_____

Integrative analysis of the mouse embryonic transcriptome

Amar V. Singh, Kenneth B. Knudsen and Thomas B. Knudsen[*]

Department of Molecular, Cellular and Craniofacial Biology, University of Louisville, School of Dentistry, Louisville, KY 40202;
Thomas B Knudsen* - Email: thomas.knudsen@louisville.edu; Phone: +502 852 4128; Fax: +502 852 4702; * Corresponding Author
received December 05, 2006; accepted January 20, 2007; published online April 10, 2007

Abstract:

Monitoring global gene expression provides insight into how genes and regulatory signals work together to guide embryo development. The fields of developmental biology and teratology are now confronted with the need for automated access to a reference library of gene-expression signatures that benchmark programmed (genetic) and adaptive (environmental) regulation of the embryonic transcriptome. Such a library must be constructed from highly-distributed microarray data. Birth Defects Systems Manager (BDSM), an open access knowledge management system, provides custom software to mine public microarray data focused on developmental health and disease. The present study describes tools for seamless data integration in the BDSM library (*MetaSample*, *MetaChip*, *CIAeasy*) using the QueryBDSM module. A field test of the prototype was run using published microarray data series derived from a variety of laboratories, experiments, microarray platforms, organ systems, and developmental stages. The datasets focused on several developing systems in the mouse embryo, including preimplantation stages, heart and nerve development, testis and ovary development, and craniofacial development. Using BDSM data integration tools, a gene-expression signature for 346 genes was resolved that accurately classified samples by organ system and developmental sequence. The module builds a potential for the BDSM approach to decipher a large number developmental processes through comparative bioinformatics analysis of embryological systems at-risk for specific defects, using multiple scenarios to define the range of probabilities leading from molecular phenotype to clinical phenotype. We conclude that an integrative analysis of global gene-expression of the developing embryo can form the foundation for constructing a reference library of signaling pathways and networks for normal and abnormal regulation of the embryonic transcriptome. These tools are available free of charge from the web-site http://systemsanalysis.louisville.edu requiring only a short registration process.

Keywords: transcriptome; mouse; expression; embryo; integrative analysis; birth defects

Background:

Animal development is fashioned by conserved signaling pathways that orchestrate morphogenesis, pattern formation, and cell differentiation - complex processes operating jointly in different parts of an embryo and in stages associated with sequential gene activation. Monitoring local and temporal changes in gene expression can provide insight into how genes and regulatory signals work together to guide development. [1] This knowledge is important for understanding the pathogenesis of birth defects and to the central problems of defining precursor target cell susceptibility and the causal mechanisms of abnormal development triggered by diverse environmental and genetic perturbations to maternal-fetal unit. [2]

Profiling gene expression on a global scale has become an important source of information for biological knowledge discovery. Despite well-known challenges confronting technology development, the analysis of global gene expression data can reveal themes in the biologically robust response patterns in gene activity. [3] Gene Expression Omnibus (GEO) is one of the main national repositories for high-information content transcript data from microarray analysis and serial analysis of gene expression (SAGE). [4] GEO has grown from 18,235 records in June 2004 to 115,415 records in December 2006, reflecting an average growth of over 100 new entries per day. A subset of this information describes the embryo proper and can be mined for major biological themes in developmental health and disease. Using keyword searches and trend analysis to mine PubMed and Medline literature databases, the public availability of embryo-based microarrays currently numbers 500-600 (564), mostly studies on mouse embryos and differentiating human cell lines.

The increasing volume of gene expression data on local and temporal states confronts the developmental biologist with the need for reference libraries and information management systems to handle optimal-scale gene-expression signatures and facilitate biological knowledge discovery. For example, Lamb et al., [5] created a prototype reference collection of gene-expression signatures from cultured human cells exposed to bioactive molecules, which serves as a platform for pattern-matching software to establish a 'connectivity map' between drugs, genes, and diseases. Another example is the integrative analysis of multi-study tumor profiles. [6, 7] The emerging database model for tumor classification based on molecular abundance profiles has implied a 67-gene core 'common transcriptional program' in multiple cancers. [8] In development-teratogenesis, a preliminary meta-analysis across microarray studies in the mouse embryo returned a gene-expression signature of 512 developmentally regulated genes of which 16% (~82-genes) changed during exposure to teratogenic agents. [2] Given the promises and pitfalls of computational methods for solving gene expression problems, automated access to a reference collection of gene-expression signatures to benchmark the programmed (genetic) and adaptive (environmental) regulation of the embryonic transcriptome is scientifically needed.

