

Development 138, 2845-2853 (2011) doi:10.1242/dev.063594
© 2011. Published by The Company of Biologists Ltd

The GUDMAP database – an online resource for genitourinary research

Simon D. Harding^{1,*}, Chris Armit², Jane Armstrong², Jane Brennan², Ying Cheng¹, Bernard Haggarty¹, Derek Houghton¹, Sue Lloyd-MacGilp², Xingjun Pi¹, Yogmatee Roochun¹, Mehran Sharghi¹, Christopher Tindal¹, Andrew P. McMahon³, Brian Gottesman³, Melissa H. Little⁴, Kylie Georgas⁴, Bruce J. Aronow⁵, S. Steven Potter⁶, Eric W. Brunskill⁶, E. Michelle Southard-Smith⁷, Cathy Mendelsohn⁸, Richard A. Baldock¹, Jamie A. Davies² and Duncan Davidson¹

SUMMARY

The GenitoUrinary Development Molecular Anatomy Project (GUDMAP) is an international consortium working to generate gene expression data and transgenic mice. GUDMAP includes data from large-scale *in situ* hybridisation screens (wholemount and section) and microarray gene expression data of microdissected, laser-captured and FACS-sorted components of the developing mouse genitourinary (GU) system. These expression data are annotated using a high-resolution anatomy ontology specific to the developing murine GU system. GUDMAP data are freely accessible at www.gudmap.org via easy-to-use interfaces. This curated, high-resolution dataset serves as a powerful resource for biologists, clinicians and bioinformaticians interested in the developing urogenital system. This paper gives examples of how the data have been used to address problems in developmental biology and provides a primer for those wishing to use the database in their own research.

KEY WORDS: Database, Gene expression, Genitourinary, *In situ* hybridisation, Microarray

INTRODUCTION

Mammalian development is complicated; each of the thousands of embryological events that operate in sequence to make a mature body typically involves the concerted action of many cells in response to a multitude of simultaneous environmental signals that regulate hundreds of different genes. Traditional gene-by-gene analyses have served developmental biology well in the era when its main goal was discovering basic developmental principles from whatever organism might illustrate a principle best. Now that the focus of many developmental biologists has moved on to understanding in detail the development of specific, clinically relevant body systems with a view to preventing or repairing malformation, the gene-by-gene approach is no longer sufficient and it is necessary to consider the genome as a whole and to take a systems biology approach (Bard, 2007; Davidson, 2009; Rumballe et al., 2010). This requires comprehensive databases of

the expression of thousands of transcripts, annotated at high resolution and made easily accessible to manual and automated queries.

The development of the genitourinary (GU) system is attracting much attention for two reasons. First, humans suffer from a wide range of congenital abnormalities of urogenital organs, which can damage this system directly (kidney failures and infertility syndromes) (Schedl, 2007) and also cause indirect damage to other organs, for example because failing kidneys can raise systemic blood pressure (Hoy et al., 2005). Second, damage to the GU tract from non-developmental causes, for example infection, is repaired poorly and there is an urgent need to develop novel approaches to GU tract organ regeneration (Unbekandt and Davies, 2009). A comprehensive description of gene expression during urogenital development will be central to understanding the mechanisms of congenital urogenital abnormalities and to the development of better ways to treat them.

GUDMAP aims to provide a fundamental description of gene expression in the developing mouse GU system, generating gene expression data that will 'enable, facilitate and stimulate research' into understanding the GU system (McMahon et al., 2008). In addition, microarray analyses and the generation of transgenic mouse strains with genetic markers will bolster the overall aim of defining molecular and cellular anatomy through developmental time. The specific focus on the GU system combined with the range of data (*in situ* and microarray across a range of developmental time points) distinguishes GUDMAP from other gene expression databases.

