Published in final edited form as:

Reprod Toxicol. 2014 June ; 45: 94–104. doi:10.1016/j.reprotox.2013.12.008.

Gene expression in teratogenic exposures: a new approach to understanding individual risk

Claudia Kappen¹ and J. Michael Salbaum²

¹Department of Developmental Biology, Pennington Biomedical Research Center, Louisiana State University System, 6400 Perkins Road, Baton Rouge, LA 70808

²Laboratory of Regulation of Gene Expression, Pennington Biomedical Research Center, Louisiana State University System, 6400 Perkins Road, Baton Rouge, LA 70808

Abstract

The phenomenon of partial or incomplete penetrance is common to many paradigms of exposure to teratogens, where only some of the exposed individuals exhibit developmental defects. We here argue that the most widely used experimental approaches in reproductive toxicology do not take partial penetrance into account, and are thus likely to miss differences between affected and unaffected individuals that contribute to susceptibility for teratogenesis. We propose that focus on the variation between exposed individuals could help to discover factors that may play a causative role for abnormal developmental processes that occur with incomplete penetrance.

The problem

Agents with developmental toxicity often cause defects or anomalies only in a fraction of the exposed individuals. Dose, time and duration of exposure, as well as biological features of the affected tissues themselves are all thought to influence developmental outcomes, such as the risk for neural tube defects [1] or long-term adverse health outcomes, such as metabolic syndrome and cardiovascular disease [2]. Susceptibility of an individual to the environmental exposure is generally believed to be determined by pre-existing genetic factors [3]. However, even in genetically identical animals, such as highly inbred strains, not all animals respond in identical fashion in many exposure paradigms. In fact, often only the minority of the exposed individuals are affected, while a large group of animals with the same exposure develop or function normally. Thus, the phenotypic outcome presents with what in genetics terms would be called "incomplete penetrance" or "partial penetrance".

Examples for phenotypes of incomplete penetrance after teratogen exposure are the heart defects and neural tube defects that occur in diabetic pregnancies in inbred mice. Although

© 2014 Elsevier Inc. All rights reserved.

Corresponding author: Claudia Kappen, Dr. rer. nat., Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, LA 70808, Tel: 225-763-2781, Claudia.Kappen@pbrc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

these defects are strikingly more frequent in mice that are hyperglycemic, not all progeny within the same litter exhibit defects, and some of the litters from a group of experimental pregnancies may even be unaffected. Furthermore, maternal diet can affect penetrance of neural tube defects in embryos that develop in diabetic FVB females [4]. Genetic factors also appear to play a role in penetrance, as in diabetic pregnancies in the inbred C57BL/6 strain embryonic defects are less frequent than in the inbred strain FVB [5], and considerably more frequent in the inbred non-obese diabetic (NOD) strain [6] (and Salbaum et al., unpublished results), an inbred mouse line with spontaneous diabetes [7]. Thus, diabetic pregnancy in inbred strains is an ideal model to dissect genetic and environmental contributions to teratogen-induced partial penetrance of anomalies.

From early genetic studies, it was proposed that partial penetrance of phenotypes reflects a threshold phenomenon [8, 9], based on Grüneberg's statement that developmental defects *"are 'quasi-continuous' characters in the sense that the underlying....basis is a continuous variable (generally not yet identified)....which is divided by a physiological threshold into normal and abnormal animals..."* [9]. In this model, the continuous variable is conceptualized as a normal distribution of data points for all individuals within a population. Individuals at the extreme fail to fulfill the requirements for proper development and exhibit a defect (Figure 1A), which -in wildtype animals- would be a very rare occurrence. Increased risk for adverse outcomes from teratogen or toxicant exposure then results from a shift in the mean of the distribution, increasing the fraction of individuals surpassing the threshold (Figure 1B). Although this model was originally formulated for categorical phenotypes (in Grüneberg's studies skeletal defects were scored as "present" or "absent"), the general considerations also apply to quantitative outcomes, or phenotypes with variable expressivity: then the extremes of the distribution correspond to the "mildest" and "severest" manifestation, or lowest and highest measurements for a given outcome parameter. The paramount goal in molecular reproductive toxicology is to identify the *"not yet identified"* continuous variable(s) that underlie the distribution, and that ultimately determine risk for adverse outcomes.

Many laboratories, including our own, have turned to unbiased approaches, such as microarray-based or sequencing-based gene expression profiling, to identify genes, molecules and pathways that are targeted in the teratogenic exposure, and that might explain detrimental outcomes. The standard approach is to compare profiles from unexposed control animals, or embryos from unexposed pregnant dams, to profiles from experimental animals or embryos that experienced the exposure conditions. However, in paradigms with incomplete penetrance, this approach may be inappropriate on both theoretical and empirical grounds, as we will argue below. Thus, we have to ask ourselves: have we been looking for the right thing ? Have we missed something important ?

The conventional approach to the interpretation of gene expression data is to search for consistent differences between a control group and an experimental group (Figure 1C) that fulfill specific statistical criteria. However, these approaches can only reveal changes that affect all (or most) samples in the experimental group, and thus cannot account for phenotypic outcomes that affect only a (minor) fraction of the individuals, such as in the case of partial penetrance (Figure 1D). Thus, it could be argued that, while changes in all

exposed animals confer some vulnerability relative to a specific outcome, the actual pathogenic triggers are present only in some exposed individuals, those at the extremes of a distribution that manifest with the abnormal phenotype or adverse outcome.

A new alternative

We propose that the traditional concept omitted an important second mechanism that can increase risk for abnormal phenotype: a greater fraction of affected individuals could also result from greater variance in the distribution (Figure 2A). In this scenario, the overall mean remains unchanged, but more individuals fall outside of the threshold. Thus, the net effect is the same as in Figure 1B, placing a larger number of individuals at risk for abnormal phenotype, and the same pattern of phenotypic outcomes would be achieved as in Figure 1D. However, the conventional frameworks of interpretation for gene expression data do not consider variability as an independent parameter. In fact, they seek to minimize it, through pooling of several individual-derived samples [10], or in the statistical framework [11–13]. Consequently, the concept of increasing variability has not been applied at the experimental level.