'Birth Defects Systems Manager' (BDSM) is a knowledge management system that provides custom software to mine public microarray data for interesting patterns across developmental stages, organ systems, and disease phenotypes. **[2, 9]** This open resource enables: consolidation of communal data and metadata relevant to developmental health and disease; interactions with current builds of national databases and data repositories; efficient algorithms for cross-species annotation of symbolic gene annotations using the NCBI sequence homology-based annotations for corresponding homologues and orthologues; specific queries across experiments to facilitate secondary analysis; and data formats interoperable with analysis software for phenetic clustering, chromosomal mapping, gene ontology classification, pathway evaluation, and network identification.

Since a comprehensive reference collection of gene-expression signatures for developing structures must be constructed from highly-distributed data, the present study was designed to empower BDSM with tools for seamless data integration: *MetaSample*, *MetaChip*, and *CIAeasy*. These tools are accessible at http://systemsanalysis.louisville.edu requiring only a short registration process to BDSM. A field test of the prototype run with published microarray data illustrates proof of concept for integrative analysis of the mouse embryonic transcriptome.

Methodology:

Dataset collections: Search of PubMed using keywords 'embryo' and 'microarray' returned 495 records of which 193 actually used the technology to study developing animal systems. GEO data sets (GDS) narrowed the list to 47 non-redundant microarray datasets, and including the keyword 'teratogen' added a few more datasets, for a total of 564 public microarrays on the embryo. Raw and/or processed microarray sample-data files and associated metadata were parsed onto the server using LoadBDSM. **[9]** BDSM currently holds 25 developmental series containing 537 samples that are derived from the public domain and 3 series containing 43 samples which are private. These data represent 15 developing organ systems, 6 chemical exposures, and 5 drug interventions across 42 development stages.

Tracking provenance: For the present study, we restricted the analysis to 160 arrays in the BDSM library based on well-annotated experiments published on normal mouse embryogenesis using an Affymetrix technology platform. These conditions are identified by GEO Series Accession number (GSE) and/or literature citation as follows: preimplantation mouse embryo (GSE1749) **[10]**; heart (GSE1479) GD10 - GD18 **[11]**; nerve (GSE972) GD9.5 – birth **[12]**; ovary (GSE1359) and testis (GSE1358) between GD11.5 – birth **[13]**; orofacial region (GSE1624) **[14]** and secondary palate **[15]** between GD13-15. The platforms for these series included: MG-U74Av2 (12488 probes), MG-U74Bv2 (12478 probes), MG430Av2 (45104 probes), MOE430Av2 (22690 probes), and MOE430Bv2 (22576 probes). Internal annotation of Affymetrix probe identifiers was performed to standardize gene labels across samples and improve cross-platform interoperability, as discussed previously. **[2]**

