

ゲノム情報を統合したコホート研究における情報管理システムの構築

● 関根 章博

京都大学医学研究科

<研究の目的と進め方>

健常者対象のゲノム・疫学コホートは、Genome-Wide Association Study (GWAS) によって同定された複数の疾患関連遺伝子群トータルの遺伝リスクの高さ（易罹患性）を前向きに証明し、さらに発症の際に、その易罹患性と相互作用する生活習慣・環境要因の種類と程度を知る有効な手段である。つまり、如何なる遺伝リスクを有する人が、如何なる生活習慣や環境要因にどの程度、どの位の期間に渡って曝されることで発症のリスクを高めるのかを知ることが可能になると考えられる。それ故、ゲノム・疫学コホートは、GWAS では成しえなかった多因子性疾患の早期診断・早期予防・早期治療の実現には不可欠なアプローチとなり、その成果は患者さんのQOLを向上させ、同時に医療費の大幅な削減に寄与することが期待される。しかしながら、研究に協力していただくボランティアの方々のプライバシー保護を十分に確保し、ボランティアの健康状態や生活習慣・環境要因の変化を長期間に渡って定期的に追跡していく必要があり、また、試験の検出力を保持し、正しい成果を導くためにも、情報の欠落を回避し、正確な情報を取得するシステム構築が不可欠である。すなわち、ゲノム・疫学コホートの実現には、1) 参加者保護を重視した匿名・連結・削除命令システムの構築、2) 個人情報管理システム、3) 欠落の少ない、正確な健診情報を取得するための定期健診システム（①受診システム、②検査・検診会社との情報授受のシステム、③検体授受のシステム）の構築、4) 欠損が少なく、精度の高い、生活習慣・環境要因の調査システム、5) 検体管理、6) 死亡・発症に関する追跡システム、7) ゲノム情報の取得・管理システム、8) データベースの構築、9) 研究者への情報開示システム等、多岐に渡る情報管理システムの構築が不可欠である。一方、京都大学医学研究科では長浜市民および市の協力を得て複数の疾患（閉塞性肺疾患・喘息・アレルギー性呼吸器疾患、慢性関節リウマチ、変形性関節症、糖尿病、高血圧、心・脳血管疾患、各種消化器がん、排尿困難、前立腺疾患）をターゲットにしたゲノム・疫学コホート（ながはま0次予防コホート事業）を診療科および基礎研究科協力の下、開始した。本事業の申請は、大学の倫理審査に加え、市の条例下（http://www.city.nagahama.shiga.jp/reiki/mokuji_bunya.html）で設立された市民、行政、市の基幹病院ならびに市の選定した有識者の参加する委員会の審査も受け、承認を得ている。そこで、本課題であるコホート研究遂行のためのシステム構築については、このフィールドを活用し、ゲノム情報を包含にしたコホート研究を支える仕組みを構築することを目的とし、ながはま0次予防コホート事業における安定稼働を持ってゴールとしたい。また、構築されたシステムは広く活用いただけるよう努力していきたいと考えている。

<2008年度の研究の当初計画>

上記目的を達成するため、2008年度の計画は、①参加者保護を重視した匿名・連結・削除命令システムの構築、②個人情報管理システム、健診システム（③受診システム、④検査・検診会社との情報・検体授受のシステム、⑤特定健診への対応）の構築、⑥生活習慣・環境要因の調査システム、⑦検体管理システム、⑧ゲノム情報の取得・管理システム、⑨データベースの構築および研究者への情報開示を目標とする。上記システムは可能な限りプログラムにて管理し、ミス、欠落等を最小限に抑え、精度維持を可能にするシステム構築を目指す。

