7. CONCLUSIONS

This chapter describes the general approach for systematically evaluating genomic data for risk assessment. This general approach is a result of refining the DBP case-study approach (see Figure 3-1). In addition, conclusions from the DBP case study, recommendations, research needs, and future considerations for applying genomic data to risk assessment are described.

7.1. APPROACH FOR EVALUATING TOXICOGENOMIC DATA IN CHEMICAL ASSESSMENTS

There were two goals of this project (see Chapter 2):

- *Develop a systematic approach that allows the risk assessor to utilize the available toxicogenomic data in chemical-specific health risk assessments performed.*
- *Perform a case study to illustrate the approach.*

The first goal was to develop an approach for evaluating toxicogenomic data in future assessments. In the DBP case study, we had the benefit of the 2006 external peer-review draft IRIS Tox Review of DBP, including data summaries and gaps. Additionally, DBP has a more extensive toxicological and toxicogenomic database than most chemicals. The DBP published literature and the draft Tox Review provided a focus to the case study on one set of endpoints (the male reproductive developmental endpoints), that occur in the lower dose range. The case-study approach (see Figure 3-1) needed refinement because the case-study chemical and process had some differences from that of a new assessment. A generalized approach (Figure 7-1) was developed for use in future chemical assessments.

The steps of the approach are

- *STEP 1: Compile the available epidemiologic, animal toxicology, toxicogenomic, and other studies.*

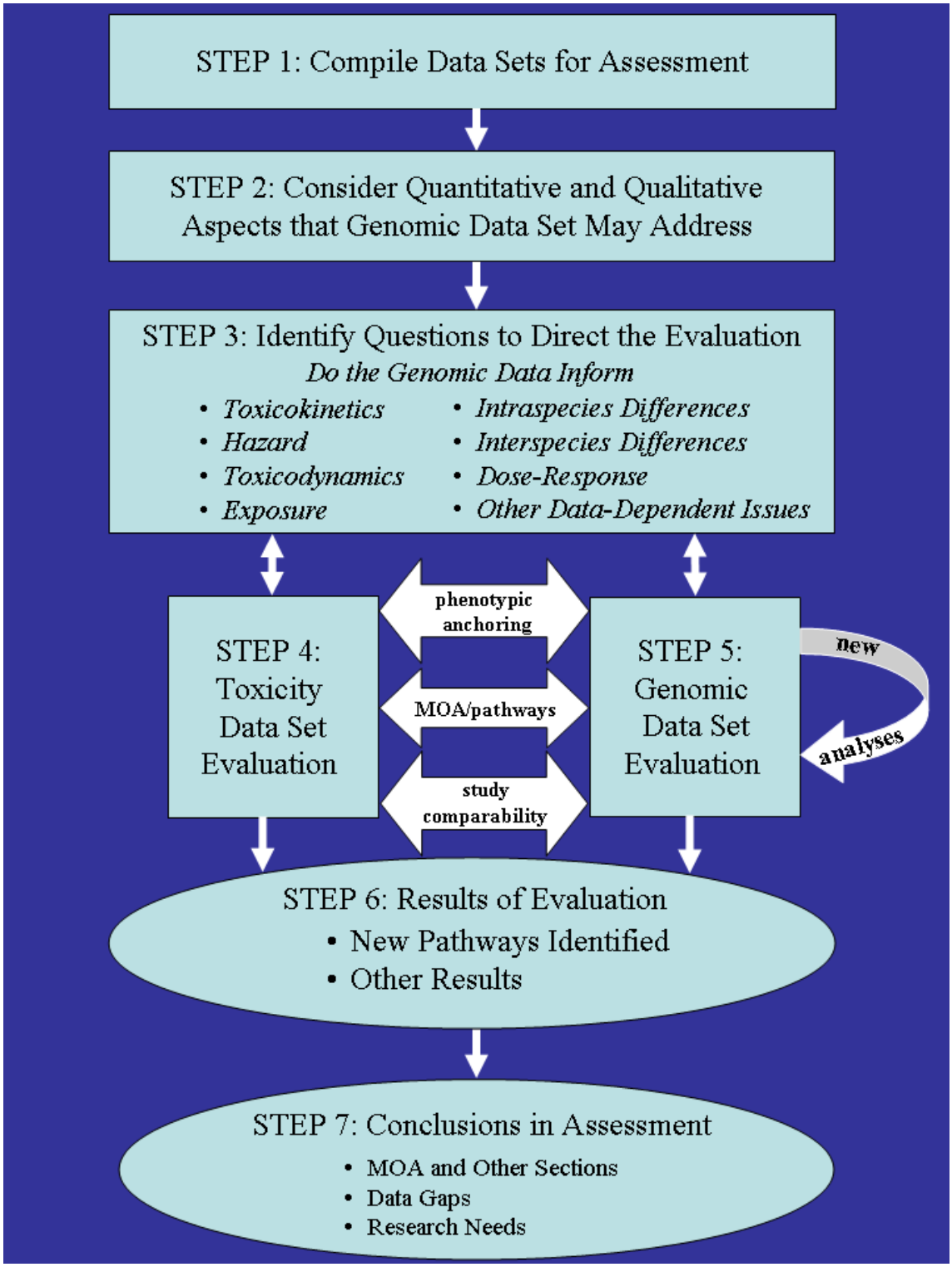

Figure 7-1. Approach for evaluating and incorporating genomic data into future chemical assessments. "Toxicity Data Set Evaluation" may include evaluation of animal toxicity data and/or human outcome data, depending on the available data for the chemical.