Data integration: Individual microarray sample-data files from the aforementioned developmental series were integrated using QueryBDSM, a module that for merging pre-processed, normalized samples from the BDSM library. Three meta-analysis tools written in PHP were designed to compare and analyze expression data across multiple chips and platforms: *MetaSample*, *MetaChip*, and *CIAeasy*. The workflow schema is diagrammed in Figure 1. Individual sample-data files are selected from the BDSM library and added to a queue for integration. The formatted input files are tab-delimited expression ratios of probes (rows) x samples (columns). QueryBDSM determines the number of distinct microarray platforms in the sample queue and merges the data as follows: if all samples come from the same microarray platform (number =1), then QueryBDSM automatically runs *MetaSample* to create a merged-table having 'columns' of normalized expression data (samples) in 'rows' derived from the platform, with unique probe identifiers (ProbeID) expanded to include GenBank accession (GeneID), and symbolic gene name (Gene Symbol). If multiple platforms are represented by samples in the queue (number >1), then QueryBDSM automatically runs *MetaChip*. *MetaChip* merges data when probe identifiers are different but represent the same annotation, such as across microarray platforms or phylogenetic species. The probe identifiers from each platform are converted to UniGene ID and then merged accordingly, with associated expression data, for those genes common across the datasets. The probes are annotated based on reverse-engineering the sequence homology-based annotations from GenBank. **[2]** For this purpose the system uses data flat files downloaded monthly from the HomoloGene and UniGene databases of NCBI. **[16]**
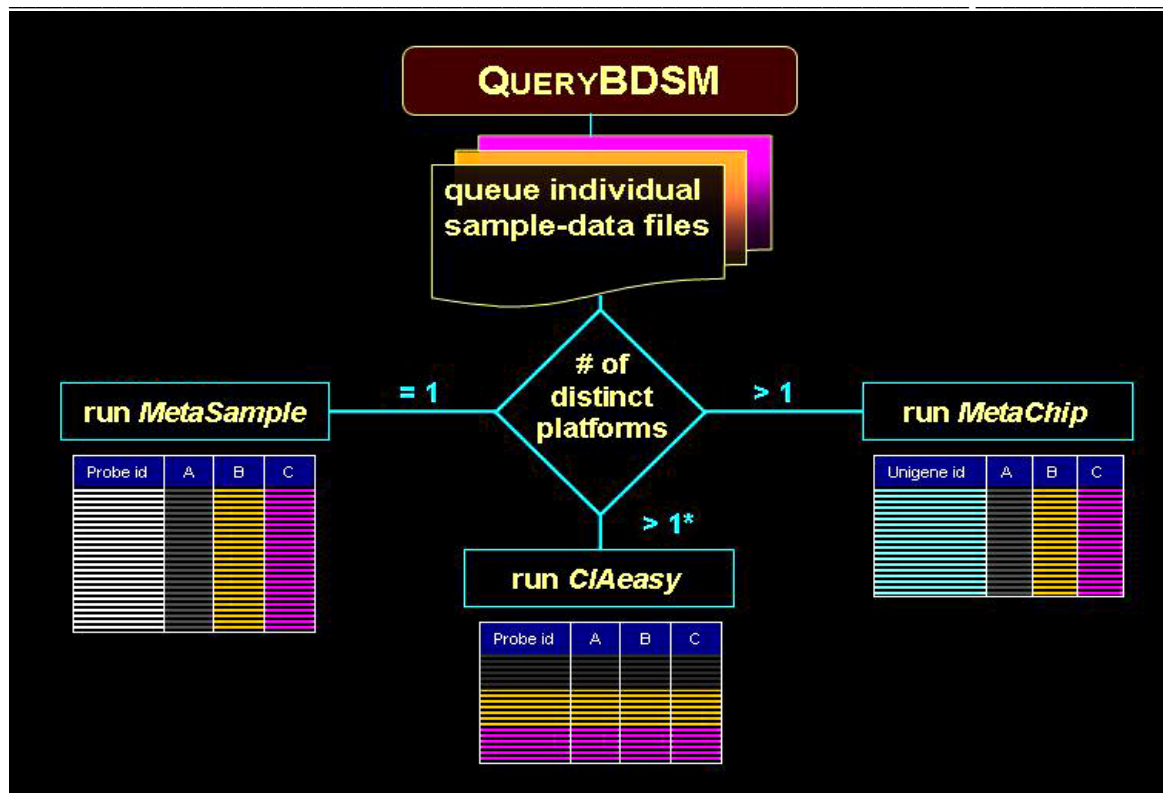
Figure 1: Workflow schema for the QueryBDSM module of BDSM. Individual files of normalized microarray data are selected from the BDSM library. QueryBDSM determines the number of distinct microarray platforms in the sample queue and merges the data as follows: if all samples come from the same platform, then *MetaSample* is used; if multiple platforms are represented, then *MetaChip* is used. *CIAeasy c*ompares joint trends in expression data for the same samples run on different platform

In contrast to *MetaSample* and *MetaChip*, automated tools for merging samples, *CIAeasy* must be specified explicitly to compare datasets for the same samples on different platforms. *CIAeasy* was created from the ADE-4 for *R* statistical computing software package **[17]** and is adapted from 'co-inertia analysis' (CIA) for microarrays. **[18]** With *CIAeasy* users can perform CIA from the BDSM web site without detailed knowledge of *R* language programming. Since samples are aligned on a common space, CIA extracts information about the joint trends in expression patterns of genes independent of probe or sequence annotation. **[18]** *CIAeasy* automatically computes successive orthogonal axes with correspondence analysis and returns the percentage of total variance explained by each eigenvector to find the strongest trends in the co-structured datasets.

Data analysis: One of the problems confronting meta-analysis is the normalcy and spread of expression data. In order to derive expression ratios from the Affymetrix data, we computed a reference denominator, averaged for each gene at the earliest stage in a developmental series. These ratios were transformed to logarithm base 2 (log2) to produce a continuous spectrum of values without biasing between up- and down-regulated genes and making the spread more normal. Each microarray set was centered to median of 0.00 and standardized by scaling to an average standard deviation of 0.5. **[2]** The merged data were imported to GeneSpring v 7.2 using the UniGene cluster ID as the unique gene identifier. The data were clustered using Pearson correlation for the gene tree and two-sided Spearman Confidence for the developmental conditions. Functional annotation used the NIH/NIAID Database for Annotation, Visualization and Integrated Discovery (DAVID). **[19]** The highest-ranking biological themes were stratified by Gene Ontology (GO) terms.

