

中間とりまとめ

～ 第2段階におけるゲノム・オミックス解析の実施計画～

平成29年1月24日

ゲノム・オミックス統合解析ワーキング

ゲノム・オミックス統合解析WG中間とりまとめ

～ 第2段階におけるゲノム・オミックス解析の実施計画～

1. はじめに
2. 大規模ゲノム解析の意義
 - 1) 海外のコホート・バイオバンクにおけるゲノム医科学研究の現状、大規模アレイ解析との比較
 - 2) 疾患関連遺伝子研究に向けた基盤の提供
 - 3) 大規模ケース・コントロールスタディでの活用
3. ゲノム解析の実施計画
 - 1) ゲノム解析のロードマップ
 - 2) 大規模SNPアレイ解析と全ゲノム情報復元
 - 3) 全ゲノムリファレンスパネルの拡充
 - 4) 全ゲノム解析の手法とその選択
4. オミックス解析等の実施計画
 - 1) 多層オミックス参照パネルの構築と展開
 - 2) 大規模メタボローム解析
 - 3) エピゲノム解析
 - 4) その他のオミックス解析等
 - トランスクリプトーム解析
 - プロテオーム解析
 - 細菌叢メタゲノム解析
5. 個別化予防に向けた統合解析のための基盤構築
 - 1) 発症リスク予測手法開発
 - 2) 統合・知識データベースと病院連携
6. 遺伝情報回付
 - 1) 個人への遺伝情報回付へ向けての検討の経緯
 - 2) 個人への遺伝情報回付のパイロット研究（単一遺伝性疾患）
 - 3) 個人への遺伝情報回付に向けての基盤整備
7. ゲノム医療実現のための人材育成
 - 1) 第1段階における人材育成の実績
 - 2) 第2段階における人材育成の計画
8. まとめ
 - (参考1) WGの構成
 - (参考2) WGの開催状況

1. はじめに

東北メディカル・メガバンク（TMM）計画は、平成23年度から平成28年度までの6年間の第1段階の取組を進めてきた。ゲノム解析における達成度としては、2,049人分全ゲノムリファレンスパネル（2KJPN）としてアレル頻度を全公開し、平成29年前半には3.5KJPNに拡充予定である。基準ゲノム（JRGv1）の公開、オミックス解析においては、コホート参加者1,008名の血漿の網羅的解析結果を「日本人多層オミックス参照パネル（jMorp）」として公表、また、コホート参加者100名の白血球2細胞（単球、CD4陽性Tリンパ球）の網羅的解析結果を「日本人多層オミックス参照パネル（iMethyl）」として公表し、平成28年度中に好中球を追加した3細胞に拡充予定である。また、東北メディカル・メガバンク統合データベースdbTMMを開発し、ゲノム医療実現推進のデータシェアリングの基盤を形成した。

平成29年度からは第2段階（平成29年度から平成32年度）に移行するが、進捗や事業を取り巻く状況の変化等を踏まえて、第2段階の推進にあたる必要がある。TMM計画第2段階におけるゲノム・オミックス解析の意義は、復興・創生期間という新たなステージに応じた被災地の健康管理等への貢献と、我が国のゲノム医療実現に向けた基盤形成およびデータシェアリングであり、これらを見据えた計画が必要である。

第2段階計画の概要

1. ゲノム医療研究の基盤構築

- ・全参加者を対象として標準ゲノムリファレンスパネルを用いて開発したアレイによるゲノム解析を行う
- ・第1段階と合わせて8千人の全ゲノム解析を行い、精度の高い日本人の標準ゲノムリファレンスパネルを構築する
- ・約7万人の参加者を対象に疾患発症と関連するマーカー探索に向けたオミックス解析を実施する
- ・収集した試料・情報の早期分譲を行うとともに、他コホート事業やバイオバンクと連携しつつ、保管試料・情報のカタログ化や横断検索できるシステムの導入など、安全性を確保しつつユーザーの利便性向上に取り組む
- ・ゲノム医療実現推進プラットフォーム事業により委託されているスーパーコンピュータ等のゲノム解析基盤を共用できる体制を構築する

2. 個別化医療・個別化予防の先導モデルの構築

- ・遺伝情報回付パイロット研究を推進する
- ・高血圧、アトピー性皮膚炎、脳梗塞等の疾患発症リスク予測手法を開発するとともに、個人毎に疾患発症リスクを回付するための手法の開発も併せて行う

3. ゲノム医療実現のための環境整備等への貢献

- ・ゲノム医療体制の構築に必要な専門人材を育成するとともに、そのキャリアパスの形成を図る。また、メディカルクラーク取得者と同程度の技能を有するデータマネージャーを育成する
- ・ゲノム医療を社会に根付かせるための取組を実施するとともに、ゲノム医療に関する倫理的諸課題の解決に取り組み、ゲノム医療を実現する素地を醸成する

2. 大規模ゲノム解析の意義

海外では、ゲノム情報等の付随した対象者の健康情報の包括的な管理、利用に向けた基盤整備、ゲノム情報等のデータシェアリングの取組及び研究基盤の整備が推進されてきている。また、がん、希少疾患・難病、未診断疾患及びファーマコゲノミクス等のゲノム研究に加え、健常人のゲノムコホートなどを通じて多因子疾患の解析も推進しつつ、ゲノム医療に向けた取組が始められている。そのためには、遺伝子と環境の相互作用による病気との因果関係を明らかにする目的のためには多数の生体試料を用いた横断的、縦断的解析が必要である。

ここでは、ゲノム研究実現のための基盤構築としての、TMM計画における大規模ゲノム解析の意義を、海外における大規模アレイ解析との比較と、オールジャパンの研究基盤の構築のための大規模コントロール（対照）データとしての活用、並びに前向きコホートにおいて発症する疾患の関連遺伝子研究に向けた基盤提供の観点から述べる。

1) 海外のコホート・バイオバンクにおけるゲノム医科学研究の現状、大規模アレイ解析との比較

2016年現在、世界中ではのべ3,500万人相当の試料について網羅的なゲノム解析データが登録されており¹、様々な遺伝的背景をもつ集団に属する数十万人レベルの大規模解析、1つの集団で得られた仮説について他集団を使って実施する多段階の検証やメタ解析等の成果が多数報告されている。UKバイオバンクでは、50万人のゲノム解析が終了したが、10万人程度のアレイ解析及び5万人程度のAdd-onコホートとしてのUK BiLEVE研究が終了した段階で、推定全ゲノム（genome imputation）データを用いた解析が実施され、肺機能（一秒率）低下及び喫煙行動に関わる遺伝的要因²、神経症的傾向³などの成果が報告され、疾患研究におけるゲノム疫学調査の重要性を明らかにした。

