

統合データベースプロジェクト研究運営委員会作業部会分科会【医学・糖鎖関係機関】議事要旨

【日 時】 平成22年10月21日(木) 14:00~17:15

【場 所】 ライフサイエンス統合データベースセンター大会議室

【出席者】 林健志(九大)、田中博(医科歯科大)、下川和郎(医科歯科大)、井戸敬介(医科歯科大)、山本洋一(阪大)、古崎晃司(阪大)、徳永勝士(東大)、小池麻子(日立製作所)、成松久(産総研)、鹿内俊秀(産総研)、豊田哲郎(理研)、田中康博(文科省)、堀田凱樹(ROIS)、米澤明憲(DBCLS)、大久保公策、高木利久(作業部会主査)

永井啓一、川本祥子、坊農秀雅、畠中秀樹、金進東、吉羽洋周、河野信、高祖歩美、坂東明日佳、八塚茂、三橋信孝、箕輪真理(以上、DBCLS)

(敬称略・順不同)

【議 事】

1. 議事要旨について(資料確認含む)

議事要旨(事前にメールで確認済み)の修正があれば本会議終了までにお申し出いただきたい。また、委員名簿の変更(JSTからの委員が変更、松原委員の職名変更、高木主査の所属変更)についてご確認いただきたい。

(高木主査より)プロジェクト(以下、PJ)周辺の近況報告を行う。半年後のPJ終了後、事業はJSTに設立される新センターへ移行する。DBCLSは技術開発を目指す。10月1日にJST新センター準備室が設立されるに伴って、高木主査はDBCLSセンター長の任を解かれ、JSTの準備室へ参加した。ただし、引き続きDBCLS特任教授として本PJを担当する。DBCLSの新センター長には米澤東大教授が就任した。準備室の情報はJSTのホームページで発信される。

2. プロジェクト平成22年度上期進捗の評価依頼について 資料2-1、2-2

文科省ライフ課田中調査員より、すでに10月15日付で作業部会の委員に依頼済みの22年度上期進捗評価について説明があった。

各機関(評価単位は機関ごと)について、9月末目標の達成状況、下期計画の妥当性、22年度当初の評価に対する意見・コメントへの対応をポイントとして、作業部会委員に相互に評価いただきたい。年度当初の評価ではすべての事業が継承するという結果だったので、確実に成果を出せるよう、建設的なアドバイスや提案をお願いしたい。

フォーマットは年度当初のものと同様(会議資料参照)。メ切りは11月10日。評価内容は無記名で集計し、まとめた情報を事後評価委員会の参考資料とする予定。

事後評価について補足すると、対応作業として実施機関向けに成果報告書(毎年提出しているものの9月末までのアップデート版でよい)の提出を来週早々依頼する予定(提出は2-3週間後)。また、中間評価の時と同様に、評価委員が実際のDBに触れる期間(12月中を予定)を設ける。評価委員の判断によりヒアリング(1年半ばを予定)が必要になることもある。最終的な事後評価内容は、JST新センターへの提言としてまとめる。

◆質疑応答◆

○PJの終了を待たずに評価することには問題は無いのか?

→4月から始まるJST新センターへ評価内容を引き継ぎたいので、今回は事後ではなく、9割方開発が終了していると思われるPJ終盤で評価を実施することとした。特に「利用する」という観点で評価していただく予定。

(高木主査より)では、本議事も利用の観点から進行する。各機関の実施内容についてはこれまでもお聞きしているので、利用に関する情報を中心に発表いただきたい。まずは中核機関の成果について簡単に紹介してもらおう。

