Agence française de sécurité sanitaire des produits de santé

Direction de l'**E**valuation des **M**édicaments **E**t des produits **B**iologiques
Département de toxicologie
Chef de département : D. Masset
Secrétaire scientifique : A. Sanh

Aids to Voluntary Submission of Genomics Data

Version 3, 22 June 2011

This Document was prepared by the Working Group « Non-Clinical Innovation »

- **Paper coordinator**
  - S. LE CROM
- **Chairman**
  - J.-R. CLAUDE
- **Members**
  - L. DOMENJOUD
  - E. FATTAL
  - J. GUILLEMAIN
  - A. GUILLOUZO
  - S. LE CROM
  - A. LE PAPE
  - S. LERONDEL
  - P. LESCUYER
  - B. MAILLERE
  - F. MOREL
  - M. PALLARDY
  - C. PINEAU
  - T. RABILLOUD
  - R. RAHMANI
- **Invited expert**
  - D. MARZIN
- **Invited members of the Afssaps**
  - D. ABDON
  - D. SAUVAIRE

Aid to voluntary submission of genomics data

Updade n°3, 22 june 2011.

GENERAL SCOPE

The scope of this document is to provide information to Applicants submitting genomics or proteomics data in the context of a voluntary submission. This information should not be considered as mandatory requirement. However this information is requested to obtain a clear understanding of what is expected from the omics data, to evaluate the technical conditions for realization of experiments and the compliance or not with Good Laboratory Practices (GLP) requirements, and finally to assess scientific interest of the data.
The current objective of this document is to encourage voluntary submissions by Applicants; there are no disqualifying responses to any of the questions. In the light of the analysis of these submissions, the future aim will be to establish a relevant guideline based upon practical experience.

FIELDS OF APPLICATION FOR THIS DOCUMENT

In this document, the following themes are covered according to the current knowledge from field specialists:

- Transcriptomics: high throughput quantitative PCR, microarrays and high throughput sequencing;
- Proteomics: 2D gel electrophoresis, mass spectrometry including targeted protein quantification using mass spectrometry (SRM/MRM) and protein microarrays.

1. APPLICANT IDENTIFICATION

The identification includes the following information: last name, first name, title and function, postal address, e-mail and phone number.

1.1. Study sponsor and director identification

1.2. If the study is performed by the Applicant:

- Study director identification

1.3. If the study is not performed by the Applicant:

- Is it performed by a public (e.g., academic) platform?
  Study director identification
- Is it performed by a Contract Research Organisation (CRO)?
  Study director identification

1.4. If the study is performed within the framework of a multidisciplinary programme, in which other laboratories are involved:

- Laboratory identifications: type of activities / Principal investigator(s) identification
- Who is the project coordinator (if different of the study sponsor)?
- Does an Advisory Committee exist (e.g., internal, independent …)?

2. PROJECT IDENTIFICATION

2.1. The aim of the study (e.g., screening, mechanistic study ….). The methodologies used and the rationale of the genomic approaches should be justified.

2.2. The place of the study in the project or in the general strategy of research or development of a product should be defined.

2.3. The aim of joint studies (e.g., biochemistry, kinetics, histology …) should be specified.

2.4. The choice of the biological material, its characteristics and the validation for its use should be defined. The experimental conditions and the study design should be described: see the MIAME and MIAPE checklists below.

2.5. Deliverables and data organisation

2.6. The statistical treatment and analysis workflow performed should be explained.

- For example, if gene or protein targets have been identified according to a functional annotation enrichment calculation, it is necessary to indicate which annotation database has been used (Gene Ontology), what statistical test has been applied and the threshold of the filters used. The fact that a technology is recent and/or expensive shall not be an excuse not to perform the necessary biological replicates and a complete statistical analysis of the results obtained.

2.7. Confidentiality, ethical considerations and data security.

- Are the data confidential? For how long?
- Is there any patent deposit request on the data? If so, what is the date of the patent deposit?
- Has an ethical committee be consulted? Are there, for example, any nominative data? Justify the absence of ethical committee advice in this case. The easy access to individual genomes, thanks to high throughput sequencing techniques, requires secure solutions on how the anonymity of the participant(s) of the study will be preserved.
- What are the solutions used for data backup and, if the data are confidential or of a strategic importance, for their safe storage.

3. STANDARDIZATION

3.1. For Transcriptomics:

- Is the study in agreement with the current MIAME requirements?
- Provide the MIAME checklist[1] for the study
- Currently the MINSEQE standard for high throughput sequencing is not sufficiently developed and documented. Meanwhile, it is necessary to keep a detailed log of the whole experiment parameters and the MIAME standard definitions is this field can be used as a model.

3.2. For Proteomics:

- Is the study in agreement with the current MIAPE requirements?
- Provide the MIAPE checklist[2] for the study.

4. VALIDATION

4.1. Study validation

- Is the Applicant or Sponsor involved in a public or private inter-laboratory validation process (e.g., ILSI-HESI …)?
- If not, provide your internal validation criteria.
- How are your procedures tested for reliability, reproducibility... (e.g., use of positive controls...)?
- When targeted protein quantification using mass spectrometry (SRM/MRM) is performed, supplementary information on the analytical methods selected is needed: process for chosing of the transitions, number of transitions per protein, method imprecision (including trypsin digestion), limit of detection, limit of quantification, …

[1] (EMEA/CHMP/20227/2004 – Guideline on pharmacogenetics briefing meetings draft).
[2] http://www.mcponline.org/misc/ParisReport.shtml

4.2. Results validation

- Are your results confirmed (e.g., RT-PCR for transcriptomics, immunoassay for proteomics) or validated by a comparison with toxicological endpoints or validated biomarkers?

5. GOOD LABORATORY PRACTICES (GLP)

5.1. Are you working in full agreement with recognized GLP requirements, only in part ("in the spirit ") or not at all?

5.2. Do you think that you will be able to become GLP compliant in the short or medium term?

5.3. Some questions regarding the GLP

- Are all the study participants' liabilities clearly defined?
- Do you have an independent Quality Assurance Unit for your study?
- Are the experimental conditions validated (Standardized Operating Procedures, SOP)? Justify the validation.
- If a modification in the initial protocol of the study occurs, is it endorsed by the Study Director and verified by the Quality Assurance Unit?
- What are your raw data, how do you use them, how and how long do you preserve them?
- Is the study report endorsed by the study Director and checked by the Quality Assurance Unit?
- In the case of a multisite program, is a multisite procedure considered or committed?
- Are the data reliable or easy retrievable?

6. DATA SUBMISSION

Provide CD/DVDs including the MIAME and/or MIAPE checklist and GPA files (GNU Privacy Assistant).
Submission in the format recommended for public databases (e.g., GEO, Array Express …) is encouraged. Indicate if these data have been used for an Expert Report?

7. DISCUSSION AND CONCLUSION

The results should be discussed in the context of up-to-date literature and a clearly identified, concise conclusion for an easy understanding by a relevant scientific assessor, however not highly and acutely specialized in the postgenomic area, shall be provided.