- *STEP 2: Consider the quantitative and qualitative aspects of the risk assessment that these data may address.*

  A thorough and systematic consideration of the types of information, in light of the available genomic data, will identify the potential utility of the genomic data and whether these data can be used quantitatively or qualitatively (see Section 3.2). The genomic data set is considered in light of whether these data could inform any risk assessment components (e.g., dose-response) and information (e.g., MOA information, interspecies TK differences) useful to risk assessment. The type of information that these data will provide to a risk assessment depends in part on the type of the available genomic studies (e.g., species, organ, design, and method). This step helps focus the genomic data evaluation and ensure that an important application is not overlooked.

- *STEP 3: Formulate questions to direct the toxicogenomic data set evaluation.*

  Questions are formulated that can direct and focus the genomic data evaluation, and thus, improve efficiency. This step is similar to a scoping exercise performed in ecological and cumulative risk assessment. Some examples of questions considered in the DBP case study were: Do the data inform the MOAs for multiple outcomes (e.g., male and female reproductive outcomes)? Do the data inform dose-response? For example, if microarray data are available, then one of the questions will likely include whether the genomic data can inform the mechanisms and/or MOAs for the chemical as microarray studies typically inform the mechanism of action of a chemical. The DBP case study describes some examples and considerations for determining the risk assessment components that may be informed by a particular genomic data set (see Section 3.3).

- *STEPS 4 and 5: Evaluate the toxicity and/or human outcome and genomic data sets.*

  The approach includes an integrated assessment of the toxicogenomic and toxicity data set to relate the affected endpoints (identified in the toxicity data set evaluation) to the pathways (identified in the toxicogenomic data set evaluation) as a method for

  Phenotypic Anchoring

  Determining the level of support for phenotypic anchoring of genomic changes to *in vivo* outcomes is critical for appropriate interpretation of genomic data for use in risk assessment. In particular, determining whether gene expression changes are associated with or in the causal pathway for an outcome of interest.

  Informing the Mechanisms of Action/MOAs

  Depending on the type of assessment performed, risk assessors may want to utilize aspects of the approach defined herein along with the MOA Framework in the EPA Cancer Guidelines (U.S. EPA, 2005a) and/or other risk assessment decision-logic frameworks for establishing MOAs.

Study Comparability

Another principle of the approach is comparing toxicity and toxicogenomic data study designs in order to identify a set of comparable studies. It is important to compare the study designs among studies. Study design aspects include the time of exposure (in light of critical windows), dose, species, strain, and time of assessment. As a result of assessing study comparability for a given data set, one can select studies for the best comparisons across the outcome and toxicogenomic genomic data sets. For example, in the DBP case study, all toxicogenomic studies were performed in the rat, and, in most cases, the testis. Therefore, the genomic data set was compared with the rat toxicity data and focused on effects in the testis. Broadening beyond the DBP example, the available toxicogenomic data are best considered in light of the toxicity or epidemiologic study data that share study design similarities with the toxicogenomic data. For example, if toxicogenomic data from human tissue or cells are available, then these data are best considered with the human epidemiologic outcome data for the chemical. However, even in the absence of comparable data in the same species, the genomic data may still be used, but with less confidence. See Chapters 4 and 5 for further details of the DBP case-study toxicity and toxicogenomic data set evaluations.

Chapter 5 includes a number of simple methods for assessing the consistency across the toxicogenomic studies. Venn diagrams have been used for illustrating the similarities and differences of DEG findings across genomic studies (see Figure 5-1). Figure 5-2 provides an excellent example of another method for assessing the consistency of findings across all types of gene expression data.

New Analyses

New analyses of toxicogenomic data may be valuable for the assessment depending on the questions asked and the nature of the analyses presented in the published studies. However, new analyses of the original data are not always needed. For instance, reanalysis may not be needed when available published data have been analyzed for application to risk assessment questions of interest. See Section 5.5 for more details of the new case study analyses methods and results, and Chapter 6 for exploratory methods development.

- *STEP 6: Describe results of evaluations and analyses to answer the questions posed in Step 3.*
- *STEP 7: Summarize the conclusions of the evaluation in the assessment.*

7.2. DBP CASE-STUDY FINDINGS

The second goal of the project was to develop a case study. The case-study findings are summarized below and the details of the case-study evaluation and analyses are presented in

Chapters 4−6 (with supplemental material in Appendices A and B). Three advantages to using DBP as the case-study chemical are as follows:

1. The temporal aspects (e.g., time of dosing and time of evaluation) could be considered because a number of well-designed studies exist.