As increasing variability of expression levels is reflected in a wider distribution curve, the mean of the wider distribution can be similar or different from the control group (Figure 2B). In either scenario it is intuitively obvious that increasing variability can account for incomplete penetrance [14–16] (when the differential gene expression scenario in Figure 1C can not). Furthermore, if variability is increased for multiple parameters, such as two or more genes -in the same animal-, the individual liability for abnormal development would increase. Experimental strategies pursued so far have not considered this at a systems-wide level. Therefore, in contrast to traditional approaches that all seek to minimize variation, we propose that focus on variability of gene expression would uncover new genes, pathways and mechanisms that are involved in abnormal development.

To illustrate how our concept impacts gene discovery and interpretation of gene expression profiling data, we offer two experimental examples: microarray-based genome-wide expression surveys from our chemical model of type I diabetes induction by injection of Streptozotocin in the FVB mouse strain (as published previously [17]), and from the non-obese diabetic mouse strain [6], in which females become diabetic spontaneously [7]. In both models, females with confirmed hyperglycemia were mated to normal males, embryos were isolated at E10.5, and phenotypically classified according to presence or absence of a neural tube defect. Controls came from normal pregnancies of non-diabetic females of the respective strains. RNA was extracted and samples were processed separately from each individual. In the chemical diabetes model (after injection of STZ into FVB females), we obtained genome-wide profiles from two controls and five diabetes-exposed embryos, of which two had a neural tube defect [17]. The original dataset was reanalyzed (with the latest version of Cyber-T [18]) to identify differentially expressed genes, applying conventional criteria of "signal present in all samples" and "statistically significant p-value" (<0.05 , after correction for multiple comparisons); this approach identified 338 differentially expressed gene probes (for details, see Methods). We then applied standard hierarchical clustering. As displayed in Figure 3A, this groups the two samples from normal pregnancies together on

one branch (green frame), different from the branch that contains the exposed individuals (red frame). However, within the exposed group, there is no distinction between embryos with (E4-NTD, E5-NTD) and without (E1-E3) a neural tube defect. Thus, the differentially expressed genes do not differ between affected and unaffected individuals within the experimental group. A parallel situation is encountered in a comparison, under identical criteria, of embryos from diabetic and non-diabetic NOD females: 705 gene probes were found differentially expressed (Figure 3B), with the 4 embryos from non-diabetic pregnancies clustering together on a branch (green frame) separate from the 8 samples from diabetes-exposed embryos (red frame); again, the 4 individuals with neural tube defects (E5-NTD, E6-NTD, E7-NTD, E8-NTD) are clustered intermingled within the diabetes-exposed group. Again, differentially expressed genes do not distinguish affected from unaffected individuals. [We know that this is independent of profiling technology since the NOD arrays were performed on the Illumina platform, and the FVB arrays on Affymetrix chips].

Because we had previously shown that exposure to maternal diabetes increases overall embryonic variability of gene expression over control conditions [14], we now sought to identify the genes that display high variation of gene expression levels within the diabetes-exposed group, and between defective and non-defective embryos. For this, we calculated a z-score for each gene probe of 3 standard deviations above and below the mean expression level for the diabetes-exposed but normally developed individuals; we consider values within this envelope as the "normal range". With this very stringent criterion in place, we then asked whether any individual measurement for a defect-affected embryo fell within or outside of this envelope. In the STZ model, this identified 333 gene probes; for 194 probes, both defective embryos were above the range, and for 139 probes, both individuals were below the range (Figure 4A, B). Analogously, for the NOD model, we detected 549 gene probes with measurements on the same side of the envelope in at least 3 neural tube defect-affected individuals; examples are shown in Figures 4C–E. By random chance, under assumption of a 3-standard deviation envelope, only 0.26% probes would be expected to fall outside of the envelope (which for the Illumina platform with 18256 probes would be predicted to be 47.5 probes). Our results clearly identify a larger than random chance number of genes with extreme gene expression levels in neural tube defect-affected individuals. These most variable gene probes have the greatest likelihood to contribute to the abnormal phenotype because they fall outside of the "normal range" in affected individuals in both experimental models.

In contrast to differentially expressed genes, genes with high variability lead to discrimination of affected and normal individuals in hierarchical cluster analysis (Figure 5). In both models, the exposed but normally developed embryos cluster together in the sample tree, and a separate branch contains the exposed neural tube defect-affected individuals. [This was the case also if we included the profiles for the non-exposed controls, which have been omitted from these Figures for clarity]. Thus, we show here that tree topologies are able to distinguish affected from non-affected individuals if we base our cluster analyses on genes with high variation of expression levels.

Next, we queried the identity of highly variable genes. We previously reported for the STZ paradigm that genes encoding transcription factors and DNA/chromatin-binding molecules

were enriched among the differentially expressed genes (by DAVID analysis [19]). This was confirmed in our re-analysis here (Figure 6A, and Supplemental Figure 1A). In contrast, the highly variable genes are enriched for genes that encode mitochondrial proteins and molecules involved in redox reactions. Genes with metabolic function were significantly enriched ($p < 0.05$; confirmed by IPA analysis; see also Supplemental Figure 1B), suggesting the possibility that embryonic energy utilization may be different between normally developing and neural-tube-defect-affected embryos in diabetic pregnancies. Interestingly, only 32 annotated genes (see Supplemental Table 1) were shared between the 293 unique annotations with differential expression, and the 325 unique annotations with highly variable expression. This highlights that differential expression criteria select mostly genes with low expression variability. Beyond these results, analysis for genes with high variability of expression between individuals reveals additional genes and pathways that are associated with neural tube defects. This conclusion is further strengthened by our results from the NOD paradigm (Figure 6B): The differentially expressed genes in this model are also strongly enriched for genes encoding transcription factors and DNA/chromatin-binding molecules (DAVID and Panther analysis; see Supplemental Figure 1C). In contrast, genes encoding signal transduction molecules were significantly enriched ($p < 0.05$) among the highly variable genes (confirmed by IPA analysis; see also Supplemental Figure 1D). This underscores again that genes with high variability generally participate in different cellular pathways than differentially expressed genes. Between 504 unique annotations for differentially expressed genes and 486 unique annotated highly variable genes in the NOD paradigm, only 31 annotations were shared in common (see Supplemental Table 2), reflecting a scenario similar to the STZ model. Thus, in the NOD model as well, the criterion of high inter-individual variability leads to discovery of new genes that may play a role in neural tube defects. Consistent with this notion is our observation that different known neural tube defect genes (as previously defined [14]) are represented among the differentially expressed and the highly variable genes, in both diabetes- exposure models (Supplemental Table 5).