Discussion:
Implementation of QueryBDSM: For proof of concept we examined samples from GEO data source GSE1391, a series describing global gene-expression profiles of the mouse embryo during preimplantation stages. **[10]** Samples included in this series represent development of the oocyte through fertilization (1-cell embryo), activation of the zygotic genome (2-cell embryo) and first differentiation (8-cell embryo) leading to divergent embryonal (inner cell mass) and trophectodermal (placental) lineages of the blastocyst. Biological replicates (3-4) arrayed at each stage used three different Affymetrix platforms: MOE430Av2 (22690 probes), MOE430Bv2 (22576 probes), and MG-U74Av2 (12488 probes). Gene-expression profiles were

_____

normalized to the 'oocyte' in each platform as the earliest stage in the series. Derived data are log2-scale expression values computed from the ratio of signals to the oocyte reference.

Using QueryBDSM, we merged datasets from the different samples to create three distinct datasets for the MOE430Av2, MOE430Bv2, and MG-U74Av2 platforms. Statistical (ANOVA) analysis, run at high stringency with Benjamini-Hochberg correction applied, returned 4417 probes (alpha = 0.0001), 1614 probes (alpha = 0.0001), and 2400 probes (alpha = 0.001) that were differentially regulated. Aside from 34 probes that overlapped between the first two datasets, different genes were detected across these diverse platforms. Hierarchical clustering revealed two basic trajectories of gene-expression in all three platforms (Figure 2). One expression cluster contained genes that increased at the 2-cell stage to the blastocyst stage, and the other cluster contained genes that decreased over these stages (not shown). These probes were mapped to the 307 reference pathways in the KEGG: Kyoto Encyclopedia of Genes and Genomes (http://www.genome.ad.jp/kegg/) library to identify metabolic themes. The top significant KEGG pathways showed concordance between platforms MOE430Av2 and MG-U74Av2 (Table 1), whereas MOE430Bv2 detected different pathways. Some pathways had marginal P-values with individual analysis that became significant in datasets joined by MetaChip (e.g., Adherens Junction, Tight Junction pathways). This illustrates the strength of the meta-analysis approach.
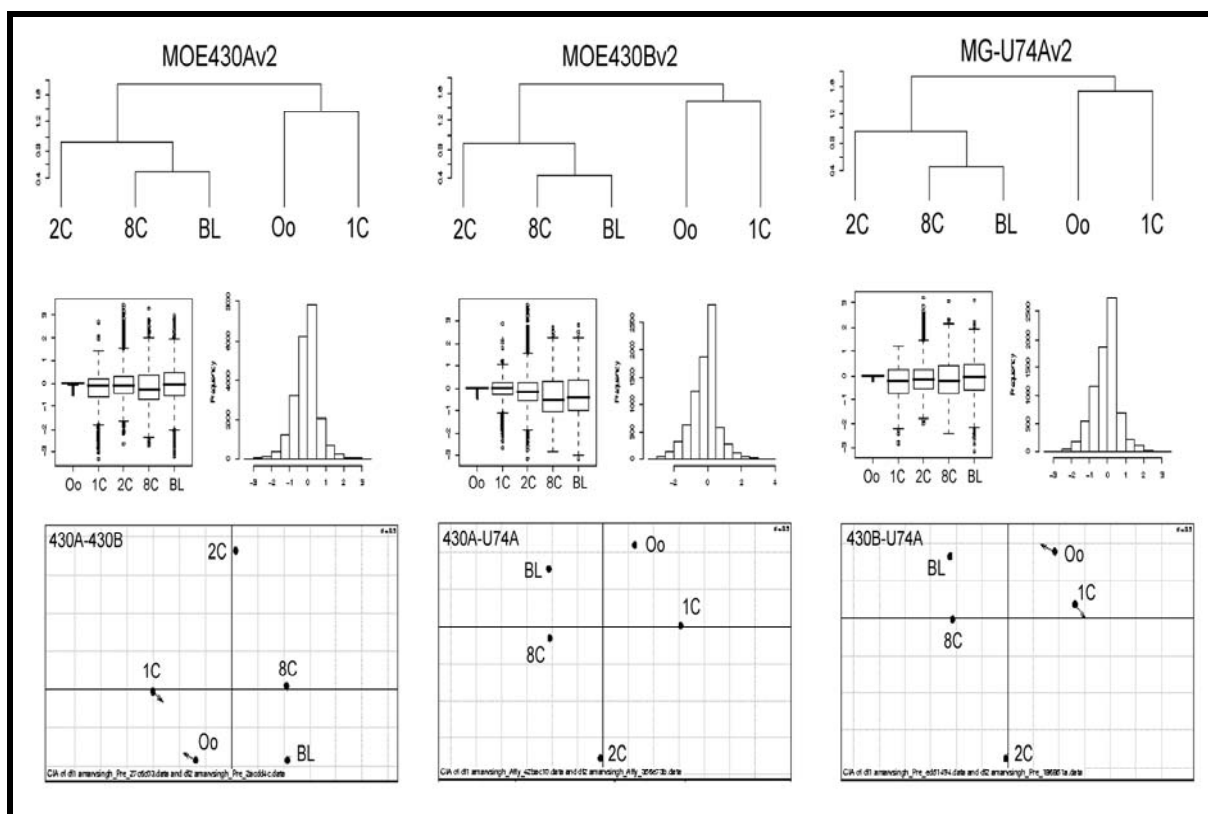

