Guidance for Industry

Pharmacogenomic Data Submissions — Companion Guidance

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the Federal Register of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the Federal Register.

For questions regarding this draft document contact (CDER) Federico Goodsaid, 301-796-1535 or (CBER) Raj Puri, 301-827-0471 .

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
National Center for Toxicological Research (NCTR)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
August 2007
Procedural

2007D-0310

GDL1

Guidance for Industry

Pharmacogenomic Data Submissions — Companion Guidance

Additional copies are available from:
Office of Training and Communication
Division of Drug Information, HFD-240
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm
and/or
Office of Communication, Training and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm.
(Tel) 800-835-4709 or 301-827-1800
and/or
Office of Health and Industry Programs
Division of Small Manufacturers Assistance (HFZ-220)
Center for Devices and Radiological Health
Food and Drug Administration
1350 Piccard Drive
Rockville, MD 20850-4307 U.S.A.
http://www.fda.gov/cdrh/ggpmain.html
Email: dsma@cdrh.fda.gov
Fax: 301.443.8818
(Tel) Manufacturers Assistance: 800.638.2041 or 301.443.6597
(Tel) International Staff Phone: 301.827.3993

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
National Center for Toxicological Research (NCTR)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
August 2007
Procedural

Draft
Not for Implementation

TABLE OF CONTENTS

I. INTRODUCTION 1

II. GENE EXPRESSION DATA FROM MICROARRAYS 2

A. RNA Isolation, Handling, and Characterization 2

1. Pre-RNA Isolation Considerations 2
2. RNA Isolation from Tissues or Cells 3
3. RNA Isolation from Whole Blood and PBMCs 4
4. RNA Storage 5
5. RNA QC 5

B. Labeling Reactions 6

C. Hybridizations for Microarrays 7

D. Fluorescence Reader Settings for Microarrays 8

E. Differentially Expressed Genes 9

F. Biological Interpretation of Lists of Differentially Expressed Genes 10

III. GENOTYPING 11

A. Genotyping Methods 11

B. DNA Isolation, Handling, and Characterization 12

C. Genotyping Report 12

IV. PROFICIENCY TESTING 13

V. GENOMIC DATA IN CLINICAL STUDY REPORTS 15

VI. GENOMIC DATA FROM NONCLINICAL TOXICOLOGY STUDIES 16

A. Expanding the Selection Process Criteria 17

B. Characterization of a Particular Compound 17

C. General Scientific Discussion 17

VII. DATA SUBMISSION FORMAT 18

A. Submission Standard 18

B. Microarray Gene Expression Data 18

C. Clinical and Nonclinical Data 19

APPENDIX I: EXPERIMENTAL SUMMARY TABLE (*EXPSUMTABLE*) 20

APPENDIX II: EXAMPLE—SUBMITTING A NONCLINICAL STUDY DATA 21

Draft
Not for Implementation

Guidance for Industry[1]
Pharmacogenomic Data Submissions — Companion Guidance

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I. INTRODUCTION

This guidance is intended to be used as a companion to the guidance *Pharmacogenomic Data Submissions* (March 2005). It reflects experience gained since the issuance of that guidance with voluntary genomic data submissions as well as with review by the FDA of numerous protocols and data submitted under investigational new drug (IND) applications, new drug applications (NDAs), and biologics license applications (BLAs). The recommendations are intended to facilitate scientific progress in the field of pharmacogenomics and to facilitate the use of pharmacogenomic data in drug development. The FDA believes that the recommendations made in this companion guidance, together with the recommendations in the March 2005 guidance, will benefit sponsors considering the submission of either voluntary genomic data submissions or marketing submissions containing genomics data. As technology changes and more experience is gained, these recommendations may be updated.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should

[1] This guidance has been prepared by the Center for Drug Evaluation and Research (CDER), the National Center for Toxicological Research (NCTR) and the Center for Biologics Evaluation and Research (CBER), in cooperation with the Center for Devices and Radiological Health (CDRH) at the Food and Drug Administration.

For the purposes of this guidance, the term *drug* or *drug product* includes human drug and biological products.

Paperwork Reduction Act Public Burden Statement: According to the Paperwork Reduction Act of 1995, a collection of information should display a valid OMB control number. The valid OMB control number for this information collection is 0910-0557 (expires 12/31/2007). The time required to complete this information collection is estimated to average 10 hours per response, including the time to review instructions, search existing data resources, gather the data needed and complete and review the information collection.

be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. GENE EXPRESSION DATA FROM MICROARRAYS

The following methodological issues should be considered when submitting gene expression data from microarrays. The recommendations made in this document apply to development of microarray data that might be submitted in support of INDs, NDAs, and BLAs. For microarray data supporting the clearance or approval of a diagnostic device, additional information beyond these recommendations may be requested.

A. RNA Isolation, Handling, and Characterization

One of the most critical steps in performing RNA-based experiments such as microarray gene expression experiments is the isolation of high quality, intact RNA. To achieve this goal and preserve sample integrity throughout the course of the experiment, some steps before and after RNA purification should be carefully planned to ensure quality during isolation and confirm high quality before use in a downstream application. A secondary goal is maximizing the yield of RNA. In addition, storage and shipping conditions of samples can influence the stability of RNA. Thus, it is very important to store the RNA under the best conditions to preserve the integrity of the sample. Finally, we recommend that standard operating procedures (SOPs) be established to ensure reproducibility of the RNA isolation method and RNA quality (e.g., see http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/docs/MAQC_Sample_Processing_Overview_SOP.pdf). The following recommendations will help achieve these goals.

1. Pre-RNA Isolation Considerations

RNA is sensitive to degradation by RNase, which is ubiquitously present in living organisms. Thus, sample-handling issues should be addressed and methods for sample handling need to be assessed to ensure that the methods and their associated metrics are suitable for the purpose to which they are applied before embarking on RNA isolation from samples. We also recommend that any work areas and equipment to be used to generate data for submission studies be dedicated specifically for RNA isolation and other RNA-related work.

<u>RNase-free reagents and disposables/glassware:</u> It is imperative to use RNase-free reagents and glassware for RNA isolation. Commercially available RNA isolation kits often provide these. It may be of value to confirm that RNase inactivation methods are functioning as expected prior to launching submission studies.

<u>RNA stabilizer(s)</u>: We recommend that the need for adding RNA stabilizing agents to samples/reagents be assessed and an appropriate RNA stabilizer be identified, and assessed for suitability in a pilot experiment.

Batch size: We recommend that the maximal batch size for sample preparation be determined to help identify and limit the time taken for the entire RNA isolation process. Establishing an upper limit for batch size will reduce problems encountered during the scaling-up process since long processing times can jeopardize RNA integrity.