しかし、網羅的な遺伝子解析実施例の81%は欧州に祖先を持つ集団由来の試料であり、欧州に祖先を持たない集団にとっては、それらの解析結果による恩恵は限定的であるとされる¹。一方、この2～3年の東アジア地域における解析数の増加は顕著であり、その多くは民族に特化したDNAアレイ解析（エスニックアレイ）と、1,000人規模の全ゲノム解析データを利用して推定（imputation）された全ゲノムデータである。これらの東アジアの国々との競争的連携の中で主導権を持ち、東アジア人のゲノムと疾患の関連解析研究を推進できるオールジャパンの研究基盤を構築するためには、第2段階早期において、15万人のアレイ解析を完遂するとともに、現状3,500人の全ゲノム解析データについて、8,000人規模を目指して規模を拡大し、日本人全ゲノム情報、特に三世代コホートの家系情報付き検体を用いた推定全ゲノムデータを得ることが最重要項目である。

一方で、TMM計画は、被災地としての研究計画として他に類がなく、また、我が国の生活習慣や食事の内容、社会生活の状況などの環境要因は、東アジアの中でも独自性が高い。ゲノム疫学研究としての成果を得るには、これらの要因と遺伝的要因の相互作用に関する前向き研究が重要であり、そのためには、10万人規模のアレイ解析が重要な意味を持つ。また、並行して推定全ゲノムデータの精度向上には全ゲノム解析数の拡大も必要である。

現在、欧州以外の地域において、一般住民におけるゲノム疫学的調査に特化して 10 万人規模のゲノム解析が実施されている国は非常に限られると予想される。TMM計画として、大規模アレイ解析による推定全ゲノム情報を蓄積し、疾患発症に関わる遺伝的要因と環境要因の相互作用を明らかにできれば、我が国のゲノム医療の基盤の形成と向上に資するところは大きい。さらに、海外におけるゲノム疫学研究との比較検証を進めることにより、地域間の遺伝的背景の差異を超えて信頼度の高い研究成果が得られることが期待される。

2) 疾患関連遺伝子研究に向けた基盤の提供

TMM計画において、DNAアレイ（ジャポニカアレイ[®]）を中心的に利用した約 15 万人の大規模アレイ解析が実現した場合には、従来の一塩基多型 SNP のみならず、多因子疾患のレアバリエーションや遺伝子・環境相互作用の効果を高精度に検査する研究基盤を提供することが可能になる。特に重要な点は、これらのデータから、多因子疾患の希少変異関連解析（レアバリエーションアソシエーション）によく用いられている集合的解析手法（Aggregation tests）のような不確実性を持つ手法を利用することなく、正面から取り組めるという点である。全ゲノム、DNAアレイ両者を組み合わせることで、以下のような展開が望める。

① ジャポニカアレイ v2 によるエビデンス増強

ジャポニカアレイ v2 では、一部の SNP を入れ替えることにより機能面の強化（過去にゲノムワイド関連解析（Genome-wide association study; GWAS）で報告があった SNP の拡張を行った点と、ヒト白血球抗原（HLA）のインピュテーション性能向上につながるマーカーの増強）を行った。この結果、従来はインピュテーションのためのアレイとしての色彩が強かったジャポニカアレイが、疾患アレイの側面も持つようになった。この変更により、これまで圧倒的に不足していた日本人での疾患関連変異のエビデンスの増強を行うことが可能になり、疾患解析拠点施設での解析をより一層推進させることができる。また、コントロール検体の数を 10~15 万程度にまで増やすことで、オッズ比が 3~5 程度の変異であれば、MAF=0.1%~0.5% までの変異の検出が視野に入る。有病率を勘案すれば、TMM計画の 15 万人コホートにおいて、高血圧、糖尿病、心疾患、脳血管疾患などが解析対象疾患の候補になる。

② 関連（アソシエーション）解析における意義と TMM計画のデータの強み

全ゲノムリファレンスパネルとアレイ解析データという 2 つのデータセットを用いることで、希少変異（レアバリエーション）を含めて、このパネルに含まれる全ゲノムバリエーション情報を、約 15 万人のコホート参加者全員について推定することが可能になる。これにより、従来の数千人程度のデータでは、あまりの低アレル頻度のために統計的検出力を維持できなかった希少変異についても、単一座位での関連解析が実現可能となる。この手法は、集合的解析手法よりもシンプルで強力な手法であると期待できる。実際に、アイスランドの deCODE ジェネティクス社では、ほぼ同レベルのデータセットを用いて、骨粗鬆症に寄与する MAF が 0.1% 周辺でオッズ比 4~5 程度の複数のレアバリエーションの捕捉に成功している⁴。

他にも、理論解析と数値シミュレーションからは、3,000人のケースに対して、TMM計画の5万人のインピュテーションデータを用いれば、MAF=0.5%の疾患関連アレルのオッズ比3程度の効果を90%以上の検出力で検出できる。また、MAF=0.1%でもオッズ比5程度の効果があれば、15万人で78%と検出力が増加し、通常の統計的仮説検定での検出力80%に接近する。

③ 家系情報の活用

TMM計画は、非血縁の地域住民コホートと家系情報付きの三世代コホートという二つの異なる特徴のコホートを持つことで、リファレンスパネルの高精度化、遺伝子型推定の精度向上などの面で、相補的かつ相乗的な効果を期待できる。これは、地域住民コホートによる遺伝的多様性の確保と、三世代コホートでの家系情報の多面的利用からもたらされる効果である。

同様のサンプリングを、ある程度の合理的な遺伝的閉鎖性を持つ集団で大規模に展開すると、各検体間に相互に中程度の血縁性が出現し始め、集団の1~2%に達すると、全検体が自身のゲノムのかなりの範囲にわたって一人以上の血縁者に連結されると予想されている⁵。通常のサンプリングでは実現できないほどの大規模な家系図相当情報を、ゲノム情報と記録された家系図から再構成できる可能性がある。実際に、前述のアイスランドの例では、家系インピュテーションによってアイスランド全国民（死亡者含む）のゲノム情報を取得し、稀少な疾患関連遺伝子の同定に成功している。

3) 大規模ケース・コントロールスタディでの活用

ゲノムワイド関連解析 (GWAS) は遺伝情報と形質情報との関連を評価する、遺伝統計学の一手法である。ヒトゲノム全体を網羅する数万~1千万箇所の一塩基多型 (SNP) のジェノタイプデータを用いて、対象形質との関連を評価する。

GWASの問題はこれらの疾患感受性多型 (SNP) の効果を総合しても、遺伝学的に推定される遺伝率 (heritability) を説明するには低いということであり、これは失われた遺伝率 (missing heritability) の問題と呼ばれる。GWAS で明らかになった^{6,7}形質や common disease と関連する SNP の効果サイズは非常に小さいことが多い(多くの場合オッズ比で 1.5 以下)。必要となるコントロール数は、疾患や形質によって異なり、コントロール数が数千で解析可能なアルコール性肝疾患 (オッズ比 2.19) から、高血圧のように 30 万まで必要な疾患まで幅がある。また、遺伝子環境相互作用解析を行うには、多くの論文・総説では、オッズ比 1.4 から 1.5 付近を疫学的に信頼できるラインとしている。年 3% の罹患率の疾患を、オッズ比 1.4 以上で遺伝子環境相互作用解析をするとすると、20 万程度のコントロール数が必要とされる⁸。また、近年一般的となってきた推定による低頻度バリエーションの関連解析には、効果サイズが大きき SNP を対象とする場合でもケース、コントロールとも大きなサンプルサイズが必要となり、単施設でのデータ収集、解析は難しくなる。