Figure 2: Cross-platform comparison of expression profiles using CIAeasy. Class comparison of secondary analysis of GSE321 data from **[10]**, using *CIAeasy*. Expression values were from statistical (ANOVA) analysis returning 4417 (MOE430Av2, alpha = 0.0001), 1614 (MOE430Bv2, alpha = 0.0001), and 2400 (MG-U74Av2, alpha = 0.001) probes. UPPER PANEL: Clustering of samples: Oocyte (Oo), 1-cell embryo (1C), 2-cell embryo (2C), 8-cell embryo (8C), and blastocyst (BL). For each platform a boxplot shows the log2 distribution of expression values for the probes at each stage (line = median, box = 75%, whiskers = 95%, and dots = outliers) and a histogram plots the cumulative distribution of probes in each expression bin. LOWER PANEL: The first two axes of ordination were projected by CIA. The horizontal axis reflects the first axis of the inertia and the vertical axis is the second axis of inertia. Circles (dots) and arrowheads (arrows) represent the projected coordinates of each platform being compared, with length of the line joining them being proportional to their divergence, e.g., shorter arrows indicate stronger joint trends between the platforms. The left compares joint trends in MOE430Av2 (dot) versus MOE430Bv2 (arrow), measure of global similarity, RV = 0.976; middle compares MOE430Av2 (dot) versus MG-U74Av2 (arrow), RV = 0.999; right compares MOE430Bv2 (dot) versus MG-U74Av2 (arrow), RV = 0.970

KEGG Pathway	Term	MOE430Av2 No. 4417	P-value	MG-U74Av2 No. 2400	P-value	MetaChip No. 3324	P-value	CIAeasy No. 4551	P-value
MMU00010	Glycolysis / Gluconeogenesis	21	0.033	21	<0.001	26 ***	<0.001	28	0.002
MMU00020	Citrate cycle (TCA cycle)	14	0.016	11	0.020	18 *	0.004	17	0.005
MMU00052	Galactose metabolism	15	0.020	--	--	14 ***	0.030	17	0.020
MMU00071	Fatty acid metabolism	18	0.027	--	--	18 **	0.083	25	<0.001
MMU00640	Propanoate metabolism	--	--	12	0.029	14 ***	0.048	18	0.016
MMU00100	Biosynthesis of steroids	--	--	7	0.038	8 ***	0.050	--	--
MMU00190	Oxidative phosphorylation	67	<0.001	57	<0.001	54 ***	<0.001	79	<0.001
MMU00193	ATP synthesis	17	0.030	13	0.038	17 **	0.085	20	0.019
MMU00230	Purine metabolism	52	0.002	40	0.002	48 ***	0.005	62	<0.001
MMU00240	Pyrimidine metabolism	36	0.001	32	<0.001	40 ***	<0.001	45	<0.001
MMU00280	Valine, Leucine and Isoleucine degradation	23	0.003	16	0.020	23 ***	0.001	31	<0.001
MMU00290	Valine, Leucine and Isoleucine biosynthesis	--	--	--	--	6 ***	0.033	--	--
MMU00310	Lysine degradation	22	0.050	--	--	22 ***	0.030	--	--
MMU00520	Nucleotide sugars metabolism	10	0.015	--	--	7 ***	0.020	13	0.049
MMU00620	Pyruvate metabolism	17	0.023	15	0.005	19 ***	0.002	24	<0.001
MMU00650	Butanoate metabolism	23	0.013	15	0.095	20 ***	0.051	28	0.004
MMU00710	Carbon fixation	11	0.033	--	--	--	--	--	--
MMU03010	Ribosome	80	<0.001	39	<0.001	22 ***	0.009	80	<0.001
MMU03020	RNA polymerase	13	0.008	13	<0.001	12 ***	0.014	16	0.001
MMU03050	Proteasome	20	<0.001	22	<0.001	21 ***	<0.001	24	<0.001
MMU03060	Protein export	--	--	--	--	7 ***	0.021	--	--
MMU04110	Cell cycle	42	<0.001	33	<0.001	49 ***	<0.001	51	<0.001
MMU04130	SNARE interactions in vesicular transport	--	--	--	--	13 ***	0.033	--	--
MMU04520	Adherens junction	26	0.050	19	0.074	26 ***	0.027	30	0.057
MMU04530	Tight junction	38	0.059	3	0.063	39 ***	0.016	--	--

