

2008380/9A

厚生労働科学研究費補助金
医薬品・医療機器等レギュラトリーサイエンス総合研究事業

国際的整合性を目指す有効性及び安全性に於ける
遺伝子発現情報の標準化に関する研究
(H19-医薬-一般-001)

平成 20 年度 総括研究報告書

研究代表者 菅野 純

平成 21(2009)年 3 月

厚生労働科学研究費補助金
医薬品・医療機器等レギュラトリーサイエンス総合研究事業

国際的整合性を目指す有効性及び安全性に於ける
遺伝子発現情報の標準化に関する研究
(H19-医薬-一般-001)

平成 20 年度 総括研究報告書

研究代表者 菅野 純

平成 21(2009)年 3 月

厚生労働科学研究費補助金

医薬品・医療機器等レギュラトリーサイエンス総合研究事業

国際的整合性を旨す有効性及び安全性に於ける

遺伝子発現情報の標準化に関する研究

(H19-医薬-一般-001)

平成 20 年度 総括研究報告書

研究代表者 菅野 純

平成 21 (2009) 年 3 月

目 次

I. 総括研究報告書

医薬品等の有効性・安全性評価に資する遺伝子発現解析の
国際的標準化に関わる研究 1
菅野 純

(資料 1) FDA's SEQC Study to Assess Second-Gen Techs,
Data Analysis for RNA Sequencing 11

(資料 2) ToxCast Program/National Center for Computational Toxicology
/US EPA 15

(資料 3) The Microarray Quality Control (MAQC) project 23

(資料 4) OECD Activities to Explore and Evaluate Regulatory Application of
Genomic Methods; Toxicogenomics 35

(資料 5) 班会議資料 51

II. 研究成果の刊行に関する一覧表 97

III. 研究成果の刊行物・別刷 99

I . 総括研究報告書

総括研究報告書

国際的整合性を目指す有効性及び安全性に於ける遺伝子発現情報の標準化に関する研究

研究代表者 菅野 純

国立医薬品食品衛生研究所・安全性生物試験研究センター・毒性部・部長

研究要旨

トキシコゲノミクス、ファーマコゲノミクス等を利用した医薬品、食品、化学工業物質等の有効性及び安全の確保に際しての mRNA 測定による遺伝子発現解析に関する国際的整合性を目指し、新しい技術、研究、手法の活用により、その標準化の手法の確立等を行う。本年度は3カ年計画の2年目であり、国内外における mRNA 測定・評価の標準化に関する動向を調査し、並行してトキシコゲノミクスにおける mRNA 標準化技術の開発・最適化を試みた。

研究分担者

油谷浩幸 東京大学 先端科学技術研究センター
教授

山口照英 国立医薬品食品衛生研究所
生物薬品部
部長

矢本敬 第一三共株式会社 安全性研究所
第七グループ長

住田佳代 住友化学株式会社 生物環境科学
研究所
主席研究員

山田 弘 (独)医薬基盤研究所 トキシコゲノミクス
インフォマティクスプロジェクト
サブプロジェクトリーダー

宇山佳明 (独)医薬品医療機器総合機構
新薬審査第三部
審査役

A. 研究目的

遺伝子発現情報を医薬品・食品・工業化学物質等の有効性及び安全性を評価するための定量的かつ高精度の指標として利用するために必要な RNA 標準化要件や関連情報を調査して、厚生労働省・日本政府が方針を定める際の基礎情報を提供する。遺伝子発現解析技術については、医薬品等の有効性・安全性評価に係る非臨床試験及び臨床試験に於いて、遺伝子発現情報の活用が急速に進展しつつある一方で、従来の遺伝子発現解析は相対的な基準が用いられてきたことから、その標準化が国際的な問題となっている。

本研究に於いては、主に米国を中心に急速に進みつつある「mRNA 測定に関わる標準化」に呼応して、国際的な動向を早急に調査し、国内での情報交換を行うことにより、遺伝子発現解析に関する標準化に向けた提言を行うとともに、国際的

標準化活動への技術的関与を検討する。

B. 方法

医薬品を始めとする人の身体が摂取し得る物質等の有効性・安全性評価に於いて、遺伝子発現情報を活用する際に課題となる遺伝子発現解析の標準化に資することを目的として、高密度マイクロアレイや定量的 PCR 等を用いた生物学的、薬理学的あるいは毒性学的研究を実施している国内の研究者と情報交換を行い、また、関連有識者、研究者を招聘し、Ad hoc 会合を行うことにより、遺伝子発現解析の標準化に関する国内の意見交換を行う。また、RNA の外部標準試薬を用いる場合に、標準試薬として求められる品質、品質管理、供給体制に関する情報収集や、国内に於ける情報交換、知識の集約を行う。