Sample collection, storage and shipping conditions: there are numerous variables that may affect sample reproducibility in microarray studies. We recommend that the impact of the following variables on RNA quality be assessed. These include:

- maximum and minimum sample dimensions
- volumes
- weights

Additional important parameters include:

- correct sampling technique per tissue/organ
- timing of sample dissection/processing time
- maximum allowed elapsed time between resection and stabilization of the tissue
- stability of specimen in transport under recommended conditions (temperature, duration, etc)

There may be other study-specific parameters to consider. For example, in oncology studies we recommend that the percent tumor in the sample be determined.

2. *RNA Isolation from Tissues or Cells*

Treatment of cells or tissue samples prior to RNA isolation and careful handling are necessary to preserve RNA. Several methods are available for successful isolation of high quality RNA. A number of reagents are also available that aid in preserving the quality of RNA. For example, an RNA stabilizer that is compatible with RNA isolation procedures may be added to the isolated tissues or cells before storing the samples. Alternatively, tissues or cells can be quickly frozen in liquid nitrogen and stored at -80 ^{0}C to prevent RNA degradation. Tissues or cells can also be homogenized in the presence of a strong denaturant that inactivates RNase, followed by freezing the homogenate at -20 ^{0}C or below. In any case, we recommend that the manufacturer's specifications be followed and that the quality of the resulting RNA be acceptable for the study. RNase-free reagents, equipment, materials, and work spaces should be used for subsequent isolation and analytical steps.

3. *RNA Isolation from Whole Blood and PBMCs*

RNA can be isolated from whole blood or from peripheral blood mononuclear cells (PBMCs). Most studies conducted so far have used the PBMCs since they are the most transcriptionally active cells in blood.[2] This fraction primarily consists of lymphocytes and monocytes. RNA isolated from PBMCs and whole blood should not be used interchangeably in the same study.

RNA isolation from whole blood: RNA may be isolated from whole blood, and this specimen type is attractive since the blood sample with the RNA stabilizer can be stored for a long time, presumably without compromising RNA quality or the stability of expression profiles under manufacturer-suggested conditions. The storage conditions and the maximum storage durations used to store whole blood samples selected should meet any acceptance criteria applicable to the selected platform. One disadvantage of RNA isolation from whole blood is that reticulocytes (immature red blood cells (RBCs)) in the specimen, while representing only 0.5-2% of the RBCs, can contribute up to 70% of the mass of mRNA in total RNA, of which globin mRNA is the major RNA. In microarray gene expression experiments, the overabundance of globin mRNA can result in failure to detect some transcripts that are of low abundance.[3] While working with whole blood specimens, the need for protocols for reducing globin mRNA from whole blood[4] or alternative methods to minimize the impact of globin mRNA on gene expression data should be considered. If it is determined that such methods are needed, ensure that they work as intended within the context of your method.

The quality of the microarray data generated from whole blood specimens can be improved by removing reticulocytes, although this often requires that blood be processed at the site of blood draw. Any manipulation of the blood sample may cause a change in the gene expression profiles of some transcripts.[5] We recommend, therefore, that a study to simulate the conditions of the preclinical or clinical blood specimen collection and manipulation to be employed be conducted, to assess the impact of key variables on the chosen method.

RNA isolation from PBMCs: RNA may be isolated from PBMCs that have been isolated from a whole blood specimen using one of several techniques. Some

---

[2] An Analysis of Blood Processing Methods to Prepare Samples for GeneChip Expression Profiling- Technical Note from Affymetrix. (http://www.affymetrix.com/support/technical/technotes/blood_technote.pdf)

[3] Fan H. (2005) The transcriptome in blood: challenges and solutions for robust expression profiling. *Current Molecular Medicine* **5**, 3-10.

[4] Debey S. et al., (2006) A highly standardized, robust, and cost-effective method for genome-wide transcriptome analysis of peripheral blood applicable to large-scale clinical trials. *Genomics* **87**, 653-664.

[5] Burczynski M.E. and Dorner A.J. (2006) Transcriptional profiling of peripheral blood cells in clinical pharmacogenomic studies. *Pharmacogenomics* **7**, 187-202.

commonly used methods include the Ficoll-Hypaque centrifugation and use of cell preparation tubes with sodium citrate. RNA isolation from PBMCs is the preferred method for many applications since the RNA is free from globin mRNA and generally gives better results on microarrays. However, it has been shown that time delays and temperature changes can affect gene expression profiles of several genes[6, 7] and therefore it is critical to isolate the PBMCs within hours of blood collection, particularly if any stabilization materials or storage conditions to stabilize the expression profiles are not used. Regardless of the method chosen, one should assure, through measurement of quality parameters, that it consistently and reliably yields RNA with acceptable performance parameters for the selected analytical method. If more than one RNA isolation method (e.g., methods from two different manufacturers) are selected, it should be ascertained that both methods give equivalent results in your system.

4. RNA Storage

For short term storage, RNA suspended in RNase-free water (with 0.1mM EDTA) or in TE buffer should be stored at -80^0 C in aliquots in non-frost-free freezers. Repeated freeze-thaws should be avoided. Generally, RNA is stable for about a year at -80^0 C under the above-mentioned conditions. For long term storage, RNA samples could be stored at -20^0 C in ethanol.

5. RNA QC

The quality of RNA samples can be monitored in several ways. The most widespread current metric is spectrophotometric analysis using the ratio of absorbance at 260nm/280nm as a measure of RNA quality and purity.[8] Two common additional methods are agarose gel electrophoresis and analysis using a dedicated RNA analyzing instrument. Considerations for RNA quality metrics include the following:

- For spectrophotometric analysis, the ratio of absorbance at 260nm and 280 nm (A_{260}/A_{280}) can be used to assess RNA purity and is typically recommended to be greater than 1.8.[9]

---

[6] Baechler E.C. (2004) Expression levels for many genes in human peripheral blood cells are highly sensitive to ex vivo incubation. *Genes and Immunity* **5**, 347-353.

[7] Debey S. (2004) Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types. *The Pharmacologics Journal* **4**, 193-207.

[8] http://arrayconsortium.tgen.org/np2/public/qualitycontrol/jsp.

[9] Dumur, C.I. et al., (2004) Evaluation of quality-control criteria for microarray gene expression analysis. *Clinical Chemistry* **50**(11), 1994-2002.

- For agarose gel analysis, generally a 1% denaturing agarose gel is used, and clearly visible 18S and 28S RNA bands are taken as measure of RNA integrity. Ideally, the intensity of the 28S band should be twice the intensity of the 18S band. Degraded RNA will have a smeared appearance and lack two clear bands.

- For dedicated RNA analysis instruments, several different metrics may be useful, e.g., presence of 18S and 28S rRNA peaks, ratio of 28S/18S bands, and percentage of total RNA represented by the rRNA peaks.[8] Specific recommendations regarding use of dedicated RNA analysis instruments and the data they generate can be found in manufacturers materials.[10]

Regardless of the method(s) chosen to assess RNA quality, it should be ascertained that the acceptance criteria for the RNA samples are consistently appropriate to yield RNA quality that is suitable for the analytical method selected. The selected RNA isolation method should minimize genomic contamination of the isolated RNA because genomic DNA could negatively affect downstream applications.