Table 1: Meta-analysis of top significant KEGG pathways altered during preimplantation mouse embryo development [1]
[1] Differentially expressed across GEO data series GSE1391 based on 3 Affymetrix platforms. Statistical (ANOVA) analysis selected significant genes with Benjamini-Hochberg correction (alpha = 0.0001 for MOE430Av2, alpha = 0.001 for MG-U74Av2, alpha = 0.05*, 0.01**, and 0.001*** for MetaChip). The number of probes or genes passing at the alpha level is given (No.). Functional annotation of the Affymetrix probe identifiers (MOE430Av2, MG-U74Av2) mapped the significant KEGG pathways (P-value, threshold set to 6 genes). Functional annotation of the 432A-U74A MetaChip used UniGene identifiers and CIAeasy used unique DAVID identifiers.

We next used QueryBDSM to merge samples from platforms MOE430Av2 and MG-U74Av2 to illustrate the *MetaChip* and *CIAeasy* tools. These chips contained 13022 and 9562 non-redundant UniGene cluster identifiers, respectively, of which 7278 were common between the two platforms when passed through *MetaChip*. Statistical (ANOVA) analysis with the same parameters as before returned 3324 genes that were differentially regulated in this data subset. The top significant KEGG pathways for the combined 430A-U74A MetaChip are concordant with the individual analysis (Table 1). *CIAeasy* was used to compare the joint trend between 4417, 1614, and 2400 probes from *MetaSample*, resulting in a high-level of similarity between these three platforms (Figure 2). Of the 6812 non-redundant probes between MOE430Av2 and MG-U74Av2 (4417 + 2400), only five probes were common. Annotating the probes which were not in common gave 4551 unique DAVID identifiers. Again, meta-analysis picked up most of the significant KEGG pathways identified in either of the singular analyses performed earlier and a few additional metabolic pathways (Table 1).

Comparative bioinformatics analysis across developing systems: An obvious limitation in combining data from several different platforms is that as more platforms are included, fewer representative genes are found to be common amongst all platforms. This problem increases when considering less comprehensive arrays, older arrays with outdated probe annotations, or arrays across animal species. Staging the entry of data from the most versatile arrays first can lessen the problem of losing information when data are combined across platforms; however, in some data mining efforts the discriminating power gained by increasing samples and conditions might outweigh the loss of information. For example, multi-platform datasets have been found to discriminate tumor classification by expression profile with as few as 25 genes. [6] For this reason it may be possible to benchmark developmental stages using a limited number of genes across many diverse platforms. To illustrate this point we used BDSM-derived data to compare expression profiles across six unrelated studies and developing systems. We constructed a virtual meta-chip for probes common to all five technology platforms represented in these studies, yielding 346 genes. Unsupervised clustering and Pearson correlation of the gene-expression profiles correctly ordered the samples first by organ system and then by developmental sequence within each system (Figure 3). Within this hierarchy commonalities and differences across organ systems were evident for the patterns of expression in subsets of genes. Unfortunately, the number of genes returned from 346 by statistical (ANOVA) analysis of each individual system, or by K-means clustering of the entire matrix, was too small for insightful functional annotation. QueryBDSM is a simple and efficient solution that can be used to construct a self-evolving reference collection of gene-expression profiles from highly-distributed data on mouse development. Although robust mapping of biological themes and pathways that are expressed at particular developmental stages is straightforward when the same technology platform is considered [2] fuzzy-clustering methods will be needed when multiple platforms are considered.

The *MetaSample* and *MetaChip* components of QueryBDSM are also available as standalone tools under the MetaBDSM module. In this way, the user can upload data outside the BDSM library. The user supplies details of the technology platform, organism, and information about the file format. Once all the required fields are entered and submitted, the files are checked for unique headers. Columns can be dropped by selecting the appropriate checkboxes. Clicking the *Continue* button combines the datasets with expression data and unique identifiers only for genes common between the datasets. These tools have been tested on Internet Explorer 6.0 or greater and Mozilla-based browsers, such as Netscape 6.0 and Firefox.

