BIOINFORMATICS ORIGINAL PAPER

Vol. 28 no. 3 2012, pages 397–402
doi:10.1093/bioinformatics/btr639

Databases and ontologies

Advance Access publication November 20, 2011

Integrating human and murine anatomical gene expression data for improved comparisons

Natalia Jiménez-Lozano[1,2,*,†] Joan Segura[3,†], José Ramón Macías[2], Juanjo Vega[2] and José María Carazo[1,2]

[1]GN7 of the National Institute for Bioinformatics (INB), [2]Biocomputing Unit of the National Centre for Biotechnology (CNB-CSIC), Darwin, 3, 28049 Madrid, Spain and [3]Computational Biology Group of the Leeds Institute of Molecular Medicine, Section of Experimental Therapeutics, Welcome Trust Brenner Building, St James's University Hospital, Leeds LS9 7TF, UK

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Information concerning the gene expression pattern in four dimensions (species, genes, anatomy and developmental stage) is crucial for unraveling the roles of genes through time. There are a variety of anatomical gene expression databases, but extracting information from them can be hampered by their diversity and heterogeneity.

Results: aGEM 3.1 (anatomic Gene Expression Mapping) addresses the issues of diversity and heterogeneity of anatomical gene expression databases by integrating six mouse gene expression resources (EMAGE, GXD, GENSAT, Allen Brain Atlas data base, EUREXPRESS and BioGPS) and three human gene expression databases (HUDSEN, Human Protein Atlas and BioGPS). Furthermore, aGEM 3.1 provides new cross analysis tools to bridge these resources.

Availability and implementation: aGEM 3.1 can be queried using gene and anatomical structure. Output information is presented in a friendly format, allowing the user to display expression maps and correlation matrices for a gene or structure during development. An in-depth study of a specific developmental stage is also possible using heatmaps that relate gene expression with anatomical components. http://agem.cnb.csic.es

Contact: natalia@cnb.csic.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on March 30, 2011; revised on October 25, 2011; accepted on November 9, 2011

1 INTRODUCTION

Different genes are expressed in various tissues at different developmental stages. Knowledge concerning the precise spatial–temporal transcript abundances of genes during development of an organism in normal as well as in diseased states provides a valuable tool for understanding the mechanisms underlying the susceptibility to complex disease.

Traditionally, gene expression has been measured using DNA microarrays and *in situ* techniques. DNA microarrays allow differential transcript levels for several genes to be measured at once (expression profiling), but information obtained from *in situ* gene expression is required for a detailed anatomical and histological snapshot of the identified genes (de Boer *et al.*, 2009). The constant increase in data and data concerning data (metadata), both from microarrays and *in situ* techniques, require organization into specialized databases. Currently, a number of diverse databases exist containing information generated using microarray techniques (ArrayExpress, Brazma *et al.*, 2003; NCBI GEO, Barrett *et al.*, 2011; Gene Expression Atlas, Wu *et al.*, 2009) and *in situ* techniques (EMAGE, Richardson *et al.*, 2010; GXD, Finger *et al.*, 2011). It has become relatively common for databases to provide *in situ* information for the implementation of new methods that allow such data to become accessible to the user in a spatial–temporal way. For example, such tools allow a user to select an anatomical location by interacting with interfaces that present sections of the organism under study at the chosen developmental stage. The combination of a gene expression data base and a spatial–temporal device to retrieve information has been named a 'gene expression atlas'. Atlases for the most relevant model species have been built including ZFIN for zebrafish (Sprague *et al.* 2006), BDGP (Tomancak *et al.*, 2007) and FlyBase (Grumbling and Strelets, 2006) for *Drosophila*, MEPD for medaka (Henrich *et al.*, 2005), EMAGE (Richardson *et al.*, 2010) and GXD (Finger *et al.*, 2011) for mouse, and HUDSEN (Kerwin *et al.*, 2010) and ArrayExpress Atlas for human (Lukk *et al.*, 2010).