Some gene expression data for the developing murine GU system are already publicly available, but they are dispersed across various resources that cover a range of species, organ systems and data types [GEO (Barrett et al., 2009), ArrayExpress (Parkinson et al., 2009), GXD (Smith et al., 2007), EMAGE (Richardson et al., 2010)]. The EuReGene database (<http://www.euregene.org>) holds

¹MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK. ²Center for Integrative Physiology, University of Edinburgh, Edinburgh EH8 9XD, UK. ³Department of Molecular and Cellular Biology and Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138, USA. ⁴Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, QLD 4067, Australia. ⁵Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Centre, Cincinnati, OH 45229-3039, USA. ⁶Division of Developmental Biology, Cincinnati Children's Hospital Medical Centre, Cincinnati, OH 45229-3039, USA. ⁷Department of Medicine, Division of Genetic Medicine, Vanderbilt University School of Medicine, Nashville, TN 37232, USA. ⁸Department of Urology and Pathology, Columbia University, New York, NY 10032, USA.

*Author for correspondence (sharding@hgu.mrc.ac.uk)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License (<http://creativecommons.org/licenses/by-nc-sa/3.0>), which permits unrestricted non-commercial use, distribution and reproduction in any medium provided that the original work is properly cited and all further distributions of the work or adaptation are subject to the same Creative Commons License terms.

in situ gene expression data for the developing [17.5 days post-coitum (dpc)] and adult mouse kidney, but not for other components of the GU system or time points of development. Although funding for the EuReGene project ended in 2009, there are plans to assimilate the data into the GUDMAP database and make it available once again. Additionally, the Kidney Development Database (Davies, 1999) (<http://golgi.ana.ed.ac.uk/kidhome.html>) holds text descriptions from published developmental studies that have a bearing on kidney development.

Here, we present an overview of the GUDMAP database. Our goals are to describe how the data are generated by expert GUDMAP consortium laboratories, how the quality of the data and its annotation are maintained by a dedicated editorial team and how the data are made available via the GUDMAP website (www.gudmap.org).

MATERIALS AND METHODS

Description and generation of GUDMAP gene expression data

The database currently holds in excess of 8700 in situ gene expression entries, covering nearly 3200 different genes. The vast majority is RNA in situ hybridisation (ISH) data, both from wholemount preparations of intact GU organs and from histological sections. Currently, 2894 unique genes have been analysed by wholemount in situ hybridisation (WISH) and 663 genes by section in situ hybridisation (SISH). Further information about the number of data entries and categories of information in the GUDMAP database can be found at www.gudmap.org/Website_Reports/Stats/gudmap_stats.html. The database also contains data from a small number of immunohistochemistry (IHC) assays, as well as in situ expression data from a small number of transgenic reporter strains.

GUDMAP currently holds ~300 individual microarray entries, which comprise samples representing specific kidney and urogenital subcompartments. This list of samples will soon include a 90-array timecourse on all the major sorted cell populations from the gonad spanning the period when its fate is decided (B. Capel and S. Jameson, Duke University, NC, USA, personal communication; data currently under curation). These were isolated using manual microdissection, laser capture microdissection (LCM) or fluorescence-activated cell sorting (FACS)-based techniques from fluorescently tagged transgenic animals (Brunskill et al., 2008). Microarray data files deposited on the website include both raw data (.CEL files) and normalised data.

GUDMAP gene expression data are being generated by a cycle of progressively more detailed studies. One approach taken has been to begin with complete gene sets (e.g. transcription factors; secreted molecules and their receptors) and perform an initial low-resolution analysis of expression in the entire GU tract using wholemount in situ screens. This is followed by prioritisation of subsets of genes for re-analysis at higher resolution using SISH, usually on the basis of an expression pattern that suggests compartment specificity in a particular organ (e.g. kidney, bladder, gonad). Another approach has been to begin with microarray analyses of specific time points or GU tissues followed by bioinformatic prioritisation of genes enriched in key temporospatial compartments for validation and further characterisation using SISH. Information from these detailed studies is used to focus further microarray assays on cells from new and existing transgenic mice carrying appropriately expressed reporters. An important component of this approach is to ensure that in situ and microarray data are fully integrated for interrogation and analysis. As the data have grown, the ontology has altered with it and our capacity to more accurately annotate entries has increased. As a result, existing entries are being re-analysed and annotated to a much higher level of detail.