さらに、国際的な動向を調査するとともに、米国等と技術的な情報交換等を進め、米国に於ける遺伝子発現解析の標準化に向けた活動への技術的関与を検討していく。

以上により、医薬品等の有効性・安全性評価に遺伝子発現情報を活用する際の条件設定等について、海外に於いて急速に標準化に向けた活動が盛んになり事実上の国際標準(デファクト・スタンダード)になる可能性があることも想定し、活動当初から当該活動に対し関与し提言等を行い、今後の標準化に向けた課題抽出や条件設定の提言等を行う。また、これにより、現在国内で検討を進めている、ファーマコゲノミクス関連のガイドライン案に資することも可能となる。

これは医薬品・医療機器等レギュラトリーサイエンス総合研究事業に於いて、安全性、有効性及び品質の評価に対して、科学的合理性を付与す

るための研究を通して保健衛生及び国民生活の質の向上に資するものとして重要である。

本年度は3ヶ年計画の2年目にあたり、情報収集を確実にを行うために、各班員は医療・創薬・化学工業製品・毒性・精度管理、行政の各方向を網羅する布陣を配し、以下の項目に焦点を当て、分担課題に対応した調査を進めた。

- I. マイクロアレイ実施経験とデータ利用(審査や研究取りまとめ、等)経験者による問題点の抽出
- II. RNA 試薬・管理に関する専門的見地からの問題点の抽出
- III. 行政対応における問題点の抽出

C. 結果

1) mRNA 測定、特にトキシコゲノミクス応用における標準化手法に関する動向(菅野)

a. 遺伝子発現(mRNA)測定技術としてのマイクロアレイ技術

遺伝子発現解析用マイクロアレイの市場シェアは変わらず Affymetrix 優位。Affymetrix Exonアレイについては、Affymetrix と共同研究ベースで定量性向上の検討を続けており、現在は Affymetrix 社内試作品の Junction アレイを利用しつつ数学的処理による対応を検討中。

b. 次世代高速シーケンサ技術について

現行製品では全 mRNA の全長を読むには速度的に性能不足で、結果的にスプライシングバリエーションの網羅的な検出が困難であるほか、一般的には前処理(ゲル切り出し等)によるデータへの影響が懸念されている。ただし利用者の印象としては前処理による影響は僅かである模様。ランニングコストは、in house であればマイ

クローレイと同程度の費用で実施可能になりつつあり、また一度に読める塩基長や読み取り速度については、今後、一層の性能向上が見込まれる。

本技術は特にマイクロ RNA 測定に適している。主要なマイクロ RNA の配列が判明した後はマイクロアレイに導入することも可能(Affymetrix, Agilent が製品を供給している)だが、シーケンス処理とハイブリダイゼーション処理とのデータ互換性については検討が必要。

次世代高速シーケンサ技術の普及後、マイクロアレイとの使い分けは結局コスト問題に落ち着くと予測される。

c. RNA 標準化に向けた国際動向

米 FDA が中心に活動している MAQC2 は、2008 年度に 2 回の face-to-face meeting を開催し、近く成果を学術論文として発表予定。ERCC も Phase IV に関連する teleconference を中心に定期的な活動を続け、Phase V に付いても準備が始まっている。

d. RNA 標準化に向けた国内動向

NIHS の Percellome Project では強制経口(単回投与・反復投与 and/or 多臓器サンプリング)実験では約 100 化合物、吸入暴露(単回暴露・反復暴露)では 10 化合物終了している。その他、個体発生時暴露についても研究が進んでいるほか、化合物暴露影響を遺伝子発現のみならず、行動影響の観点からも評価を開始している。データベースは一部公開済み(6 化合物、2006/6~)で、登録ユーザー数は延べ 102 名となっている。また現行システムは H21 年度に使用終了し、RNA 標準化にも活用可能な高機能のユーザーインターフェイスを実現すべく、

in house 開発に移行する計画。新システム移行後はより積極的なデータ公開を目指している。

データ解析技術としては独自に開発した tmf Clustering と RSort の活用により、効率よく全遺伝子を網羅的に解析することが可能となった。また Phenotype independent の暴露条件(低用量条件)での遺伝子発現情報を既知情報(Pubmed, OMIM)と照合することで、毒性機序における重要な分子の解明や毒性発現に関与するシグナルネットワークの抽出に成功した例(サリドマイド)もある。

H21 年度以降は、構築済みの Percellome データベースを活用し、バイオインフォマティクス開発研究に注力する予定(SBI 北野らとの共同研究を予定)。なおバイオインフォマティクスによるネットワーク抽出にはモデル設定が肝要であり、まずは既存データを参考に設定の最適化を進める。