<2008年度の成果>

2008年度は順調に進んでいる。以下に、構築したシステムの概要を説明する。

①匿名・連結・削除命令システム：個人情報保護は本事業遂行に伴う最優先課題である。よって、個人情報保護は、技術的・物理的・人的・組織的安全措置を念頭に、二重匿名化にて構築した。つまり、全ての情報および情報の元となる検体が集約する健診現場には、多くの関係者（行政関係者、医師、京大研究者、健診・検査会社関係者、医療機器関係者）が立ち入るため、個人情報（氏名、所在、生年月日等）を含まない、匿名化番号による受診とした。また、健診によって得られた全ての情報は、まず再匿名化管理者に集め、再匿名化した上で、関係研究者に開示する仕組みとした。この際、個人情報管理者のみが、個人—匿名化連結情報を所持し、これを長浜市に設置、一方、京大には匿名化—再匿名化連結情報を取り扱える再匿名化管理者を現状1名設置し、全ての健診情報、検体をまずその再匿名化管理者に提出し（健診現場で取得できる情報は鍵付きケースにて直接提出、検査・健診会社も再匿名化管理者に結果報告を行い、それ以外の者には報告しないことを義務付けている）、再匿名化した情報のみを診療・ゲノム情報管理者が整備をして、研究に活用する仕組みとした。なお、健診現場に立ち入る一部の管理者（個人情報管理者、健診・検査会社、再匿名化管理者）以外は、一切の情報の持ち帰りの手段を禁じた。それ故、全ての情報を個人情報に連結するには、長浜市の個人情報管理者による個人—匿名連結と、京大の再匿名化管理者による匿名化—再匿名化連結の、両者が連動したときのみ可能となる。基本的にはこの連動は行なわないが、例外は特定健診の個人への返却、および問合せ・削除命令の時に限定しており、それ以外は、両施設の最高責任者の許可を必要とする。この二重匿名化にて個人情報（長浜市）、匿名化番号管理（健診現場）、再匿名化番号管理（研究現場）の切り離しが可能となり、厳重な個人情報保護（再遺伝子検査による個人特定を行なわない限り）になると考えている。また、各情報は複数の施設による管理に加え、ネット非接続下、アクセスには登録、指紋照合、パスワードによって盗難にも対応している。よって、健診現場の流れは、まず同意を得、個人情報を長浜市担当者が取得した後、直ちに匿名化を行い、受診は匿名化番号で行なう。健診直後に持ち帰る情報および健診・検査会社の報告は、匿名化番号でまず再匿名化管理者に提出し、再匿名化した情報を研究に資する。この運営に際し、プログラムによる1) 匿名化システム、2) 個人—匿名化連結システムを構築し、長浜市に運営を任せた。また、プログラムによる3) 再匿名化システム、4) 匿名化—再匿名化連結システム、5) 情報削除システムを構築した。

②個人情報管理システム：個人情報入力支援システムを構築し、長浜市に運営を任せた。入力支援システムは、ミスを回避するため、郵便番号の選択入力により町名まで、生年月日の入力により年齢を自動発行する等の、簡便かつ入力ミスを減じるシステムとした。また、個人—匿名化連結情報と連動させ、匿名化番号を入力することで個人の宛名を封書にプリント、あるいは、個人名を入力することで匿名化番号をプリントできるシステムとしたので、個人情報管理者が中身を見ない方式で検査結果の個人への返却（次年度の死亡・発症追跡にも活用）、個人からの関係部署への問合せを可能にした。このシステムを用いて、特定健診項目の

みを抜粋し、参加者に報告することで、コホート参加を他者に知られること無く、特定健診保健指導を受けることも可能となる。また、受診者の匿名化リストのみを自動発行する仕組みとし、全受診者の紛失確認の基盤情報とした。

③受診システム：現状、健診現場にネット環境が整備されていないため、全ての検査を受診したか否かは、受診表（紙ベース）で実施している。情報の紛失か、欠損（予定量の採血ができず途中で断念等）を知るために、受診表は各検査毎のチェック方式とし、スキャナーによって電子化、受診していない検査をプログラムで知りえる方式とした。但し、今後はネット環境が設置された健診施設を利用できる可能性があるため、建物のネット環境を活用し、受診状況を把握するシステムも構築した。また、殆どの検査は、電子化データ（生化学検査、心電図、血圧等）で受理できるが、従来、紙ベースで情報入力していた歯科検診についてはタッチパネルによる入力支援システムを構築し、健診終了時には電子化を可能とした。

④健診・検査会社との情報授受のシステム：健診・検査会社からの結果報告は、匿名化番号（健診・検査会社は個人情報を所持していない）にて受理する。この際、報告には、検体、測定に関するコメント欄（溶血、測定レンジ外のため希釈測定等）、測定キットのレンジ、判定基準値を届け、電子化データとして受領する形式とした。この方式の採用により、結果の信憑性を確保し、また、研究者からの簡単な問合せに対応できるシステムを構築した。また、②の受診者リストおよび③の受診状況と照合し、全結果報告ならびに全保存検体の欠落を検証（紛失を迅速に見つけ出す）できるシステムとした。

⑤特定健診への対応の構築：特定健診結果も④と同様に行なうが、国保連合会へのxmlフォーマットによる報告が義務付けられているため、健診会社からの匿名化番号による報告を受けた後、xmlファイルに個人情報を②の個人情報管理システムから追加記載するシステムを構築した。また、特定健診の報告の一部に質問表による項目（喫煙するか？等）が含まれるため、次項の質問表から特定健診項目に該当する部分を抜粋し、健診会社に報告するシステムを構築した。

⑥生活習慣・環境要因の調査システム：研究対象とする疾患が多いため、それに伴って回答が約770箇所にも及ぶ。そのため、欠落箇所、矛盾結果を目視で判定することは困難であり、後日、参加者に再確認することも困難となる。そのため、健診受付時に事前に回答をいただいた質問表（1次回答）を提出いただき、他の検査受診中にスキャンし、回答のない質問、矛盾する回答をweb上で再質問できる仕組みとした。また、再質問にはコンピュータに不慣れな参加者、意図が分かりづらい質問への回答をサポートする人材を配置した。そのサポートスタッフが、ボランティアの依頼を受け、既に回答した質問も再確認するための質問抽出もweb上で行えるようにし、回答の精度を維持する仕組みとした。また、質問表の中に含まれる特定健診（国保連合会提出）部分のみを抽出し、健診が終了するまでに、入力が正しく行なわれているかをweb上で確認できる仕組みを構築した。これにより、得られた回答（2次回答）は現状ほぼ完全で、1次回答の際に脱落あるいはミスしていた平均約7%の質問をレスキューした。また、1次回答と2次回答の差分や回答変更を調べることで、参加者には回答しにくい質問の抽出にも成功したので、調査質問を設定した担当者にフィードバックし、回答しやすい改良を促すことも可能にした。