It is intriguing that in both mouse strains, despite differences in the etiology of maternal diabetes, the embryonic transcriptional response involves genes that encode DNA/ chromatin-binding and transcription factor molecules. Thus, epigenetic regulation of gene expression plays a role in both paradigms. Yet, despite this similarity, the identities of misregulated genes are different between both strains: only 17 genes are common to the differentially expressed groups from both models (Figure 6C, Supplemental Table 3), and 14 to the high variability groups of genes (Figure 6D, Supplemental Table 4). Thus, in each strain, there is a unique response, which could reflect the different modes of diabetes induction, differences in levels of hyperglycemia and other metabolic parameters, differences in pregnancy progression, or differences in genome structure and responsiveness. Consistent in both experimental paradigms, however, is that the exposure to maternal diabetes causes higher variability in gene expression, and that genes with high variability are able to distinguish neural tube defect-affected individuals from their normally developing littermates, while conventionally discovered differentially expressed genes are not successful in discriminating the abnormal from normal phenotypes. Therefore, we

propose that the highly variable genes more suitably mirror, and possibly explain, the partial penetrance of the neural tube defect phenotype.

In this work, we have demonstrated an approach in which extremes of gene expression can be attributed to defective development after teratogen exposure, and by which individuals at risk for abnormal phenotype can be classified via gene expression signatures. In this way, focus on inter-individual variability can discriminate normal from aberrant development. We therefore propose that the concept of variability holds excellent promise for the discovery and investigation of factors that precipitate partial penetrance of developmental abnormalities. Thus, while the past focus on common differences in gene expression was productive in identifying pathways that predispose to susceptibility, it neglected factors that trigger pathogenesis at the individual level. So, our answer to our initial question is: yes, we are now challenged to understand the unique factors that contribute to risk for abnormal development in each affected individual.

Yet, before we can formulate conditions under which this paradigm could have power to predict risk for abnormal development, we need to consider the limitations of the current experimental models and interpretive framework:

Limitations related to the datasets

The sample sizes used here do not allow us to test whether the expression values for a given gene are normally distributed. The statistical models for interpretation of expression profiling datasets typically require this assumption, but rigorous tests with sufficient sample numbers have not been conducted to date. We find that, even with pooling of samples from exposed -but normally developed- embryos in the NOD model, the overall variability of gene expression levels was increased with exposure to maternal diabetes. Thus, at least some of the variability is associated with exposure and not solely with the presence of a developmental defect. This confirms our earlier observations of increased variability in diabetes-exposed normal morphology embryos in the STZ model [14]. Under assumptions of a classical normal distribution, an envelope of 3 standard deviations would include >99.7% of all expected datapoints. Strikingly, measurements from individual NOD embryos with neural tube defects exhibited even higher variation, exceeding this high stringency criterion. Intuitively, one would expect that with larger sample sizes per group, standard deviations and coefficients of variance would decrease. However, analyzing variability in 4 different human gene expression sets from different tissues and diseases, Ho et al. showed that larger sample sizes also facilitate detection of variability that can be obscured by small sample number [20]. A second aspect that is impossible to evaluate with our current sample sizes is whether gene expression variation follows our prediction of a wider normal distribution, or whether it is due to a shift from unimodal (one peak) to a bimodal (two peaks) or otherwise skewed distributions [21] (see Figure 7 for illustration). We previously observed -in an unrelated experimental paradigm- that even in genetically identical animals without detectable anomalies, deviations of gene expression measurements from normality are possible [22]. Applying various algorithms to assess such skewing, Hellwig et al. were able to detect of deviation from unimodal distribution in a large dataset (n=194) from a breast cancer patient cohort, establishing prognostic value for treatment responses and

outcomes [16]. Similar studies in developmental/teratological/toxicological paradigms will be required to determine the power of variability-based approaches for detection and prediction of individual risk for abnormal phenotypes.

The definition of variability can influence the outcomes

In the past, we used the coefficient of variation to evaluate variability [14], but this criterion tends to favor discovery of genes with low expression levels (Kappen and Salbaum, unpublished observations). Here, for ease of graphical presentation of the concept of inter-individual variation, and to detect variability independent of expression level, we used a Z-score envelope of three standard deviations from the mean to identify genes and samples with high deviation from the mean. As would be expected, in addition to the examples shown in Figure 4, we also detected probes that were outside of the envelope in only 3 or 2 of 4 the affected individuals in the NOD model (not shown). Also expected was that some probes exhibited higher variability in controls than in experimental samples, but this was the case only for a small fraction of all genes surveyed (see Methods). These aspects are dependent on the definition of the envelope, and therefore subject to experimenter criteria and statistical cut-offs. The comparatively stringent criterion of 3-fold standard deviation was chosen heuristically with the purpose to craft a clear and compelling argument in our presentation, but other criteria could also be appropriate to the structure of the empirical dataset. For example, Vedell et al., explored variability defined as the maximum fold-change between expression levels of the same gene in different individuals [13]. The most suitable parameters for gene discovery by variation ideally should be evaluated in simulations and validated through bootstrapping approaches.

It is important to note here that we assume that variation from technical aspects -e.g. pipetting accuracy, reagent quality, and instrument calibration- would affect all samples to the same degree, and therefore should not differ in extent between exposed and non-exposed, or normal and abnormally developing individuals. This can be ascertained in any given dataset through bootstrapping and post-hoc simulations as well. If technical variation can be held minimal, the major source of variation would be biological differences between individuals, namely the biological variation that determines whether an exposed individual will exhibit the adverse outcome or not.

A fundamental assumption in comparisons between control and exposed samples is that levels of mRNA/cell and mRNA/tissue are similar. However, this was recently called into question: In cells with higher levels of c-myc activity, which was associated with greater cell size, mRNA production was significantly elevated [23]. Our data do not implicate c-myc as an either differentially expressed or highly variable gene in our models. Yet, other cell cycle regulators are found among differentially regulated genes, with Skp2 common among differentially expressed genes, and Cdc23 common among high variability genes, in both experimental models. We also did not observe statistically significant effects of exposure on embryo size at E10.5, which would be expected to differ if cell size or proliferation were affected. Thus, we currently do not have evidence to address the role of general rates of transcriptional activity in our results. Intriguingly, in cultured cells, individual cell variations in global rates of transcription were linked to ATP availability and mitochondrial content

and activity [24]. At least in HeLa cells, inter-cell variability was reduced by anti-oxidants, suggesting a possible strategy for experimental modulation and interventions.