2. A causal connection (i.e., a high degree of phenotypic anchoring) between gene expression changes for some of the steroidogenesis pathway genes with a number of the male reproductive developmental effects has been well-established.

3. Two well-established MOAs for DBP have been defined at the molecular level. A number of endpoints resulting from *in utero* DBP exposure have MOAs that have not been identified or established, thus allowing for a query of the genomic data for possible additional MOAs.

7.2.1. MOA Case Study Question: Do the DBP Genomic Data Inform Mechanism(s) of Action and MOA(s)?

In the DBP case study, we found that toxicogenomic data did inform the TD steps of the mechanisms of action and MOAs. The available genomic and other gene expression data, hormone measurement data, and toxicity data for DBP were instrumental in establishing two of its MOAs: (1) a decrease in fetal testicular T, and (2) a decrease in *Insl3* expression. A decrease in fetal testicular T is the MOA responsible for a number of the male reproductive developmental effects in the rat. The genomic and other gene expression data identified changes in genes involved in steroidogenesis and cholesterol transport, providing evidence for the underlying basis for the observed decrease in fetal testicular T. A decrease in *Insl3* expression is one of the two MOAs responsible for the undescended testis effect, and this MOA is well-established from the results of RT-PCR and *in vivo* toxicology studies. RT-PCR studies identified reduced *Insl3* expression after *in utero* DBP exposure (Wilson et al., 2004) as an MOA for agenesis or abnormalities in the gubernaculum, effects that are not seen after exposure to anti-androgens (i.e., chemicals that affect T synthesis or activity). These results provided support for the *Insl3* MOA as a second well-established MOA for the male reproductive developmental effects of DBP.

The rodent reproductive developmental toxicity data set is robust, having a high quantity and relatively high quality of studies. Additionally, there are a number of rodent toxicity studies that used similar study designs (e.g., dose, species, strain, timing of exposure) as some of the

toxicogenomic studies. This aspect of the DBP data set is exceptional for the case study, allowing for the establishment of the relationship between dose, pathways, and outcomes. We evaluated the rodent reproductive toxicity data set for low incidence and low-dose findings but due to data limitations (see Chapter 4), no new findings were identified. We also evaluated the male reproductive developmental toxicity data set for effects that currently do not have a well-established MOA (see Chapter 4). The testes outcomes were the focus of the case study because the DBP toxicogenomic studies were all performed on testicular tissue. Five effects in the testes effects associated with DBP exposure that do not have well-defined MOAs were identified in this evaluation.

The toxicogenomic and other gene expression studies, including nine published RT-PCR and microarray studies in the rat after *in utero* DBP exposure (Plummer et al., 2007; Bowman et al., 2005; Liu et al., 2005; Thompson et al., 2005; Lehmann et al., 2004; Thompson et al., 2004; Wilson et al., 2004; Barlow et al., 2003; Shultz et al., 2001), were evaluated. The review of the toxicogenomic data set focused on an evaluation of the consistency of findings from the published studies, both across microarray studies and all gene expression data, and on whether any additional pathways may illuminate the unexplained endpoints. The evaluation of the published literature found that the gene level findings from the DBP genomic studies (i.e., microarray, RT-PCR, and protein expression) were highly consistent in both the identification of DEGs and their direction of effect.

New analyses of the Liu et al. (2005) microarray study were performed because we were interested in performing a complete pathway analysis of these data (which had not been the purpose of the published study). These evaluations (see Chapter 5) indicate that there are a number of pathways affected after *in utero* DBP exposure; some of these pathways are related to new MOAs that are distinct from the reduced fetal testicular T or the *Insl3* signaling MOAs. The Liu et al. (2005) DBP data set was reanalyzed using two different methods, the SNR and REM, both using a statistical cut-off of $p < 0.05$. Each method identified the steroidogenesis and cholesterol transport pathways, corroborating prior study conclusions. Each analysis also identified putative new pathways and processes that are not associated with either *Insl3* or steroidogenesis pathways; some were similar across analytical methods and some were different. The common pathways identified between the two methods (see Table 6-3) fall into eight processes (characterized by Ingenuity): cell signaling, growth and differentiation, metabolism,

transcription, immune response, cell adhesion, hormones, and disease. Among these, 54 putative new pathways that are not related to the two known MOAs, reduced T or *Insl3* expression, were identified. Further, a subset of pathways (e.g., WNT signaling and cytoskeleton remodeling) was identified in our analysis that had not previously been identified in the published literature for DBP. One or more of these putative new pathways may be related to the toxicity endpoints without identified MOAs in the rat testes, but additional hypothesis testing studies are needed. Evaluating the genomic and toxicity data sets together provided information on potential, heretofore unexplored, MOAs.

There are a number of possible reasons for the differences in findings between our reanalysis and the published analysis of the Liu et al. (2005) data. These include but are not limited to

- The analyses had different purposes. Liu et al. (2005) was interested in determining whether there is a developmental phthalate genomic signature. The purpose of our analysis was to identify all affected pathways.
- In the four years since the study was published, gene and pathway annotation has increased.