### 3. DBCLS の成果概要 (永井)

◇資料説明◇ 資料 3-1、3-2

9 月末の達成状況を見ると当初想定した目標は達成した。これらを踏まえ、PJ 終了時の目標として、経産省 MEDALS との共通化、統合検索のプロトタイプ構築、DB 受け入れ 45 件を設定した。ポータルサイトで紹介している各種サービスの利用状況を資料 3-2 にリストアップした。ポータルサイト全体では、ユニーク訪問者 4.8 万、ページビュー 100 万(いずれも H22.1-9 の月間ピーク値)である。サービス毎では、横断検索、日本語コンテンツ公開、新着論文レビュー、DNA DB 総覧、BodyParts、統合 TV へのアクセスが多く、その他開発系サービスにも一定のアクセスがあるので、それらが浸透しつつあることが分かる。

◆質疑応答◆

(特になし)

### 4. 平成 22 年度 9 月末時点進捗状況について

▶ 産総研糖鎖医工学センター (成松、鹿内)

◇資料説明◇ 資料 5-1、5-2

今年度だけでも糖鎖に関する複数の DB を構築し、既開発分と合わせて当初目標を上回るコンテンツを整備できた。これらの DB の検索インタフェース(IF)として、キーワード検索、統合検索(開発中、年度末公開予定、一部 DBCLS がサポート)、構造検索(糖鎖構造)、構造検索(糖関連化合物、未公開)を開発している。さらに糖鎖合成に関する DB を整備して、糖鎖化合物統合 DB を構築する予定である。この DB システム(JCGGDB)をベースにアジア各国の拠点となる研究機関と交渉を開始しており、アジア連携も模索中である。

◆質疑応答◆

○DB は英語化されているのか？必要なスタッフはいるのか？

→日英で作る。中国語版も作る。スタッフの中には英語、中国語に対応できる者がいる。

○ユーザー数はどのくらいか？

→資料 5-1 に掲げた数値は各 DB のページビュー。ユニーク ID 数は 500 程度。

○利用者は糖鎖研究者が中心か？

→現在はそうだが、今後は周辺にも広がると考えている。糖鎖のコミュニティ自体が 500~600 名くらい。統合検索など、使い勝手を良くすれば、糖鎖専門家以外にももっと広がると思う。

○それぞれの DB に対応する IF が同じようできて違うところもあるようだが。

→最終的には統一したい。それぞれの DB に部分的に特化した過渡的なものなので、差異がある。

○米国と日本でカバーしているものはどのくらい共通でどのくらい違うか？

→JCGGDB に入っている Wet のデータは産総研の自前データである。日本固有の得意な分野の技術もあるので、その部分はユニーク。米国で集めているのは米国内の研究機関のものなので、オーバーラップは無い。テキストマイニングで抽出した情報はソースが同じなので重なるはず。

○データ交換フォーマットなどは国際的な取り決めがあるか？

→そのような調整はこれから。そのイニシアティブをとるためにも 3 極構造の中にアジアの中心として日本が早く参加する必要あり。

○糖鎖構造の表現については、他の化合物と共通なものがあるのか？

→単糖のシンボルは 3 極共通。米国のコンソーシアムの提案に基づいている。修飾情報などは Edit できる

ようになっており、化合物データへのコンバーターもある。

○PJ の目標として公開の次には共有化を進めるといものがあるが、多様なデータは一括ダウンロード (DL) 可能か？ライセンスの扱いはどのようになっているか？

→ライセンスはフリーにしたい。まだ全体の一括 DL はできないが、将来的にできるようにしたい。ただし、年度内は個別の DB 構築で手いっぱいなので、DL が可能になるのは、年度が明けてからになると思う。

➤ 九州大学 (林)

◇資料説明◇ 資料 6

我が国の GWAS データの精度情報の提供が不十分であるので、それを解消するために他機関で取得された情報について、QC を実施することだけを目的としている。現在国内で大規模なデータ解析を実施している機関はいくつかあるが、中立的な立場で QC を行っている第 3 者機関が無い。GWAS の QC とは何か？国際的にスタンダードとなっているツールを使って、パイプラインを構築し、その説明をホームページに記載している。QC だけ行うので、倫理的な問題は無いと考えている。7 つのデータセットについて、医薬基盤研と東大(統合 DB プロジェクト)に利用申請し、4 つのデータについてはすでに解析を終了した。残る 3 つのデータについては現在未取得だが、今後取得して解析したい。