日本人の 2 型糖尿病の GWAS 研究においては、研究の規模拡大に応じて検出される SNP の数は検体数の対数に比例して増大するが、症例数と対照数では対照数のほうが検出される

SNP数との相関は強い。この点からも対照数を増加させる戦略は有用性がより高いことが予想される。

ゲノム研究の対象となる疾患として次のような疾患が考えられ、既報等から推定されるコントロールは数万以上必要なものが多いと考えられる。今回のコホート研究においては、大規模ケース・コントロールスタディで用いられるコントロールを提供することが可能となる。

TMM計画における対象疾患

1. 頻度の高い疾患及び量的形質；コントロール数、数千～15万以上
 - ・ 糖尿病⁹、脳卒中（脳出血、脳梗塞）¹⁰、高血圧症¹¹、冠動脈疾患（狭心症、心筋梗塞）¹²、慢性閉塞性肺疾患、アレルギー疾患、骨粗鬆症、脂質異常症、気管支喘息、肝炎、難聴、関節リウマチ、緑内障などの疾患
 - ・ 血圧、血中脂質（コレステロール、中性脂肪）、空腹時血糖、腎機能（血中クレアチニン、eGFR）などの量的形質
2. 精神疾患；コントロール数、数万以上
統合失調症¹³、認知症、双極性障害、うつ病など
3. 頻度の高い癌；コントロール数、2万以上
前立腺癌¹⁴、子宮内膜癌¹⁵、大腸癌、肺癌、乳癌、前立腺癌など

以上のことから、TMM計画における大規模ゲノム解析のデータは、日本の他の疾患コホートに対して、十分な数のコントロール検体を提供し、それらにおける疾患研究を加速することが示唆される。また、他の精度高い診断が付いている疾患コホートと連携することで、精度高いケースコントロール研究を展開することも可能となる。

3. ゲノム解析の実施計画

1) ゲノム解析のロードマップ

第2段階におけるゲノム解析は、引き続き、全国の研究者によるゲノム医療研究推進のための基盤情報となることを目的として、全ゲノム解読（アレイ情報含む）と情報解析を進める。同基盤情報を、速やかにデータシェアリングを通じて公開し、全国の研究機関が利用可能となることを推進する。また、有病率等を踏まえ、希少遺伝性疾患から多因子疾患までのさまざまなコントロールとして価値のある情報を計画的に取得することで、国内外の種々の疾患研究での利用促進と成果の創出を行う。

両大学においても、同ゲノム基盤情報を活用するとともに、地域住民コホート、三世代コホートそれぞれの重点疾患罹患者のゲノム情報拡充を行うことで、原因同定に向けた解析と基盤形成を実施する。解析は、前向きコホート調査からのアウトカムが取得でき、成果を創出できる順に年度毎に目標を設定し実施する。なお、各年度の疾患同定に有効な場合には、三世代コホート調査の特長を最大限活用し、家系解析を実施する。

全ゲノム解析とSNPアレイ解析については、取得できる情報の質および役割が異なる部分があり全体でバランスをもって情報化していく必要があるが、大規模解析を進める観点か

ら SNP アレイ解析を優先的に行う。全ゲノム解析と SNP アレイ解析の双方とも、できるかぎり 1 検体あたりの情報の質を下げることなく人数規模、解析のアウトカムを最大限にできるように、費用の面も含め勘案した計画とする。状況によっては民間に情報取得を委託することも検討する。

両大学においても、これらの基盤を用いて疾患関連遺伝子研究等に取り組むが、さらに症例が必要な場合には、適宜国内外のコホートと連携することで重点疾患を中心とした因子の解析を進める。

SNP アレイ解析と全ゲノム解析の進め方については、2)、3) で述べる。

2) 大規模 SNP アレイ解析と全ゲノム情報復元

国内におけるゲノム医療研究推進のためのゲノム基盤情報形成の観点から、15 万人規模のコントロールパネルとして、①40 歳以上の成人パネル（主に 60 歳以上）、②40 歳以下の成人パネル、③その他のパネル（1 万トリオパネルなど）を構築することが重要である。多因子疾患は 2 型糖尿病、高血圧など年齢層が高いコントロールが有効であることから、①の成人パネルを優先して構築し、②40 歳以下の成人パネルは、若年発症の疾患や妊娠期間中の疾患解析のためのコントロールとして、③1 万人トリオパネルなどは、三世代コホート調査に関連する疾患解析に有用である。

TMM 計画のコホートにおいて発症する疾患関連遺伝子解析のための基盤形成については、順次発症してくると考えられる①抑うつ症状、脳卒中、心筋梗塞、狭心症の各因子などの探索、②妊娠関係疾患（妊娠高血圧症候群、新生児低出生または他罹患、産後うつ）、③追跡調査を行うことで可能となる動脈硬化、呼吸機能変化、血圧上昇、④小児の疾患（注意欠陥・多動性障害（ADHD）、自閉スペクトラム症、先天性代謝異常他）の順序で、これらの疾患の解析に利用できる検体についてゲノム解析を行う。なお、ある疾患の解析の観点から検体を選択した場合においても、他の疾患に対してはコントロールデータとして活用することが出来るため、適切な検体を選択することにより、コントロールパネルと疾患関連遺伝子解析のための基盤形成を同時に実現可能である。

アレイ解析結果については、全ゲノムリファレンスパネルを活用した全ゲノム情報復元（インピュテーション）を実施し、提供する。第 2 段階では同パネルの人数規模を 8,000 人規模にすることで 0.1% までのバリエーションを高精度に復元することが可能となり、これらのバリエーションと疾患との関連探索が可能となる。

なお、疾患解析においては、全ゲノム解析などの疾患解析で取得される情報を組み合わせて提供することで、疾患のレアバリエーションを精度高く復元することができるようになり、解析の効率も向上するものと期待される。

3) 全ゲノムリファレンスパネルの拡充

ゲノム情報基盤の構築に向け、第 2 段階において 4,000 人の全ゲノム解析を実施し、最終的に 6,000 万個以上のバリエーションを含む 8,000 人規模の全ゲノムリファレンスパネルを構築することを目標とする。検体の選択においては、より適切な全国の基盤情報となるように、

アンケートの出身地情報（両親、祖父母含む）等の情報を用いることで、できる限り全国のハプロタイプを網羅していく。また、*de novo*などの解析に有用なトリオ解析も含めて実施する。

疾患関連遺伝子解析のための基盤形成については、これとは別に、毎年度 100 人以上の全ゲノム解析を実施することで、重点疾患に関連する疾患・形質の対象者が保有する新規レアバリエントを重点的に収集することができる。これにより、全ゲノムリファレンスパネルではカバーできない、当計画の重点疾患に関連するレアバリエントを効率良く収集することを目指す。各年度の重点疾患の考え方については、SNPアレイ解析のものと同じとする。