Repeated identification of DEGs and pathways via different analysis methods provides an additional level of confidence regarding the importance of "common" DEGs and pathways. However, it is important to note that the lack of repeated identification of a gene or pathway does not necessarily indicate a lack of biological importance for these genes or pathways.

We also asked whether there were appropriate data to develop a temporal gene network model, a sequence of the gene and pathway interactions over time, for DBP. Using the data from Thompson et al. (2005), the only time-course study available at the time of the project, changes in gene expression and pathways were modeled (see Figure 6-5). Two limitations of these data are that (1) the exposure interval was at the tail end of the critical window of exposure, GD 18, a time that most consider too late to induce the full spectrum of male reproductive developmental effects; and (2) the duration of exposure and developmental time were not aligned because all animals were sacrificed on GD 19 (i.e., the 1 hr time point was the *latest* in development; see Sections 5.3.1 and 6.2.2 for more discussion). The more recent study of Plummer et al. (2007) could provide more appropriate data for building a temporal and spatial network model as both

time-course of exposure over the critical window of development and microdissection of the testis cell types were employed in their study.

7.2.2. Interspecies MOA Case Study Question: Do the DBP Genomic Data Inform Interspecies Differences in TD?

Human gene expression data are not available for DBP. Therefore, the case study used information on interspecies similarities of the affected pathways from other available data and methodologies. We explored the interspecies (rat-to-human) differences in the TD part of MOA, focusing on the steroidogenesis pathway underlying one of the DBP MOAs, the decrease in fetal testicular T MOA. Comparisons of the steroidogenesis genes and pathway were performed to evaluate cross-species similarity metrics using three approaches: (1) protein sequence similarity; (2) pathway network similarities; and (3) promoter-region conservation (see Chapter 6). Results from all three approaches indicate that steroidogenesis pathways are relatively highly conserved across rats and humans and, thus, qualitatively, the rat and human mechanisms for steroidogenesis are highly similar.

These results further corroborate what is known about the similar roles for androgens during normal male development in both rats and humans. However, the data sources used for all three approaches have gaps in the knowledge bases. The pathway network diagramming data source is not of high enough quality or comprehensive enough to utilize quantitatively. In fact, it is difficult to use any of the three new lines of evidence to quantitatively inform the relative sensitivity to DBP across species. It is possible that the small differences across species have a strong penetrance, leading to significant differences in the specific enzymes that may become more sensitive to DBP and thus, affecting T production. We further considered whether some steroidogenesis genes are of higher relative importance and, thus, should be weighted higher in a cross-species assessment of the steroidogenesis pathway. The initiating event for DBP action in the male reproductive developmental outcomes has not been established. However, some information about the rate-limiting steps for steroidogenesis, in the unperturbed scenario, is available (reviewed in Miller, 2008). Some studies have identified CYP11A1 (also called P450SCC) as a limiting enzymatic step for T production (Omura and Morohashi, 1995; Miller, 1988). However, the available information on kinetics reflects the unperturbed state because the rate-limiting step was defined in assays without DBP exposure. Additionally, the rate-limiting step information is limited in scope to steroidogenic enzymes and not all upstream activities

leading to T production, such as STAR, a protein that impacts the availability of cholesterol (by transporting cholesterol to the inner mitochondrial membrane for cleavage by P450SCC) for T production. Thus, there is no *a priori* knowledge to argue for placing more weight on a particular gene leading to T production.

Because there are some questions as to the reliability of the data used to generate the pathway comparisons used for each species and the relative importance of individual steroidogenesis enzymes, there is no basis on which to transform a measure of conservation to a quantitative measure of sensitivity. While the confidence in these cross-species comparisons of the steroidogenesis pathway were not high enough to use the findings quantitatively, for the DBP example, the findings do add to the WOE suggesting that the role of T in male fetal development in rats and humans is well-conserved. These methods, however, when based on high-quality data, could be applied quantitatively to future chemical assessments. Further, the exploratory methods for developing metrics for cross-species pathway similarities described in this document (see Chapter 6) could be developed further and validated in the future for quantitative use in risk assessment.

7.2.3. Application of Genomic Data to Risk Assessment: Exploratory Methods and Preliminary Results

Chapter 6 describes exploratory methods and preliminary results for analyzing genomic data for risk assessment application, developing a DBP gene network model, and measuring cross-species differences for a given pathway.

None of the DBP genomic studies were designed with the application to risk assessment in mind. Methods for analyzing microarray and other -omic data were originally developed for screening purposes (i.e., designed to err on the side of false positives over false negatives). For risk assessment application, genomic analytical tools are needed that are different from those used in screening that can reliably separate signal from noise. In traditional pathway level analysis, first, DEGs are identified by a statistical filter, and second, significant genes are mapped to their respective pathways. Typically, the presence of three affected genes (DEGs) within a pathway is the cut-off for identifying a particular pathway. Depending on the number of genes that map to any given pathway, the role of the pathway can be over- or underestimated. To overcome this problem, we explored using the pathway activity level method (calculating *PALp*) that identifies affected pathways in the single step. This method ranks pathways based on

the expression level of all genes in a given pathway and shows promise for use in risk assessment and further validation is underway.