2)トキシコゲノミクスの標準化に関する調査研究(山田)

a. 国内の研究動向:トキシコゲノミクス・インフォマティクスプロジェクト(TGP2)の活動状況

TGP2 では RNA 標準化のための基本的な技術評価の一環として、Affymetrix GeneChip を用いて遺伝子発現を測定した時の施設内・施設間での相関について評価を行った。サンプルは、acetaminophen を 0、300、1,000mg/kg の用量で単回投与し、24 時間後に採取したラットの肝臓を用いた。採取した肝臓のホモジネートを TGP2 参加企業に配布し、各施設での SOP に従って遺伝子発現データを取得した。取得したデータは標準化した後、相関係数、t 検定、

Fold Change 等、各種手法により遺伝子発現の解析を行った。その結果、どの施設でも施設内での変動は小さく、多少の SOP の差異があっても殆ど遺伝子発現データに影響を与えることなく、良好な施設間での相関が示されることが確認された。検討結果の中からは、"absent"のプローブセットを除外すると相関が非常に改良されること、発現データを log₂ 変換することにより薬物影響の施設間差が小さくなること等が明らかとなった。現在、Agilent マイクロアレイを用いて同様の検討を進めており、平成 21 年度中には成果をまとめる予定である。

b. Extended OECD/IPCS Advisory Group on Toxicogenomics (Utrecht, Netherlands)会議

Extended OECD/IPCS Advisory Group on Toxicogenomics 会議は、オランダのユトリヒトにおいて平成 20 年 6 月 19 日から 20 日にかけて開催され、約 40 名の参加者による討議が行われた。今回の会議の主題は、1) US ToxCast Program の Phase I 及びその他の国における high throughput/high content screening に関する情報共有、2) molecular screening に関するレギュラトリーサイエンスからの潜在的な要望と応用方法についての確認、3) Molecular Screening Project に対する参加国及び関係機関の今後の連携のあり方や次のステップに向けたに合意の形成であった。

i) Potential Regulatory Needs and Applications of the Molecular Screening

EPA から molecular screening の利点として、使用動物数の削減やコスト低減、評価に要する時間の短縮があげられ、パスウェイあるいは毒性メカニズムの解釈に有用であると報

告された。ToxCast の進捗状況については、Phase I で 14 種の HPV 化合物について評価を実施しており、Phase II では HPV、MPV 化合物の他にナノマテリアルについても計画されるとのことであった。

ii) Report from the U.S. ToxCast Program

ToxCast Program での主要な取り組みは、毒性メカニズムの解明とその成果を応用した HTS assay 開発、ケミカルライブラリーのスクリーニング、ToxCast が導き出す結果と in vivo データとの比較とのことである。今後の検討課題としては、特に代謝物の評価が取りあげられていた。

iii) Report from member countries and

stakeholders on Advances in HTS & HCS for Prioritizing Environmental Chemicals

情報共有として、以下の報告があった。

- Toxicogenomics in the Netherlands: Jos Kleinjans, Maastricht University, Netherlands
- Assessing Exposure-Dose-Toxicity Relationships within the EPA's ToxCast Program: Mel Anderson, The Hamner Institute
- The rat two-generation reproductive toxicity study: Aldert Piersma, Netherlands
- National Toxicogenomics Program in National Institute of Toxicological Research: Sue Nie Park, NITR/KFDA
- CALUX® bioassays for human biomonitoring: Peter Benish, Biodetection Systems BV.
- Transgenic Xenopus as a small model organism for detecting endocrine disruption:

Barbara Demeneix, Muséum National
d'Histoire Naturelle, France.

iv) Initial discussion about the Molecular
Screening Project

Molecular Screening Project における対象
化合物の選択基準に関する議論が行われ、
ヒトでのデータが取得可能なものの重視や、
US、カナダ、EU で進行中のレギュラトリプロ
グラムに即したものとといった意見が提起され
た。対象となる毒性パスウェイとしては、主要
な代謝系の他、PPAR-alpha、Thyroid
Receptor、Sensitization/Immunotoxicity とい
ったものが挙げられた。

2 日目の会議では、次のステップに向けて以
下のサブグループを組織することが提案され
た。

- Thyroid Signaling
- Cancer Epigenetics
- PPAR alpha
- Sensitization/Immunotoxicity
- Databases
- Chemical Nomination
- Developmental and Reproductive Effects

全体を通して、まだ Molecular screening に
genomics データを活用する提案はなく、現段階
では、むしろ混乱するので積極的には利用し
ないとの考えが大勢を占めていた。

3) RNA 等、基準物質の精度管理に関する調査
研究(山口)

RNA 製品の製造・配布(販売)における品質維
持・安定供給等の技術的な問題は解決済みであ

るが、RNA 標準化のための RNA 製品を独自に開
発・販売するような大きな動きはない。国際的なコ
ンソーシアムや行政側から標準プロトコール/試
薬の情報が提供されるのを待っている状況であ
る。