⑦検体管理システム：検査会社および健診会社から回収された保存検体（血清、血漿、DNA）は、全て2次元バーコード付きのチューブに分注して保存する。この分注を半自動化で行なうシステムを構築し、パンキング装置保存に適應する形式でリスト化する仕組みとした。

⑧ゲノム情報の取得・管理システム：2008年度のコホート健診は11月28日に開始した。そのため、ゲノムDNAを抽出依頼した検査会社から回収している状況下のため、ゲノム情報の取得は未だ実施していない。但し、本年度2月末には、参加協力者が少なく

とも1,200を超え、健診も一段落するので、それを待ってタイピングを開始する予定である。なお、GWASのシステムは既に研究組織内に構築されていて、稼働している。また、実用的な問題から、すぐに全DNAのゲノムワイドタイピングを実施することは困難であるため、既にGWASによって報告された疾患関連候補遺伝子についてタイピングを行なって前向き証明を開始する予定である。これまでに報告されている疾患関連候補遺伝子座に関するタイピングには、Invader法およびTaqMan法を用いる予定であり、オーダーフォームの作成システム、タイピング実施のためのシステム、タイピング結果の処理システム、タイピング精度を検証するシステムは既に構築済みである。現在、候補遺伝子座の選出を行なっているため、検体がある程度回収されるのを待って、年度内にはタイピングを開始する。なお、遺伝情報を管理するサーバーはネット非接続環境下で、管理者の指紋照合ならびにパスワードでのみアクセスする方式を採用し、7重の施錠下、入退出管理下に設置する準備は完了している。また、診療情報とは別のサーバー内に保管することとしてある。

⑨データベースの構築および研究者への情報開示：全ての情報は電子化し、情報の種別により、いくつかのサーバーに分割し、情報管理者は目的の情報をSQLにて迅速に取得できるようデータベースを構築している。なお、情報を利用する各研究者は登録制とし、登録者のみがアクセスできるよう、指紋およびパスワード照合によってエンターする仕組みとした。現状、情報開示のPCは、ネット非接続下で、設置も1箇所制限している。また、PC上での情報解析は可能であるが、情報は紙や電子媒体にコピーできない状況としてある。現在、大学全体のネット環境の責任部署とセキュリティ確保を前提に討議中のため、開示の方法はボランティア保護を確保しながら、研究者が活用しやすくなるよう改良を検討している。

総括：本年度目標としたシステム構築はプログラム管理にて、ほぼ予定通り達成している。紙媒体、目視、マニュアル操作による情報管理を可能な限り排除した本システム構築により、2007年に実施したパイロット試験に比較して、顕著に欠落やミスの少ない、精度の高い結果取得が可能になった。今後は、ながはま0次予防健診を実施しながら、構築したシステムの問題点を抽出してさらなる改良をする。また、本年度に候補遺伝子座のタイピングをスタートし、目標達成する。

<国内外での成果の位置づけ>

国内外にはいくつかのコホート事業が存在するが、情報管理に関する詳細は表面化しにくく、他のコホート事業との正確な比較は難しい。憶測の域を出ないが、本研究が実現した欠損値やミスの少ない、精度維持したシステム構築は少ないと予想している。それ故、コホート事業の立ち上げに苦勞されている他施設には協力していきたいと考えている。また、健常者を対象としたコホート事業は検出力の問題から、多数のボランティア協力が必要とされ、この点からも、他コホート事業との連携は重要となる。次年度には、より高度なシステムを構築できるので、積極的に他コホート事業とシステム利用に関する討議を進め、また、評価を受ける所存である。また、人間間における遺伝的階層化を考慮し、まずは国内コホート事業との連携を優先したい。

<達成できなかったこと、予想外の困難、その理由>

唯一、研究者への情報開示は制約を受けるため（前記⑨）、個人情報保護を確保しつつ、研究者が扱いやすいよう改良を加えたい。

<今後の課題>

システム全体を構築するためには、1) 死亡・発症の追跡システム、2) 住所・氏名等の変更に関する追跡システム、3) 匿名化下における、2回目以降の健診結果の連結、4) 発症に関連する遺伝情報と生活習慣の相互作用の解析システム構築、の課題が残っていて、次年度の目標としたい。

<成果公表リスト>

リストはなし。コホート研究立上げを予定している班員と、システムの利用について討議を開始している。