B. Labeling Reactions

In genomic submissions, it is important that sponsors use a labeling system that has been documented to perform well on a given manufacturer's array. It is critical that the sponsor begin the labeling process with high-quality RNA-free of contaminants that might affect the labeling efficiency or introduce labeling bias, as compromised RNA quality will affect subsequent steps of sample processing and ultimately lead to poorer quality microarray data. We recommend that the use of accepted quality measures (18S/28S ratios) be included in this report and that RNA samples prepared for labeling be of comparable quality.

We recommend the use of consistent methods of target labeling throughout the particular study or studies that will be analyzed as a group since dissimilar microarray data could be obtained when kits from different manufacturers or different types of labeling kits are used. If there is any change in a critical component in the labeling kit (kit manufacturer, key enzyme or reagent), we recommend that it be tested to demonstrate comparability of the data generated prior to being used with samples analyzed as an arm of a study. We recommend that reagent lot acceptance criteria be developed to ensure the reproducibility of labeling reactions.

The use of standard operating procedures (SOPs) is encouraged, and we recommend that operators be fully trained on all protocols prior to processing of samples for the study. Equipment should be on an appropriate maintenance schedule and the laboratory environment maintained in accordance with the manufacturer's recommendations.

[10] Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, Mueller O, Schroeder A. and Auffray C. (2005) Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Research* **33**(6), e56.

236 The development of QC or intermediate labeling steps is highly recommended. If any
237 intermediate QC step indicates a problem and the RNA is of reasonable quality, the labeling
238 process can be repeated to produce higher quality input material for hybridization to the
239 microarray chip. In addition, it is recommended that reagents be stored under appropriate
240 conditions. Use of controls and reference standards are recommended to verify consistent
241 performance throughout the labeling procedure.
242
243 We recommend the use of validated standard operating procedures (SOPs) addressing all aspects
244 of sample collection, storage, and sample and array processing to generate microarray data, and
245 all operators should be fully trained on all protocols prior to initiating the study. It is also
246 advisable to establish appropriate maintenance schedules for all equipment, and ensure that the
247 laboratory environment is maintained in accordance with the SOPs.
248
249 **C. Hybridizations for Microarrays**
250
251 You should include pertinent information on reproducibility and accuracy of array hybridization
252 in your submission package. In the absence of widely accepted QA/QC control metrics for DNA
253 microarray technologies or consensus on how to establish the reliability of the results obtained
254 from a DNA microarray experiment, we recommend you establish and assess internal control
255 metrics for quality and reliability. For example, some organizations have used QA/QC pass/fail
256 filters to eliminate outlier arrays and some array manufacturers recommend thresholds for certain
257 platform-specific QC measurements.
258
259 Currently, the ERCC (External RNA Controls Consortium)[11] and MAQC (MicroArray Quality
260 Control Consortium) groups are developing spike-ins and reference standards, which may be
261 useful in evaluating the quality of a particular microarray experiment when available. Another
262 recent effort has produced a pair of reference RNA pools for use with rat DNA microarrays that
263 allow accuracy, reproducibility, and dynamic range assessments.[12] Conceptually, this strategy
264 could be used to produce reference materials for any organism, including human. Until such
265 independent resources are widely available and consensus quality standards are developed and
266 implemented by the microarray community, carefully adhering to the microarray manufacturer's
267 recommended procedures offers the best current practice at this time. Detailed protocols have
268 been prepared by major DNA microarray manufacturers and posted on the MAQC Web site.[12]
269 Because the microarray field is evolving, it is important to note that manufacturers occasionally
270 change probe sequences and protocols, reflecting continuing improvements to this technology.
271 Regardless of the source of quality control materials and methods, we recommend you describe
272 how you selected those that you use, and how you determined that they were acceptable for your
273 purposes.
274
275 We recommend that the following be clearly outlined in a figure:
276

---

[11] External RNA Controls Consortium. (2005) The External RNA Controls Consortium: a progress report. *Nature Methods* **2:** 731 - 734.

[12] http://edkb.fda.gov/MAQC/

- Microarray chip details

A key information in the submission package is the information of micoarray chips used. There are at least two different categories of microarray chips, commercial chips and customer chips (array manufactured by sponsor, or contractors).

1. If commercial array chips are used in the study, sponsor should provide the following information: the name of manufacture, type of array, lot #, manufacture date (or expiration date), and array QC parameters (QC tests performed by vendor).

2. If customer chips are used for the study, the sponsor should provide: manufacturing protocol, documents from vendors if any materials purchased from commercial resources, QC thresholds, and QC testing results.

- Microarray experimental design details

We recommend you include sample processing and labeling (e.g., were samples processed in the same batch or different batches; was the same procedure used for all samples, technical replication, biological replication and other appropriate information).

- How data were generated and analyzed

One approach would be to start with how the primary data were obtained (e.g., laser scanner settings, software settings for image acquisition). We recommend you explain how the data from individual microarrays were combined and the normalization method and then provide data filtering, data analysis, statistical tests, and other appropriate information.

D. Fluorescence Reader Settings for Microarrays

Microarray technology uses a multi-step process in which variability at each step must be reduced to maximize the probability of detecting changes that arise from biology and not from experimental artifact. Scanners used to collect the microarray signals are a potential source of variability in data derived from this technology. Recent publications have pointed out the importance of optimal reader settings for obtaining high-quality microarray data.[13] The signal readout system is often thought of as a black box that quantitates the signal from each DNA microarray spot. The measurement of the abundance of RNA species by DNA microarray technology assumes a linear relationship between the signal read-out from the scanner and the dye concentration, which is further assumed to be linearly correlated with transcript abundance in the RNA sample.

Each array system, scanner type, and signaling dye combination, may have its own linear dynamic range, which changes with voltage gains. Important recommendations for scanners that

[13] Shi, L., Tong, W., Su, Z., Han, T., Han, J., Puri, R.K., Fang, H., Branham, W.S., Chen, J.J., Xu, Z., Harris, S.C., Hong, H., Xie, Q., Perkins, R.G., and Fuscoe, J.C. (2005). Microarray scanner calibration curves: characteristics and implications. BMC Bioinformatics 6 (Suppl 2):S11.

will help minimize technical variability and improve consistency of data collection include the following:

1. Calibration of scanners as recommended by the manufacturers
2. Routine use of standardized scanner reference materials for calibrations to allow for characterization of concentration-dependent read-outs
3. Attention to scanner settings (e.g., laser power and voltage gain). Specifically, we recommend that scanner settings be set to maximize the linear dynamic range.
4. Keeping the scanner laser power and voltage settings constant during a study. Note that some scanners are not tunable, so that this source of variability is eliminated.
5. If the dye-intensity to signal output relationship is defined, possible corrections when signals fall outside of the linear dynamic range, thus reducing variability in the very high or very low signal range
6. Submission of scanner setting and calibration information as part of the submission package

E. Differentially Expressed Genes

Specific genes sets derived from microarray experiments can be proposed as genomic biomarkers for a specific endpoint in a defined context. Such specific gene sets should be reproduced upon review if the analysis protocol is identical to that reported by the sponsor. The sponsor should include in the submission a clear description of the steps, parameters, and algorithms leading to the list of differentially expressed genes list in the genomic submission.