全ゲノムリファレンスパネルは、ヒトの参照配列との個人毎の差をバリエントごとに集計した情報であるが、ヒトの参照配列および情報解析技術が日々向上していることを勘案し、国際参照配列や日本人の保有する配列を加えたより適切な参照配列に対する情報解析を実施することで、さらに精度の高い全ゲノムリファレンスパネルに改訂する。第1段階では、長鎖型シーケンサーを用いてヒト白血球抗原（HLA）のリファレンスパネルの構築を行ったが、疾患に関連する他の難読領域のハプロタイプについても同様に実施する。また、シーケンス解析手法の至適化を通して挿入変異の効率的な収集も行い、全ゲノムリファレンスパネルの質的向上も推進する。

年度毎に改訂する全ゲノムリファレンスパネルを国内はもとより国外の研究者が利用できるように公開することで、希少疾患の原因因子の同定に向けたさらなる絞込みや、多因子疾患のエクソーム解析や全ゲノム解析の結果との比較解析による原因領域の同定などの活用を促進する。

4) 全ゲノム解析の手法とその選択

HiSeqX の登場により、ヒト全ゲノム解析（WGS）の解析コストは大幅に低下した。HiSeqX シリーズが、ヒト WGS 専用という縛りのもとに他の HiSeq シリーズと差別化し、試薬が低価格に設定されているためである。そのため、TMM計画において整備した HiSeq2500 の WGS コストのおよそ3分の1で、リード長 150bp、深度 30x の解析が民間企業への委託で可能である。一方、HiSeqX シリーズは、オーダードアレイというクラスター形成で新規の手法を採用、その性質により長いインサート長のライブラリDNAのシーケンスに対応していない。

TMM計画では、HiSeq2500 において 550bp インサート長ライブラリを使用した解析を行った結果、150~250bp の挿入変異や縦列型反復配列 Short Tandem Repeat（STR）に対して、高い検出力があることを見出した。そのため、当該ライブラリをスタンダードとして採用している。また、平成 28 年からリード長を 259bp に延長し深度を 25x とした改良版プロトコルを採用した。この新規プロトコルでは、解析コストを3割削減、しかしSNPの検出力はほぼ同定度に保ちつつ、挿入変異の検出力がさらに向上していることが、予備的な結果より明らかとなった。

これまで、多くの挿入変異やSTR、更にはコピー数多型（CNV）が、疾患発症の原因として同定されている。これらの変異やCNVの効率的な検出には、PacBio などの長鎖リ

ード解析が選択肢となるが、コスト面から多検体の解析には適していない。TMM計画で採用したリード長 259bp、深度 25x の解析は、数百ベースの挿入変異の収集の現実的な選択肢である。今後、同プロトコルに加え、さらに新しい解析プロトコルを導入するなど継続して改善を進めることで、他のゲノムコホートと、データの質の点で差別化が期待できる。

以上の技術動向を踏まえ、TMM計画においては、疾患感受性に関与する日本人特有の挿入変異を網羅することは、全ゲノムリファレンスパネルの質的向上の重要な柱であり、全ゲノム解析は、シーケンサーの機器開発の状況を見据えながら、基本的に内部実施で行う。

4. オミックス解析等の実施計画

1) 多層オミックス参照パネルの構築と展開

①多層オミックス解析の意義

現在の医療は、エビデンスに基づいた医療 evidence based medicine (EBM) であり、仮説主導型の「平均的な患者」を対象としてデザインされてきた均一的な医療である。しかし一般人口集団は多様な対象者から構成されており、平均的な事象を個々に客観的に適応するのは難しい。今後の個別化予防・個別化医療に向かうためには、個人の様々な違いを反映したデータ主導型の医療へと研究を展開していく必要がある。そのためには、ゲノム情報に加えて、エピゲノム、トランスクリプトーム、メタボロームなどの生体分子情報を得て、表現型の違いを正確に定義する必要がある。また、その結果をもとに GWAS を行うことで、ゲノム解析における失われた遺伝率 (missing heritability) を補完し、経時的な変化の解析にも貢献すると考えられる。例えば、RNA 発現に個人差のある遺伝子と共分離する SNP を同定することにより、当該遺伝子の発現調節に関する分子機構を解析することや、疾病に関連する代謝物の細胞内動態とよく相関する遺伝子発現パターンを見出すことで、疾病の標的分子や疾病発症に関連する遺伝子情報を得ることに活用することができる。

②日本人多層オミックス参照パネルの構築と公開

以上のような多層オミックス解析データの重要性にも関わらず、国際的に見ても集団ベースでの多層オミックス解析のデータ資源は稀である。特に、健常人由来の生細胞でゲノム解読情報を有するものは本計画のものが唯一と考えられる。

そのため、TMM計画では、多層オミックス解析のデータベースである「日本人多層オミックス参照パネル」の構築及び公開を進めてきており、第1段階においては、メタボローム解析、プロテオーム解析をベースとした「jMorp」、エピゲノム解析、トランスクリプトーム解析をベースとした「iMethyl」という2種類の日本人多層オミックス参照パネルを構築、公開した。

第2段階においては、2) 以下で述べるように、メタボローム解析、エピゲノム解析、トランスクリプトーム解析などの多層オミックス解析を行い、データを拡充する。また、これらの解析結果をゲノム情報も統合した日本人多層オミックス参照パネルを構築し、個別化予防・個別化医療のプラットフォームとなる研究基盤として提供する。

ここで、TMM計画において作成された高精度ゲノム解読情報を持つ不死化リンパ球は、エピゲノム、トランスクリプトーム、メタボロームなどの生体分子の網羅的解析技術を適用することで遺伝子機能を包括的に探る多層オミックス解析を集団ベースで実施可能な、極めてユニークなバイオリソースであることから、これを、統合的な日本人多層オミックス参照パネルの構築に活用する。

なお、メタボローム解析、エピゲノム解析、トランスクリプトーム解析は1検体当たりの解析コストが異なることから、それぞれの解析規模については、コストも考慮して設定する。

2) 大規模メタボローム解析

個人の代謝環境（メタボローム）はセントラルドグマの最下流にあり、遺伝要因と環境要因のエンドポイントであり、重要である。このため、様々な遺伝要因・環境要因と疾患発症などの表現型との因果関係を解明する上で、最も効果的な解析対象である。第1段階の成果をふまえて、第2段階ではメタボローム解析により疾患の原因となる遺伝要因と環境要因の相互作用を明らかにすることのできる研究基盤の構築を目標とする。関係する要因がゲノム以上に膨大で、かつ複雑に絡み合うため、従来のゲノムを主とした症例対照研究と同等以上のサンプル数の解析が必要となる。

このため、TMM計画の解析基盤の中で特に定量性が高いNMR法を最大限に活用して、メタボローム解析の対象を大幅に拡充する。さらに、高感度な質量分析法については、対象を絞って活用することで詳細な代謝物情報を取得し、ゲノム情報や調査表情報等との関連解析に利用できるようにする。また、これらの解析データを活用することで、追跡調査・詳細二次調査の情報を元に疾患発症の病態解析とバイオマーカー探索を行い、各種遺伝要因や環境要因が個人の体質に与える影響を解析する。