この 3 つのデータについては、審査はしたが、審査の対象ではないという結論が出されたと理解しており、その場合東大グループの判断にゆだねられていると思っているが、現時点では出せないという報告を貰っている。QC しか実施しないのでそもそも審査の対象にもならないと理解しているのだが、とにかくデータの受領が無いのが現実なので申請中ということで記載した。

<東大グループより>データアクセス委員会では申請を許可するに必要な条件が整っていない(研究グループの承諾が取れない、IC が外部共有を許していない等)ため、データを渡すことはできないと判断した。しかし、九大の実施内容は PJ 内部での QC パイプラインの検討なので、結果共有は PJ 内部に留めるなど、一定の条件のもとであれば使用を許可する方策があるのではないかと考えている。

◆質疑応答◆

○PJ 終了までに行うパイプラインの拡充とは？

→主成分解析の際に入れるべき Reference を入れていないので、それを入れる。

○交渉も始めていないものも含めて、国内で終了している GWAS 解析は他にどのくらいあるのか？

→理研の分は公開されていない。医薬基盤研で公開されている GeMDBJ と文科省の GWAS-DB の中からある程度の規模のものについて申請した。実際にはもっとたくさん実施されているが、公開されているものはこの二つの場所にあるもののみである。

○他の未公開データについて実施者に公開を促すというのは重要だが本 PJ の課題とは別の問題である。

○論文公開あるいはデータを Submit するときはある程度 QC すべきということになっているのでは？

→レフェリーは要求していると思うが、雑誌によってレベルが異なり、いい加減なものもあると思う。QC 値をつけるという世界的なルールにはなっていない。

○ある程度はどこでもやっているはずだが、すべてのエラーや問題点を排除するのはできない。

→QC されているものでもある程度の差があるということが今回示された。このことは、既解析のデータであっても、気をつけて使わなくてはいけないということを示していると考える。

➤ 東京大学 (徳永、小池)

◇資料説明◇ 資料 7-1

9 月末までの目標について、内部作業としては達成されているが、公開はあくまで論文発表後なので、未公

開のものが多い。また公開できるものも統計的情報のみとなる。また、個人ごとの Genotype 情報や生データまで提供できるものは限られている。一部のデータについては、Genotype 情報や生データまで提供できるよう改めて手続きを進めてもらっており、データの公開状況には段階があることをご理解いただきたい。PJ 終了までには搭載作業をさらに進める。Resequencing については実験情報だけでは不十分なので、文献データの登録も進めている。以下、公開 DB の実装状況について報告。

◆質疑応答◆

○CNV データは何処で取得されたものか？CNV のデータは方法論からして、最近大きく変化があるので、そのような解釈で書いておかないと誤解を生むと思うのでご配慮いただきたい。

○肝炎の GWAS について、非常にクリアな結果が出ているが、その要因は何か？

→日本人の遺伝的要因と臨床サンプルの収集がきちんとしていたことがあると思う。

○薬剤の応答性についてはデータが出しやすく、結果がはっきりしているので、注目する医師は多いと思う。

○本 DB はスタディ単位で登録されているブラウザで個人単位のデータを見られるわけではないが、スタディ単位で見たい人はどういう目的を持ってこのサイトを利用するのか？