Organisation of GUDMAP data in the database

GUDMAP entries

The GUDMAP database is organised around genes, anatomical parts of the GU system and developmental stages. The primary identifiers are as follows: for genes, Mouse Genome Informatics (MGI) gene IDs; for

anatomical parts, GUDMAP ontology terms with eMouse Atlas Project (EMAP) IDs; for stages, Theiler stage (TS) numbers (Theiler, 1989). Gene symbols and gene names are updated every few days from MGI.

Each in situ entry in the database contains the expression data for a single gene assayed with a defined probe in a single sex at a single Theiler stage.

GUDMAP recognises that it is an important feature to be able to provide a full description of the probes used in database entries to enable replication by the community. Therefore, all methodologies for probe design and the sequence of the specific probes used or generated are indicated in each entry in the database. The probe section of an entry page is highlighted in Fig. 4C. A large number of WISH and SISH entries have been hybridised with PCR-generated riboprobe sequences, as previously described (Georgas et al., 2009; Georgas et al., 2008). A link is available under 'Resources' to 'UQ GUDMAP – Probe Design' (<http://uqgudmap.imb.uq.edu.au/tools.html>) (Thiagarajan et al., 2011). 'UQ GUDMAP Tools' is provided and maintained by GUDMAP consortium members at the University of Queensland, Australia, and provides access to the tools used by this group for the design of PCR primers for the generation of locus-specific riboprobes for ISH analysis.

Each microarray entry contains gene expression data for tissues or cell populations sampled at a single developmental stage. Hence, there are multiple entries for many genes, reflecting differences in tissue/organ stage or type. The GUDMAP microarray data follow the same MIAME-compliant conventions as used in GEO (Barrett et al., 2009) and all entries are secondarily submitted to GEO. Microarray entries are called 'samples' and are grouped into 'series' that are essentially experiments that comprise groups of samples in a study.

GUDMAP ontology and annotation

GUDMAP has developed a high-resolution ontology to describe the anatomical structure of the developing murine GU tract (Little et al., 2007). Written as a partonomic, text-based, hierarchical ontology for both the embryonic and postnatal stages, it has been developed as an expansion of the existing EMAP ontology (Baldock et al., 2003; Bard et al., 1998; Davidson and Baldock, 2001) (<http://www.emouseatlas.org/>). Encompassing TS17-28 of development, (representing ~10 dpc to sexually mature adult), the ontology denotes structures that can be identified histologically (<http://www.gudmap.org/Resources/Ontologies.html>). Anatomical structures are displayed as part of one or more structural parents. For example, at TS23, the early proximal tubule is part of the renal cortex, which in turn is part of the metanephros (kidney). This organisation allows easy and flexible use of the ontology for annotation and searching.

The anatomy ontology is based on the standard Theiler developmental stages. The use of stages (e.g. TS23) rather than developmental time (e.g. 15 dpc) gives more precision to the ontology and thus to annotated gene expression entries in the database, although for the majority of entries developmental time is also specified. This is because embryos, even those in the same litter, can vary in developmental rate so that at the same developmental time several developmental stages can be represented. The GUDMAP ontology is currently being incorporated into the existing EMAP ontology, which is used by other resources such as the mouse Gene Expression Database (GXD) (Smith et al., 2007) and EMAGE (Richardson et al., 2010). To this end, each term in the GUDMAP ontology is given a unique EMAP ID.

Each data-producing laboratory uses the anatomy ontology to annotate gene expression. For a given gene, expression in anatomical structures is annotated as present, not detected, uncertain or not examined. Pattern information can be applied to refine the annotation using a controlled vocabulary of terms such as regional, spotted, graded and ubiquitous (for definitions, see the GUDMAP glossary at http://www.gudmap.org/Help/Glossary.html#Expression_Pattern). The hierarchical structure of the ontology can be used to infer expression between terms with 'part_of' relationships (http://www.obofoundry.org/ro/#OBO_REL:part_of) (Smith et al., 2005). Thus, annotations of differing resolution can be easily accommodated and when searches are performed inferences can be made up or down the hierarchy (http://www.gudmap.org/Help/Glossary.html#Inferred_Annotation).