4) 臨床における mRNA 測定の標準化に関する動
向(油谷)

臨床領域での遺伝子発現解析技術応用は依
然 discovery に限定されている。DNA 解析技術の
応用は、がん治療などいくつかの利用例がある。
DNA サンプルであれば、臨床現場においても試
料取得を安定して実施することが可能であり、先
ずは mutation 等を対象としたテーラーメイド医療
が主流となるだろう。

尚、MAQC のグループメンバーの一部が、シ
ークエンス技術に関しても同様の検証を行うため
のコンソーシアム SEQC を立ち上げた。DNA シー
クエンスだけでなく、高速シークエンス技術による
遺伝子発現(mRNA)測定についても対象として
いるので、SEQC の動向にも注視すべきである。

Preclinical な遺伝子発現解析技術の応用につ
いて、ヒトを対象とするならば iPS や in vitro でのア
プローチとなるが、その場合、遺伝子発現解析プ
ロトコール自体の標準化よりも、むしろ評価実験
に用いる細胞株調整など培養系の標準化のほう
が困難であろう。

5) 創薬における mRNA 測定の標準化に関する動
向(矢本)

a. 製薬企業における遺伝子発現解析の現状

第一三共株式会社安全性研究所における取り
組みを具体例として引用しつつ報告する。

i) サンプル調整およびRNA取り扱いの標準化

第一三共株式会社安全性研究所では他社と同様に組織からの総 RNA 抽出には Qiagen 社 RNeasy キットおよび Invitrogen 社 Trizol 試薬を、また RNA の品質チェックに関しても国内外で広く普及している Agilent 社の Bioanalyzer を用いて検討を行っており、実験プロトコールおよび RNA の品質検査に関しては、事実上標準化が進んでいると考えられる。

ii) 定量的 RT-PCR 法、マイクロアレイ解析の標準化状況

定量的 RT-PCR 法に関しては、検量線サンプル数、検量線の傾きの幅と相関係数、ハウスキープ遺伝子の選択などで議論が必要であり、またマイクロアレイ解析に関してはアレイプラットフォームに起因するデータの相違や試験情報、サンプル情報の注釈付けに関する規程の曖昧さに加え、製薬企業で広く普及している Affymetrix 社の GeneChip システムに限っても、cRNA 伸長反応の確認、プローブレベルデータの解析法の選択 (MAS5, GCRMA など)、統計処理法、データベース間のデータ互換性を高めるためのデータ解析法等に関して、さらに議論が必要と考えられる。

第一三共株式会社安全性研究所ではマイクロアレイデータの薬物安全性評価のために、グルタチオン枯渇を評価するための毒性バイオマーカー遺伝子の選抜、および複数の毒性バイオマーカー遺伝子リストを用いた毒性エンドポイントごとのスコア計算・活用に関する解析に取り組んでいる。各社で行っている同様の研究成果が広く紹介され、遺伝子発現解析の標準化が進むことによって、こうした基礎データが薬

物安全性評価に際して有用な情報となり得る。

b. マイクロアレイデータに基づいた毒性バイオマーカーのバリデーション

本件に関する取り組みは欧米が先行しているが、日本においても何かモデルケースを一つ作って産官学共同研究を実施すべきである。このような経験は薬物安全性評価のためのトキシゲノミクスデータの生成・比較解析等の標準化に貢献し、製薬企業、規制当局双方に意義があるだろう。

6) 化学工業製品の安全性確保における mRNA 測定の標準化に関する動向(住田)

a. mRNA 測定

化学工業界における製品の安全性確保に関して、mRNA 測定は依然として統一的な標準化のレベルに至っていない。現状、試料からの total RNA 抽出、抽出された total RNA の品質確認、それに続くマイクロアレイや定量的 RT-PCR を用いた遺伝子の発現量測定、データの品質確認は、国内外で広く普及している機器を用い、また、提供されているプロトコールあるいは基準に従って実施しており、局所的な標準化はある程度進んでいると考えられる。しかしながら、アレイプラットフォーム間でデータが異なる、あるいは、アレイプラットフォームが同一であってもデータ処理法などの違いによりデータが異なるといった場合があり、より詳細に実験ステップ毎の基準を設定し、統計処理法やデータ解析法等に関して、さらなる標準化が必要である。

b. 遺伝子発現解析技術

一方、遺伝子発現解析技術は、社内で蓄積

された遺伝子発現データをベースとした開発化合物の安全性評価において活用されつつある。具体的には、開発早期における化合物の毒性予測や診断、毒性発現機構の解明、毒性既知化合物との作用同等性や作用差別化が主なものである。

