

統合ゲノム医科学情報研究拠点の形成

① 計画の概要

ゲノム解析技術の飛躍的な発展により大規模ゲノム配列解析計画が国内外で進行中である。またゲノム配列のどの塩基多型・変異がいかなる生物学的機能・疾患関連性を持つかを解析する大規模計測技術も進展し、大規模かつ多岐にわたる生命ビッグデータも急速に蓄積している。本研究では、最も情報の集約化が進み喫緊の成果が期待される小児希少疾患ゲノム・がんゲノム統合情報解析をモデルに他の生活習慣病も対象にして、臨床情報・ゲノムデータ、先端計測データについてビッグデータを集約する。集積データを最先端情報解析技術を用いて統合的に解析することにより、疾患の発症機構の全体像を解明する。特に生活習慣病においては複雑な環境因子、遺伝因子の機能的相互作用が重要であり、ゲノム配列だけではなく、遺伝子発現情報、腸内細菌叢との相互作用を含めた多角的統合解析が特に重要となる。

データの集約と利活用については、本格稼働を開始する学術情報ネットワーク SINET を活用する。SINET を通じて、厚労省ががんゲノム医療実現プロジェクト臨床情報の集約拠点であるがんゲノム情報管理センター（C-CAT）およびバイオバンクネットワーク等、我が国の大規模情報集約拠点とのデータ連携体制を構築する。これらの情報集約拠点が、主として、ゲノム配列自体あるいは付随する臨床情報の収集を対象とするのに対し、本計画ではそれらにさらに先進計測技術による大規模生体関連情報を加えて、より高次の意義抽出を目指す。日本人ゲノムデータの集約拠点であるバイオサイエンスデータベースセンター（NBDC）とも連携して、個人情報の機微性にも留意した枠組みの構築を実現する。全体として、我が国の主要ゲノム情報資源を結ぶ情報解析技術の集約型の学術コアの形成を目指すものである。創出される情報学的資源はこの情報中核から SINET を通じてさらに広く全国の研究者に提供される。

② 学術的な意義

ゲノム医学研究では高精度の臨床情報とゲノム情報の統合解析が重要で、物理的・情動的な疾患リソースの果たす役割が大きく、高度医療機関を中核としたネットワーク構築が重要となる。本課題研究者による大規模ゲノム解析の成果としては、東京大学医科学研究所がバイオバンク・ジャパン（BBJ）（研究代表者村上善則）リソースを基盤に、理化学研究所と共同で多数の疾患についてリスク因子となるゲノム多型を解明し Nature 姉妹誌だけでも 50 本以上の論文を発表している。同様に、辻省次らが AMED 「希少・難病分野の臨床ゲノム情報統合データベース整備」で、宮野悟らが文科省新学術領域「システム癌新次元」で、松原洋一（国立成育医療研究センター）、松本直通（横浜市立大学）らが AMED 難治性疾患実用化研究事業で、間野博行（国立がん研究センター）、松田浩一（東京大学医科学研究所）らが次世代がん医療創生研究事業で、それぞれの対象疾患における成果を結実させている。同定された新規疾患関連多型・変異は、多くの疾患研究者コミュニティに利用されて共同研究へと発展し、また、製薬企業等でのデータ利活用により新規創薬シーズとして研究が飛躍的に加速されている。実際、東大医科研ヒトゲノム解析センターは長年、これらのデータ集約、利活用の拠点として、また独自の情報解析手法開発の拠点として我が国を牽引してきた。また、同大学新領域創成科学研究科は大学院教育を担い、ゲノム医学とバイオインフォマティクスの融合分野に精通した人材育成により該当分野を牽引している。しかし、近年のゲノム関連研究の予想を超えて急速な拡大、特に産出データの多様化、高度化は、さらなる解析環境の深化と人材の育成を求めている。本研究課題により実現されるデータ集約拠点において、基礎、応用研究の正の循環をさらに加速させる。