The particulars of the experimental models

In both experimental models for exposure to maternal diabetes, the large number of genes probes with highly variable expression is significantly different from expectations for random events: in fact, we have demonstrated that even with high stringency criteria, increased variability can be detected for a substantial fraction of total genes surveyed. These results highlight that variability of embryonic gene expression is an inherent parameter in our diabetes-exposure paradigms. However, it is not known to what extent this could be generalized to other teratogen exposures, or to genetic models with incomplete penetrance. For example, in a mouse transgenic model in which the *Hoxc8* gene was overexpressed in the FVB inbred strain [22], we did not find evidence for increased variability in transgene-expressing animals compared to controls ([25] and unpublished observation). On the other hand, a mutation in the chromatin remodeling gene *Smarcc1* (encoding Baf155) was found to be associated with higher variability of gene expression, although this conclusion was based on smaller sample numbers (Niswander laboratory and Salbaum, unpublished observations). One could surmise that in paradigms in which the teratogen has a known molecular target, such as retinoic acid binding to retinoic acid receptors, or TCDD binding to the aryl hydrocarbon receptor, the specificity of such interactions makes pleiotropic effects less likely, although partial penetrance is also observed in some of these models [26, 27]. Analyses of multiple teratological/toxicological paradigms from the perspective of variability are required to determine whether increased variation is associated only with specific exposures or whether it is a characteristic hallmark of all environmental exposures. We here argue that the aspect of variability has been neglected in the search for pathogenic mechanisms in the past, and that it could serve as a paradigm shift for teratogenic exposures with partial penetrance outcomes.

Causes or consequences?

It should be noted that in this study, phenotypic outcome needed to be known for classification of NTD-affected and unaffected individuals. Thus, we had to analyze our samples at a time (gestational day 10.5 in the mouse) when the phenotype was patently manifest. Thus, we are aware that the increased variability we detect in phenotypically abnormal individuals could be a consequence of the developmental defect, rather than it's cause. In order to have a better chance at discovery of causative factors, we will have to conduct analyses at earlier timepoints, before or during neural tube closure. Billington et al. identified variability in p53 activity as a possible basis for incomplete penetrance of craniofacial defects in mutants for the Twisted gastrulation gene [28], but similar to our study, their molecular analysis used samples from embryos with overt defects. For earlier stages, when phenotype cannot yet be unequivocally ascertained, analytical approaches will have to be applied that can define patterns of gene expression in subsets of individuals in the absence of overt phenotype, and link them to risk for future abnormal development. Such approaches are currently being developed for gene expression profiling datasets in the cancer field [20, 29] and in neurobiology [30]. For example, in Schizophrenia, modules of co-expressed and co-regulated genes were detected that distinguished disease from control

cases [31], and variability of gene expression was associated with altered network connectivity in Schizophrenia and Parkinson's Disease [30], as well as in depression [32]. Thus, analysis of expression variances holds promise in reproductive toxicology as well, for discovery of biomarkers for exposure, for understanding of molecular mechanisms in pathogenesis after exposure, and for assessment of risk for abnormal development. The ultimate goal for such approaches is to identify markers that can move risk assessment from the subgroup and population level to a predictive/prognostic mode at the level of the individual.

Finally, we recognize the challenges for functional validation of the contribution of genes discovered on the basis of variability. Single gene knockout or transgenic experiments may not appropriately reproduce partial penetrance. They will also be insufficient if risk for partial penetrance, or the degree of partial penetrance, is determined by more than one factor/gene, or by deregulation of entire subsets of genes. Nonetheless, as more comprehensive information on genes involved in a given phenotype becomes available -the list of known neural tube defects genes includes some 400 genes by now [14, 33] (and Salbaum unpublished)-, we should come into a position that will allow us to assess the relative contribution of each gene to the phenotype of interest, as well as the respective gene's role in partial penetrance and variable expressivity. Furthermore, we can expect that experimental approaches will become available to move from the reductionist single-gene focus to a systems biology perspective that unravels Grüneberg's "*generally not yet identified*" [9] molecular contributors to individual risk for birth defects and abnormal development.

Relevance of the new paradigm

This emphatic consideration of molecular variability is not only of theoretical but also of biological significance, because it has implications for prevention of undesired consequences from teratological/toxicological exposures: If the exposure increases overall variation in gene expression, whether as main effect or in combination with deregulation of transcription factors or signaling pathways, potential interventions will have to be aimed at reducing such variation, rather than at specific molecules/pathways. Interestingly, we have observed that supplementation of folic acid prior to conception was associated with reduced variation of gene expression in uterus ([34], and unpublished observations), but we do not currently have information whether this scenario also applies to the developing embryo. Another important implication of inter-individual variation is that the risk for abnormal development could be associated with different risk factors in different individuals. This notion is illustrated in Panel E of Figure 4, where a given affected individual may have measurements within the "normal range" for some genes, and outside of the envelope for others (for example E6-NTD in green) for different genes in each affected individual (compare to E7-NTD in purple, or E8-NTD in orange). From a single-gene perspective, these cases would be dismissed as inconsistent with "a mechanistic role". However, from a systems perspective, high variation at a subset of genes could potentially affect the overall activity of a particular signaling pathway [35], or critical cell fate decisions. The prevailing concept of a "multifactorial threshold model" for the etiology of neural tube defects [36] certainly accommodates the theoretical possibility that each NTD individual may be its own

unique case -epigenotype- at the molecular level. Again, this would have important implications for prevention strategies.

Examples for association of individual-specific (also called idiosyncratic [37]) molecular variations with phenotypically divergent outcomes have been reported in other systems: Investigating intestinal cell specification in *C. elegans*, Raj et al., showed that variations of gene expression levels can cause alternative phenotypic outcomes, implicating a threshold model in which "mutations compromise mechanisms that normally buffer such variability" [15]. In *C. elegans* Tbx9 mutants with incomplete penetrance of larval defects, two independent mechanisms were found to influence phenotype penetrance, namely compensatory gene expression and buffering by Hsp90 [38]. In any given individual, these two pathways can vary independently, generating unique epigenotypes, and different phenotypes. Similarly, diet-related variability of DNA methylation in mouse liver was targeted to a unique repertoire of loci in each exposed individual [37]. In fact, already at the blastocyst stage in the mouse, individuals may be epigenetic "mosaics" based upon variability of DNA methylation [39] and gene expression [40], and hence, each individual can present as a unique epigenotype. What is currently unknown is whether there are certain genetic loci that are constitutionally more tightly controlled [41], or more or less responsive to perturbations at the molecular, cellular and organismal level [42]. In yeast, some coordinated fluctuations were found for genes with high variability, suggesting that there may be specific "regulons" of variability [43]. Interestingly, in this model, such "regulons" were enriched for pathways associated with stress response, amino acid synthesis, and mitochondrial regulation. The possibility that developmental variability in gene expression could be linked to cellular energy status [44, 45] has important implications for understanding the risk for developmental defects after exposures to under-nutrition and in over-nutrition conditions, such as maternal diabetes [46, 47] and obesity [48].