具体的な解析計画は以下の3点である。

a. 機能未知遺伝子及び効果不明のゲノム多型の機能解析

メタボローム解析の対象数を拡張し、さらにゲノム情報とのGWASを行うことで、機能が不明な未知遺伝子や効果が不明なゲノム多型の機能解析を行う。TMM計画では既に512名のメタボロームGWAS解析により複数の一塩基多型が血中代謝物に影響を与えることを同定した¹⁶。また、海外の大規模メタボローム解析では、GWAS解析の成果に基づき、機能未知遺伝子の機能同定につながった研究も報告されている¹⁷。第2段階では詳細二次調査を受けた対象者を中心に最大5万人規模のメタボローム解析を実施し、ゲノム情報とのGWAS解析を行うことで、機能未知遺伝子や効果不明なゲノム多型の機能解析を行う。

b. 症例対照研究による病態解析及びバイオマーカー探索

第2段階に行われる追跡調査、詳細二次調査において、疾患を発症した参加者を中心にメタボローム解析を行い、病態解析とバイオマーカー探索を行う。具体的には、発症が早期に判明する三世代コホート参加者で想定される疾患（妊娠高血圧症候群、低出生体重、産後うつ等。発症率合計約20%）を優先解析対象にし、妊婦の血液のメタボローム解析を行う。

また、三世代コホート参加者以外でも早期に疾患発症を確認できたコホート参加者を対象にメタボローム解析を行う。これらにより、最大2万人規模の解析を実施する。

c. 環境要因が身体に与える影響の定量的解析

ヒトの各種疾患は様々な遺伝的要因と環境要因が複雑に関与して発症するが、血中代謝物は各個人の体質を最もよく反映するマーカーであり、双方に敏感に反応する。このため遺伝的要因と環境要因が各個人に与える影響を、メタボローム解析を行うことで定量的に見積もることが可能となる。例えば、喫煙や飲酒などは中長期的に血中代謝物に影響を与えることが知られており、喫煙でのプリン分解経路代謝物やビリルビンなどの変動¹⁸、飲酒での多数の代謝物が変動することが報告されている¹⁹。同様に、摂取量や摂取期間も代謝物解析により従来のアンケート情報だけでは難しかった定量的かつ客観的な評価ができる可能性が高い¹⁸⁻²¹。そこで、第2段階では環境要因が体質に与える影響を客観的かつ定量的に解析する手法を、代謝物情報（上記計画aとbで解析した代謝物情報、合計最大7万人）と調査票情報等との関連解析を行うことで確立する。

3) エピゲノム解析

人体を構成する細胞は一部の特殊な細胞を除き同一のゲノム情報を有するが、ヒストン修飾やCpGにおけるDNAメチル化などのエピゲノムが遺伝子発現を制御することで、数百を超える多彩な細胞・組織を生み出している。また、エピゲノムは生活習慣やストレスなどの影響を受けることが知られており、被災を受けた方々が多数参加しているTMM計画で経時的なエピゲノム解析を行なうことは個別化予防・個別化医療の観点からも非常に重要である。

第1段階においては、コホート血液検体からの細胞分取方法の確立、DNAメチル化状態の中期的な安定性の検討、検体処理バイアス補正方法の確立、DNAメチル化キャプチャ法による廉価かつ高精度な解析手法の開発、などのコホート研究におけるDNAメチル化解析の基盤構築を達成するとともに、転写因子結合やヒストン修飾などについても feasibility study として小規模な条件検討を行ってきた。

そこで、第2段階においてはこれらの成果を踏まえて以下について進める。

a. 機能未知遺伝子及び効果不明のゲノム多型の機能解析

DNAメチル化は他のオミックス情報と異なり、組織特異的制御の他に発生初期のエピゲノム制御が固定される例も知られている。すなわち血液のDNAメチル化情報が精神疾患などの他の組織に起因する疾患発症に関わる遺伝子制御を反映するサロゲートマーカーとなりうるということが知られている。そこで、2,000例のDNAメチル化解析を行い、mQTL（DNAメチル化と関連する遺伝子座）を同定することで機能未知遺伝子や効果不明なゲノム多型の機能解析を行う。

b. 症例対照研究による病態解析及びバイオマーカー探索

生活習慣やストレスがエピゲノムに変化を引き起こすことが知られている。そこで、a. で測定するDNAメチル化とともに、第2段階にて行なう詳細二次調査で収集する血液検体のDNAメチル化解析を行なうことで疾患関連バイオマーカーの探索を行う。

被災地住民を含む3,000例のキャプチャ法によるDNAメチル化解析を行い、ストレスや生活習慣と関連するCpG部位（メチル化マーカー）の同定を目指し、個別化医療の実現を推進する。

4) その他のオミックス解析等

①トランスクリプトーム解析

TMM計画では、これまで数千人規模の全ゲノム解析を実施し、日本人集団における希少多型を含む全多型を全ゲノムリファレンスパネルとして公開してきた。これらの成果は、日本人に最適化したSNPアレイの作製に利用されるとともに、クリニカルシーケンスのリファレンスとして活用されている²²。一方、全ゲノムリファレンスパネルの中で非コード領域に見出された多型の機能アノテーションが、第2段階の課題となっている。これまでGWAS解析で同定されてきた疾患感受性多型の8割程度が、非コード領域に見出された事実は、この課題の重要性を示している。

そこで、第2段階においては、第1段階で蓄積した、数千人規模のRNA検体、不死化リンパ球などのトランスクリプトーム解析に適した検体を活用し、トランスクリプトーム解析による多型の機能アノテーションに取り組む。

a. 機能未知遺伝子、及び効果不明のゲノム多型の機能解析

第1段階で実施した100検体の単球、CD4陽性Tリンパ球の3層オミックスデータを用いた予備的解析から、約5万のeQTL（遺伝子発現と関連する遺伝子座）を同定した。この中には喘息や脳梗塞、高血圧との関連が知られる遺伝子の新規eQTLも含まれていた。そこで、TMMで収集した全血、及びPBMCを用いた約8,000例のトランスクリプトーム解析を実施することで100万規模のeQTLを同定し、機能未知遺伝子や効果不明なゲノム多型の機能解析を行う。

b. 生理機能検査値及びコホート情報との関連解析

地域支援センターやサテライトにて収集し、a. で実施した血液のトランスクリプトーム解析の結果と詳細な生理機能検査値やコホートデータとの関連解析を実施することで、生活習慣と関連して発現量が増減する遺伝子などの疾患マーカーとなり得る遺伝子を同定する。さらに追跡調査などで収集した罹患情報と遺伝子発現変化を捉えることで個別化予防の実現に資する疾患マーカーの同定も目指す。

c. 不死化細胞を用いた薬剤マーカー探索と高度なトランスクリプトーム解析

不死化細胞の基底状態のオミックス情報を取得し、薬剤暴露等による遺伝子発現変化トラ

ンスクリプトーム解析で捉えることで薬剤マーカーを探索するとともに、様々な環境の細胞に対し長鎖シーケンサー、strand specific RNA-seq アレル特異的発現解析、スプライスバリエーションの解析などによるトランスクリプトーム解析を実施することで高度な eQTL 解析や関連解析の手法などの技術開発を行なう。