Having well-structured gene expression information relating to model organisms organized into dedicated databases, it is necessary to integrate and make this diverse set of information accessible in a harmonized manner. Several initiatives have emerged that investigate the possible correlation of global patterns among orthologous genes from different species (Lu *et al.*, 2009; Ravasi *et al.*, 2010), some of them including Web Platforms. However, problems related to the proper handling of developmental stage information and anatomy mapping are still pending. Therefore, gene expression blending from various organisms requires a specific integration effort at three levels: orthologous genes finding, developmental stage mapping and anatomy mapping. The first level evolved alongside sequencing initiatives. However, the requirements for the second and third levels have recently been identified, with efforts oriented toward combining data from

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

various organisms. Let us consider, as examples, the approach followed by two mature databases, 4DXpress (Haundry *et al.*, 2007) and XSPAN (Luger *et al.*, 2004). They represent two partial solutions, in that 4DXpress tackles the issue of gene expression data blending, whereas XSPAN addresses the issue of anatomical ontology mapping, with no synergy between these approaches. Indeed, XSPAN faces the issue of linking tissues in the anatomical ontologies of the main model species (human, mouse, *Drosophila* and *Caenorhabditis elegans*) without taking advantage of ontology mapping to combine information from an organism's devoted gene expression resources. However, in 2008 4DXpress contained *in situ* hybridization data from >16 000 genes from four organisms (*Drosophila*, Medaka, zebrafish and mouse). Its main drawback is that integration has been reached exclusively using gene orthology, and there has been no anatomical mapping among various organisms.

To build aGEMv3.1, a developmental stage and anatomy mapping have been carried out. Moreover, aGEM merges gene expression information from various organisms using different technologies: *in situ* techniques and microarrays. Microarray data for human and mouse is derived from the BioGPS data base (Wu *et al.*, 2009), whereas *in situ* experiments are stored in HUDSEN (Kerwin *et al.*, 2010) and Human Protein Atlas (Uhlen *et al.*, 2010) for human and EMAGE (Richardson *et al.*, 2010), GXD (Finger *et al.*, 2011), GENSAT (Sayers *et al.*, 2011), EUREXPRESS (Diez-Roux *et al.*, 2011) and ABA (Jones *et al.*, 2009) for mouse. The information displayed in aGEM 3.1 is presented in an intuitive manner, allowing the complementation of relatively scarce data on humans with mouse data.

Currently, aGEM provides a friendly and intuitive display of gene expression information for >1 450 000 and 278 000 gene–structure pairs for mouse and human, respectively. To our knowledge, aGEM is the first tool integrating data from two different experimental techniques: microarrays and *in situ* experiments. Therefore, the aGEM Platform is in demand for cross-species meta-analysis, and is currently in use by >200 external users who have actively contributed to its development and testing since the first release in 2009. Future releases will concentrate on incorporating more human expression resources, as the number of human gene–structure pairs is somewhat limited.

2 SYSTEM AND METHODS

The integration process described in the previous section demanded extensive analysis of anatomical gene expression data resources in order to map the information into a tailored-made aGEM integrative data model, combining information concerning genes, anatomical structures and developmental stages. For genes, MGI and ENSEMBL identifiers were used as the standard for mouse and human, respectively. Orthology mapping was taken from Mouse Genome Informatics (http://www.informatics.jax.org/reports/homologymap/mouse_human.shtml).

In this cross-species framework, anatomical structures were integrated in three steps. First, a representative ontology for each organism was chosen (standard-like ontology) and anatomical terms from remaining ontologies were mapped to them. In this study, the EMAP ontologies for mouse embryo [from Theiler stage (TS) 1 to 26] and MA ontology for adult (TS28) were considered to be the standard-like ontology for mouse, while eVOC ontologies [from Carnegie Stage (CS) 1 to 20, plus adult] were the standard-like ontology for human. Secondly, anatomical vocabularies from BioGPS-mouse, ABA and GENSAT databases were aligned with terms from EMAP and MA ontologies for mouse. EUREXPRESS database was populated using EMAP ontology, and therefore no alignment was necessary. Similarly, human vocabularies from HUDSEN, Human Protein Atlas and BioGPS were mapped with eVOC-human ontology. Once all integrated mouse databases use the EMAP and MA language and human databases use the eVOC-human language, the third step consists of identifying homologies among terms from the two standard-like ontologies. This has enabled the mouse and human gene expression information to be integrated. The choice of eVOC-human ontology as the standard for human was intentional as a trade-off between a much desired mapping simplicity at the expense of a somehow limited spatial resolution, and allows us to take advantage of EMAP/eVOC-mouse and eVOC-mouse/eVOC-human mappings provided by Kruger *et al.* (Fig. 1 and Supplementary Material). These mappings have allowed gene expression to be integrated across species, in that correspondence among mouse and human terms is established by eVOC-mouse terms having a total of 1460 mouse–human common structures.