Gene network models can be very useful for understanding the temporal sequence of critical biological events perturbed after chemical exposure, and thus, useful to a risk assessment. We developed a method for developing a gene network model for DBP based on the available data. The availability of one time-course study (Thompson et al., 2005) enabled our group to model the series of events that occurred between exposure to DBP and the onset of reproductive outcomes. However, given the limitations of the Thompson et al. (2005) study design, we could not determine the genes and pathways affected by DBP exposure earliest in the critical window from this study. However, the exercise allowed us to develop methods for analyzing time-course data for use in gene network modeling.

We also explored the use of three different methods to assess rat-to-human conservation as metrics that may inform the interspecies differences for one MOA, reduced fetal testicular T (Section 7.2.2). More work in the area of cross-species metrics is needed. Efforts to address the challenges in using similarity scores to quantitatively estimate the human relevance of an MOA are ongoing (Section 6.3).

7.2.4. Application of Genomic Data to Risk Assessment: Using Data Quantitatively

This case study was limited to qualitative uses of genomics in risk assessment due to the absence of dose-response, global gene expression studies (i.e., microarray studies) for DBP. EPA and the larger scientific community working with genomics are interested in methods to use genomic data quantitatively in risk assessment. There is one dose-response RT-PCR study that, although not a genomic (i.e., not global) study, was considered for use quantitatively in risk assessment (Lehmann et al., 2004; see Table 7-1). Some strengths of the Lehmann et al. (2004) study include the following:

- The study includes multiple doses ranging from low to high.
- Some of the genes assessed in this study were first identified in microarray studies, providing a relatively high level of confidence in the connection between the expression of some of the genes and particular outcomes, as well as demonstrating reproducibility across studies. For example, findings for *Star* gene expression are reproduced across protein expression, RT-PCR, and microarray studies.

Table 7-1. DBP dose-response progression of statistically significant events illustrated with a subset of precursor event data (steroidogenesis gene expression, T expression) and *in vivo* endpoints with the reduced T MOA

	0.1 mg/kg-d	1 mg/kg-d	10 mg/kg-d	30 mg/kg-d	50 mg/kg-d	80 mg/kg-d	100 mg/kg-d
Precursor event[a]	↓ *Hsd3b*	↓ *Hsd3b* ↓*Scarb1*	NC in gene exp. NC in [T]	ND for gene exp. NC in [T]	↓*Scarb1* ↓*Hsd3b* ↓*Star* ↓*Cyp11a1* ↓[T]	ND for gene exp.	↓*Scarb1* ↓*3β-HSD* ↓*StAR* ↓*P450scc* ↓[T]
In vivo endpoint						↑ incidence of absent, poorly developed, or atrophic testis and underdeveloped or absent epididymis[b]	Retained nipples and areolae[c]

NC, no statistically significant change; ND = not determined (Lehmann et al. [2004] did not test 80 mg/kg-d).

[a]Lehmann et al. (2004).
[b]NTP (1991).
[c]Mylchreest et al. (2000).

However, there are a number of issues in applying these dose-response RT-PCR data with confidence to BMD modeling. These limitations include the following:

- Some of the gene expression changes are not reproducible. For example, *Kit* was observed to be significantly altered in the Lehmann et al. (2004) study but was not observed to be significantly reduced after *in utero* DBP exposure in a microarray study (Liu et al., 2005) utilizing the Affymetrix gene chip, yet *Kit* is on the Affymetrix rat chip.

- The relationship between statistical and biological significance is not known for these gene expression data. For example, the expression of *Hsd3b* mRNA is statistically significantly altered at lower doses than a statistically significant T decrease was observed. Thus, Lehmann et al. (2004) argued that the changes in *Hsd3b* at 0.1 and 1.0 mg/kg-d were not biologically significant. Alternatively, *Hsd3b* gene expression changes could be a precursor to T level changes in time and thus, be a valid precursor event. It is also not known whether changes in the expression of a single or multiple steroidogenesis genes would lead to a significant alteration in T and the phenotype.

- Interlitter variability could not be characterized from the Lehmann et al. (2004) data because the RT-PCR data were collected on five individual pups representing four to five litters per treatment group (i.e., ~1 pup/litter). In order to have appropriate data for BMD modeling, litter mean values calculated from a study with a greater sample size and multiple litters are needed to allow characterization of intra- and interlitter variability. The use of the litter as the statistical unit is generally agreed upon because of the high variability in gene expression for pups within one litter (Barlow et al., 2003).

We concluded that the available dose-response RT-PCR data for DBP are not of sufficient quality due to the lack of information about interspecies variability. Additionally, there is not sufficient knowledge about the biological significance of a gene expression change (and the level of change that is biologically significant), for one or a subset of genes, that would invariably lead to a reduction in T and in turn, lead to the observed male reproductive developmental outcomes.