Other tools are available to assemble data when the platform is same, such as Microarray data assembler. [20] This Excel-based program inherits Excel's limitation from file sizes and number of samples (256 columns and 65,000 data points) whereas the *MetaSample* and *MetaChip* tools do not have this limitation. These tools create temporary tables in the Oracle database and join them using the functionality of Oracle before putting it into a text file, reducing restrictions on the number of samples and size of files. Although users can theoretically combine 100 files at a single time, it is not recommended to load more then 25 files at a time.

Conclusion:

The representation of experimental samples as developmentally contiguous groups is expected to yield a novel mosaic view of gene-expression signatures and genetic dependencies. Although sufficient data exists for data-mining efforts to begin, the ultimate goal of an unabridged reference collection must be viewed as a long-term effort. Regarding the embryo, a search of OVID MedLine using keywords 'embryo' OR 'teratogen' (136,146 records) AND 'microarray' OR 'SAGE' (16,906 records) returned 343 total records. At the current rate of 564 microarrays per 343 publication records (factor = 1.64), the trajectory of embryo-based microarray publications projects GEO to hold in excess of 1,476 microarrays relevant for embryogenesis or teratogenesis by the year 2010.

As studies unravel gene-expression signatures, the key principles in teratology – namely, chemical effects on biological mechanisms, dose-response relationships, factors underlying genetic susceptibility, stage-dependent responses, and maternal influences, can be framed in a systems biology context to address an 'experience database' for ranking pathways and networks by strength of association with anatomical landmarks and developmental abnormalities. [2] The BDSM resource would parallel efforts toward molecular diagnostics in cancer biology (http://www.oncomine.org/), which includes data sets profiling human tumor samples. [21] Since interpreting gene-expression signatures in birth defects will be predicated on posterior (prior) knowledge about developmental health and disease, an important payoff from this bioinformatics effort is to recognize and characterize how these biological states emerge from adaptation or adverse regulation of the embryonic transcriptome.
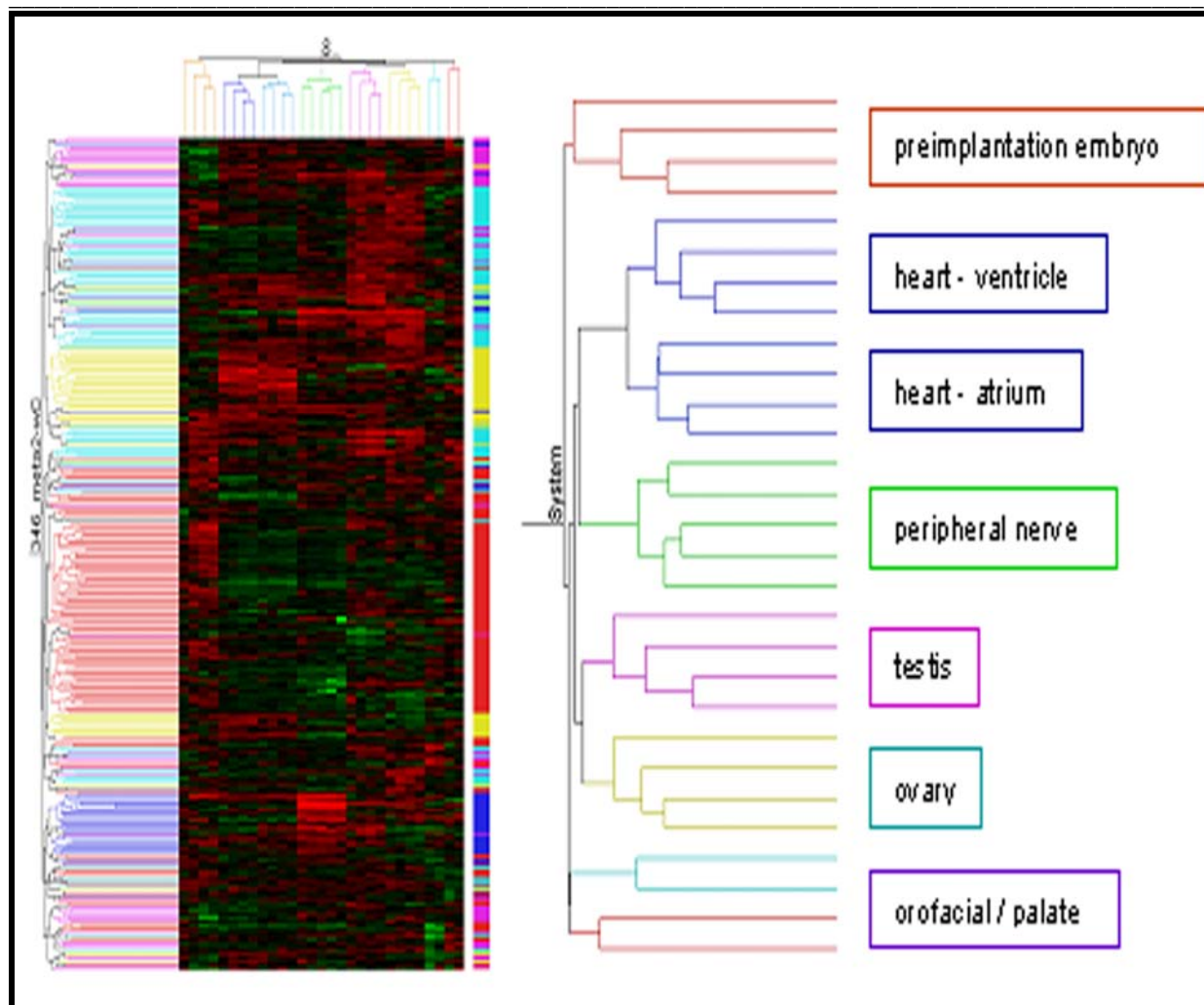
Figure 3: Hierarchy of molecular phenotypes in developing mouse embryos. In five GEO platforms (160 samples), 346 genes were found differentially expressed in BDSM profiles: preimplantation mouse embryo (GSE1749) **[10]**, heart (GSE1479) GD10 - GD18 **[11]** nerve (GSE972) GD9.5 – birth **[12]**; ovary (GSE1359) and testis (GSE1358) GD11.5 – birth **[13]**; orofacial (GSE1624) **[14]** and palate **[15]** between GD13-15. Each developing organ system was properly ordered by its natural chronology, based on the 346 gene-expression signature and an unsupervised clustering algorithm with Pearson correlation. Gene colors are mapped by K-means clustering (6 sets).