Fig. 1. GUDMAP database search page. The main access point to the GUDMAP gene expression data. Users can access the data via simple queries (A) or by browsing through the in situ and microarray data (B). It is also possible to use more advanced anatomy queries, accessed via the Boolean anatomy button (C). A quick search tool is accessible from all website pages, enabling fast queries in the gene, anatomy, accession ID and function categories (D).

Data curation

The GUDMAP Editorial Office (EO) provides assistance with the data submission process and checks the data for completeness and compliance, including definitive checking of the identity of probes used in ISH. Full details of the EO protocols can be found at <http://www.gudmap.org/Research/Protocols/EO.html>. The EO ensures that data are immediately made publicly available and easy to access on the web interface. Additionally, the EO helps to maintain open networks of communication between consortium members and users. For example, as far as time allows, editors are on-hand to assist users with searches that are more complex than those provided on the standard interface.

RESULTS

Using GUDMAP

The starting point for using GUDMAP is the gene expression database homepage (http://www.gudmap.org/gudmap/pages/database_homepage.html) (Fig. 1). The database is organised around genes, anatomical terms and developmental stages (see Materials and methods, GUDMAP entries) and can be queried by gene (via gene symbol), by anatomy (via anatomy term), by molecular function (via Gene Ontology annotation) and by accession ID for genes, probes and entries (MGI, Entrez, Ensembl, etc.). More complex anatomical queries can be made using a 'Boolean anatomy search'. Researchers can also browse through the in situ, transgenic and microarray data in the database. The website also provides tutorials that describe how the GU system develops in mice and organ summary pages that provide an overview of gene expression in individual structures in the GU system. The website also contains a series of demonstrations (<http://www.gudmap.org/Help/Demos.html>) and a step-by-step tutorial (http://www.gudmap.org/Help/Using_GUDMAP_Tutorial.html), which illustrate how to use the GUDMAP website and database to answer relevant and interesting biological questions, demonstrating the functionality and value of GUDMAP, as well as being useful user guides.

Finding expression data for a given gene

To find where a gene is expressed during development, a simple gene query can be used to return the 'expression summary' for that gene (Fig. 2). The gene expression summary gives access to microarray expression profiles (Fig. 3), in situ expression entries (Fig. 4), in situ image galleries (Fig. 5) and disease associations for that gene. Figs 1-5 illustrate how to execute a simple query and introduce the main features/pages of the GUDMAP expression database.

Using the Boolean anatomy search to find genes that mark a structure of interest

From the expression database homepage (Fig. 1), it is possible to make a simple 'query by anatomy'. This searches for entries that have either a direct or inferred annotation to a particular anatomical component of interest. To perform a more complex anatomical query, the Boolean anatomy search can be used. This tool can search for database entries or genes across multiple anatomical components, across a range of developmental stages and with particular patterns and locations of expression. For example, it can be used to search for the co-expression of genes in components of interest or to identify genes that might mark a particular structure. In the latter case, to find genes with expression restricted to the ureteric tip, a Boolean search would look for genes with 'present' expression in the ureteric tip, but 'not detected' expression in adjacent structures (ureteric trunk, cap mesenchyme, etc.). A demonstration of the use of the Boolean anatomy search in this way is given on the website (http://www.gudmap.org/Help/Demos.html#Demo_2; http://www.gudmap.org/Help/Using_GUDMAP_Tutorial.html#Boolean).

GUDMAP transgenic mouse strains

The project has generated (and continues to add to) a resource of novel transgenic mouse strains carrying genetic markers that are either currently available or will become so in due course (Fig. 6).