Potentially, multiple biological mechanisms exist by which variability in gene expression could be mediated: our work has implicated altered transcription factor expression and activity [17], and altered chromatin modifications [49]; altered rates of RNA synthesis and stability may contribute [23, 50], and perturbed regulation by micro-RNAs has been shown to affect neural tube closure [51]. Altered contribution of different cell types to morphogenesis of embryonic structures could also play a role (Salbaum et al., manuscript submitted). Future experimental approaches will have to investigate to what extent individual and common factors contribute to the individual risk for aberrant development. Clearly, as illustrated in this manuscript and in a variety of other model systems [52], increased consideration of and focus on inter-individual variability could stimulate new avenues of research in risk assessment, teratology and reproductive toxicology, and would have important implications for disease prevention.

METHODS

Mouse models

We published microarray profiling results from the STZ model of diabetic pregnancy in FVB mice previously [17]. We used the data from Experiment I reported in that study for re-analysis in this present manuscript. The two embryos with NTDs exhibited defective neural

tube closure in the trunk neural tube, with size of the lesion spanning the length of several somite pairs. The NOD strain of mice develops diabetes spontaneously [7], with onset between 10 – 17 weeks of age in females. Mice were obtained from The Jackson Laboratory and blood glucose levels were monitored weekly. Once females exhibited hyperglycemia (>250 mg/dL blood glucose), they were mated to non-diabetic NOD males; non-diabetic females of the respective age served as the control group. Embryos were isolated at gestational day 10.5 (E10.5) and assessed microscopically for developmental defects, particularly neural tube defects. Incidence of neural tube defects in the NOD model was 39.1% in our hands, consistent with the original publication [6]. Of the four embryos with NTDs, three had defective neural tube closure in the trunk region, and one had a completely open neural tube. No defects were observed in pregnancies of non-diabetic dams in either model.

RNA preparation and microarray assay

RNA was prepared as described previously [53], from individual embryos. Details for the STZ model were as published; RNA samples came from individual embryos and were processed for microarray analysis on the Affymetrix platform. The results are publicly available (GEO accession number GSE9675). RNA samples from the NOD strain were prepared as before. Equal amounts of RNA from 4 embryos from independent pregnancies were pooled, and

Annotation for function and pathway analysis

Molecular function of encoded gene products was annotated based upon GO terms, and where GO terms were ambiguous, MGI (<http://www.informatics.jax.org/>) and Gene Cards (<http://www.genecards.org/>) information was supplemented. Pathway information was gathered from multiple sources, including DAVID [54] (<http://david.abcc.ncifcrf.gov/>), Panther [55] (<http://www.pantherdb.org/pathway/>) and Ingenuity Systems (<http://www.ingenuity.com/products/ipa>), using default parameters and Bonferroni correction for multiple testing. Only the top scoring categories from each analysis were considered here.

Identification of genes with high variability of expression

We used two approaches to detect gene expression variability: One was to rank genes by coefficient of variation (CV = standard deviation from the mean divided by mean expression level) across all diabetes-exposed samples, including neural tube defect-affected individuals. In the STZ model, the top 200 most variable genes exhibited variation over 36.33% from the mean. Similarly, in the NOD model, the top 200 genes most variable genes displayed over 36.48% from the mean. Given sufficient sample numbers available in the NOD model, we found greater coefficients of variation for 16705 probes (out of 18256 total detected) in all diabetes-exposed embryos, whereas the coefficients of variation were greater for only 1551 probes in the normal control group. Even when the CVs were compared for normally developed embryos only, there still remained 12631 probes (75.6% of 16705) with increased variation in diabetes-exposed over control embryos. Thus, exposure to diabetes, even in morphologically normal embryos, is associated with increased variation of gene expression levels. When CVs for all individual NTD-affected embryos together were compared to the CVs for the group of samples from diabetes-exposed but normally developed embryos, 8193

gene probes (64.9% of 12631) exhibited increased variability in NTD samples. Thus, samples from NTD-affected embryos contribute substantially to the variation. However, due to the mathematical transformation (dividing by the mean expression level), CV values will be larger for genes with low expression levels. Thus, they can serve for gene discovery only when additional assumptions are being made or additional criteria applied.

In our second approach, we therefore evaluated the extent by which individual measurements deviate from the mean, without regard to expression level. This was done using the Z-score, which is a well-established criterion for normally distributed data. For each gene probe, we calculated the mean expression level and standard deviation for the 4 samples from diabetes-exposed normally developed embryos, and calculated the Z-score as 3 standard deviations from the mean in either direction. Mathematically, this predicts that more than 99.7% of all observations should occur within the Z-score envelope. We then asked whether any measurements for individual NTD-affected embryos were outside of this envelope. We detected 549 probes as highly variable in NTD-affected embryos, fulfilling the criterion that at least 3 individuals fall outside on the same side of the envelope, which is a significantly ($P < 0.05$) greater number of probes than can be randomly expected (0.26% of $18256 = 47.5$ probes). This concept of deviation from an envelope is visually intuitive and was therefore chosen for graphical representation of high variability in gene expression.

Hierarchical clustering

Datasets for differentially expressed or highly variable gene probes, as defined above for each experimental model, were entered into Cluster 3 [56] (<http://rana.lbl.gov/EisenSoftware.htm>), normalized by log transformation, and median polished as described in the Cluster 3 manual (<http://bonsai.hgc.jp/~mdehoon/software/cluster/cluster3.pdf>). Hierarchical clustering was performed with the default settings for Euclidian distance and complete linkage. Java TreeView [57] (<http://jtreeview.sourceforge.net/>) was used to visualize the results, with identical color settings for all analyses (yellow: increased expression over mean; blue: decreased expression from the mean). All figures and graphs were assembled for publication in Canvas (version 9.0.2; ACD Systems).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to thank Dr. N. Arbour Delahaye for NOD sample preparation, the PBRC Genomics Core for Illumina microarray assays and initial post-hybridization processing of results, Dr. C. Kruger for preparing GEO submissions and performing IPA and Panther analyses, and Drs. C. Kruger, J. Rahnenführer, and J. Volaufova for discussions. Funding for this work was provided in parts by Eunice Kennedy Shriver National Institute of Child Health and Human Development grants R01-HD037804, with Supplements S1 and S2 (to C.K.), R01-HD055528, with supplement S1 (to J.M.S.), the Peggy M. Pennington Cole Endowed Chair in Maternal Biology and Pennington Biomedical Research Foundation. The PBRC Genomics Core Facility is supported in parts by COBRE (P20-GM103528) and NORC (P30-DK072476) grants from the National Institutes of Health.