References:
[01] E. H. Davidson & D.H. Erwin, *Science,* 311:796 (2006) [PMID: 16469913]
[02] A. V. Singh, *et al.*, *Reprod. Toxicol.,* 19:421 (2005) [PMID:15686875]
[03] E. Marshall, *Science,* 306:630 (2004) [PMID: 15499004]
[04] T. Barrett, *et al., Nucleic Acids Res.,* 33:D562 (2005) [PMID: 15608262]
[05] J. Lamb, *et al., Science,* 313:1929 (2006) [PMID: 17008526]
[06] G. Bloom, *et al., Am. J. Pathol.,* 164:9 (2004) [PMID: 14695313]
[07] D. R. Rhodes & A. M. Chinnaiyan, *Nat. Genet.,* 37:S31 (2005) [PMID: 15920528]
[08] D. R. Rhodes, *et al., Proc. Natl. Acad. Sci.,* 101:9309 (2004) [PMID: 15184677]
[09] K. B. Knudsen, *et al., Reprod. Toxicol.,* 20:369 (2005) [PMID:15923107]
[10] F. Zeng, *et al.*, *Dev. Biol.,* 272:483 (2004) [PMID:15282163]
[11] M. Tanaka, *et al., Cold Spring Harb. Symp. Quant. Biol.,* 67:317 (2002) [PMID:12858555]

_____

[12] J. Buchstaller, *et al., J. Neurosci.,* 24:2357 (2004) [PMID:15014110]

[13] C. L. Small, *et al., Biol. Reprod.,* 72:492 (2005) [PMID:15496517]

[14] P. Mukhopadhyay, *et al., Birth Defects Res.,* 70:912 (2004) [PMID:15578713]

[15] N. L. Brown, *et al., Dev. Growth Differ.,* 45:153 (2003) [PMID: 12752503]

[16] D. L. Wheeler, *et al., Nucleic Acids Res.,* 34:D173 (2006) [PMID: 16381840]

[17] J. Thioulouse *et al., Stat. Comput.,* 7:75 (1997) [PMID: 8670621]

[18] A. C. Culhane, *et al., BMC Bioinformatics,* 4:59 (2003) [PMID: 14633289]

[19] G. Dennis, *et al., Genome Biol.,* 4:P3 (2001) [PMID: 12734009]

[20] R. Anbazhagan, *Bioinformatics*, 19:157 (2003) [PMID: 12499309]

[21] D. R. Rhodes, *et al., Neoplasia,* 6:1 (2004) [PMID: 15068665]