例えば住友化学生物環境科学研究所では、現在、毒性予測や診断においては、発がん性予測が先行しており、発がん性予測モデルの構築や病理診断マーカーの探索を行い、総合的な開発判断早期化の一助としている。利用可能なデータベースを拡張することにより、予測精度がさらに向上すると考えられることから、データベース間の互換性を高めるためにも実験プロトコルとデータ処理法の両面で統一的な標準化が必須であると思われる。遺伝子発現データは、開発化合物が呈する毒性に関して、毒性回避へのフィードバックやプロアクティブな対応のための発現機構解明やその作用同等性の評価に有用な情報となりつつある。また、毒性研究で主に使用されるラットやマウスよりもさらにヒト遺伝子との類似性が高い霊長類であるマモセットのマイクロアレイを独自に作製し、毒性学的により精緻な評価を目指している。

7) 医薬品審査過程における mRNA 情報の標準化に関する動向(宇山)

a. ファーマコゲノミクス(PGx)に関する国際動向

i) ICH では、「医薬品開発に関連したゲノムバイオマーカー適格性確認のための資料を規制当局に提出する際の様式」(E16)についての検討が、平成 20 年 6 月から正式に開始され、現在ドラフト案の作成が進められており、平成

21 年夏頃には、ドラフト案が公表される予定である。

ii) 米国 FDA では、ゲノム薬理を利用した臨床試験のデザイン、データ解析と用量設定あるいは添付文書における記載方法等についてガイドライン案の作成が進められている。また欧州 EMEA では、バイオマーカーの評価に関する助言制度を公式に立ち上げることが検討されている。

b. PGx に関する国内動向

日本においては、平成 20 年 4 月に「DNA チップを用いた遺伝子型判定用診断薬に関する評価指標」が、平成 20 年 9 月には「ゲノム薬理学を利用した治験について」が、それぞれ厚生労働省から通知され、日本におけるゲノム薬理学を利用した治験等の基盤が整備されつつある。また、(独)医薬品医療機器総合機構(PMDA)では、ゲノム薬理学及びバイオマーカーに関する対面助言制度の新設が検討されており、平成 21 年 1 月からパブリックコメントの募集が開始された。さらに PMDA では、FDA と EMEA との間で合同で進められているバイオマーカーの評価会議に、今年度もオブザーバーとして引き続き参加している。

以上のように、PGx に関連する行政的な取り組みは着実に進められており、今後も、海外での動向等を注視しながら、本邦での取り組みを検討していく必要がある。

E. 結論

国内外における mRNA 測定・評価の標準化に関する動向を調査し、並行してトキシコゲノミクスにおける mRNA 標準化技術の開発・最適化を試みた。

本年度、国際的な mRNA 標準化団体(米 MAQC、米 ERCC 等)は公式な成果発表(論文発表)を行わなかったが、標準化に関する議論はまとまりつつあり、内部では論文作成および次期計画策定が進んでいる。また MAQC の一部メンバーを中心に次世代シーケンサ技術を利用した RNA 測定技術の評価・標準化を目指す動きが活発化し、2008 年末には公式に SEQC のメンバー招集が実施された。国内では産学官の各領域において測定・解析技術面で更なる進捗があった。トキシコゲノミクス大規模データベース構築も順調に進んでおり、同時に膨大なデータから特徴的な反応を網羅的かつ自動的に抽出する技術基盤も整い実用段階を迎えつつある。行政面では ICH E16(International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use、Genomic Biomarkers Related to Drug Response: Context, Structure, and Format of Qualification Submissions)などの国際的な協力体制下にて対応が進められており、現在は規制当局への提出フォーマット等を策定中である。国内でも(独)医薬品医療機器総合機構において、審査資料として遺伝子発現データが提出された場合の対応検討が進められており、近々バイオマーカーオリフィケーションに特化した相談制度を設ける予定となっている。

本研究により得られた成果は、医薬品等の有効性・安全性評価に遺伝子発現情報を活用する際の標準プロトコール作成に向けた課題抽出や条件設定検討に直結する。特に厚生労働行政の医薬品・医療機器等レギュラトリーサイエンス総合研究事業に於いて、安全性・有効性及び品質の評価に科学的合理性を付与するための研究を強力にサポートし、これを通じて保健衛生及び国民生活の質の向

上に資する。

国内外共に、RNA 測定技術の標準化に必要な技術の開発および行政の対応準備は着実に進んでおり、来年度も調査を継続し、MAQC,ERCC 等の Academia side および ICH E16 等の Regulatory side の動きを考慮しつつ、必要に応じて実務的な研究班・研究体制の立ち上げを関係当局(厚生労働省医薬食品局審査管理課等)と協議する。

F. 健康危険情報

特になし

G. 研究発表

1. 論文発表

Sanosaka T, Namihira M, Asano H, Kohyama J, Aisaki K, Igarashi K, Kanno J, Nakashima K. Identification of genes that restrict astrocyte differentiation of midgestational neural precursor cells. *Neuroscience*. (2008) 155(3):780-8.