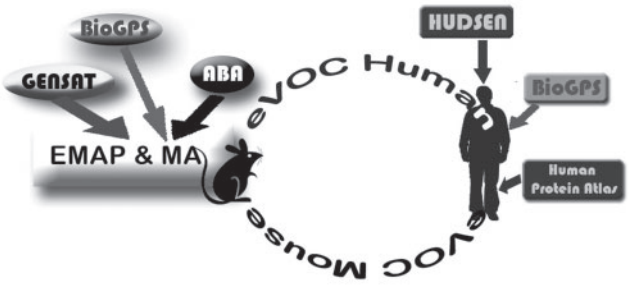

Fig. 1. Ontology mapping paradigm. Mouse vocabulary databases were mapped to EMAP and MA ontologies, and human vocabularies to the eVOC-human ontology. Mouse and human terms were mapped through EMAP/eVOC-mouse and eVOC-mouse/eVOC-human alignments provided by Kruger *et al.*

The alignment process was accomplished by a pattern matching script to find equal terms, and then one-to-one comparisons were performed for the non-automatically matched terms. This manual mapping process provided a list of analogues, or alternatively, terms similar to those of the chosen ontology. Analogue terms are defined as ways of referring to the same anatomical entity. For instance, cerebrum (ABA vocabulary) is analogous to telencephalon (EMAP/MA ontologies). When no analogue terms were identified, matching was achieved by similarity. In the case of the mouse, specific vocabularies of the central nervous system (as ABA and GENSAT) describe finer details than the chosen standard ontology (EMAP and MA). In such cases, these terms were mapped to the corresponding general EMAP and MA terms. However, although three kinds of relationships have been identified during the mapping process (equal, analogue and similar) no differences have been made among them during the internal aGEM 3.1 working.

The same procedure was applied for human, where BioGPS-human, Human Protein Atlas and HUDSEN databases were mapped to eVOC-human ontology.

EMAP ontology was developed in the framework of the Edinburgh Mouse Atlas Project (a partnership between the MRC Human Genetics Unit at Edinburgh University and the Jackson Laboratories, USA), and it is presented as a hierarchy of anatomical terms for each developmental stage, displaying the structural relationships between the anatomical entities within each stage and during development (Baldock and Davidson, 2008). Jackson Laboratories further extended the existing EMAP ontologies to the Mouse Adult ontology (MA, http://www.informatics.jax.org/searches/AMA_form.shtml; Hayamizu *et al.*, 2005) corresponding to TS28.

eVOC was developed by the South African National Bioinformatics Institute (http://bioportal.bioontology.org/ontologies/44302/?p=terms; note: to see mouse eVOC ontology terms display Mouse Development menu)

Table 1. Equivalences between Theiler stages and Carnegie stages (http://www.ana.ed.ac.uk/anatomy/database/humat/MouseComp.html)

Theile stages (Mouse)	Carnegie stages (Human)
1	1
2	2
4	3
6	4
8	5
9	6
10	7
11	8
12	9
13	10
14	11
15	12
16	13
17	14
18	15
19	16
20	17
21	18 & 19
22	20
28	Adult

and comprises orthogonal controlled vocabularies for mouse and human (eVOC ontologies; Kelso *et al.*, 2003), providing powerful means for comparisons between these two model organisms. eVOC ontologies are simplified representations describing the temporal and spatial distribution of developmental human and mouse anatomies (Kruger *et al.*, 2007).

Time is important in anatomy, and relationships between time windows (developmental stages) are crucial in this integrative Platform. aGEM uses human and mouse stage mapping provided by Jonathan Bard (Table 1).