7.3. LESSONS LEARNED

The lessons learned from the case study are grouped by research needs that are useful to research scientists and recommendations that are useful to risk assessors.

7.3.1. Research Needs

7.3.1.1. *Data Gaps and Research Needs: DBP*

There are some research needs that would be very useful, specifically for a DBP risk assessment including the following:

1. *Develop a gene network model for DBP using the Plummer et al. (2007) data.* This data set would be an excellent source of temporal and spatial gene expression information because one of its studies includes three time intervals, thus covering the entire critical window for male reproductive outcomes, and a second study used microdissection of the cord and interstitial cells of the testis. This study was not modeled because it was not published until after the modeling work had been completed, and we had not obtained the data. By comparing gene expression, Plummer et al. (2007) hypothesized the MOA underlying the gonocyte and LC effects.

2. *Perform microarray studies in male reproductive tissues, other than the testis, affected by DBP in order to understand the similarities and differences in DBP-affected pathways across reproductive organs and tissues in the male rat.* Bowman et al. (2005) performed such a study in the WDs, but studies in other male tissues are needed.

3. *Perform microarray studies in human tissues (either cell lines or from aborted male fetal tissue), along with parallel in vitro and in vivo studies in rats for validation and comparison.* Such data would provide valuable information about interspecies differences in TD sensitivity. Some human studies found an association between *in utero* phthalate exposure and newborn male reproductive developmental measures (Main et al., 2006; Swan et al., 2005) that indicate human relevance for some of the DBP effects observed in male rat studies.

4. *Perform well-designed proteomic and metabolomic studies to understand the effects of in utero DBP exposure on the function of expressed proteins and on cellular metabolites.* These data may provide complementary data to the available transcriptomic data, which could yield some new insights.

5. *Perform genomic studies to identify early, critical, upstream events as a means to identify the initiating event for DBP's action in the testis.* This would require performing studies much earlier in gestation, at the beginning of sexual differentiation. In addition, such studies may require greater sensitivity regarding gene expression change identification because a statistically significant change may be greater than a biologically significant change. If identified, the initiating event could be utilized in the risk assessment, thereby reducing uncertainty.

6. *Perform genomic studies to understand whether the female reproductive tract malformations after DBP exposure have common or different MOAs with the male development reproductive effects.* This line of research would identify pathways affected in the developing female reproductive tracts after early gestational DBP exposure.

7. *Compare the affected DEGs and pathways between the phthalates with and without developmental effects could be useful for a cumulative risk assessment of the developmental phthalates.* All of the data from the Liu et al. (2005) data set could be utilized to evaluate this issue. Further, evaluating consistency of findings across chemicals in the same MOA class that do and do not produce the same set of effects could be useful for improving specificity of the pathway and MOA findings for DBP.

8. Studies to distinguish affected genes and pathways that may be compensatory vs. those in the causal pathway for DBP-toxicity.

7.3.1.2. *Research Needs for Toxicity and Toxicogenomic Studies for Use in Risk Assessment*

EPA and the larger scientific community are interested in methods to use genomic data quantitatively in risk assessment. This case study was limited to qualitative uses of genomics in risk assessment due to the absence of dose-response global gene expression studies for DBP. This is the case for many chemicals as multiple dose studies are very costly. However, multiple dose microarray or other global gene expression studies are needed (see Table 7-2). Such studies need to be designed properly such that the identification and interpretation of lower dose effects is possible. Gene expression changes at the lower dose may not be affected in every organ, tissue or cell sample assessed. High single dose microarray studies have been performed such that all organs are affected and one can assess a smaller sample size than for a dose-response study. In a dose-response study including low- to high doses, the sample size per dose group would need to be high enough to increase statistical power (i.e., the detection of gene expression changes when only a few animals are affected). For example, if an endpoint is affected in 20% of the animals at lower doses, then the sample size for microarray studies must be large enough to identify the affected animals (with affected gene expression). Perhaps the highest priority study is one that assesses global gene expression and toxicity endpoints of interest as components of the same experiment; the organ or tissue of interest would be collected at the appropriate age in one group of animals and a second group would be followed through to evaluation of the endpoint of interest. In this manner, such a study would generate data that could define the relationships between dose, time of exposure, gene expression level changes, pathway level changes, and *in vivo* changes.