Niida A, Smith AD, Imoto S, Tsutsumi S, Aburatani H, Zhang MQ, Akiyama T. Integrative bioinformatics analysis of transcriptional regulatory programs in breast cancer cells. *BMC Bioinformatics*. (2008) 9:404.

Ishiguro A, Toyoshima S & Uyama Y, Current Japanese regulatory situations of pharmacogenomics in drug administration, *Exp Review Clin Pharmacol*, 1: 505-514, 2008

森和彦、宇山佳明: 国際共同治験の基本的
考え方について, 医薬品研究, 39: 557-575,
2008

2. 学会発表

菅野 純、トキシコゲノミクス(Percellome
Project)を基盤とした分子毒性学の展開の試
み、第35回日本トキシコロジー学会学術年会、
2008年6月26日、東京、口演

北嶋 聡、菅野 純、トキシコゲノミクスによる毒
性評価法の高精細化、第35回日本トキシコ
ロジー学会学術年会、2008年6月27日、東京、
口演

H. 知的財産所有権の出願・登録状況(予定も含 む)

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

FDA's SEQC Study to Assess Second-Gen Techs, Data Analysis for RNA Sequencing

January 06, 2009

By Julia Karow

- Type size: - +
- Email
- Print
- RSS Feed

The US Food and Drug Administration is organizing a new research study to "objectively assess the technical performance of different next-generation sequencing technologies" for DNA and RNA analyses, and to evaluate the pros and cons of various data analysis methods, according to a notice published in the *Federal Register* last month.

The study, called Sequencing Quality Control, or SEQC, is open to the research community and is expected to be completed by the end of this year. According to its organizers, the project is "a natural extension" of the MicroArray Quality Control, or MAQC, project that the FDA has been running since 2005 in order to evaluate the performance of microarray and related platforms for gene expression analysis.

The FDA is interested in conducting the new study because it expects the new sequencing technologies "to be adopted by the pharmaceutical and medical industries for advancing personalized nutrition and medicine," according to the notice. SEQC will "help prepare FDA for the next wave of submission of genomic data generated from the next-generation sequencing technologies."

According to Leming Shi, a SEQC organizer and an FDA researcher, there have been "serious discussions" for more than a year among MAQC participants about a performance assessment of the new sequencing technologies and data analysis methods for RNA sequencing.

Several MAQC members over the last two years or so have already been sequencing the two reference RNA samples from the first phase of the MAQC project on second-generation sequencing platforms (see *In Sequence* 4/22/2008). In addition, Shi and his colleagues have received inquiries from scientists outside of MAQC about the possibility of such a study.

The first phase of the MAQC project — MAQC I — measured gene expression levels in two standardized RNA samples on seven microarray and three other platforms, including qPCR, at three independent test sites. Results from that study, which involved participants from 51 organizations, were published in *Nature Biotechnology* in 2006. The second phase of the project, which includes participants from 60 groups, is focusing on data analysis and predictive models. Its results are currently being prepared for publication.

SEQC, which the organizers consider the third phase of MAQC, was formally launched at a meeting at an FDA campus in Silver Spring, Md., on Dec. 16 and 17 that hosted more than 40 researchers interested in the project. Among them were representatives from Illumina, Life Technologies, Roche's 454 Life Sciences, and Helicos BoSciences.

"We predict that, like previous phases of MAQC, the impact of data analysis will be the more interesting part of SEQC."

Researchers and reviewers from several FDA centers are also participating in SEQC, according to Shi, a computational chemist at the FDA's National Center for Toxicological Research in Jefferson, Ark. The study is currently seeking additional participants, and interested parties can submit requests until Jan. 9. Besides vendors of sequencing technologies, institutions "interested in the generation, management,

analysis, and interpretation" of sequencing data are invited to join, according to the FDA. The study design is expected to be finished next month, and researchers will start collecting RNA sequence data on the two MAQC reference RNA samples this spring. These samples will be spiked with external RNA controls of known sequence and abundance, according to Shi.

The two MAQC reference RNA samples were "a natural choice" for benchmarking RNA sequencing data in SEQC, Shi said in an e-mail message, because "a huge amount of expression data has already been collected" on them. "In fact, all major sequencing players have been using the two RNA samples internally for quality control and protocol optimization purposes," he said.

Each sequencing platform will be tested at three sites using the same set of RNA samples, and the results will be compared with those from the MAQC I study.