Information concerning gene orthology, anatomical and stage mapping has been stored and is accessed through a Perl library developed at the Biocomputing Unit (Supplementary Material).

Release 2.0 of aGEM was entirely devoted to mouse (Jiménez-Lozano *et al.*, 2009). However, one of the most relevant features of the current 3.1 release is the existence of human gene expression information (HUDSEN data base for CS12-23, BioGPS and Human Protein Atlas for adult) that, together with the new 'Cross Module', allows cross-species analysis to be carried out by comparing the gene expression information from these two organisms intuitively. Moreover, new functionalities have been added including heatmaps and cluster maps, and the capability to download them.

3 ALGORITHM

Information extracted from the integrated databases has been organized following a simple aGEM tailor-made data model. aGEM consists of 11 tables; nine correspond to the nine integrated human and mouse databases (EMAGE, GXD, GENSAT, ABA, BioGPS mouse, EUREXPRESS, BioGPS human, HUDSEN and Human Protein Atlas) and two (named mouse and human core tables) contain all gene–structure pairs with corresponding statistical information collected from the individual tables. Last aGEM complete database update was carried out on July 2011, and it is regularly updated at least once per year.

Diversity in source databases is at the ontology/vocabulary level and at the gene expression annotation level. Each data base refers to the strength of gene expression in a non-numerical way, with the exception of BioGPS. To compute statistical information from the databases, heterogeneity of gene expression annotations in the various databases had to be addressed. For integration purposes, expression levels from source databases were summarized in three levels, ranging from 0 (low expression) to 2 (high expression). aGEM can be queried by gene/s or anatomical structure/s. Retrieved statistical information is displayed in four ways: expression maps, correlation matrices, hierarchical clusters and heatmaps. The first three modes are displayed independently of the fixed parameters (gene, structure or developmental stage). However, heatmaps are computed only when a developmental stage is chosen.

To compute expression maps, the more frequent value from the core tables for each gene–structure pair is extracted. In other words, the final expression value retrieved for each developmental stage corresponds to the value obtained most often among tables. Correlation among genes in a structure or among structures for a given gene is computed using the Kendal tau rank correlation coefficient.

Hierarchy of gene clusters or anatomical structure clusters is addressed using aGEMv3.1 by means of an R module. Hierarchical clusters constitute an intuitive way to display relationships among genes expressed in a chosen structure or among structures where a gene is expressed. Heatmaps describe the expression pattern in the chosen developmental stage. Heatmaps are accompanied by a gene cluster and a structure cluster. To avoid grouping into clusters of gene–structure pairs where the expression is unknown, just genes with known expression in at least 50% of the structures and structures where there is available expression data for more than 50% genes are considered in the heatmaps. The idea behind this rule is avoiding to group genes and structures with no information among them. Due to differences between human and mouse ontologies, there is a lack of synchronization between human and mouse heatmaps. Therefore, to ease users in the process of human and mouse comparisons, an additional feature that applies the aforementioned filtering process to mouse and human simultaneously has been implemented. The results are synchronized mouse and human heatmaps.

4 IMPLEMENTATION

The architecture of the Platform is based on Perl modules that query, retrieve and process information from an MySQL data base (DBI Perl). Javascript issues the dynamic web display and R generates heatmaps and hierarchical clusters.

Queries to the aGEM Platform are translated into SQL queries, mouse and human MySQL core tables are retrieved, and results are organized by three Perl modules (one for mouse, one for human and a third one for cross) and displayed in a tab arrangement using Javascript. The contents in each tab are organized into tables (Fig. 2) that link to information on annotations and expression maps that are launched on the fly by Perl modules.

aGEM 3.1 can be queried by gene and anatomical structure. The first type of query can be carried out by using the ENSEMBL identifier, common gene symbol, MGI identifier or UniProt accession number. Querying by anatomical structure is simplified by displaying the terms from the chosen developmental stage (Carnegie Stage for human or Theiler Stage for mouse) in hierarchical trees.