③ 国内外の動向と当該研究計画の位置づけ

世界各国で多数の自国民についてその個人のゲノム配列を網羅的にカタログ化する試みが進められている。これにより国民の潜在的な健康リスクの把握、時には民族に特徴的な病態・薬剤への応答等を踏まえた個人に最適化された治療戦略の実践を目指す。これは、医療関連分野だけでなく広くバイオ関連研究分野の産官学の研究・産業応用において、将来にわたって自国利益の確保を目指す国家事業である。英国” Genome England”、米国” US ALL”、中国からここ数年の間に登録されるゲノム配列は 200 万件を超える。日本でも東北メディカルメガバンク機構において標準日本人ゲノムのカタログ化が進められ、BBJ では 3,000 人の全ゲノム配列情報を含む 27 万人の生体試料、関連情報の蓄積が完了している。GA4GH 等ゲノムデータを利活用する国際的な枠組み整備が急速に進んでいる。急速なデータ蓄積に際して全ての国にとり喫緊の課題は、いかにしてこれらの情報を集約化と先端計測技術を加えた統合的情報解析を可能にするか、その計算機環境の整備である。取り組みはその国の実情に即して異なるが、我が国ではその対応への具体的な策案が未定である。

④ 実施機関と実施体制

東京大学医科学研究所，同医学部附属病院，同大学院医学系研究科，同大学院新領域創成科学研究科，同先端科学技術研究

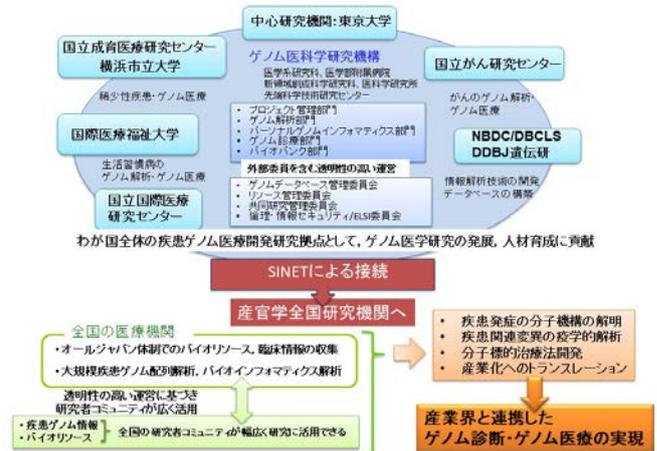


図1. 実施機関と実施体制

センター、同定量生命科学研究所が中核機関となる。直接的には、これらの部局に他の全学部局を加えた「ゲノム医科学研究機構」が実務を担当する。大学外参画機関の役割分担として、NBDC と国立遺伝学研究所 (DDBJ)、および統合データベースセンター (DBCLS) では、機微性に留意したデータベースの構築、情報解析技術の開発を担う。対象疾患の観点から、国立がん研究センターはがんを、国立成育医療研究センター・横浜市立大学は稀少性難病を、国立国際医療研究センターは生活習慣病等の多因子疾患を、国際医療福祉大学は神経難病についての解析を担う。また、SINET との接続は、東京大学未来社会協創推進本部との連携のもとで行う。いずれの研究機関においても研究科・部局レベルでの合意が得られている。

中核組織として運営にあたる東大ゲノム医科学研究機構では、プロジェクト運営部門で統括と8つの分担研究機関との緊密な連携、アウトリーチ活動を、疾患リソース部門では2附属病院を中心にゲノム DNA、血液、病理組織等の収集・管理を、ゲノム解析部門では大規模ゲノムデータの集約解析と情報のデータベース化、維持管理を行う。データベース部門では個人情報保護に務めながら研究者コミュニティによる活用推進を図り、制限付きアクセスや認証・秘匿化によるデータ提供を、ゲノムインフォマティクス部門では、多角的データの統合解析を可能とする情報解析技術の開発を行う。共同研究管理部門では、全国の研究者による共同利用・共同研究を推進し、倫理・情報セキュリティ委員会ではゲノム情報、臨床情報の扱い等に関する ELSI の検討と情報・セキュリティを担当する。ゲノム診療検討委員会は精密ゲノム医療推進と実施に関する課題の検討を行う。

⑤ 所要経費

計算機環境の整備費を中心に、シークエンサーを含む最低限度の計測機器設備費および人件費として、総計で 183 億円を計上する。その内訳を以下に示す。

- ・次世代シーケンサー、その他の先端計測機器として 5 億円
- ・ゲノムインフォマティクスのための計算機として 90 億円
- ・ゲノム関連解析データ発生費用として 50 億円
- ・リソースの収集・管理システムとして 10 億円
- ・人件費 50 人 x 7年 28 億円