②プロテオーム解析

第1段階では、血漿などの生体試料の質量分析技術（網羅的解析法と標的定量法）を確立するとともに、コホート参加者 501 人の解析で検出された 256 種の血漿タンパク質の検出頻度と検出ペプチドのアミノ配列を、公開情報基盤 jMorp として全国の研究者に提供した。

第2段階では、これらの成果を応用・展開することにより、全国の研究者による、疾患バイオマーカーの探索・検証などの研究開発を支援する。具体的な取組は以下の通りである。

a. 共同研究への積極的な対応

TMM計画のコホート検体を用いた共同研究において、バイオマーカー探索・検証などを目的とするプロテオーム解析を実施する。

b. 試料分譲への対応

プロテオーム解析を目的とするコホート検体の分譲要請に対応して、貴重で限りある生体試料の有効活用の観点から、可能な場合は当機構においてプロテオーム解析を実施してデータを取得し、試料の代わりにデータとして分譲する。

c. 不死化細胞を用いた薬剤マーカー探索

第1段階における検討により、コホート検体を用いた網羅的プロテオーム解析はスループットの面で多検体の実施が困難であることが判明したため、全ゲノムリファレンスパネルのアノテーションの充実を目的として、ゲノム解析が実施されたコホート参加者の株化細胞を対象に、多層オミックス解析の一環としてプロテオーム解析を実施する。

③細菌叢メタゲノム解析

近年、口腔細菌叢が歯周病、う歯などの口腔疾患のみならず、糖尿病、動脈硬化、関節リウマチなど様々な全身性疾患と関連することが報告されている。TMM計画の前向きコホート調査において、口腔細菌叢とこれら疾患との関連を縦断的に解析することにより、疾患バイオマーカーあるいは介入標的となりうる細菌種を同定できる可能性がある。

第1段階では、コホート参加者約 2.5 万人（地域住民コホート約 1.7 万人、三世代コホート約 0.8 万人）を対象に、歯科検診の際に細菌叢メタゲノム解析のための口腔検体（歯垢、唾液、舌苔）を収集した。第2段階においては、引き続き歯科検診と検体収集を行なうとともに、これら検体を用いた細菌叢プロファイリングを行う。これらの解析結果は、個別化予防・個別化医療の実現と発展のための公開情報基盤として、全国の研究者に提供する。

これらの情報を活用して期待される成果としては、様々な疾患の危険因子あるいは逆に抑

制因子となる細菌種の同定、細菌叢プロファイリングに基づく疾患リスク判定法・予防法の開発、疾患予防・治療のための介入標的となりうる細菌種の同定などが挙げられる。

解析手法としては、多検体を比較的安価に解析可能な、16Sリボソーム遺伝子解析を中心とするが、一部のコホート参加者を対象とする症例対照研究などでは、全メタゲノムショットガン解析の実施も検討する。

5. 個別化予防に向けた統合解析のための基盤構築

1) 発症リスク予測手法開発

地域住民コホートで対象とする疾患のほとんどは common disease であり、疾患により比率は異なるが、発症には遺伝的素因と環境要因、更にはそれらの相互作用が影響する。遺伝的素因の影響力は、双生児や家系構成員を用いた研究により推定されているが、前述の通りこれまでの大規模 GWAS で同定された疾患感受性多型の効果を統合しても、推定された遺伝率に到達しない。

そこでTMM計画では第1段階より、新規の発症リスク予測法の開発を進めてきた。東北大学では、多数のバリエーションからのリスク要因の抽出、更には、膨大な組み合わせに上る遺伝子・遺伝子相互作用や遺伝子・環境相互作用からリスク要因を柔軟に抽出するために、統計的機械学習手法の一種である高次元変数選択（スパースモデリング）の枠組みを利用したゲノムリスク予測法の開発を行った。独自に開発したSmooth-threshold multivariate genetic prediction (STMG) 法では、GWASなどでのバリエーション選択時に用いられるゲノムワイド有意水準について、数学的な工夫を行い、選択されたりリスク因子候補間の交絡を調整した上で、予測モデルを構築できる。実際に、この手法をアルツハイマー病公共次世代シーケンサー（NGS）データに適用し、通常のPolygenic model (PGM) よりも高い精度を示すリスク予測モデルの構築に成功した²³。

他方、GWASが疾患等と関連する多型を同定することを目的とするのに対し、Polygenic model (PGM) は測定した多型のすべてを線形混合モデルに適用することで高精度にリスクを算出することを目的とした手法である。岩手医科大学では、独自に開発した性年齢情報を調節変数として利用できるiwate polygenic model (iPGM) を開発し、国内コホート・バイオバンクと連携し数万検体を利用することで、脳梗塞の遺伝的リスクを算出できる新規予測法を確立した²⁴。

第2段階においては、上記の手法を中心にした予測モデルの構築、オミックス情報や古典的リスクを考慮した発症リスク法の開発などを行う。特に、前向きに取得された環境データを利用することで、網羅的な遺伝子・環境相互作用解析を行い、これまで勘案が不可能であった相互作用成分を取り込んだリスク予測モデルの構築を行う。従来のゲノム多型のみならず、NGSによるレアバリエーション、環境要因、遺伝子・環境相互作用を丁寧にリスク予測モデルに取り込む。また、文献情報などの大規模付随情報、メタボロミクス・プロテオミクスなどのバイオマーカー情報など多様な情報を深層学習などの人工知能技術を用いて分析することで、従来の線形手法では実現できなかったリスク予測を可能にする。

2) 統合・知識データベースと病院連携

第1段階では、健康調査情報および解析情報を統合する統合データベース dbTMM を開発し、全国の研究者が所定の手続きを経てこれらの情報を利用できるようになった。また、dbTMM により、ゲノム・オミックス解析情報及び健康調査情報に基づくコホート対象者の精確な層別化が可能となった。これは、ゲノム医療実現推進に必要なデータシェアリングのための基盤構築の一環となるものである。

第2段階では、ゲノムコホートの詳細二次調査、追跡調査にともない、dbTMM に収載する情報の種類及び量を拡大する。特に、詳細二次調査、追跡調査によって得られる時系列のデータの統合・機能強化を行い、高度化する。また、病院医療情報の利活用による追跡調査（電子的追跡調査）として、調査実施地域内の基幹病院等と連携し、コホート参加者の診療情報を収集する。これにより、診療情報、ゲノム・オミックス解析情報、生活習慣・環境暴露等の健康調査情報を統合した疾患登録、フェノタイピングを行い、それらの結果を統合データベースに収載することでデータシェアリングを行う。