According to Shi, SEQC will also solicit proposals on sequencing other RNA samples "if they are deemed 'interesting' and certain references/standards are available for comparing results." In addition, "we are also interested in applications other than gene expression," which will be "addressed separately," he said, without providing further details.

Besides comparing new sequencing platforms, SEQC will also evaluate a variety of bioinformatic solutions for analyzing and handling the data. "We will not limit ourselves to any predetermined approaches," Shi said. "In fact, we welcome SEQC participants to explore many different methods for sequence mapping and assembly, and to compare the resulting data against each other or with the 'truth' embedded in the reference RNA samples."

Results from this evaluation may be especially interesting, he suggested. "We predict that, like previous phases of MAQC, the impact of data analysis will be the more interesting part of SEQC." After collecting sequence data this spring, SEQC plans to analyze the results over the summer and submit a manuscript for publication by the end of the year. "Participants expressed the need to work with a tight timeline because of the evolving nature of the sequencing technologies," according to Shi. Like MAQC, the new study will provide a "neutral environment" in which participants can "openly discuss ideas, debate scientific issues, and share expertise," which is "critical to cultivate a new generation of users of the new technologies," Shi said.

More information for those interested in participating in SEQC can be found here.

TITLE: FDA's SEQC Study to Assess Second-Gen Techs, Data Analysis for RNA Sequencing | In Sequence | Sequencing | GenomeWeb
DATE: 2009/03/27 15:43

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration

[Docket No. FDA-2008-N-0644]

SEQC—The Sequencing Quality Control Project

AGENCY: Food and Drug Administration, HHS.

ACTION: Notice of solicitation.

SUMMARY: The Food and Drug Administration (FDA) is soliciting volunteers to participate in the SEQC (Sequencing Quality Control) project to objectively assess the technical performance of different next-generation sequencing technologies in DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) analyses and to evaluate the advantages and limitations of various bioinformatics solutions in handling and analyzing the massive new data sets. The SEQC project is a natural extension of the MicroArray Quality Control (MAQC) project (<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maq/>) and is being coordinated by the FDA. This project is open to the public. Vendors of next-generation sequencing technologies and institutions interested in the generation, management, analysis, and interpretation of the resulting sequence data are welcome to participate.

DATES: Requests to participate in the SEQC project at the National Center for Toxicological Research (NCTR) should be submitted on or before 4:30 p.m., CST, January 9, 2009, or be postmarked on or before January 9, 2009.

ADDRESSES: Requests to participate in the SEQC project should be sent to Leming Shi, National Center for Toxicological Research, Food and Drug

Administration, 3900 NCTR Rd., Jefferson, AR 72079, 870-543-7387, FAX: 870-543-7854; e-mail: leming.shi@fda.hhs.gov.

SUPPLEMENTARY INFORMATION: FDA's Critical Path Initiative (<http://www.fda.gov/oc/initiatives/criticalpath/>) identifies pharmacogenomics as a key opportunity in advancing medical product development and personalized medicine. FDA has issued the "Guidance for Industry: Pharmacogenomic Data Submissions" (<http://www.fda.gov/cder/guidance/6400fnl.pdf>) to facilitate scientific progress in the field of pharmacogenomic data integration in drug development and medical diagnostics.

Microarrays represent a core technology in pharmacogenomics and toxicogenomics; however, next-generation sequencing technologies promise to provide some unique advantages in DNA and RNA analyses and are expected to be adopted by the pharmaceutical and medical industries for advancing personalized nutrition and medicine.

The SEQC project, with broad participation from scientists and reviewers within FDA and collaborators across the public, academic, and private sectors, is expected to help prepare FDA for the next wave of submission of genomic data generated from the next-generation sequencing technologies.

Dated: December 17, 2008.

Jeffrey Shuren,

Associate Commissioner for Policy and Planning.

BILLING CODE 4160-01-S



ToxCast™ Program

Predicting Hazard, Characterizing Toxicity Pathways, and Prioritizing the Toxicity Testing of Environmental Chemicals



Introduction

In 2007, EPA launched ToxCast™ in order to develop a cost-effective approach for prioritizing the toxicity testing of large numbers of chemicals in a short period of time. Using data from state-of-the-art high throughput screening (HTS) bioassays developed in the pharmaceutical industry, ToxCast™ is building computational models to forecast the potential human toxicity of chemicals. These hazard predictions will provide EPA regulatory programs with science-based information helpful in prioritizing chemicals for more detailed toxicological evaluations, and lead to more efficient use of animal testing.