⑥ 年次計画

2019 年度：大規模ゲノム関連データについてのデータ集約・解析拠点および関連情報解析技術開発拠点を整備、あわせて計測設備整備を行う。短鎖高出力次世代シーケンサー 1 台、長鎖シーケンサー 3 台を導入する。一細胞解析機器、イメージング機器等、先端計測機器の整備も行う。構築される計算機環境を SINET へ接続し、初年度からデータシェアの実践を図る。SINET を通じた国際的なデータシェアを図る枠組みの開発も行う。一方でデータ資源、あるいは生体試料自体の収集・管理拠点を整備し、統合的大規模リソースとしての拠点整備を開始する。ゲノム多様性データベースについて、研究者コミュニティが利用できるようなアクセスポリシーを定め IRB 承認の手続きを進める。

2020 年度～2025 年度：個人ゲノム関連データの利活用について直近の成果が期待される分野（がん、難治疾患）についてモデル事業を開始する。統合ビッグデータにさらに臨床情報を付加した解析を実施し、疾患発症・重症化や予後に関連したゲノム多様性、背景にある生物学的分子機序を明らかにし、治療法開発のためのシーズ開発、産業界へのトランスレーションを進める。また、これらの解析を担う若手人材の育成を実践する教育プログラムを開始する。これには現在、新領域創成科学研究科において民間 12 社と実践している「データサイエンス育成プログラム”DSTEP”」を母体とする。さらに多くの疾患分野へ対象を拡張しつつ、データシェアによる研究者コミュニティによる利用を開始する。これと同時に、診療へのトランスレーションを開始する。専門研究者、医師、患者に大規模ゲノム解析の結果を伝える専門家（ゲノムトランスレータ）、遺伝カウンセラーなど、接続的人材の育成も行い医療機関における新規の職種として実現していく。分野の拡大により研究期間内に雇用された人材は期間終了後も継続的に雇用が可能となる。

2020 年度～2025 年度：個人ゲノム関連データの利活用について直近の成果が期待される分野（がん、難治疾患）についてモデル事業を開始する。統合ビッグデータにさらに臨床情報を付加した解析を実施し、疾患発症・重症化や予後に関連したゲノム多様性、背景にある生物学的分子機序を明らかにし、治療法開発のためのシーズ開発、産業界へのトランスレーションを進める。また、これらの解析を担う若手人材の育成を実践する教育プログラムを開始する。これには現在、新領域創成科学研究科において民間 12 社と実践している「データサイエンス育成プログラム”DSTEP”」を母体とする。さらに多くの疾患分野へ対象を拡張しつつ、データシェアによる研究者コミュニティによる利用を開始する。これと同時に、診療へのトランスレーションを開始する。専門研究者、医師、患者に大規模ゲノム解析の結果を伝える専門家（ゲノムトランスレータ）、遺伝カウンセラーなど、接続的人材の育成も行い医療機関における新規の職種として実現していく。分野の拡大により研究期間内に雇用された人材は期間終了後も継続的に雇用が可能となる。

⑦ 社会的価値

本研究拠点から生産される日本人独自の要素を含む生体・臨床情報は、研究者コミュニティや産業界に広く利用され経済の活性化につながる。即ち、本研究基盤が今後長年にわたり、日本のすべての医学・生物学研究の基盤となって学術の発展に貢献するとともに、その成果が疾患の原因解明による革新的な予防、診断、治療法の開発、創薬研究や新規医療技術の確立を促し、その産業基盤と雇用を確立する。これは、該当分野への全体的な国民の理解を促進するとともに、同分野を将来にわたって牽引する若手人材を誘致の育成につながると思っている。本計画はその正の循環を回す最初のステップとして位置づけられる。該当する SDG 項目としては、SDG3:健康と福祉の増進に貢献する；医療費削減や医薬品の輸入超過を抑制する上でも社会的価値が高い；SDG4:質の高い教育；今後長期間にわたるバイオ関連人材の育成の基盤となる；SDG9:産業と技術革新の基盤についても人材育成も含めて、疑いなく必須の社会基盤である、等が挙げられるが、我が国の風土に適した社会的価値はより広範囲にわたる。

⑧ 本計画に関する連絡先

村上 善則（東京大学・医科学研究所）



図2. 研究計画の概要