Different analysis protocols may yield dissimilar lists of differentially expressed genes, and these cannot be justified solely through a biological interpretation if they are to be proposed as genomic biomarkers. To the extent that these genomic biomarker sets become part of a decision-making process in drug development or therapeutic applications, we recommend that transfer of genomic biomarker sets from microarrays to other platforms (such as quantitative RT-PCR) be attempted only after the sponsor concludes that these differentially expressed genes are sensitive, specific, and reproducible.

Sources of variability in microarray data leading to the step in which the differentially expressed gene list is determined may be minimized by following the recommendations in this document. To determine which genes are in fact differentially expressed, a number of factors need to be considered that may have confounding effects:

- The application of platform-specific flags
- Rejection criteria for low-intensity transcripts
- Rejection criteria for outlier hybridizations
- Platform-specific normalization protocols
- Data analysis protocol for selection of differentially expressed genes

361 There is no consensus at this time regarding the appropriate choices for each of these factors.
362 The sponsor should exercise care in how parameters and protocols are chosen for each of these
363 factors and should consult current literature regarding efforts to reach a consensus.[14, 15, 16, 17, 18,
364 19]
365
366 In principle, several analysis protocols can be used to determine lists of differentially expressed
367 gene lists for a sufficiently large number of technical and biological replicates. In practice,
368 constraints on the number of technical and biological replicates are likely to be the norm in
369 genomic submissions. For example, technical replicates are constrained by the minimum amount
370 of RNA needed to hybridize each biological sample. Both clinical as well as preclinical samples
371 may have major constraints in the total amount of RNA available from each biological sample.
372 Biological replicates are constrained by the total number of subjects to be included in a study.
373 We recommend that these constraints be considered in the selection of analysis protocols for the
374 determination of differentially expressed genes.
375
376 **F. Biological Interpretation of Lists of Differentially Expressed Genes**
377
378 Once the list of differentially expressed genes has been generated via a variety of statistical and
379 analytical tools, the next step in the process should be to interpret the biological meaning of gene
380 expression changes and determine whether biological pathways may be of functional relevance
381 to the mechanism of drug action, or may be correlated to safety and/or efficacy.
382
383 A number of questions should be addressed at this point, including, for example:
384 • Are genes from a particular pathway or set of pathways significantly overrepresented in
385 the list?
386 • How many pathways are affected?
387 • Can the mechanism of action be inferred from the functions of the pathways altered or
388 from the pattern of expression across the genes within these pathways?

---

[14] Simon R. Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. (2006) *J Natl Cancer Inst.* **98**(17):1169-71.

[15] Simon R. (2006) A checklist for evaluating reports of expression profiling for treatment selection. Clin *Adv Hematol Oncol.* **4**(3):219-24.

[16] Dobbin KK, Simon RM. (2007) Sample size planning for developing classifiers using high dimensional DNA microarray data. *Biostatistics.* **8**(1):101-17.

[17] Varma S, Simon R. (2006) Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics.* **7**:91.

[18] Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, Zhang L, Mei N, Chen T, Herman D, Goodsaid FM, Hurban P, Phillips KL, Xu J, Deng X, Sun YA, Tong W, Dragan YP, Shi L. (2006) Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol.* **24**(9):1162-1169.

[19] Canales RD, Luo Y, Willey JC, Austermiller B, Barbacioru CC, Boysen C, Hunkapiller K, Jensen RV, Knight CR, Lee KY, Ma Y, Maqsodi B, Papallo A, Peters EH, Poulter K, Ruppel PL, Samaha RR, Shi L, Yang W, Zhang L, Goodsaid FM. (2006) Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol.* **24**(9):1115-22.

- What is the tissue specificity of the pathways and the gene function in relation to biological processes?
- What are the magnitude and/or pattern of the alteration in a particular pathway in relation to treatments with other compounds (related or unrelated) with known pharmacological or toxicological properties?

At present, no single tool can be used to find answers to all these questions, but a combination of tools can be used to address a particular question of interest as thoroughly as possible. To this end, a variety of analytical platforms are available, either free on the Web or via purchase of a commercially available product.

An overlap of the biological interpretations obtained with two or more different databases can facilitate a consensus on what the interpretation should be. However, this is not always the case. Consensus can be hindered by many factors including, but not limited to, absence of information on the compound of interest in the reference databases or a lack of annotation for particular pathways of interest. For example, subsets of genes may be placed in specific pathways in one system, but they may not be represented in the same pathways in another pathway analysis tool, or genes may not have been evaluated in a particular platform. In pathway analysis databases, the information may differ depending on which content is extracted from the literature and how that extraction is performed (whether automated or by manual curators). In addition, a critical distinction is whether all information is extracted, or if only the information supported by direct experimental evidence included in the publication is extracted. We recommend heavy reliance on the literature and on reference databases to extract functional information on specific gene lists and generate hypotheses on the biological significance of the relevant set of genes.

We also recommend that the biological significance of gene sets proposed by a sponsor be accompanied by a standard set of information that will enable recapitulation of the analysis and assessment of the validity of the interpretation by regulatory reviewers. In addition, we recommend that the gene sets proposed by sponsors should be validated by other conventional techniques, such as Q-PCR, or RT-PCR. Such information should include, but not be limited to:

- Type of database used for annotation, including vendor name
- Methods and approaches (cut-off, statistical tests) used to identify over-represented pathways within the database
- References used to justify any user-defined annotation
- A summary by the sponsor of the interpretation of the pathway annotation results

III. GENOTYPING

A. Genotyping Methods

Genetic differences among individuals occur in a variety of forms, from alterations in chromosomal arrangement or copy number to single base-pair changes. Much of the genetic variation currently used in pharmacogenetics occurs at the level of individual genes (e.g., drug metabolizing enzymes) on a scale ranging from single base-pair changes to entire gene

duplications or deletions. Examining genomic DNA is often the most reliable and practical method for characterizing genetic variation, although methods based on protein or mRNA expression levels can be preferable in some situations, such as when determining treatment-sensitivity of cancer or viral infection. Many methods are currently available for characterizing DNA variations, and new methods are rapidly being developed.