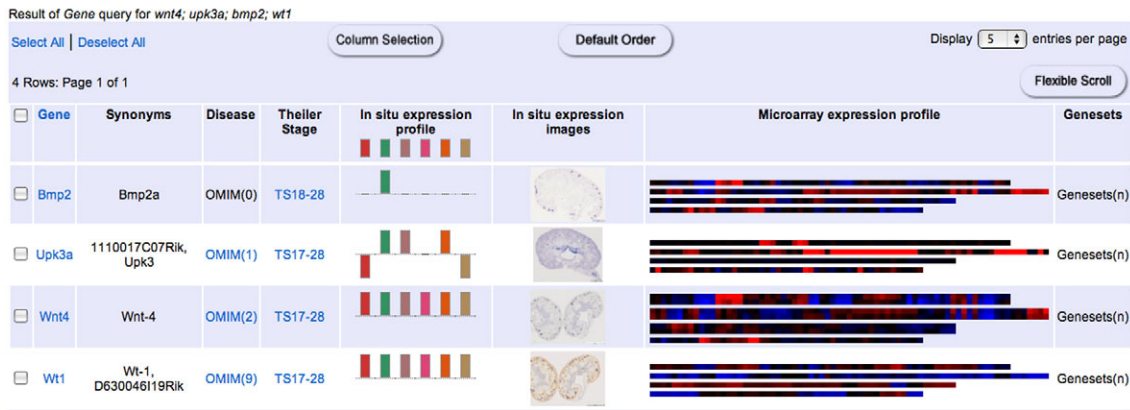


Fig. 2. GUDMAP gene expression summaries. A query for a gene, or list of genes, can be used to find where a gene(s) is expressed during development. The result of such a query returns a gene expression summary for that gene or genes. In this example, a query for *Wnt4*, *Upk3a*, *Bmp2* and *Wt1* has returned a list of four gene expression summaries, one for each gene. These summaries give an overview of the in situ and microarray gene expression data available for a gene. The microarray data are summarised as miniature heatmaps in the 'Microarray expression profile' column, which can be clicked to link to a more detailed heatmap page (see Fig. 3). The in situ data are presented as a simple bar chart, with the six coloured bars each representing an anatomical focus group [mesonephros (red), metanephros (green), lower urinary tract (purple) and early (pink), male (orange) and female (brown) reproductive systems]. A bar above the line indicates that the gene is present in that anatomy group, a bar below indicates that the gene is not detected. The bars link to lists of in situ entries (see Fig. 4). All in situ images for the gene are summarised in the image gallery (see Fig. 5), to which there is a link from the thumbnail in the 'In situ expression images' column. Expression summaries of many genes can be viewed at once, aiding quick comparisons of expression.

Strains currently posted are: Id3, Upk1b, Klf3, Osr2, Sox18, Cyp11a1, Ifitm3, Akr1b7, Upk3a, Tmem100 and S100b. These transgenic strains have been chosen mainly on the basis of in situ gene expression results in order to facilitate more detailed gene expression analyses as well as functional studies in the GU system. Verification, preliminary characterisation and information about availability are given on the website (<http://www.gudmap.org/Resources/MouseStrains/index.html>). All transgenic mice strains generated by the GUDMAP consortium that are currently available can be ordered (<http://jaxmice.jax.org/index.html>).

Tutorials of genitourinary development and organ summary pages

To help interpret GUDMAP data, a description of GU development in the mouse is given on the tutorial pages (<http://www.gudmap.org/About/Tutorial/Overview.html>). In addition, the specific development of individual components of the GU system is described on the organ summary pages (http://www.gudmap.org/Organ_Summaries/index.html), along with links to expression data for these components. These pages are supplemented with schematic diagrams (Kylie Georgas, The