By integrating the KEGG pathways data base in aGEMv3.1, the user can query using a set of genes involved in a given

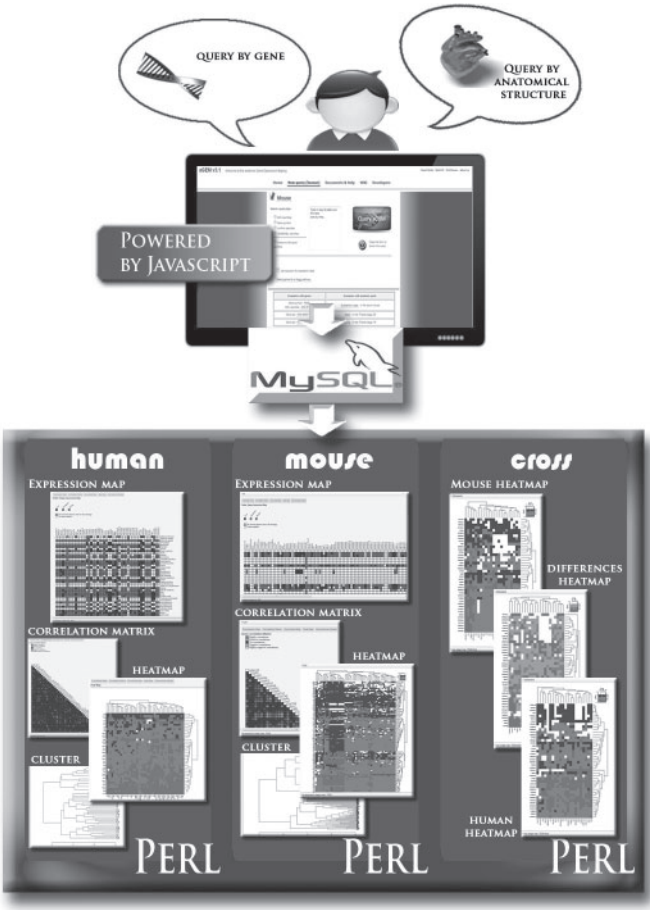

Fig. 2. aGEMv3.1 architecture.

process. This utility is very useful as it allows gene expression differences and similarities to be compared in physiological and disease states. Emphasizing the disease state, information from OMIM (human diseases) and MTB (tumours in mice) databases is displayed associated with genes in the results interface, when available.

Whether the user queries by gene or structure, the output information is presented in a format that allows the user to identify the experiments from which gene expression information has been inferred. The system also displays expression maps and correlation matrices for a gene or structure during development. Heatmaps and hierarchical cluster representations provided by aGEM allow in-depth study of the expression profile for a specific developmental stage, and this information can be downloaded in text format.

Navigation through the Platform is eased by the documentation and help sections, where the user can access a manual and tutorial describing all the outputs in depth, together with the answers to the most frequently asked questions. Furthermore, additional examples concerning mouse and human are located in the main query page.

Independently of the chosen species for the query (mouse or human), the system will display information from the organisms separately (Mouse Data tab and Human Data tab), and display cross information (Cross Data tab). To calculate co-expression among genes, aGEM takes into account the variable space (anatomical structure) and the variable time (developmental stage). Cross-species meta-analysis can be divided into two types: co-expression meta-analysis and expression meta-analysis. Co-expression meta-analysis calculates whether genes co-expressed in one species are co-expressed in another (Lu *et al.*, 2009), and aGEM3.1 allows the user to identify if genes co-expressed in mouse are co-expressed in human during development by selecting a given anatomical structure in the Cross Data section. Analysis of similarity between expression profiles of homologous genes in various species (expression meta-analysis) is gathered for each common developmental stage by selecting the 'Cross Map' or the 'Cross Heatmap' options.