Table 7-2 describes some of the priority research needs for toxicogenomic studies for developmentally toxic chemicals, including DBP. First, appropriate time-course gene expression data over the critical window, using a small subset of genes whose altered expression is linked to

Table 7-2. Research needs for toxicogenomic studies to be used in risk assessment

Purpose	Study needed
1) Develop a gene network model	Exposure time-course microarray data.
2) Improve pathway analysis statistical power	Number of replicates increased.
3) Use of toxicogenomic data to inform toxicokinetics in dose-response analysis	Genomic and toxicity studies with same study design: Generate TK data in relevant study (time, dose, tissue), and obtain relevant internal dose measure to derive best internal dose metric.
4) Use of toxicogenomic data in dose-response analysis	Multiple doses in microarray studies in parallel with toxicity studies for phenotypic anchoring.
5) Determining the degree of phenotypic anchoring; informing MOAs (see Figure 3-4)	Similar study design characteristics for genomic and toxicity studies (i.e., dose, timing of exposure, organ/tissue evaluated). This includes assessing whether genes and pathways are due to compensatory mechanisms and/or general toxic responses.
6) Assess intraspecies differences	A study assessing multiple doses across rat strains (e.g., Wistar vs. SD); endpoint and microarray component of the study.
7) Assess interspecies differences	A study to assess whether different species with similar pathways (genes and sequence of steps) have a similar sensitivity to a given chemical. The findings could potentially enhance the utility of toxicogenomic data to aid species extrapolation in risk assessments.
8) Appropriate statistical pathway analysis methods for use in risk assessment	Further comparisons and evaluations of different methods.
9) Screening and categorizing chemicals by MOA in risk assessment (e.g., cumulative risk assessment)	Genomic (transcriptomic, proteomic, and/or metabolomic) signatures can be particularly useful for screening and categorizing chemicals by MOA in risk assessment.

the outcome of interest, would be very relevant for developing a regulatory network model. These studies need to be carefully designed based on the information on the critical window of exposure and the relationship to the particular outcome of concern. Second, the statistical power of pathway-analysis methods for global expression techniques, including microarrays,

proteomics and metabolomics, could be improved by designing and performing studies with more replicates. Thus, variability would be better characterized. Third, it would be helpful to design genomic studies that could inform both TK and dose-response (see Table 7-2, #3 and #4).

Performing genomic and toxicity studies with similar designs would provide useful information. These studies would be designed at the most relevant time of exposure, include low to high doses, and assess the relevant tissues. Relevant internal dose measurements could be obtained on which to base the internal dose metric. These studies, employing genomic and toxicity studies of comparable designs, would provide information about the relationship of dose, gene expression, and outcome, and thus, could potentially be used in dose-response analysis. Studies with both a toxicity and toxicogenomic component would obviously require assessment of a large sample size to be informative. These same studies could be used to inform MOAs (Table 7-2, #5) and could be adapted to comparing species (Table 7-2, #6). Regarding quantitative measures of intraspecies and interspecies differences, it should be noted that the same information which is necessary for quantitative assessment of interspecies differences (Section 7.2.2) may be useful for characterizing intraspecies variability, and vice versa. In particular, factors that explain or predict interstrain differences in rodent sensitivity to DBP, such as those noted between Wistar and SD rats, may be hypothesized to contribute to human variability. Further, toxicologically important interstrain differences identified from the toxicogenomic data could be an excellent data source for investigating whether they are also important for modulating interspecies sensitivity.

Finally, further development and comparison studies to identify appropriate statistical pathway analysis methods for use in risk assessment are needed (Table 7-2, #8). It is important to note that such studies require research funding and laboratories with expertise in both genomics and toxicology. Research needs for toxicity studies that would improve the utility in risk assessment are also described in Table 7-3. As was noted for the DBP case (see Chapter 4), complete reporting is necessary for studies that are intended for use in risk assessment.

7.3.2. Recommendations

Based on the lessons learned from performing the DBP case-study exercise, we developed some recommendations or best practices for evaluating genomic data in new assessments. The approach includes systematic consideration of

Table 7-3. Research needs for toxicity studies for utilizing toxicogenomic and toxicity data together in risk assessment

Study aspect	Research need
Study design	Exposing animals during optimal developmental stage/time (i.e., for the critical window).
	Assessing outcome at optimum developmental stage/time for that outcome.
	Parallel study design characteristics with toxicogenomic studies (i.e., dose, timing of exposure, organ/tissue evaluated) to obtain comparable toxicity and toxicogenomic studies to aid in determining relationships between gene expression changes and outcomes.
Reporting	Individual animal data to aid identification of low incidence effects, correlate gene expression changes and outcomes, and characterize intraspecies variability.
	All endpoints that were evaluated (independent of whether the outcome was positive or negative).

the genomic data for whether they could inform risk assessment steps, identification of questions to direct the evaluation, and evaluation of the genomic data and toxicity data to assess phenotypic anchoring. In addition, we have some specific recommendations. The first two recommendations are straightforward and could reasonably be performed by a risk assessor with basic training in genomics data evaluation and interpretation while the third recommendation requires expertise in genomic data analysis methods for implementation. The recommendations are presented below:

1. *Evaluate the genomic and other gene expression data for consistency of findings across studies to provide a WOE evaluation of the affected gene expression and pathways.* Some simple methods, such as using Venn diagrams and gene expression compilation approaches, can be applied to risk assessment. When evaluating the consistency of toxicogenomic data findings, it is advantageous to include all available gene expression data (single gene, global gene expression, protein, RNA) because single gene expression

techniques have been traditionally used to confirm the results of global gene expression studies and because single gene expression data add to the database.