REFERENCES

1. Wallingford JB, Niswander LA, Shaw GM, Finnell RH. The continuing challenge of understanding, preventing, and treating neural tube defects. *Science*. 2013; 339:1222002. [PubMed: 23449594]
2. Fox DA, Grandjean P, de Groot D, Paule MG. Developmental origins of adult diseases and neurotoxicity: epidemiological and experimental studies. *Neurotoxicology*. 2012; 33:810–816. [PubMed: 22245043]
3. Cortessis VK, Thomas DC, Levine AJ, Breton CV, Mack TM, Siegmund KD, et al. Environmental epigenetics: prospects for studying epigenetic mediation of exposure-response relationships. *Hum Genet*. 2012; 131:1565–1589. [PubMed: 22740325]
4. Kappen C, Kruger C, MacGowan J, Salbaum JM. Maternal diet modulates the risk for neural tube defects in a mouse model of diabetic pregnancy. *Reprod Toxicol*. 2011; 31:41–49. [PubMed: 20868740]
5. Pani L, Horal M, Loeken MR. Polymorphic susceptibility to the molecular causes of neural tube defects during diabetic embryopathy. *Diabetes*. 2002; 51:2871–2874. [PubMed: 12196484]
6. Otani H, Tanaka O, Tatewaki R, Naora H, Yoneyama T. Diabetic environment and genetic predisposition as causes of congenital malformations in NOD mouse embryos. *Diabetes*. 1991; 40:1245–1250. [PubMed: 1936588]
7. Leiter EH. The genetics of diabetes susceptibility in mice. *FASEB J*. 1989; 3:2231–2241. [PubMed: 2673897]
8. Fraser FC. The William Allan Memorial Award Address: evolution of a palatable multifactorial threshold model. *Am J Hum Genet*. 1980; 32:796–813. [PubMed: 7446524]
9. Grüneberg H. Genetical Studies on the Skeleton of the Mouse – IV. Quasi-continuous Variations. *J Genet*. 1952; 51:95–114.
10. Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet*. 2006; 7:55–65. [PubMed: 16369572]
11. Demissie M, Mascialino B, Calza S, Pawitan Y. Unequal group variances in microarray data analyses. *Bioinformatics*. 2008; 24:1168–1174. [PubMed: 18344518]
12. Thomas R, de la Torre L, Chang X, Mehrotra S. Validation and characterization of DNA microarray gene expression data distribution and associated moments. *BMC Bioinformatics*. 2010; 11:576. [PubMed: 21092329]
13. Vedell PT, Svenson KL, Churchill GA. Stochastic variation of transcript abundance in C57BL/6J mice. *BMC Genomics*. 2011; 12:167. [PubMed: 21450099]
14. Salbaum JM, Kappen C. Neural tube defect genes and maternal diabetes during pregnancy. *Birth Defects Res A Clin Mol Teratol*. 2010; 88:601–611. [PubMed: 20564432]
15. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. *Nature*. 2010; 463:913–918. [PubMed: 20164922]
16. Hellwig B, Hengstler JG, Schmidt M, Gehrmann MC, Schormann W, Rahnenführer J. Comparison of scores for bimodality of gene expression distributions and genome-wide evaluation of the prognostic relevance of high-scoring genes. *BMC Bioinformatics*. 2010; 11:276. [PubMed: 20500820]
17. Pavlinkova G, Salbaum JM, Kappen C. Maternal Diabetes alters Transcriptional Programs in the Developing Embryo. *BMC Genomics*. 2009; 10:274. [PubMed: 19538749]
18. Kayala MA, Baldi P. Cyber-T web server: differential analysis of high-throughput data. *Nucleic Acids Res*. 2012; 40:W553–W559. [PubMed: 22600740]
19. Sherman BT, Huang da W, Tan Q, Guo Y, Bour S, Liu D, et al. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics*. 2007; 8:426. [PubMed: 17980028]
20. Ho JW, Stefani M, dos Remedios CG, Charleston MA. Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics*. 2008; 24
21. Wang J, Wen S, Symmans WF, Pusztai L, Coombes KR. The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Inform*. 2009; 7:199–216. [PubMed: 19718451]