Table 2. List of genes extracted from Ingenuity and Panther databases that are involved in HD

Mouse gene ID	Human symbol	Mouse gene ID	Human symbol
MGI:1306796	*APAF1*	MGI:101931	*SLC1A2*
MGI:1924182	*ARFIP2*	MGI:95805	*GRB2*
MGI:99702	*BAX*	MGI:95821	*GRIN2B*
MGI:88139	*BCL2L1*	MGI:1261831	*HAP1*
MGI:88145	*BDNF*	MGI:1099804	*HIP1*
MGI:96646	*JUN*	MGI:96433	*IGF1R*
MGI:88264	*CAPN2*	MGI:96623	*ITPR1*
MGI:1261423	*CASP8*	MGI:97321	*NGF*
MGI:96544	*CASP1*	MGI:1345181	*PACSIN1*
MGI:1312922	*CASP12*	MGI:1926119	*PDK1*
MGI:1277950	*CASP9*	MGI:104897	*REST*
MGI:101765	*CDK5*	MGI:102788	*RPH3A*
MGI:2388633	*CLTC*	MGI:700011	*SH3GL3*
MGI:104726	*CPLX2*	MGI:107157	*SIN3A*
MGI:88562	*CTSD*	MGI:1277151	*SNCA*
MGI:107745	*DCTN1*	MGI:98372	*SP1*
MGI:892995	*DNAJC5*	MGI:109355	*STX1A*
MGI:107750	*DYNC1I2*	MGI:101838	*TBP*
MGI:95290	*EGF*	MGI:98731	*TGM2*
MGI:95294	*EGFR*	MGI:1925141	*UBE2S*
MGI:95752	*GLS*		

5 DISCUSSION

The usefulness of aGEM is discussed using a specific example related to a well-known pathology, illustrating in this way the type of analyses that aGEM provides.

Huntington's disease (HD) is a rare neurodegenerative disorder of the central nervous system characterized by abnormal movements, behavioural and psychiatric disturbances, and dementia.

A set of 41 genes related to HD extracted from Ingenuity and Panther databases have used as input for aGEMv3.1 (Table 2). Heatmaps for mouse at TS28 and human adult have been obtained and analysed by using the option of synchronization, where expression information regarding non-common human and mouse anatomic structures has been removed (Fig. 3). Synchronized heatmaps show gene expression information for a total of 22 common human and mouse structures (Table 3) and 14 genes. No-synchronized heatmaps can be found in the Supplementary Material. Each position in the heatmap represents the expression strength of a gene in a specific structure. Heatmaps are especially valuable in that they allow for an intuitive clustering representation of the variable depicted in them. Consequently, in our heatmaps the clusters

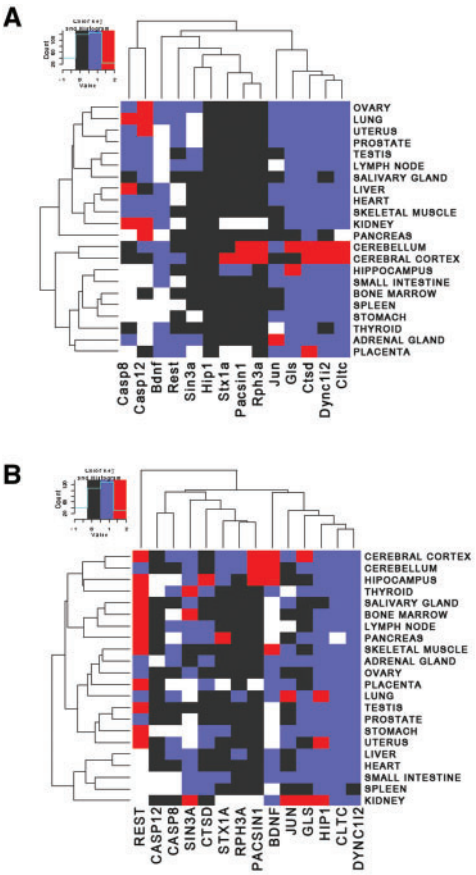

Fig. 3. Synchronized heatmaps for mouse (**A**) and human (**B**). Colors indicate the expression level that ranges from black (no expression) to red (strong expression), with blue being medium expression. White locations mean that there is no data available.