他バイオバンクとの連携のため、インターネットから利用可能な統合データベース dbTMM-pub およびカタログを開発し、このシステムを国内の主要なバイオバンクの横断検索に活用できるようにする。この際、日本医療研究開発機構によるバイオバンクの共通項目の策定に協力する。これにより、「貯めるだけでなく、活用されるバンク」を目指してユーザーの利便性を向上させることができる。

また、dbTMM を用いた大規模データ解析・人工知能解析の結果や研究成果等を、多因子疾患の遺伝要因・環境要因に関する知識として集積する構造化知識データベース kbTMM を開発する。大規模データ解析・人工知能解析には、スパースモデリングなどを用いて変数選択を行い、深層学習をはじめとした機械学習技術を用いる。こうして得られた解析結果や研究成果等の集積した知識により、dbTMM の高度な利用と精確なフェノタイピングを支援する。dbTMM に収載されることになる約 15 万人の症例については、超高次元ベクトルの変数選択とクラスタリングにより類似症例検索を可能とし、より高度な層別化を支援する。この類似症例検索は疾患の発症予測への応用が可能である。

構造化知識データベースの開発にあたっては、RDF やオントロジー (HPO¹ 等) に基づき相互運用性を高める。kbTMM を用いて、多因子疾患の遺伝情報回付等ゲノム医療の支援を行う。

1. Human Phenotype Ontology (HPO ; 人間表現型オントロジー) : 人間の病気の原因となる表現型について標準化された用語を提供すること

統合・知識データベースの形成により、ゲノム情報等の大規模データ解析・人工知能技術の利活用と集積した知識の支援により、層別化された試料・情報を提供できるバイオバンク、「インテリジェント・バイオバンク」へと進化させる。

6. 遺伝情報回付

1) 個人への遺伝情報回付へ向けての検討の経緯

TMM計画の目的とする個別化医療・個別化予防の実現のためには、個人への遺伝情報の返却（回付）は必須である。コホート研究においては、説明同意文書において個人への遺伝情報を回付する要件を定め、その準備が完了したら改めて対象者に連絡し同意を得ることとしている。しかしながら特に現状の研究において、その研究参加者個人への遺伝情報の回付は、我が国では例がなく世界的にもほとんど行われていない²⁵。

特にTMM計画における地域の住民を対象にした疾患コホートのように、医療機関が関わらない、いわゆる健康な方（ostensibly healthy）な方を対象にして、かつ対象者数も大規模な住民コホートでの回付の前例はない。また、海外では、様々な検討が始まっているが²⁶、コンセンサスのあるガイドライン、方針のようなものは未だない。一方、倫理的側面では、参加者への裨益（有益性）とともに、研究の文脈であることより“不利益がない”という視点は非常に重要である。

そこで、第1段階では、東北大学及び岩手医科大学が、平成27年2月に遺伝情報回付検討タスクフォースを立ち上げて検討を開始し、抽出した種々の検討課題を報告書にまとめた。その後、我が国のバイオバンク研究者、遺伝関連学会との意見・情報交換を経て、同報告書は平成27年5月に開催された遺伝情報等回付検討委員会に提出された。本検討委員会では、回付に関してのポリシー、方向性について議論が行われた。

2) 個人への遺伝情報回付のパイロット研究（単一遺伝性疾患）

遺伝情報等回付検討委員会においては、個人への遺伝情報回付の基盤を得るための課題点の検証や抽出を目的とする、段階的な研究の必要性が議論され、パイロット研究を行うこととなった。ただし、生活習慣病などの多因子疾患の回付については、その臨床的妥当性、臨床的有用性、また actionability（対応介入方法）の知見が、現時点では不十分であることからパイロット研究の回付疾患の対象とせず、単一遺伝性疾患が選択された。

本研究では、個人への遺伝情報の回付における心理社会的側面の把握、技術的・手続き的な課題の検証、医療との連携、遺伝カウンセリング体制の構築への知見などを得て、将来の回付にあたっての基盤整備に活かすことを目的としている。対象疾患を「家族性高コレステロール血症」とし、コホートの健康調査結果より表現型（脂質異常症）を有している方に研究の呼びかけを行った。参加希望者については、遺伝講習会受講後、遺伝情報回付の再同意を取ることとしている。なお、本研究では、家族性高コレステロール血症の原因である3つの遺伝子（LDLR、PCSK9、APOB）のみのゲノム情報を回付し、対象者および血縁者に重大な影響を与える偶発的所見が見いだされる可能性は想定していない。

3) 個人への遺伝情報回付に向けての基盤整備

第2段階においては、引き続き遺伝情報回付パイロット研究を推進する。併せて、TMM計画における個人への遺伝情報回付の基盤を得るため、複数のパイロット研究を計画する。

加えて、実際の遺伝情報の回付だけではなく、①疾患に関連する遺伝情報回付に関わる地域の医療関係者に対して、どのような意識や態度があるのか、受診された際にどのような資源や援助が必要かを明らかにする調査、②生活習慣病などの多因子疾患の発症リスクを知った際の理解についての調査、③回付後の行動に関する調査などを計画する。

そして、コホート調査により得られ、統合・知識データベースに格納される生活習慣や環境因子を利用し、ゲノム解析、オミックス解析によって得られたデータをもとに開発した精緻な遺伝的リスク予測手法を用いて、遺伝情報回付に向けての基盤整備を目指す。

7. ゲノム医療実現のための人材育成

1) 第1段階における人材育成の実績

コホート調査における参加者リクルートの支援を行う人材である、ゲノム・メディカルリサーチコーディネーター (GMRC) については、看護師、臨床検査技師、一般適性者等を被災地域で採用し、必要な教育、研修を実施して試験に合格したものを GMRC と認定した。現在までに、199 名の ToMMo GMRC 認定者が在職している。(実働数 155 名)

遺伝カウンセラーについては、4 名が東北大学大学院のコースを修了し、認定遺伝カウンセラー (CGC) の資格を取得した。同コースには現在 3 名が在籍している。臨床遺伝専門医については、新たに 3 名が認定され、現在 5 名が在職している。バイオインフォマティク・生命情報科学者については、6 名の博士課程、17 名の修士課程レベルの学生の教育を行っている。また、2 名の民間研究員を受け入れ、博士号の取得をめざした研究指導を進めている。

2) 第2段階における人材育成の計画

第2段階においては、診療情報等に基づく正確な表現型情報を抽出することが第1段階以上に重要であるため、メディカルクラーク取得者と同程度の技能を有する GMRC やデータマネージャー (TMM医療情報コーディネータ) の育成を行う (目標: 80 人)。

バイオインフォマティク・生命情報科学者については、主としてゲノム解析をする研究者と、医療情報とゲノム情報を合わせて解析する研究者の両者が必要となる。こちらに関連し、TMM計画で収集される最先端の大規模な情報を最大限活用した、On the Job Training による情報解析技術のトレーニング及び最先端の研究者の招聘による新規情報解析技術の習得を進めることで、実践型の人材育成を行う。また、国外や産業界とも協力することにより、多面的な育成や多角的なキャリアパスの形成を行う。