In its first phase, ToxCast™ is profiling over 300 well-characterized chemicals (primarily pesticides) in over 400 HTS endpoints. These endpoints include biochemical assays of protein function, cell-based transcriptional reporter assays, multi-cell interaction assays, transcriptomics on primary cell cultures, and developmental assays in zebrafish embryos. Almost all of the compounds being examined in Phase 1 of ToxCast™ have been tested in traditional toxicology tests, including developmental toxicity, multi-generation studies, and sub-chronic and chronic rodent bioassays. ToxRefDB, a relational database being created to house this information, will contain nearly \$1B worth of toxicity studies in animals when completed. ToxRefDB is integrated into a more comprehensive data management system developed by NCCT called ACToR (Aggregated Computational Toxicology Resource), that manages the large-scale datasets of ToxCast™. ACToR is comprised of several independent data repositories linked to a common database of chemical structures and properties, and to tools for development of predictive HTS and genomic bioactivity signatures that strongly correlate with specific toxicity endpoints from ToxRefDB. These ToxCast™ signatures will be defined and evaluated by their ability to predict outcomes from existing mammalian toxicity testing, and identify toxicity pathways that are relevant to human health effects.

The second phase of ToxCast™ will screen additional compounds representing broader chemical structure and use classes, in order to evaluate the predictive bioactivity signatures developed in Phase I. Following successful conclusion of Phases I and II, ToxCast™ will provide EPA regulatory programs an efficient tool for rapidly and efficiently screening compounds and prioritizing further toxicity testing.

ToxCast™ Navigation

- [Introduction](#)
- [Chemicals](#)
- [Assays](#)
- [Information Management](#)
- [Partnerships](#)
- [Contractors](#)
- [Presentations](#)
- [Publications](#)
- [News](#)
- [Data Analysis Summit](#)

Data Set	Description	Download	Publication
Landscape Chemicals and Assays	Supplemental data , including the set of chemicals and data sources described in the paper "The Toxicity Data Landscape of Environmental Chemicals". (More Information)	Download (15.5 MB, ZIP)	Judson et al. (2008) " The Toxicity Data Landscape for Environmental Chemicals " Environmental Health Perspectives doi:{10.1289/ehp.0800168}

The Toxicity Data Landscape for Environmental Chemicals

This paper, published in *Environmental Health Perspectives*, is authored by staff of the EPA Offices of Research and Development, Pesticide Programs, Pollution Prevention and Toxics, Water, Science Coordination and Policy and the Great Lakes National Program Office. We surveyed the major types of chemicals regulated by the EPA and compiled a non-redundant list of these chemicals to provide candidates for the ToxCast screening and prioritization program (<http://www.epa.gov/ncct/toxcast>). The chemicals considered are high- and medium production volume chemicals (HPVs and MPVs), pesticide and antimicrobial active compounds, pesticidal inerts, candidate air and water pollutants, potential persistent or bioaccumulative chemicals, chemicals proposed for testing in assays of the EPA Endocrine Disruptor Screening Program (EDSP), chemicals tracked by the Toxics Release Inventory (TRI) Program, and chemicals evaluated as part of the Integrated Risk Screening System (IRIS). The final list contained 9,912 unique chemicals.

This paper also evaluated how much data were available to help evaluate the toxicity of these chemicals. Data were compiled from over 200 public sources into a central database called ACToR (Aggregated Computational Toxicology Resource, <http://actor.epa.gov/actor>). Sources of data included the EPA, FDA, NIH, other state and federal governmental agencies in the U.S., Canada, Europe and Japan, WHO, academic groups, industry and non-governmental organizations. Types of data collected included chemical identity and structure, physical-chemical properties, *in vivo* toxicology, chemical use levels, and regulatory information.

The conclusions of this paper are:

1. Of the 9,912 chemicals on this list, at least limited acute hazard data was publicly available on 66%. Conversely, no toxicology data was available on 34%.
2. Data on specific disease endpoints was more limited: Carcinogenicity: 26%; Developmental Toxicity: 29%; Reproductive Toxicity: 11%; Genotoxicity: 28%
3. This analysis helps define the set of chemicals that can be used to train and validate models linking *in vitro* assays with *in vivo* animal toxicity, as part of the ToxCast program.

While toxicity data on many existing chemicals is insufficient, there is a growing realization that the traditional means of comprehensive laboratory animal testing is not practical and cannot keep pace with the demand. This is driving the need for faster, cheaper, and scientifically-superior methods for screening, prioritizing and testing large inventories of chemicals. In some cases, the Agency uses existing information such as exposure, predictive computer models (e.g., SAR/QSAR), acute or *in vitro* data to prioritize chemicals for level of concern and perhaps further evaluation. The EPA ToxCast program is developing systematic approaches for screening and prioritization using *in vitro* testing methods developed largely by the pharmaceutical industry.

For more information on this paper, ACToR, or the EPA ToxCast program, contact: Richard Judson (judson.richard@epa.gov, 919-541-3085) or David Dix (dix.david@epa.gov, 919-541-2701).