22. Kruger C, Talmadge C, Kappen C. Expression of folate pathway genes in the cartilage of Hoxd4 and Hoxc8 transgenic mice. *Birth Defects Res A Clin Mol Teratol.* 2006; 76:216–229. [PubMed: 16586448]
23. Loven J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, et al. Revisiting global gene expression analysis. *Cell.* 2012; 151:476–482. [PubMed: 23101621]
24. das Neves RP, Jones NS, Andreu L, Gupta R, Enver T, Iborra FJ. Connecting variability in global transcription rate to mitochondrial variability. *PLoS Biol.* 2010; 8:e1000560. [PubMed: 21179497]
25. Kruger C, Kappen C. Microarray analysis of defective cartilage in Hoxc8- and Hoxd4-transgenic mice. *Cartilage.* 2010; 1:217–232.
26. Ghyselincx NB, Dupe V, Dierich A, Messaddeq N, Garnier JM, Rochette-Egly C, et al. Role of the retinoic acid receptor beta (RARbeta) during mouse development. *Int J Dev Biol.* 1997; 41:425–447. [PubMed: 9240560]
27. Luo J, Sucov HM, Bader JA, Evans RM, Giguere V. Compound mutants for retinoic acid receptor (RAR) beta and RAR alpha 1 reveal developmental functions for multiple RAR beta isoforms. *Mech Dev.* 1996; 55:33–44. [PubMed: 8734497]
28. Billington CJ Jr, Ng B, Forsman C, Schmidt B, Bagchi A, Symer DE, et al. The molecular and cellular basis of variable craniofacial phenotypes and their genetic rescue in Twisted gastrulation mutant mice. *Dev Biol.* 2011; 355:21–31. [PubMed: 21549111]
29. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 2012; 40:4288–4297. [PubMed: 22287627]
30. Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, et al. Variance of gene expression identifies altered network constraints in neurological disease. *PLoS Genet.* 2011; 7:e1002207. [PubMed: 21852951]
31. Torkamani A, Dean B, Schork NJ, Thomas EA. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res.* 2010; 20:403–412. [PubMed: 20197298]
32. Gaiteri C, Sibille E. Differentially expressed genes in major depression reside on the periphery of resilient gene coexpression networks. *Front Neurosci.* 2011; 5:95. [PubMed: 21922000]
33. Harris MJ, Juriloff DM. An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. *Birth Defects Res A Clin Mol Teratol.* 2010; 88:653–669. [PubMed: 20740593]
34. Salbaum JM, Kruger C, Kappen C. Mutation at the folate receptor 4 locus modulates gene expression profiles in the mouse uterus in response to periconceptional folate supplementation. *Biochimica Biophysica Acta.* 2013/ May 09.:1653–1661. ed2013.
35. Kim KH, Sauro HM. In search of noise-induced bimodality. *BMC Biol.* 2012; 10:89. [PubMed: 23134773]
36. Fraser FC. Some underlooked properties of the multifactorial/threshold model. *Am J Hum Genet.* 1998; 62:1262–1265. [PubMed: 9545400]
37. Li CC, Cropley JE, Cowley MJ, Preiss T, Martin DI, Suter CM. A sustained dietary change increases epigenetic variation in isogenic mice. *PLoS Genet.* 2011; 7:e1001380. [PubMed: 21541011]
38. Burga A, Casanueva MO, Lehner B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature.* 2011; 480:250–253. [PubMed: 22158248]
39. Cirio MC, Martel J, Mann M, Toppings M, Bartolomei M, Trasler J, et al. DNA methyltransferase 1o functions during preimplantation development to preclude a profound level of epigenetic variation. *Dev Biol.* 2008; 324:139–150. [PubMed: 18845137]
40. Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, et al. Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One.* 2011; 6:e21208. [PubMed: 21731673]
41. Pritchard C, Coil D, Hawley S, Hsu L, Nelson PS. The contributions of normal variation and genetic background to mammalian gene expression. *Genome Biol.* 2006; 7:R26. [PubMed: 16584536]

42. Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. *Science*. 2012; 336:183–187. [PubMed: 22499939]
43. Stewart-Ornstein J, Weissman JS, El-Samad H. Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol Cell*. 2012; 45:483–493. [PubMed: 22365828]
44. Lu C, Thompson CB. Metabolic regulation of epigenetics. *Cell Metab*. 2012; 16:9–17. [PubMed: 22768835]
45. Wellen KE, Thompson CB. Cellular metabolic stress: considering how cells respond to nutrient excess. *Mol Cell*. 2010; 40:323–332. [PubMed: 20965425]
46. Freinkel N. Diabetic embryopathy and fuel-mediated organ teratogenesis: Lessons from animal models. *Horm Metabol Res*. 1988; 20:463–475.
47. Salbaum JM, Kappen C. Diabetic embryopathy: a role for the epigenome? *Birth Defects Res A Clin Mol Teratol*. 2011; 91:770–780. [PubMed: 21538816]
48. Carmichael SL, Rasmussen SA, Shaw GM. Prepregnancy obesity: a complex risk factor for selected birth defects. *Birth Defects Res A Clin Mol Teratol*. 2010; 88:804–810. [PubMed: 20973050]
49. Salbaum JM, Kappen C. Responses of the embryonic epigenome to maternal diabetes. *Birth Defects Res A Clin Mol Teratol*. 2012; 94:770–781. [PubMed: 22786762]
50. Lin CY, Loven J, Rahl PB, Paranal RM, Burge CB, Bradner JE, et al. Transcriptional amplification in tumor cells with elevated c-Myc. *Cell*. 2012; 151:56–67. [PubMed: 23021215]
51. Maller Schulman BR, Liang X, Stahlhut C, DelConte C, Stefani G, Slack FJ. The let-7 microRNA target gene, *Mlin41/Trim71* is required for mouse embryonic survival and neural tube closure. *Cell Cycle*. 2008; 7:3935–3942. [PubMed: 19098426]
52. Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. *Science*. 2005; 309:2010–2013. [PubMed: 16179466]
53. Kruger C, Kappen C. Expression of cartilage developmental genes in *Hoxc8*- and *Hoxd4*-transgenic mice. *PLoS One*. 2010; 5:e8978. [PubMed: 20126390]
54. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. [PubMed: 19131956]
55. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 2013; 41:D377–D386. [PubMed: 23193289]
56. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics*. 2004; 20:1453–1454. [PubMed: 14871861]
57. Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics*. 2004; 20:3246–3248. [PubMed: 15180930]

Highlights

- ▶ Many teratogen paradigms produce incomplete penetrance, with only some individuals affected.
- ▶ Comparisons between normal and exposed individuals can't explain partial penetrance.
- ▶ We propose differences between exposed subjects as key factors in susceptibility.
- ▶ We illustrate this concept for two models for diabetes-induced neural tube defects.
- ▶ A focus on variability offers a new perspective to gene discovery in teratogenesis.

Figure 1. Paradigms with partial penetrance: incongruity between approach and outcome

Legend: **(A)**: In normal conditions, almost all individuals in the population fulfill the required threshold(s) and very few animals ever exhibit the abnormal phenotype, hence the small fraction of affected individuals (red). Phenotype penetrance is close to 0%. **(B)**: The teratogenic exposure (orange curve) moves the mean of the distribution in a direction that increases the number of affected individual in the population of all exposed individuals. The fraction of affected individuals (large red hatched area) determines the penetrance, which, in the extreme, could reach 100%. **(C)**: Traditional approaches in gene expression profiling identify differentially expressed genes by virtue of a statistically significant difference between the mean of a control group's gene expression values (black) and the mean of the gene expression values for a group of experimental samples (green and red). For simplicity, only one direction of change is depicted here, namely lower mean expression levels in the exposed group; an analogous situation would be present when the expression levels for a given gene would be higher in experimental samples than in controls. **(D)**: The typical outcome from exposure to the teratogenic condition (orange) is that a fraction of individuals exhibits the abnormal phenotype (red) while other exposed individuals undergo normal development (green). The dashed line represents the threshold between normal and abnormal development. It is visually obvious that the discovery paradigm in C cannot resolve molecular differences between individuals with normal and abnormal phenotype, because the mean for the experimental group includes normal and affected individuals.