Table 3. List of anatomical structures common to mouse and human

Common human and mouse anatomic locations	
Adrenal gland	Pancreas
Bone marrow	Placenta
Cerebellum	Prostate
Cerebral cortex	Salivary gland
Heart	Skeletal muscle
Hippocampus	Small intestine
Kidney	Spleen
Liver	Stomach
Lung	Testis
Lymph node	Thyroid
Ovary	Uterus

highlight similarities in gene expression among genes (gene cluster) or among structures (structure cluster). A given structure cluster groups anatomical locations where similar expression patterns have been shown for the input genes. In much the same way, a given gene cluster groups those genes whose expression is comparable in all the studied structures. This fact is exemplified in Figure 3, where it can

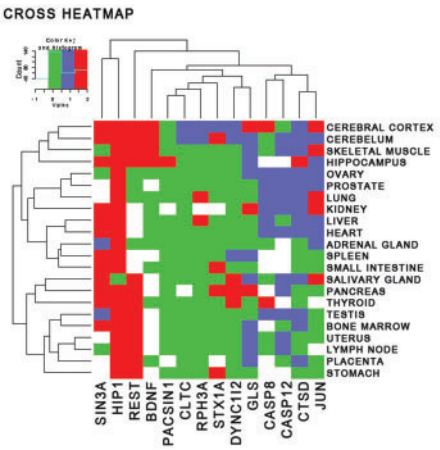

Fig. 4. Expression pattern differences between mouse and human for a set of orthologous genes related to HD displayed as a heatmap (green: same expression, blue: mouse > human, red: human>mouse).

be seen, as an example, that Casp8 and Casp12 genes have similar expression patterns in the studied structures, both for human and for mouse, as an illustration of anatomical structure clustering, cerebral cortex and cerebellum belong to the same family in the structure cluster in mouse as well as in human.

Close examination of the synchronized mouse and human heatmaps obtained from the aGEM Platform (Fig. 3) shows expression in three anatomical structures belonging to the central nervous system: cerebral cortex, cerebellum and hippocampus, belonging the two firsts to the same cluster. However, the pattern of gene expression observed in the heatmaps is not restricted to central nervous system structures; for example, expression can be observed in the endocrine system (e.g. pancreas, adrenal gland, thyroid, placenta); exocrine system (salivary gland); haemolymphoid system (bone marrow and lymph node); reproductive system (ovary, uterus, testis); digestive system (stomach, small intestine); renal system (kidney); skeletal muscle; cardiovascular system (heart); and respiratory system (lung) (Fig. 3). While some of these structures have a clear role in some HD-associated pathologies (weight loss, heart failure, skeletal muscle wasting, pneumonia, impaired glucose tolerance and testicular atrophy), others could not be clearly related.

Cross heatmaps provided by aGEM display the global patterns of tissue-specific expression of orthologous genes, allowing matches or mismatches between patterns in human and mouse to be easily identified (Fig. 4). In this case, the colors in the heatmap do not denote expression strength, but differences in this variable between mouse and human: green means same expression for mouse and human, blue implies a high expression in mouse than in human and red indicates a high expression in human than in mouse. As an illustration, *HIP1* gene expression is clearly greater in human than in mouse in all the common structures excepting salivary gland. As it was commented before, the structure and gene clusters summarize the heatmap information by highlighting set of structures or genes with similar expression difference patterns between mouse and human. In the example shown, differences in the expression of *PACSIN1* and *CLTC* genes are well conserved between mouse and human for the studied set of anatomical structures.

The example presented here aims to give the user an outlook of the expression patterns in the different anatomical locations for a concrete set of input genes related with a given pathological process. The same procedure could be employed with any other set of genes involved in any pathological or physiological process. Indeed, HD was used in this article only as a test case to attract user's attention on the usefulness and applicability of aGEM3.1. The test case does not intend to provide the basis for the demonstration of a concrete hypothesis, but to guide users in the kind of analyses that can be carried out with the described Platform. Therefore, we encourage researchers to launch aGEM with the set of genes they are working on. Hopefully, aGEM will highlight some common features among these genes that could be very helpful in the context of their scientific researching.

ACKNOWLEDGEMENTS

Professor R. Baldock and the EMAP team are acknowledged for supporting the present work. We sincerely thank the staff of EMAGE, GXD, GENSAT, ABA, KEGG, BioGPS, Human Protein Atlas, EUREXPRESS and the OMIM databases for free data base access and maintenance, and particularly Susan Lindsay for HUDSEN data base access, and Adele Kruger and Winston Hide for eVOC ontologies. We acknowledge Jesús Cuenca for useful help in aGEMv3.1 performance improvement.