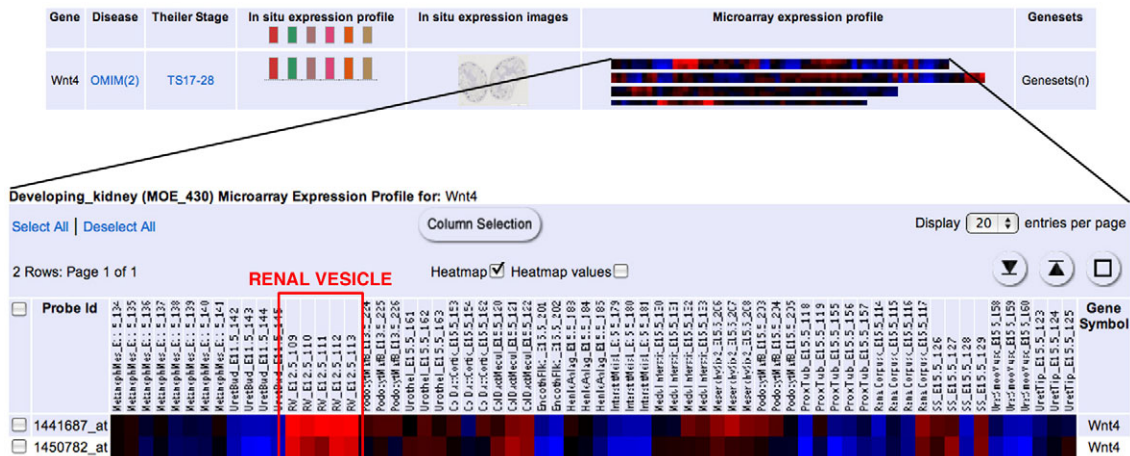


Fig. 3. GUDMAP microarray expression profiles and heatmaps. Microarray expression profiles can be accessed through the link on a gene expression summary. Shown is a microarray expression profile for *Wnt4* across components of the developing kidney in the form of a heatmap display. Each probe set of *Wnt4* is displayed in an individual row of the heatmap display, with each column representing an individual microarray sample, each corresponding to a component of the developing kidney. Microarray expression profiles are available for components of the developing kidney, lower urinary tract and reproductive system for the MOE_430 chip and for the developing kidney for the ST_1 chip. The heatmap displays a colour intensity based on the log₂ RMA normalised expression values relative to the median value for each row (probe set). Values greater than the median are red, those less than the median are blue, and values close to the median are black. In this example, it can be seen that for probe 1441687_at there is a strong indication that expression is present in the renal vesicle. For more information on GUDMAP protocols to normalise cDNA microarray CHP files, see http://www.gudmap.org/Help/Microarray_Help.html.

Sections From GUDMAP In Situ Entry Page (GUDMAP:8202, *Wnt4*, TS23, metanephros)

Images

A

Metanephros. Anti-Calb1 antibody visualised with DAB in orange and *Wnt4* RNA visualised with BM purple in blue.

Nephrogenic zone & cortex. *Wnt4* RNA signal in early nephrons. Note that Calb1 protein and *Wnt4* are absent from the connecting tubule of the ureteric tip on right.

Nephrogenic zone and cortex. Calb1 protein absent from renal connecting tubules.

Nephrogenic zone and cortex. Calb1 protein and *Wnt4* absent from connecting tubule between the ureteric tip (Calb1) and renal vesicle expressing *Wnt4* at lower left.

Nephrogenic zone. *Wnt4* in renal vesicle on left and pretubular aggregate on the right of the Calb1 expressing ureteric tips. Both *Wnt4* and Calb1 are absent from the connecting tubule.

Nephrogenic zone and cortex. *Wnt4* in early nephrons. Note S-shaped body at centre.

Probe

C

Probe Name: MGI:1997489
 Clone Name: RIKEN clone 2B10012H18
 Sequence: Fully Sequenced. Accession number for part sequence: AK012727.1
 Gene: Symbol: *Wnt4*
 Name: wingless-related MMTV Integration site 4 (MGI:98957)
 5' primer sequence: GCCTAGCCTTCACAGTCC
 3' primer sequence: CGATGTTAATACGACTCACATATAGGCTTGAGGACCCAAAACCAA
 5' primer location: 672
 Curator Notes: Editor's Note: Sequence analysis using Blast suggests no significant similarity between the 3' primer sequence and the specified accession ID
 Origin of Clone used to make the Probe: Strain:
 Tissue:
 Probe Type: RNA
 Type: antisense
 Labelled with: digoxigenin
 Visualisation method: alkaline phosphatase + BM purple

Expression Mapping

View annotated components as a list Show annotation under groups

Expression Strengths Key:

- Present (unspecified strength)
- Present (strong)
- Present (moderate)
- Present (weak)
- Uncertain
- Not Detected

Expression Patterns Key:

- Graded
- Regional
- Spotted
- Ubiquitous
- Restricted
- Single cell
- Contains note

