのものに情熱をもち、それを目的・使命とするバイオインフォマティクスの育成が必要である。

医学系研究拠点が、国内外のバイオインフォマティクスを目指す若者を柔軟かつ積極的に受け入れる環境を整備し、研究現場において研究支援の技術者から、独立した研究者(PI)を志向する人材まで、幅広く受け入れ、育成・雇用するための、キャリアパス設定と、それに基づく、さらなる優れた若手人材を惹き付けるロールモデルを生み出す努力が必要である。

- 理工系分野(数学、情報学、物理学、工学等)における若手研究者へのアプローチ:

少なくとも現在の教育を受けた医学・生物学の実験系研究者をバイオインフォマティクスに育成するよりも、数学への抵抗が少ない人材に医学・生物学を教える方が障壁が少なく、時間・コストも少なく済むと考えられる。上記キャリアパス設定・ロールモデル創出においては特に、理工系分野の若手研究者を疾患研究に惹き付け、そこで活躍する状況を創出するとともに、有為な人材への情報発信・啓発が必要である。

- 広く産学官を視野に入れた、多彩で柔軟な雇用形態の創出:

しかしアカデミアにおけるキャリアパスの構築は短期的には難しく、またポストも非常に限られることから、必ずしもアカデミアによる直接雇用にこだわらず、プロジェクト毎に、産学官から広く優れた人材を確保する仕組みや、NBDC 等に国としてのバイオインフォマティクス研究・教育拠点を作り、若手研究者支援・派遣事業や医学系研究拠点等との共同研究を展開することなどを検討すべきである。産学官連携は、臨床シーケンシング等によるゲノム医療の実用化が世界的に見込まれることから、産業育成としての意味もある。

- ヒトゲノムデータを解析するバイオインフォマティクスの認定・研修制度の検討:

ゲノム医学領域のバイオインフォマティクスには今後ますます、試料・情報提供者に対する厳格な人権・個人情報保護の責務が求められる。バイオインフォマティクスの認定・資格制度の設計は、優れた人材の発掘・育成を促進し、職責に応じた適切な雇用条件をより安定して確保することにもつながる可能性があり、検討すべき事項の一つと思われる。

4. 生活習慣・環境要因と相関する候補遺伝子多型解析

4-1) 遺伝子-環境交互作用について

- 大規模分子疫学コホート研究における意義:

JPHC-NEXT 等の新規分子疫学コホートにおいて、将来実施されるであろう多角的・多目的なゲノム解析の中心が、ゲノム領域全体を網羅するような大規模解析となることは、ほぼ間違いない。このような網羅的解析がもたらす膨大な知見により、多因子疾患である生活習慣病の遺伝学的な原因究明は大幅に進むものと期待される。しかし、遺伝要因は容易に修飾し得ない特徴を持っているので、遺伝要因と交互作用をもって疾病の発生・発症に関与している生活習慣・環境要因を明らかにしていくことは、新規分子疫学コホートから得られた遺伝学的な知見を疾病の予防に繋

げるためにも重要である。

遺伝子-環境相互作用の検討においては、ゲノム網羅的アプローチに加え、疾患メカニズムなどの知見に基づく仮説による候補遺伝子アプローチを組み合わせることにより、双方の特長を生かした効率の良い検討が可能となる。

- サンプルサイズに関する課題：

ゲノム網羅的アプローチは仮説に依存しないアプローチであり、今まで知識・常識を超えた発見につながる可能性があるが、検定の多重性が問題となり、特に相互作用検討では単独の分子疫学コホート研究では一般に検出力不足である。一方、仮説に基づき、少数の SNPs を解析対象とする候補遺伝子アプローチにおいては、同じ症例数であればゲノム網羅的アプローチよりも高い検出力が期待できる。しかし、候補遺伝子アプローチであっても一般に相互作用の検出力は高くなく、オッズ比 1.5 程度の遺伝子-環境相互作用を統計学的有意に検出するには、少なくとも数十万人規模のコホートサイズが必要である(下図参照)。