Funding: National Institute for Bioinformatics (www.inab.org), a platform of 'Instituto de Salud Carlos III', Industrial and Technological Development Centre (CDTI) under the AMIT Programme (CEN-20101014), RESOLVE (UE CE:FP7-202047), Ministerio de Ciencia e Innovación (BIO2010-16566), Biostruct-X (FP7-Infrastructures-2011-1), Centrosoma 3D (CSD2006-00023).

Conflict of Interest: none declared.

REFERENCES

Baldock,R. and Davidson,D. (2008) The Edinburgh Mouse Atlas. *Anatomy Ontologies for Bioinformatics, Principles and Practice*. Springer, London, pp. 249–265.

Barrett,T. *et al.* (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.*, **39**, D1005–D1010.

Brazma,A. *et al.* (2003) ArrayExpress–a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.*, **31**, 68–71.

de Boer,B.A. *et al.* (2009) More than a decade of developmental gene expression atlases: where are we now? *Nucleic Acids Res.*, **37**, 7349–7359.

Diez-Roux,G. *et al.* (2011) A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.*, **9**, e1000582.

Finger,J.H. *et al.* (2011) The mouse Gene Expression Database (GXD): 2011 update. *Nucleic Acids Res.*, **39**, D835–D841.

Grumbling,G. and Strelets,V. (2006) FlyBase: anatomical data, images and queries. *Nucleic Acids Res.*, **34**, D484–D488.

Haudry,Y. *et al.* (2008) 4DXpress: a database for cross-species expression pattern comparisons. *Nucleic Acids Res.*, **36**, D847–D853.

Hayamizu,T.F. *et al.* (2005) The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biol.*, **6**, R29.

Henrich,T. *et al.* (2005) MEPD: a resource for medaka gene expression patterns. *Bioinformatics*, **21**, 3195–3197.

Jiménez-Lozano,N. *et al.* (2009) aGEM: an integrative system for analyzing spatial-temporal gene-expression information. *Bioinformatics*, **25**, 2566–2572.

Jones,A.R. *et al.* (2009) The Allen Brain Atlas: 5 years and beyond. *Nat. Rev. Neurosci.*, **10**, 821–828.

Kelso,J. *et al.* (2003) eVOC: a controlled vocabulary for unifying gene expression data. *Genome Res.*, **13**, 1222–1230.

Kerwin,J. *et al.* (2010) The HUDSEN Atlas: a three-dimensional (3D) spatial framework for studying gene expression in the developing human brain. *J. Anat.*, **217**, 289–299.

Kruger,A. *et al.* (2007) Simplified ontologies allowing comparison of developmental mammalian gene expression. *Genome Biol.*, **8**, R229.

Lu,Y. *et al.* (2009) Cross species analysis of microarray expression data. *Bioinformatics*, **25**, 1476–1483.

Luger,S. *et al.* (2004) Cross-species mapping between anatomical ontologies: terminological and structural support. In *Proceedings of Bio-Ontology Workshop 30 July 2004 at International Conference on Intelligent Systems for Molecular Biology (ISMB), Glasgow*.

Lukk,M. *et al.* (2010) A global map of human gene expression. *Nat. Biotechnol.*, **28**, 322–324.

Ravasi,T. *et al.* (2010) An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*, **140**, 744–752.

Richardson,L. *et al.* (2010) EMAGE mouse embryo spatial gene expression database: 2010 update. *Nucleic Acids Res.*, **38**, D703–D709.

Sayers,E.W. *et al.* (2011) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **39**, D38–D51.

Sprague,J. *et al.* (2006) The zebrafish information network: the zebrafish model organism database. *Nucleic Acids Res.*, **34**, D581–D585.

Tomancak,P. *et al.* (2007) Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.*, **8**, R145.

Uhlen,M. *et al.* (2010). Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.*, **28**, 1248–1250.

Wu,C. *et al.* (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.*, **10**, R130.