→自分が発見した SNP が他の集団で相関がみられるかを見たいということだと思う。

○とすると疾患名などで検索するということか？他に大きな DB としては何があるか？

→Wellcome Trust と NIH のものが大きい。

○それらと同じような情報が見られると思っていいか？データに到達するまでの操作数が多いが。

→大体同じ様式である。いずれも詳細データに 1 クリックで到達するのは難しい。

○PJ 終了時までの追加目標の 5 件については、すでにめどがたっているものか？

→東大でタイピングを引き受けていて年度末までには登録になりそうなものと、すでにデータの登録について声をかけているところで 5 件程度はあると考えている。

○その結果、我が国の(理研以外の)データは網羅できるレベルになっているか？

→自分が承知しているもののうち 8-9 割はカバーできるのではないかと思う。

○当初より本課題の目標は、理研とそれ以外という区分けでいうと、それ以外の部分を網羅できるものである。網羅性の観点からすると、人類遺伝学会での論文投稿条件とするなど、積極的なアプローチの仕方もあるのではないか？今後のことを考えると PJ 終了までに+5 で良しとするよりももっと積極的な仕掛けをつくるといったストラテジーの見直しも必要ではないか？認知度を上げるというのはあくまで手段である。

○以前、人類遺伝学会でそのような議論をするという話もあったと思うが現在の状況は？

→来週の人類遺伝学会において、関連の委員会で提案をする予定。ただ、対象の DB としては統合 DB だけを指定することはできない。どこかに Deposit するという形で著者の Instruction に盛り込みたい。

○国際的にも Genotype データまで DB に Deposit することはまだ難しいと思うが。

→統計データについて格納することを考えている。

○Deposit されているのは、統計データか、個別の Genotype あるいは生データか？

→Genotype データまではなかなか出してもらえないので、統計データがほとんど。ただ、人によっては Genotype まで提供してもいいという人もいる。実際には頻度データをいただいて、こちらで統計データを計算している。

○理研からのデータはもらえないのか？論文から取れないのか？

→理研は理研で DB を作るのだからこちらには提供しないというお話だった。理研の検体採取時の同意書が旧式だったので、DB への登録等が盛り込まれていないことも提供できない原因の一つ。また、論文情報からデータを入力しているのは Resequencing のデータで、GWAS の統計情報はすべてが論文に掲載されていないので、ピックアップされた数値のみを掲載しても意味がないが、統計情報の全てを公共 DB に Deposit した

くてはならないというルールがないので、入手する手段がない。

○理研 DB の責任者も変更になったので、改めて交渉できる可能性があるかもしれない。

○データは研究者のものか、患者のものかという議論もある。患者に利益が行くのに役立たせるべきところを、研究者が押さえているというのはおかしい。

○Funding をするところがデータの提供を指導すべき。同意書の古さを盾に取っているのではないか。

○集めたデータを基にさらに高度な研究ができたという例はあるか？

→いまのところ、担当機関では解析を実施していない。タイピングデータをそっくり出してもらったところはまだないし、もしやるとしても現在の状況では提供者の許可が必要になる。

○再解析可能な個人のタイピングデータのデポジトリができるというめどはあるのか？

→データアクセス委員会としては同意を取る時から対応すべきとしており、今後可能になっていくはず。現在は、個別の研究グループで再解析している段階。コントロールデータではそのような動きが出てきている。今後類似の疾患で異なるプラットフォームで解析されたデータが蓄積されれば、相乗効果が出ると思う。

○やはり説明同意書が甘いとなかなか共有することができない。Funding Agency が同意書のひな型を指定するくらいにならないといけない。

## ▶ 東京医科歯科大学グループ

◇資料説明◇ 資料 8-1、8-2（医科歯科大）、資料 8-3、8-4（阪大）

【医科歯科大】（田中）

異なる疾患間を検索できると横断的統合、分子階層（分子から臨床まで）的に検索できる縦断的統合について実施し、データの公開のための倫理規定ガイドラインを整備することが目標。横断的統合としては、シーラスを利用したプロトタイプも検討中。縦断的統合について病態（臨床、病理）と遺伝子の相関を見る解析機能などを実装している。ガイドラインについては、東大グループの方針も参考に調整中。