B. DNA Isolation, Handling, and Characterization

Whole blood is commonly used for the extraction of genomic DNA in clinical research settings. Blood collection tubes generally use anticoagulants such as EDTA, CPD, ACD, Citrate or Heparin. DNA in a blood sample is susceptible to degradation unless properly stored. Although manufacturers of blood collection tubes usually recommend appropriate storage conditions for optimum stability, we recommend you ensure that these conditions yield DNA that is suitable for your assay, for example, by checking for the presence of full-length DNA.

When DNA is isolated from blood, carryover of contaminants such as salts, phenol, ethanol, heme (in blood DNA isolation), and detergents from conventional purification procedures can inhibit performance of DNA in downstream applications. In addition, contamination with the anticoagulant heparin impairs amplification by PCR.[20, 21] Potential for contamination and interference in isolation procedures should be assessed, and procedures for avoiding these should be implemented where necessary.

Although DNA is a relatively stable molecule, it should be stored carefully. Degradation of DNA can have a major effect on any results obtained, generating errors that are both quantitative and qualitative. There are several factors that can result in DNA degradation including introduction of enzymatically active nucleases, acid hydrolysis, and degradation due to repeated freeze-thaw cycles. You should implement DNA handling and storage procedures that limit these and any other factors that could affect DNA quality. For example:

- Avoid exposure of DNA solutions to nucleases that may be present on lab equipment or in reagents;
- Store DNA at a slightly alkaline pH (e.g., Tris EDTA buffer) once isolated;
- Maintain long-term storage of DNA at -20^0 C or at -80^0 C.
- Freeze sample in aliquots to reduce freeze-thaw degradation

C. Genotyping Report

We recommend that the following information be included in the genotyping report, regardless of the genomics submission type (see the *Pharamcogenomic Data Submissions* guidance for regulatory requirements):

- Description of assay platform or methodology

[20] Smythe et al., BMC Infectious Diseases, 2002, 2:13.

[21] Yokota et al., Journal of Clinical Laboratory Analysis, 1999, 13: 133 – 140.

- Samples studied, including demographics and sample size justification for genotype/clinical phenotype correlation and adequate coverage for ethnic/racial groups; include expected allele frequency in different populations
- Alleles measured and correlation with metabolic status designation
  - For metabolizing enzymes, how EM (extensive metabolizer), PM (poor metabolizer), IM (intermediate metabolizer), or UM (ultra rapid metabolizer) are determined
  - Sample test report
  - For new genes, correlation between gene variant and encoded protein activity
- Whether the assay was performed in a CLIA-certified lab or research lab

IV. PROFICIENCY TESTING

High-quality data are the foundation for deriving reliable biological conclusions from a microarray gene expression study. However, large differences in data quality have been observed in published data sets when the same platform was used by different laboratories.[22, 23] In many cases, poor quality of microarray data was due not to the inherent quality problems of a platform but to the lack of technical proficiency of the laboratory that generated the data. Such a systematic *procedural failure* in a laboratory is much more serious than randomly failed hybridizations that lead to outlying arrays, because the laboratory may not recognize that it has a procedural failure problem.

The Agency recommends that sponsors provide data that will enable FDA reviewers to objectively evaluate the competency of the laboratory that generated the data in a genomic submission. Many studies report quality control metrics or use standards to provide internal assessments of microarray data. This information is useful for confirming the technical ability to reproducibly perform a given assay within an individual study.

In addition to within-laboratory testing, an assessment of the overall competence of a facility can be performed through inter-laboratory comparisons, such as proficiency testing. Laboratory proficiency can be monitored through a number of approaches.

- RNA sources

Two FDA-led initiatives have developed and characterized reference RNA samples for proficiency testing. Mixed tissue pools of rat RNA samples have been designed with known

[22] Shi L, Tong W, Goodsaid F, Frueh FW, Fang H, Han T, Fuscoe JC and Casciano DA (2004) QA/QC: challenges and pitfalls facing the microarray community and regulatory agencies. *Expert Rev Mol Diagn* **4**:761-77.

[23] Shi L, Tong W, Fang H, Scherf U, Han J, Puri RK, Frueh FW, Goodsaid FM, Guo L, Su Z, Han T, Fuscoe JC, Xu ZA, Patterson TA, Hong H, Xie Q, Perkins RG, Chen JJ and Casciano DA (2005) Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics* 6 Suppl 2:S12.

differences in tissue-selective genes[24] and have been used in the first proficiency testing program for microarray laboratories.[25] In addition, the MicroArray Quality Control (MAQC) Project[26] developed two human reference materials and extensively tested them on multiple gene expression platforms. Data from both initiatives have been deposited in public databases, and the RNA samples used in the MAQC project are now commercially available for use by laboratories to assess ability to reproduce MAQC data.

- Experimental design for proficiency testing

Most RNA-based genomic assays are designed to detect differentially expressed genes or profiles. A proficiency testing program for these assays could be centered on testing of replicates of two biologically different samples with known differences in transcript abundance, in order to measure the ability to repeatedly detect differential gene expression. For example, the laboratory could plan to process three or more replicates of sample A (labeled A1, A2, and A3) and three or more replicates of sample B (labeled B1, B2, and B3), to evaluate the within-laboratory repeatability both in terms of repeatable intensity measurements and repeatable detection of differential gene expression. If multiple laboratories provide data generated using the same RNA samples and the same platform, site-to-site reproducibility and comparability of sites to detect differences in expression can be assessed. We recommend that laboratories use a proficiency testing program, and that the testing be repeated throughout the year so that multiple data sets from the same laboratory can be compared to confirm the consistency of the laboratory's performance over time.

- Laboratory compliance

The Agency encourages microarray facilities to adhere to the good laboratory practices outlined in 21 CFR 58. Laboratories may also wish to obtain CMS/CLIA certification if the microarray data have potential clinical or diagnostic applications. All CLIA-compliant assays require repeated data comparisons with other providers to verify the competency of individual laboratories. Participation in a proficiency testing program would fulfill this CLIA requirement.

---

[24] Thompson KL, Rosenzweig BA, Pine PS, Retief J, Turpaz Y, Afshari CA, Hamadeh HK, Damore MA, Boedigheimer M, Blomme E, Ciurlionis R, Waring JF, Fuscoe JC, Paules R, Tucker CJ, Fare T, Coffey EM, He Y, Collins PJ, Jarnagin K, Fujimoto S, Ganter B, Kiser G, Kaysser-Kranich T, Sina J and Sistare FD (2005) Use of a mixed tissue RNA design for performance assessments on multiple microarray formats. *Nucleic Acids Res* **33**:e187.

[25] Reid LH *et al.* (2006). Proficiency testing program for microarray facilities (in preparation). http://www.expressionanalysis.com/proficiency_test.html.

[26] Shi L, Reid LH *et al* (2006) MicroArray Quality Control (MAQC) Project: A comprehensive survey demonstrates concordant results between gene expression technology platforms. *Nat Biotechnol* **24**(9), 1151-1161.