認定遺伝カウンセラーについては、関係学会との連携や遺伝情報回付パイロット研究の結果を踏まえて、多因子疾患のカウンセリングに対応するために必要な人材像を明確化するとともに育成を行う。

多因子遺伝子病を中心とする遺伝子診療や個別化医療を実践するため、臨床遺伝専門医制度の規定に相当する遺伝医療に関する知識を持った医師 (TMMゲノム医療スペシャリスト) の養成をめざす。第2段階中に、東北メディカル・メガバンク機構の教員、ToMMo クリニカ

ル・フェローについて数名の育成を目指す。

ゲノム医療研究に重要となる人材として、医療情報とゲノム情報をあわせながら、セキュリティに配慮して統合的に情報管理・解析ができるゲノム医療情報技術者・研究者の育成を目指す。

ゲノム医療体制の構築に必要な認定遺伝カウンセラー、TMM医療情報コーディネータ・メディカルクラーク、データマネージャーのキャリアパスの形成を図る。

8. まとめ

TMM計画は、第1段階においてコホート15万人リクルートの達成、ゲノム解析における2,049人分全ゲノムリファレンスパネル(2KJPN)、基準ゲノム(JRGv1)の公開、オミックス解析における1,008人分の血漿の「日本人多層オミックス参照パネル(jMorp)」の公開、100人分の白血球2細胞の網羅的解析結果「日本人多層オミックス参照パネル(iMethyl)」の公開を行ってきた。これらにより、ゲノム医療実現推進のデータシェアリングの基盤を形成した。

平成29年度からの第2段階においては、創生という新たな段階に向けて、我が国のゲノム医療実現に向けた基盤形成とデータシェアリングを進める。また、東北大学及び岩手医科大学においても、個別化予防・個別化医療の実現を目指し、ゲノム情報、多層オミックス解析のデータを用いた、データ主導型の医療への基盤研究を推進するとともに、発症リスク予測手法開発等の次世代医療に向けた研究を推進する。さらには、日本人多層オミックス参照パネルを構築、公開することで、健康、未病、病気の各段階をとらえられる個別化予防・個別化医療プラットフォームの基盤形成を目指す。

(参考1) WGの構成

主査：榊 佳之	静岡雙葉学園 理事長
高木 利久	東京大学大学院理学系研究科 教授
油谷 浩幸	東京大学先端科学技術研究センター 教授
小原 收	かずさDNA研究所 副所長
木川 隆則	理化学研究所生命システム研究センター 生体分子構造動態研究チーム チームリーダー
菅野 純夫	東京大学大学院新領域創成科学研究科 教授
榊原 康文	慶應義塾大学理工学部 教授
徳永 勝士	東京大学大学院医学系研究科 教授
中川 英刀	理化学研究所統合生命医科学研究センター ゲノムシーケンス解析研究チーム チームリーダー
夏目 徹	産業技術総合研究所創薬分子プロファイリング研究センター 研究センター長
松本 直通	横浜市立大学大学院医学研究科 教授
山田 亮	京都大学大学院医学研究科附属ゲノム医学センター 教授
吉田 輝彦	国立がん研究センター研究所遺伝医学研究分野 分野長

(参考2) WG開催状況

第1回 平成28年11月21日(東北大学東京分室)

- 東北メディカル・メガバンク計画の第2段階の推進にかかる基本方針
(中間とりまとめ)
- 全国WGの運営について
- 第2段階計画実施計画案について

第2回 平成29年1月24日(東北大学東京分室)

- ゲノム・オミックス統合解析WG中間とりまとめ(案)

文献

1. Popejoy AB, Fullerton SM: Genomics is failing on diversity. *Nature* 538:161–164, 2016
2. Wain LV, Shrine N, Miller S, et al: Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* 3:769–81, 2015
3. Smith DJ, Escott-Price V, Davies G, et al: Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatry* 21:749–57, 2016
4. Styrkarsdottir U, Thorleifsson G, Sulem P, et al: Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* 497:517–20, 2013
5. Kong A, Masson G, Frigge ML, et al: Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet* 40:1068–75, 2008
6. Manolio TA, Collins FS, Cox NJ, et al: Finding the missing heritability of complex diseases. *Nature* 461:747–53, 2009
7. Eichler EE, Flint J, Gibson G, et al: Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11:446–50, 2010
8. Manolio TA, Bailey-Wilson JE, Collins FS: Genes, environment and the value of prospective cohort studies. *Nat Rev Genet* 7:812–20, 2006
9. Fuchsberger C, Flannick J, Teslovich TM, et al: The genetic architecture of type 2 diabetes. *Nature* 536:41–7, 2016
10. Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 15:695–707, 2016
11. Hoffmann TJ, Ehret GB, Nandakumar P, et al: Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* 49:54–64, 2017
12. Nikpay M, Goel A, Won HH, et al: A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 47:1121–30, 2015
13. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 49:27–35, 2017
14. Eeles RA, Olama AA, Benlloch S, et al: Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* 45:385–91, 391e1–2, 2013
15. Cheng TH, Thompson DJ, O'Mara TA, et al: Five endometrial cancer risk loci identified through genome-wide association analysis. *Nat Genet* 48:667–74, 2016
16. Koshiba S, Motoike I, Kojima K, et al: The structural origin of metabolic quantitative diversity. *Sci Rep* 6:31463, 2016

17. Suhre K, Shin SY, Petersen AK, et al: Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 477:54–60, 2011
18. Prasad G, Jones B, Schmidt E, et al: Global metabolomic profiles reveal differences in oxidative stress and inflammation pathways in smokers and moist snuff consumers. *Journal of Metabolomics* 1, 2015
19. Harada S, Takebayashi T, Kurihara A, et al: Metabolomic profiling reveals novel biomarkers of alcohol intake and alcohol-induced liver injury in community-dwelling men. *Environ Health Prev Med* 21:18–26, 2016
20. Jaremek M, Yu Z, Mangino M, et al: Alcohol-induced metabolomic differences in humans. *Transl Psychiatry* 3:e276, 2013
21. Xu T, Holzapfel C, Dong X, et al: Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med* 11:60, 2013
22. Ueta M, Sawai H, Shingaki R, et al: Genome-wide association study using the ethnicity-specific Japonica array: identification of new susceptibility loci for cold medicine-related Stevens–Johnson syndrome with severe ocular complications. *J Hum Genet*, 2017
23. Ueki M, Tamiya G: Smooth-Threshold Multivariate Genetic Prediction with Unbiased Model Selection. *Genet Epidemiol* 40:233–43, 2016
24. Hachiya T, Kamatani Y, Takahashi A, et al: Genetic Predisposition to Ischemic Stroke: A Polygenic Risk Score. *Stroke*, 2016
25. Haukkala A, Kujala E, Alha P: The return of unexpected research results in a biobank study and referral to health care for heritable long QT syndrome. *Public Health Genomics* 16:241–50, 2013
26. Knoppers BM, Zawati MH, Senecal K: Return of genetic testing results in the era of whole-genome sequencing. *Nat Rev Genet* 16:553–9, 2015