2. *Perform BMD modeling on high-quality RT-PCR dose-response studies of genes known to be in the causal pathway of an MOA or outcome of interest.* Obtaining a BMD and BMDL is a useful starting point for both linear low-dose extrapolation and reference value approaches. We are not indicating which approach is appropriate to take for making predictions about the potential risk below the BMD or BMDL. "High quality" is defined in this context as a well-conducted study that assessed enough animals and litters for sufficient statistical power for characterizing the mean responses and the variability (interlitter and intralitter).

3. *Perform new analysis of toxicogenomic data in cases when the new analysis is likely to yield new information that would be useful to the risk assessment. Examples include:*

    - Perform a new pathway analysis in order to identify all affected pathways or other risk assessment applications. When the available published microarray studies have been conducted for purposes (e.g., basic science, pharmaceutical development) other than risk assessment, it may be useful to reanalyze the data for risk assessment purposes. Information about all affected pathways may contribute to an understanding of the mechanisms and MOAs.

    - Identify the genes and pathways affected over a critical window of exposure if global gene expression time-course data are available. Specifically, by developing a gene network over time, it may be possible to identify the earliest affected genes and/or pathways, which in turn may represent the earlier or initiating events for the outcome of interest.

7.3.3. Application of Genomic Data to Risk Assessment: Future Considerations

A number of the issues that emerged in evaluating the DBP genomic data set are relevant to using genomic data in risk assessment in general. Some issues regarding the use of genomic data are to the same as for the use of precursor information in risk assessment, regardless of the technique used to gather the information. Two outstanding questions are

- *How is the biologically significant level of change in a precursor marker determined?* And, specifically for toxicogenomic data, *what are the key genes (i.e., a key gene, a handful of genes associated with the outcome of interest, a genomic signature) whose altered expression leads to an adverse outcome?* Currently, decisions about the degree of change of a precursor event tend to be based on statistical significance because data to address biological significance are typically lacking (as is the case for T levels and male development of the testis). Genes are identified as DEGs in microarray studies based on statistical-significance criteria that may not reflect biological significant changes (i.e., identified genes may not be biologically meaningful while unidentified genes may be

meaningful). This point is also relevant to the question: *What pathway analysis methods are most appropriate for risk assessment?* As noted in Section 5.5, it is difficult to know whether one has identified the biologically relevant DEGs and pathways. Statistically significant changes and repeated findings of the same genes and pathways across studies and using different analytical methods, while providing corroboration, do not necessarily provide a greater confidence regarding biological significance of these genes and pathways over other genes and pathways. Further, there is a bias towards the well-annotated genes as biologically significant when, in fact, the unannotated genes could be of greater importance.

- *What are the requirements for linkage of precursor events to in vivo endpoints?* Studies to assess the relationship between the gene expression and outcomes are needed to establish a causal connection. It is important to note that DBP has two well-established MOAs and strong phenotypic anchoring of some gene expression changes, which is not typical.

There are also a number of technical issues in utilizing microarray data in EPA risk assessments that have not fully been surmounted. The primary technical issue is the validation of the reproducibility of microarray study results. Reproducibility depends on biological sample preparation, interlaboratory (presumably related to operator and protocol differences), intralaboratory (presumably related to operator differences), and batch and platform variability. The results of the MAQC-I project (see Chapters 2 and 5) revealed that reproducibility was achieved when using the same biological sample. This is very encouraging for using microarray data in risk assessment. However, biological sample variability still needs to be addressed in order that protocols and details of the underlying reasons for the variability can be understood. MACQ-II and III projects are underway to address additional technical issues (see Chapter 2).

A number of the issues stem from the complexity of the data output from the global expression techniques (e.g., microarrays, proteomics, and metabolomics). This is in part a training issue. To address the training needs, the EPA Risk Assessment Forum held introductory and intermediate level training in genomics in 2007, and the FDA held genomics training (http://www.fda.gov/cder/genomics/Default.htm). However, it would be advantageous for organizations that perform risk assessments to embark on further training of risk assessors to enable them to perform analyses of microarray and other genomic data analysis techniques, and to understand the issues in applying traditional analytical methods to risk assessment.

If additional case studies are performed using the approach outlined in Figure 7-1, then we recommend a chemical whose exposure leads to both cancer and noncancer outcomes to

explore use of these data for multiple outcomes, as well as the impacts on the different risk assessment paradigms and processes (e.g., cancer vs. noncancer). Further, performing case studies on data-rich and data-poor chemicals would aid in further evaluating the approach described herein. For instance, performing a case study on a chemical with dose-response data and on a chemical with human polymorphism data would address issues in evaluating these types of data for risk assessment, allowing further refinement of the approach.

The approach for utilizing toxicogenomic data in risk assessment outlined in this document may be applied to other chemical assessments. This document advances the effort to devise strategies for using genomic data in risk assessment by defining an approach, performing a case study, and defining critical issues that need to be addressed to better utilize these data in risk assessment. This case study serves as an example of the considerations and methods for using genomic data in future risk assessments.