547 **V. GENOMIC DATA IN CLINICAL STUDY REPORTS**
548
549 There are many possible sources of data for genomic data submissions. Genomic data from
550 clinical studies may result from microarray expression profiling experiments, genotyping or
551 single-nucleotide polymorphism (SNP) experiments, or from other evolving analytical
552 methodologies pertaining to drug dosing or metabolism, safety assessments, or efficacy
553 evaluations. Genomic data may also be reported from studies where other data are also reported,
554 such as with efficacy or safety data from clinical or nonclinical studies. However, these data can
555 be reviewed only if the content of the clinical data report included in the submission contains
556 ***sufficient detail*** regarding the sample selection.
557
558 The following describes FDA's current thinking about what data should be submitted with
559 genomics data in a submission to the Agency (including a voluntary submission). Regulatory
560 applications for these data are described in detail in FDA's *Pharmacogenomic Data Submissions*
561 guidance in the context of different algorithms for the submission of pharmacogenomic data
562 consistent with FDA requirements for INDs, NDAs, and BLAs, as well as for Voluntary
563 Genomic Data Submissions (VGDS). Throughout the following discussion, we suggest that you
564 refer to the *Pharmacogenomic Data Submissions* guidance for in-depth background on this
565 discussion.
566
567 In all genomic submissions, a full clinical study report is very helpful to Agency reviewers. The
568 report should provide a clear explanation of how the critical design features of the study were
569 chosen as well as enough information on the plan, methods, and conduct of the study to
570 eliminate ambiguity in how the study was carried out. The report with its appendices should also
571 provide individual patient data relevant to pharmacogenomics, including demographic and
572 baseline data, and details of analytical methods such as validation reports to allow replication of
573 the critical analyses. It is also particularly important that all analyses, tables, and figures carry
574 clear identification of the set of patients from which they were generated.
575
576 To improve the usefulness of the submission, we recommend that the content of the clinical
577 section describing a genomic experiment contain the following information:
578 – Title page
579 – Table of contents
580 – Synopses and summary of findings
581 – Background and scientific rationale
582 – Primary and secondary study objectives
583 – Study design, sample collection and storage, and pharmacogenomic methods
584 – Clinical study protocol, including minimally[27]:
585 - inclusion and exclusion criteria
586 - demographic data
587 - listing of individual experimental measurements by patient, including
588 pharmacokinetic/pharmacodynamic datasets and lab results; and explanation of
589 missing data

[27] ICH guidance *E3 Structure and Content of Clinical Study Reports.*

590 - disposition of patients
591 - protocol deviations
592 - individual adverse events or laboratory abnormalities
593 - pharmacogenomic and other biomarker datasets as necessary
594 - correlation between clinical and pharmacogenomic data
595 - discussion of the comprehensiveness of genomic data in all randomized patients
596 and the impact of loss of such information, if any, on inferences and associations,
597 especially with outcomes
598 - additional discussion and conclusions
599 –References and any supplementary materials.
600
601 The specific sequence and grouping of topics may change if alternatives are more logical for a
602 particular study. The *Pharmacogenomics Data Submissions* guidance and other Agency
603 regulations and guidance contain detailed discussions on specific regulatory requirements.
604
605 The preferred ***submission standard*** for clinical data is the Clinical Data Interchanges Standards
606 Consortium (CDISC) Study Data Tabulation Model (SDTM) standard. Please see the FDA Data
607 Standards Council Web site[28] for more information on the standard.[29]
608
609
610 **VI. GENOMIC DATA FROM NONCLINICAL TOXICOLOGY STUDIES**
611
612 Genomic data can be collected in nonclinical studies, such as toxicogenomic studies. This
613 section describes how to submit nonclinical toxicology data with a genomic data submission.
614 How the data should be submitted depends on the purpose of the submission. Three general
615 types of submissions can be identified:
616
617 • The first type of submission might have the objective of expanding the selection process
618 criteria (i.e., screening to aid in the selection of a lead compound for clinical
619 development or to eliminate compounds with certain characteristics).
620 • The second might present the characterization of a particular compound.
621 • The third might present a general scientific discussion that might not be related to the
622 development of a compound and/or compound class.
623

[28] See http://www.fda.gov/oc/datacouncil/.

[29] The SDTM can be obtained from the CDISC Web site at http://www.cdisc.org/models/sds/v3.1/index.html .

SDTM Implementation Guides:

- *The Study Data Tabulation Model Implementation Guide (SDTM-IG) for clinical study data* can be obtained from the CDISC web site at: http://www.cdisc.org/models/sds/v3.1/index.html
- *The Study Data Specification for submitting SDTM datasets to CDER* can be obtained at http://www.fda.gov/cder/regulatory/ersr/Studydata-v1.1.pdf

PK/PD data submission should be in SAS.XPT-compatible format.

A. Expanding the Selection Process Criteria

When a submission is intended to expand the selection process criteria and precede the development of a compound (i.e., screening for lead compounds or to eliminate certain characteristics), we recommend the inclusion of the following information:

1. General narrative about the objective of the submitted application, brief narrative about the compound(s), intended use, and mechanism of action
2. Objective of the submitted study with its experimental design (treatment, duration, replicates, drug formulation, route of administration, rationale for dose selection). As applicable, information about species, strain, sex, genetic background, age, weights, developmental stage, organ/tissue where sample originated, cell type can be included. We recommend that a brief description of sample handling, storage and preparation methodology also be included.
3. Toxicology parameters including clinical pathology (serum chemistry and hematology) and histopathology data consistent with STP guidelines (Toxicologic Pathology, 32, 126-131 (2004)), preferably in an electronic format). When applicable, the correlation between pathology findings and genetic variation or gene or protein expression should be explained.
4. Correlations of individual animal data to genetic variation or gene or protein expression should be explained.
5. Pharmacokinetic parameters and ADME properties of the compound should be provided if known. When applicable, correlation between pharmacokinetic findings and genetic variation or gene or protein expression should be highlighted.
6. Reference should be made to scientific and analytical methods for genetic variation or gene or protein expression, including genotyping or expression profiling methods, statistical methods, and software packages used.

B. Characterization of a Particular Compound

If the intent of a submission is to characterize a particular compound, it is generally recommended that the toxicology portion of the submission be reported in a similar format to a toxicology report. These reports follow the good review practices template (Section 4.1 m (1 to 6)). If the template is not used, a copy of the study protocol should accompany the line listings and generally include clinical signs, mortality, body weight, food consumption, hematology, clinical chemistry, urinalysis, gross pathology, organ weights, histopathology, and pharmaco/toxicokinetics (as available) with a full tabulation of data suitable for detailed review. These data contain line listings of the individual data points, including laboratory data points, for each animal along with summary tabulations of data points. A copy of the study protocol is expected to accompany the line listings.