Figure 2. A new paradigm for incorporating partial penetrance into experimental approaches
 Legend: **(A)**: Instead of shifting the mean of the distribution, the exposure leads to greater width of the distribution, which places more individuals beyond the threshold. Note that a similar fraction of individuals would be affected as in Figure 1C. **(B)**: Experimentally, high variation would manifest in a wider spread of measurements from exposed individuals, which may or may not be accompanied by a shift of the mean for the group, although some individuals exhibit dramatic deviation from the mean. Individuals at the extreme of the spectrum would be expected to be affected (red), while individuals within a range comparable to unexposed controls would develop normally (green). As in Figure 1, only one direction of change is depicted; a priori, it is equally likely that individuals with high gene expression levels could be affected. Instead of using mean values for the experimental group as the criterion for discovery, the variance within the experimental group becomes the primary criterion for discovery. It is intuitively visually obvious that the discovery paradigm in B is able to resolve molecular differences between individuals with normal and abnormal phenotype, as the affected individuals come to lie at the extremes of their group.

Figure 3. Lack of discrimination between affected and unaffected on the basis of differential gene expression

Microarray datasets (Affymetrix platform for A; Illumina platform for B) were analyzed with the latest version of Cyber-T to identify differentially expressed genes, applying conventional criteria: all samples have a detectable signal for the genes under consideration and the mean expression levels in the control group differ from the mean for the exposed group by a statistically significant p-value (<0.05 , after correction for multiple comparisons). The datasets were median polished according to the instructions for Cluster 3, and hierarchical clustering was performed with Euclidian distance and complete linkage settings. A: Heatmap for 338 differentially expressed gene probes in the STZ model (yellow: higher than median; blue: lower than median). The two individuals from normal pregnancies cluster together on one branch (C1 and C2, green frame), different from the branch that contains the exposed individuals (E1-E5, red frame). There is no distinction between embryos with (E4-NTD, E5-NTD) and without (E1-E3) a neural tube defect within the exposed group. B: Hierarchical clustering applied to 705 differentially expressed gene probes in the NOD model. Four samples from non-diabetic pregnancies cluster together on a branch (C1-C4, green frame) separate from the 8 samples from diabetes-exposed embryos (E1-E8, red frame); the 4 individuals with neural tube defects (E5-E8) are clustered intermingled within the diabetes-exposed group. There is no distinction between affected individuals (E5-NTD, E6-NTD, E7-NTD, E8-NTD) from unaffected embryo samples (E1-

E4) within the exposed group. [We know that this is independent of profiling technology since the NOD arrays were performed on the Illumina platform, and the FVB arrays on Affymetrix chips]. The sample tree topologies are not to scale, but for both models those branchpoints between affected and unaffected exposed samples are located deep in the sample tree, indicative of inefficient resolution of relationships.

Figure 4. Identification of genes expressed with high variability among the exposed embryos

Legend: For each gene probe, we determined the mean level of expression in normally developed embryos from diabetic pregnancies, and the standard deviation from the mean. A z-score was calculated for each gene probe of 3 standard deviations above (Z SD3+) and 3 standard deviations below (Z SD3-) the mean; we defined values within this envelope as the "normal range", compatible with normal development. For each NTD-affected individual, we determined the "fold-change" of individual deviation from the mean for each gene probe (Y-axis), and filtered the dataset for those probes where individual fold-change was greater than the Z-score envelope; Z-scores were sorted from lowest to highest along the X-axis. A, B: In the STZ model, 333 gene probes with high individual variability were identified; for 194 probes (A), both individuals (blue=E4-NTD; red=E5-NTD) had values above the "normal range", and for 139 probes, both defective embryos scored below the "normal range", as defined by the Z-score envelope. C-E: In the NOD model, 4 individuals were affected by neural tube defects (turquoise=E5-NTD, green=E6-NTD, purple=E7-NTD, orange=E8-NTD). Depicted here as examples are gene subsets for which at least three NTD-affected individuals had measurements on the same outside of the envelope. C: all 4 individuals fall below the "normal range" (32 genes); D: all 4 individuals are outside of the envelope, 3 below, 1 above (87 genes), and E: 3 individual measurements below, 1 individual datapoint inside the envelope (228 genes). Additional configurations were detected (such as 4, or at least 3 individuals above the range) but are not shown for clarity and brevity.

Figure 5. Efficient discrimination between normally and abnormally developed individuals by virtue of high variability of gene expression levels

Legend: Hierarchical cluster analyses for diabetes-exposed E10.5 samples from embryos with and without NTDs (yellow: higher expression, blue: lower expression level); samples from normal pregnancies were omitted for clarity. A: Clustering with gene probes that are highly variable across all diabetes-exposed samples from the STZ model (the probes shown in Figure 4, Panels A and B) achieves grouping of morphologically normal embryo samples together (green frame), clearly distinct from NTD-affected individuals (red frame). B: In the NOD model, highly variable gene probes (those shown in Figure 4, Panels C-E) support clustering of phenotype-affected individuals together (red frame), distinct from samples from normally developed diabetes-exposed embryos (green frame). The NTD samples were equally well discriminated even when 500 randomly selected gene probes were added to the cluster analysis to simulate noise (results not shown).

Figure 6. Gene Discovery by Expression Profiling: Variability adds a new Dimension for Understanding Partial Penetrance

Legend: Annotation for gene names and function was derived from GO terms, DAVID analysis, GeneCards and the Mouse Genome Informatics tools. Probes without annotation and duplicates for the same gene were eliminated. Confirmation of enriched pathways was obtained by Panther and IPA analysis. A: Differentially expressed (green) and highly variability gene subsets (red) in the STZ model (solid lines) overlap by ~12% (gray shaded area in the Venn diagram). Sizes of circles and overlap areas are scaled proportionate to the number of unique annotations. B: In the NOD model (broken lines), both gene subsets overlap by ~6%. C: Limited overlap of differentially expressed genes between both experimental models. D: Limited overlap of highly variable genes between both experimental models.

Figure 7. Shapes of data distribution - unimodality and bimodality in gene expression

Legend: It is generally assumed that, with large enough sample numbers, expression levels for any gene follow a normal distribution as depicted in A. Increased variability then is associated with a flattened unimodal distribution (blue line). B: Bimodal distribution with a proportion of affected individuals. For illustration purposes, only increased expression levels are shown.