コホートサイズ	症例数	最小検出オッズ比
40,000	約 340	2.38
100,000	約 850	1.75
250,000	約2,100	1.44
500,000	約4,200	1.30

- 症例数は、追跡期間を10年間、症例発生割合を年間0.001、追跡不能割合を年間0.03と仮定して算出した。
- 最小検出オッズ比は、症例:対照 = 1:2、有意水準を0.05、検出力を0.80、リスク環境の頻度を0.20、リスク遺伝子のアレル頻度を0.20、遺伝モデルを優性遺伝型と仮定して算出した。

4-2) Mendelian randomization 法について

- 大規模分子疫学コホート研究における意義：

Mendelian Randomization 法は、遺伝子型を操作変数として交絡要因による影響を小さくすることにより、バイオマーカーなどの因子と疾病との間に因果関係があるか否かについて検討する手法である。近年の GWAS の進歩により、曝露に関連する SNPs が明らかになりつつあることから、今後の大規模分子疫学コホート研究の中での活用例の増加が期待される手法である。

- 適用にあたっての注意点および課題：

十分な検出力の確保のために、操作変数としてバイオマーカーなど曝露と相関が強い SNP を選択することが重要である。それに加え十分なサンプルサイズで検討する必要がある。

まだ分子疫学コホート研究への活用例が多くないことから、適用条件の確認および結果の解釈について、さらなる実証的評価が必要である。

4-3) リスク予測モデルについて

- 大規模分子疫学コホート研究における意義：

コホート研究には通常の症例対照研究では困難な疾病の絶対リスクを評価できるという研究デザイン上の特徴があるので、JPHC-NEXT 等の新規分子疫学コホートから得られるゲノム情報は、疾病の本態解明・因果関係評価に資する研究だけでなく、生活習慣・環境要因等と組み合わせることで個人の絶対リスクを予測するような疾病の個別化予防に繋がる研究に活用されることが望まれている。

- サンプリング割合の推定方法の改善の必要性：

コホート内症例対照研究およびケース・コホート研究デザインにより絶対リスクを推計するためには、サンプリング割合で重み付けする解析手法を用いる。ケース・コホート研究は、基本的にランダムサンプリングであるのでサンプリング割合は単純であるが、コホート内症例対照研究は、複数のマッチング変数の組み合わせにより決まるため、その推定方法がいくつか知られており、結果に影響する要因の一つであるため、より良い推定方法の開発が求められる。

- 統計モデルによるアウトカムの予測能の評価方法の検討の必要性:

リスク予測モデルの判別能等の評価は、C 統計量等の統計学的指標が用いられるが、Genetic Risk Score などのバイオマーカー予測能の臨床的意義を評価するには不十分であることが指摘されている。そこでモデルの予測能を検討する際には、C 統計量のほか、Integrated Discrimination Improvement (IDI)、Net Reclassification Improvement (NRI)、Hosmer Lemeshow の検定などの統計量を用いることが望ましく、さらにより適切な指標の開発も必要である。

- 予測モデルの外的妥当性の検証の必要性:

一つの分子疫学コホート研究から導かれたリスク予測モデルが一般化されるまでには、その外的妥当性に関して、独立した分子疫学コホート研究のデータを用いて検証されなければならない。そのためには、我が国において、相互検証可能な独立した分子疫学コホート研究が複数存在する必要がある。

5. 大規模分子疫学コホート研究におけるゲノム解析のデザイン・サンプルサイズの検討

- 大規模分子疫学コホート研究において生体試料の分析を行う研究のデザインに対する提言:

数十万から 100 万人規模の大規模分子疫学コホート研究において、生体試料を用いて疾病アウトカムとの関連を検討する際に、コホート対象者全員の生体試料を分析することは、特に費用・時間の面で効率が悪い。そこでコホート研究において生体試料の分析を伴う研究を効率的に行う研究デザインとしてコホート内症例対照研究とケース・コホート研究の 2 つのデザインが知られおり、いずれも一部の対象者のみを解析対象とすることで生体試料分析にかかる費用・時間が節約できる。これらのコホート内症例対照研究とケース・コホート研究の 2 つのデザインには、それぞれ長所・短所があることから、その特徴を踏まえて、研究目的、生体試料の種類、分析項目に応じて併用していくことが大切である。

具体的には、ゲノム解析を主体とする場合、あるいは臓器横断的な解析を行う場合はケース・コホート研究デザインを中心に、また血漿分析を主体とする場合はコホート内症例対照研究デザインを中心に計画することが望ましい。

III. 他研究との統合

1. 調査票情報の統合に関する提言

調査票情報の正確性の指標としての妥当性研究について、また、「データ統合のための妥当性研究」の結果に基づき、データ統合の方法および将来の統合を見据えた調査票のあり方について提言としてまとめる。

- 妥当性研究の必要性:

生活習慣情報の中には調査票で把握することが困難であるために、情報の正確性の保証のために妥当性研究が必要な場合がある。例えば、日本人の食生活は多種多様であり、食習慣の把握を目的として作成された調査票から計算された食品群・栄養素摂取量はあくまでバイアスを内包した推計値である。妥当性が保証された調査票に基づいて食習慣を把握した研究であることが統合参画の可否において要点となるため、食事記録調査などをゴールドスタンダードとして設計した妥当性研究を実施することが重要である。自身で妥当性研究を行わない場合は、妥当性が保証された調査票、調査項目を採用することが望ましい。

- 新規に開始する研究への提言:

妥当性が保証された調査票、調査項目を採用するか、新規に作成する場合は独自に必要なに応じ妥当性研究を実施する。時期の設定や例示の有無などの質問形式や、選択肢の設定も少なからず回答に影響するため、他研究の調査票の構造を精査し、統合可能となるような配慮が必要である。

- すでに開始している研究への提言:

必要に応じ妥当性研究を実施することが望まれる。JPHC-NEXT 調査票(調査項目)あるいは、日本多施設共同コホート研究(J-MICC Study) 調査票(調査項目)を採用している場合は「データ統合のための妥当性研究」において示された方法論(希薄化因子 λ を用いた補正式)に基づき食品群・栄養素データの統合が可能である。その他の調査票を採用している場合も、妥当性研究を実施していれば、同様の方法論に基づき補正式を算出の上、統合可能である。喫煙、家族歴などのカテゴリカルデータについては項目により、また、選択肢のまとめ方により統合可能性が異なるため、留意する必要がある。

2. 生体試料情報統合に関する提言

生体試料の情報を統合するには、収集・処理・保管の過程が標準化された生体試料を統合し、統合された生体試料から標準化された手段で情報を得るのが最善の方法と考えられる。個別に構築が進められている現在進行中の分子疫学コホート研究で、今から生体試料の収集・処理・保管の過程を標準化するのは極めて困難であるため、その類似点と相違点を明らかにし、将来の統合に向けた課題を整理しておく必要がある。また、情報を有効利用する観点から、既得の生体試料情報を統合する場合も想定される。既得の生体試料情報は、その由来が異なれば、標準化すべき生体試料の収集・処理・保管の過程や情報を得る手段が異なっている可能性が高く、統合するにあたり何らかの相互補正が必要になってくる。

- JPHC-NEXT、J-MICC Study、東北メディカル・メガバンク計画との統合について:

現在国内で構築が進められている主な分子疫学コホート研究である JPHC-NEXT と J-MICC Study に、疾病罹患前の生体試料を収集している東北メディカル・メガバンク計画を加え、血液試料の収集・処理・保管の過程を比較したところ、血液試料の収集・処理・保管の過程は、3 研究間でほぼ共通しており、採血管の種類、採血条件、採血後の保存方法、採血から処理・保管までの時間、長期保管の方法と言った重要な点で一致が見られた。血液処理の過程で遠心機の回転数が異なっている点は、将来の統合で大きな問題点にならないと考えられるが、統合を行う生体試料情報毎に検討されるべきであろう。尿試料の収集・処理・保管の過程は、試料収集を行って