C. General Scientific Discussion

When a submission contains data to support a general scientific discussion that is not necessarily related to the development of a compound and/or compound class, the minimal amount of nonclinical data to be submitted should be similar to the previously described scenarios.

However, it is up to the sponsor to provide adequate information to clarify and support the scientific issues discussed. The data submitted will probably not be detailed, but we recommend that it be tabulated in a form that will be concise and adequately descriptive for the specific purpose of the submission.

VII. DATA SUBMISSION FORMAT

A general description of clinical and non-clinical data associated with genomic data submissions is included in Sections III and IV of this guidance.. This section provides details on electronic data submission formats for genomic and associated non-clinical or clinical data.

A. Submission Standard

For any type of genomic data submission, we encourage you to submit the data electronically in a tab-delimited file conforming to the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) standard or the Standard for Exchange of Nonclinical Data (SEND) SDTM format per the CDISC guidelines (http://www.cdisc.org/).[30]

B. Microarray Gene Expression Data

When a microarray gene expression experiment is included in a genomic data submission, both raw and normalized gene expression data as well as the gene lists that are used to support the biological conclusions in the submission should be submitted electronically.

- *Raw data* – It is recommended that one file be submitted per array. For example, CEL files would be submitted for the Affymetrix GeneChip platform, while the tab delimited spreadsheet format could be used for other platforms with the gene ID (e.g., GenBank Acc#, manufacturer ID), in the first column.
- *Normalized data* – It is recommended that one file be submitted per array. The tab delimited spreadsheet format should be used with the gene ID (e.g., GenBank Acc#, manufacturer ID) in the first column.
- *Gene lists* – Lists of genes supporting a biological interpretation in the submission should be included. Probeset IDs in each array should identify each entry in these lists. The lists should be submitted along with parameters such as fold change and p-value for each gene of interest in a tab delimited format.
- Besides the parameters mentioned above, the gene lists (or Results) submission should also include following information:
    - software used for data analysis

[30] More information can be found at FDA Data Standards Council Web site, http://www.fda.gov/oc/datacouncil/. The *Standard for Exchange of Nonclinical Data (SEND) Implementation Guide for Animal Toxicology Studies* can be obtained from the CDISC Web site at: http://www.cdisc.org/models/send/v2.3/SENDV2.3ImplementationGuide.pdf.

- filtering conditions (such as intensity filter, spot flag filter, spot size filter, and detection call filter)
- normalization method selected for data analysis (there are several different normalization methods available, such as median, Lowess, and housekeeping gene normalizations)
- methods selected for statistical analyses.

In addition to the data files, an experimental summary table (called ExpSumTable, Appendix I) should be prepared to summarize the key experimental parameters investigated in the microarray study. The experimental parameters should be prepared in accordance to the MIAME (Minimum Information About a Microarray Experiment) guidelines.

C. Clinical and Nonclinical Data

The Study Data Tabulation Model (SDTM) that encompasses both CDISC and SEND has been developed to guide the organization, structure, and format for both clinical and nonclinical data submissions. For genomic data submissions, clinical and nonclinical data should be prepared in accordance to the SDTM. CDISC/SEND organizes the study data under the concept of *domains*. Each domain summarizes a collection of observations with a topic-specific commonality. At this point, we ask that each domain be prepared as a separate file in a tab-delimited format. Appendix II provides examples of data formatted for a nonclinical data submission.

APPENDIX I: EXPERIMENTAL SUMMARY TABLE (*EXPSUMTABLE*)

The ExpSumTable summarizes key experimental parameters investigated in a microarray study. The first three columns are required. The first two columns provide the subject ID (e.g., animal ID) and Array ID respectively. The microarray raw data file is specified in the third column. The remaining columns provide the key experimental parameters that could be used to group array data for analysis. Sponsors should consider including parameters in the ExpSumTable useful in data analysis.

SubID	ArrayID	File	dose(ppk)	Tissue	Chemical	...
1	Ctl_1	Ctl_1.cel	0	Liver	Corn Oil	...
2	Ctl_2	Ctl_2.cel	0	Liver	Corn Oil	...
3	Ctl_3	Ctl_3.cel	0	Liver	Corn Oil	...
4	Ctl_4	Ctl_4.cel	0	Liver	Corn Oil	...
5	Ctl_5	Ctl_5.cel	0	Liver	Corn Oil	...
6	Ctl_6	Ctl_6.cel	0	Liver	Corn Oil	...
7	Ctl_7	Ctl_7.cel	0	Liver	Corn Oil	...
8	Ctl_8	Ctl_8.cel	0	Liver	Corn Oil	...
9	Ctl_9	Ctl_9.cel	0	Liver	Corn Oil	...
10	Ctl_10	Ctl_10.cel	0	Liver	Corn Oil	...
11	Ctl_11	Ctl_11.cel	0	Liver	Corn Oil	...
12	Treat_1	Treat_1.cel	10	Liver	Cmpd_1	...
13	Treat_2	Treat_2.cel	50	Liver	Cmpd_1	...
14	Treat_3	Treat_3.cel	100	Liver	Cmpd_1	...
15	Treat_4	Treat_4.cel	10	Liver	Cmpd_2	...
16	Treat_5	Treat_5.cel	50	Liver	Cmpd_2	...
17	Treat_6	Treat_6.cel	100	Liver	Cmpd_2	...
18	Treat_7	Treat_7.cel	10	Liver	Cmpd_3	...
19	Treat_8	Treat_8.cel	50	Liver	Cmpd_3	...
20	Treat_9	Treat_9.cel	100	Liver	Cmpd_3	...
21	Treat_10	Treat_10.cel	10	Liver	Cmpd_4	...
22	Treat_11	Treat_11.cel	50	Liver	Cmpd_4	...
23	Treat_12	Treat_12.cel	100	Liver	Cmpd_4	...

740 **APPENDIX II: EXAMPLE—SUBMITTING NONCLINICAL STUDY DATA**
741
742 The preparation of nonclinical study data included in a genomic data submission is illustrated
743 through the hypothetical example below. You can find more details on data preparation in the
744 SEND format at: http://www.cdisc.org/models/send/v2.3/SENDV2.3ImplementationGuide.pdf.
745
746 The objective of the example experiment is to identify gene expression patterns that might be
747 related to liver toxicity. Ten rats were used in the study, five for control and five dosed by oral
748 gavage with Drug X in a 6-day repeated-dose experiment. Microarray gene expression and
749 clinical pathology data were reported for each rat in the study. For the genomic data submission,
750 domains 1-6 are required. Refer to the SEND implementation guide noted above regarding
751 which domains apply to the study. It is important to use a short name starting with the two-letter
752 domain code for the column names (variables).