【阪大】（古崎）

多施設での情報共有、ロールモデルの提示を目指している。「統合」の意味するものを再度定義し、医科歯科との分担についても整理した。9月末にはシノニム検索機能等、システムの改良作業を完了した。プログラムについてもすでに可能なものはすべて公開。基本的には厚労省の難病に関する情報収集のフォーマットに従って、投薬について項目を追加した形なので、他の疾患でもある程度汎用性のあるものと考えている。

### ◆質疑応答◆

○遺伝子発現について、肝臓がんの比較対象組織は何を用いているか？

→いくつかの組み合わせ（例：血管浸襲の有無、2年以内の再発の有無、癌部と非癌部）がある。

○何と何を比べれば、病理の比較になるのか？

→がん浸潤とキナーゼについては相関が見える。そのほか、いわゆるマーカーとサイクルの関係。

○厚労省が決めたフォームに記入された情報が臨床研究・教育支援センターに集められているのか？

→紙で集められている。最初の収集単位はおそらく県だと思う。

○そのデータを電子的に収集するシステムを構築したということか？

→もともとパーキンソン病について、紙で提出する情報をそのほかの付随情報とともにオンラインで入力することのできるシステムを作って、協力病院に配布していた。ひと手間かけて電子情報にさせていただくと同時に、提出用の書類が作成できるように。

○臨床研究・教育支援センターではそれをすべて紙で収集しているのか？それらは利用できるのか？

→そうである。申請により閲覧することはできるはず。

○シノニム辞書とはパーキンソン病に特化したものか？米国などにあるものを翻訳したのか？

→自前で構築のためのシステムを作り、構築したものである。Feasibility としてパーキンソン病で作った。標準化されているかどうかは未検証。

○臨床オミックス解析について、まだプロトの段階か？オリジナルの提案か？がん以外でも使えるか？

→オリジナルに、分子的、臨床的、病理的な情報をつなげることを試みた。がん以外にも、いくつかの種類の情報があれば使える。

○ヒートマップはXY軸をユーザーが変更できるのか？

→比較する群によって、クラスタリングの結果をインタラクティブに変更できる。

○研究運営委員会で（国立がんセンターの）吉田先生が、医学データについては Federation 型（実際のデータは個々の研究者が持ち、研究についての情報などを中央が集約する形）がいいのではないかと提唱されていたが、形を決めて生データを含めてすべて集めるのとどちらがいいか？

→米国ではプライバシーについての規制が厳しいので、中央には出しているものだけ出す（所在情報など）という方式によって統合しているケースが多いと思う。

○マウスのフェノタイプも共通部分だけ出して、個別のデータはそれぞれが保有するという形で連携が取れている。そういった話し合いをやるのがいいのでは？その時に理研の先生も呼んで、現実的に出せる物は何かを一緒に考えるのがいいのでは。

→中央デポジトリで形を決めて、出せる人が出すというのが現実的かもしれないと思う。

○臨床オミックスについて、原因－結果が見えるものではない。何に役に立つのかがよく見えないので、それが見えるようにしてほしい。例えば診断とか。そうすれば出す人のインセンティブにもなるのでは。

→パスウェイ情報などと関連付けるなど、検討している。次にはそのようなものが見えるようにしたい。

○臨床の治療方針にどう役に立つのか、が重要。そのためには臨床データをもっと細かく入れる必要がある。データを集める体制も要検討。対象疾患によっても、必要な情報の粒度が異なる。

○DB はメッセージ発信の場なので利用者がだれかを明確にして考えないといけないと思う。

○そういう問題意識もある程度やらないと出てこない。この PJ でやってみて見えてきたので、次の段階を考えるとときに何かしら役立つ情報があるはず。

## 5. その他

事務局より、シンポジウムご協力へのお礼、シンポジウム発表ポスターデータ送付(CC ライセンス付与)のお願い、ユーザー評価開始の案内と対応へのお礼があった。

(17 : 15 終了)