いる JPHC-NEXT と東北メディカル・メガバンク計画との間でほぼ一致していた。統合を行う生体試料情報によっては、試料の収集・処理・保管の過程により厳しい条件が付く場合があるので、試料の収集・処理・保管を行った日時等、生体試料に付随する情報を記録しておくことは重要と考えられる。現在国内で構築が進められている主な分子疫学コホート研究である JPHC-NEXT と J-MICC Study、疾病罹患前の生体試料を収集している東北メディカル・メガバンク計画との間では、将来の生体試料情報統合が十分可能と考えられる。

- これから構築される新規分子疫学コホートに望まれること:

これから構築される新規分子疫学コホートでは、生体試料の収集・処理・保管の過程を先行する上記 3 研究に合わせて標準化し、生体試料の収集・処理・保管を行った日時等、生体試料に付随する情報を記録しておくことが、将来の生体試料情報統合のために望まれる。

- 既得の生体試料情報を統合する場合の課題について:

分子疫学コホート研究において由来の異なる既得の生体試料情報を統合する目的は、疾病との関連を検討することにあるので、①生体試料情報そのものを相互補正する方法だけでなく、②生体試料情報と疾病との関連を相互補正する方法が考慮される。①②何れの補正方法でも、標準化された生体試料情報と疾病との関連に概ね近い関連を得ることが可能だが、①②何れの補正方法でも、標準化された生体試料情報と疾病との関連から「系統的に」外れた関連を導き出し、てしまう可能性があることを、限界点として認識しておく必要がある。

おわりに――我が国における大規模分子疫学コホート研究の構築を実現させるために

わが国では、21 世紀に入ってから、徐々に大規模分子疫学コホートの構築が開始されているが、現在大規模でも J-MICC Study の十万人規模である。また国立がん研究センターを中心に平成 22 年より開始した次世代多目的コホート研究(JPHC-NEXT)は、10 万人を目標に構築を進めているが、近隣アジア諸国では、30 万人から 100 万人を目標とした分子疫学コホート構築が進んでおり、目標としている 10 万人は、これらに対処していくのに十分な大きさとはいえない。このような状況の中、国民の健康の維持・増進に必須の研究基盤として、国際的評価にも耐えうる質の高い大規模分子疫学コホート研究の構築が求められている。

今回、「準備段階研究」として位置づけられている本プロジェクトにおいて、実現可能性の高い標準プロトコールと、既存の分子疫学コホート研究における情報統合の方法論が提示された。これらの方法論を用いて、本プロジェクトで立ち上げたコホートを含む新規コホートの維持・構築の推進、それらに JPHC-NEXT、J-MICC Study および東北メディカル・メガバンク計画などの既存の分子疫学コホートを加えた、オールジャパンの健常人コホートを確保するために「日本大規模分子疫学コホートコンソーシアム」の立ち上げを提案する。すでに JPHC、JPHC-NEXT、J-MICC Study の 3 研究において約 20 万人分の生体試料が保管されていることから、これらの既存大規模分子疫学コホート研究をもとにコンソーシアムを構築することが、国際的に後れを取ることなく効率的かつ経済的に大規模分子疫学コホート研究の構築を実現させる鍵である。また、このようなコンソーシアムの構築を健康・医療分野の国家戦略として行うことにより、日本の成長戦略におけるイノベーション創出につながる基盤構築が可能となる。また上記提言に挙げた、大規模分子疫学コホート研究の推進をはじめとする健康・医療分野の問題にとどまらない、社会制度にもかかわる諸問題について、俯瞰的な立場で戦略的に取り組むことにより積極的な解決が期待される。