753
754 Domain 1: *Study Design Summary*
755

SSPARMCD	SSPAR	SSVAL	SSSEQ
STTYP	Study Type	Repeat Dose Toxicity	1
LBNAM	Laboratory Name	Company XYZ	2
LBLOC	Laboratory Location	City, State	3
SPECIES	Species	Rat	4
STRAIN	Strain	Sprague-Dawley	5
DESIGN	Study Design	Parallel	6
TRMSAC	Terminal Sacrifice Period	1-6 days	7
GLPTYP	GLP Type	FDA	8
QARPT	QA Report	Yes	9
DURDOS	Duration of Dosing	6 days	10
STTITL	Study Title	6-Day Oral Toxicity Study in Male Sprague-Dawley Rats treated with a drug	11
ALTSTDID	Alternate Study Id	Submission ID 123456	12
SENDVER	SEND Version	2.3	13
STDTC	In-Life Start Date	5/1/2001	14
ENDTC	In-Life End Date	7/1/2001	15

756
757
758 Domain 2: *Subject Characteristics*
759

USUBJID	ARMCD	SCTESTCD	SCORRES	SCTEST	SCSTRESC	SCSEQ
1	1	SEX	Male	Sex	Male	1
2	1	SEX	Male	Sex	Male	2
3	1	SEX	Male	Sex	Male	3
4	1	SEX	Male	Sex	Male	4
5	1	SEX	Male	Sex	Male	5
6	2	SEX	Male	Sex	Male	6
7	2	SEX	Male	Sex	Male	7
8	2	SEX	Male	Sex	Male	8
9	2	SEX	Male	Sex	Male	9
10	2	SEX	Male	Sex	Male	10

Domain 3: *Group Characteristics*

ARMCD	GCTESTCD	GCORRES	GCTEST	GCSTRESC	GCSEQ
1	GRPNAM	Low-dose	Group Name	Low-dose	1
1	CTLGRPDFL	N	Control Group Flag	N	2
2	GRPNAM	Control	Group Name	Control	3
2	CTLGRPDFL	Y	Control Group Flag	Y	4

Domain 4: *Exposure*

USUBJID	ARMCD	EXTRT	EXTRTV	EXDOSE	EXDOSU	EXDOSFRQ	EXDOSFRM	EXDOSTOT	EXROUTE	EXDUR	EXGRPID	EXSEQ	STDY*	ENDY*
1	1	X	Labrafil	10	mg/kg	once daily	liquid	10	oral	P6D	1	1	1	6
2	1	X	Labrafil	10	mg/kg	once daily	liquid	10	oral	P6D	1	2	1	6
3	1	X	Labrafil	10	mg/kg	once daily	liquid	10	oral	P6D	1	3	1	6
4	1	X	Labrafil	10	mg/kg	once daily	liquid	10	oral	P6D	1	4	1	6
5	1	X	Labrafil	10	mg/kg	once daily	liquid	10	oral	P6D	1	5	1	6
6	2	Vehicle	Labrafil	0	mg/kg	once daily	liquid	0	oral	P6D	2	6	1	6
7	2	Vehicle	Labrafil	0	mg/kg	once daily	liquid	0	oral	P6D	2	7	1	6
8	2	Vehicle	Labrafil	0	mg/kg	once daily	liquid	0	oral	P6D	2	8	1	6
9	2	Vehicle	Labrafil	0	mg/kg	once daily	liquid	0	oral	P6D	2	9	1	6
10	2	Vehicle	Labrafil	0	mg/kg	once daily	liquid	0	oral	P6D	2	10	1	6

*General SDTM timing fields, always permissible (see 2.2.5 of the SDTM document at http://www.fda.gov/cder/regulatory/ersr/Studydata-v1.1.pdf)

Domain 5: *Clinical Pathology*

USUBJID	CPTESTCD	CPSTRESN	CPSTRESU	CPSPEC	CPCAT	CPSCAT	CPTEST	CPSTRESC	CPORRES	CPORRESU	CPGRPID	CPSEQ	DY*
1	MONO	0.484	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.484	0.484	10E+9/L	1	1	6
2	MONO	0.418	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.418	0.418	10E+9/L	1	2	6
3	MONO	0.429	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.429	0.429	10E+9/L	1	3	6
4	MONO	0.447	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.447	0.447	10E+9/L	1	4	6
5	MONO	0.471	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.471	0.471	10E+9/L	1	5	6
6	MONO	0.441	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.441	0.441	10E+9/L	2	6	6
7	MONO	0.407	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.407	0.407	10E+9/L	2	7	6
8	MONO	0.448	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.448	0.448	10E+9/L	2	8	6
9	MONO	0.408	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.408	0.408	10E+9/L	2	9	6
10	MONO	0.418	10E+9/L	Blood	HEM	Chemical Analysis	Monocyte count	0.418	0.418	10E+9/L	2	10	6
1	ALT	57	U/L	Blood	CHEM	Chemical Analysis	ALT	57	57	U/L	1	11	6
2	ALT	44	U/L	Blood	CHEM	Chemical Analysis	ALT	44	44	U/L	1	12	6
3	ALT	42	U/L	Blood	CHEM	Chemical Analysis	ALT	42	42	U/L	1	13	6
4	ALT	39	U/L	Blood	CHEM	Chemical Analysis	ALT	39	39	U/L	1	14	6
5	ALT	45	U/L	Blood	CHEM	Chemical Analysis	ALT	45	45	U/L	1	15	6
6	ALT	39	U/L	Blood	CHEM	Chemical Analysis	ALT	39	39	U/L	2	16	6
7	ALT	40	U/L	Blood	CHEM	Chemical Analysis	ALT	40	40	U/L	2	17	6
8	ALT	40	U/L	Blood	CHEM	Chemical Analysis	ALT	40	40	U/L	2	18	6
9	ALT	39	U/L	Blood	CHEM	Chemical Analysis	ALT	39	39	U/L	2	19	6
10	ALT	38	U/L	Blood	CHEM	Chemical Analysis	ALT	38	38	U/L	2	20	6

* General SDTM timing fields, always permissible (see 2.2.5 of the SDTM document at http://www.fda.gov/cder/regulatory/ersr/Studydata-v1.1.pdf)

Domain 6: *Microscopic Findings*

USUBJID	MITESTCD	MITEST	MIORRES	MISTAT	MIREASND	MIGRPID	MISEQ	DY
1	LIVER	Liver	Normal			1	1	6
2	LIVER	Liver	Normal			1	2	6
3	LIVER	Liver	Normal			1	3	6
4	LIVER	Liver	Normal			1	4	6
5	LIVER	Liver	Normal			1	5	6
6	LIVER	Liver	Mild periportal vacuolation			2	6	6
7	LIVER	Liver	Mild periportal vacuolation			2	7	6
8	LIVER	Liver	Mild periportal vacuolation			2	8	6
9	LIVER	Liver	Mild periportal vacuolation			2	9	6
10	LIVER	Liver	Mild periportal vacuolation			2	10	6