B

JavaScript Tree Menu

- mouse (EMAP:25788)
 - organ system (EMAP:7428)
 - visceral organ (EMAP:7436)
 - urinary system (EMAP:8217)
 - mesentery (EMAP:8224)
 - metanephros (EMAP:8226)
 - renal capsule (EMAP:8237)
 - nephrogenic zone (EMAP:27724)
 - nephrogenic interstitium (EMAP:27731)
 - cap mesenchyme (EMAP:27738)
 - pretubular aggregate (EMAP:27745)
 - ureteric tip (EMAP:27752)
 - ureteric tree terminal branch excluding tip itself (EMAP:31048)
 - renal cortex (EMAP:8236)
 - renal vesicle (EMAP:27831)
 - comma-shaped body (EMAP:27837)
 - upper limb of comma-shaped body (EMAP:27843)
 - lower limb of comma-shaped body (EMAP:27849)
 - renal connecting segment of comma-shaped body (EMAP:31712)
 - S-shaped body (EMAP:27855)

Fig. 4. GUDMAP in situ expression entry pages. The components of a GUDMAP database in situ entry page, as shown for GUDMAP:8208, which is for *Wnt4* in the metanephros at Theiler stage (TS) 23 (http://www.gudmap.org/gudmap/pages/ish_submission.html?id=GUDMAP%3A8208). The full entry page contains details of the submission information, in situ images, expression mapping (annotations), probe and specimen details. (A) The images section contains thumbnails of the images provided for the entry. These link to the original images so that they can be viewed in detail using the image viewer (see Fig. 5). (B) The expression mapping section shows the anatomical annotations for the entry. In this example, the annotations indicate that *Wnt4* is expressed in the comma-shaped body (CSB), with strong expression in the upper limb of the CSB and weak expression in its lower limb. Each annotation can give an indication of expression strength, along with any pattern information and any additional notes. (C) The entry page for the submission also includes full details of the probe used in the assay. This includes the probe type, labelling methods, primer locations and sequence, and has links to the probe details at MGI (where applicable). These details are given to enable replication by the community.

University of Queensland, Australia) that serve to illustrate the developing components of the mouse GU system across different stages. For users seeking to identify genes that have expression patterns that are similar to those of their own interest, these schematics provide a simple visual representation that relates the spatial distribution of gene expression to anatomical ontology terms that can be readily searched in the database.

Disease and phenotype data

Further development of the GUDMAP resource has involved obtaining disease-gene associations and building these data into the GUDMAP database architecture. This enables disease data to be integrated with the existing gene expression data, enhancing GUDMAP as a research tool for GU development and disease. A separate part of the web interface (http://www.gudmap.org/gudmap_dis/index.jsp) allows expression data to be accessed by searching for genes that are associated with a disease or phenotype of interest, or by finding genes that share a similar pattern of phenotype or disease association with a gene of interest. Disease-gene associations are taken primarily from Online Mendelian Inheritance in Man (OMIM; <http://www.ncbi.nlm.nih.gov/omim>) (Amberger et al., 2009), published by the National Center for Biotechnology Information (NCBI), and phenotype associations are taken from MGI (<http://www.informatics.jax.org/>

[phenotypes.shtml](http://www.informatics.jax.org/phenotypes.shtml)). Disease and phenotype associations are determined for all genes that have in situ data in GUDMAP. Disease-gene associations are obtained in two ways: directly from NCBI and by text-mining OMIM entries for orthologous human gene symbols. If a gene symbol is present, an association between the disease and the gene is assumed and the entry can be assessed manually to confirm the nature of the association. Full details of the methods used to obtain these associations are described on the GUDMAP website (http://www.gudmap.org/gudmap_dis/Dis_Info.jsp).

Submission of data to GUDMAP

Until recently, only data from the GUDMAP consortium have been accepted for publication in the database. Now it is possible to enter data from other sources. If you would like to submit data, please contact the GUDMAP EO (gudmap-editors@gudmap.org); high-resolution SISH data will be given priority.

DISCUSSION

Key outcomes of GUDMAP data analysis

The data contained within the GUDMAP database have already proven valuable in advancing the understanding of morphology and morphogenesis within the GU tract. Analyses have led to the identification of key markers of specific time points or compartments

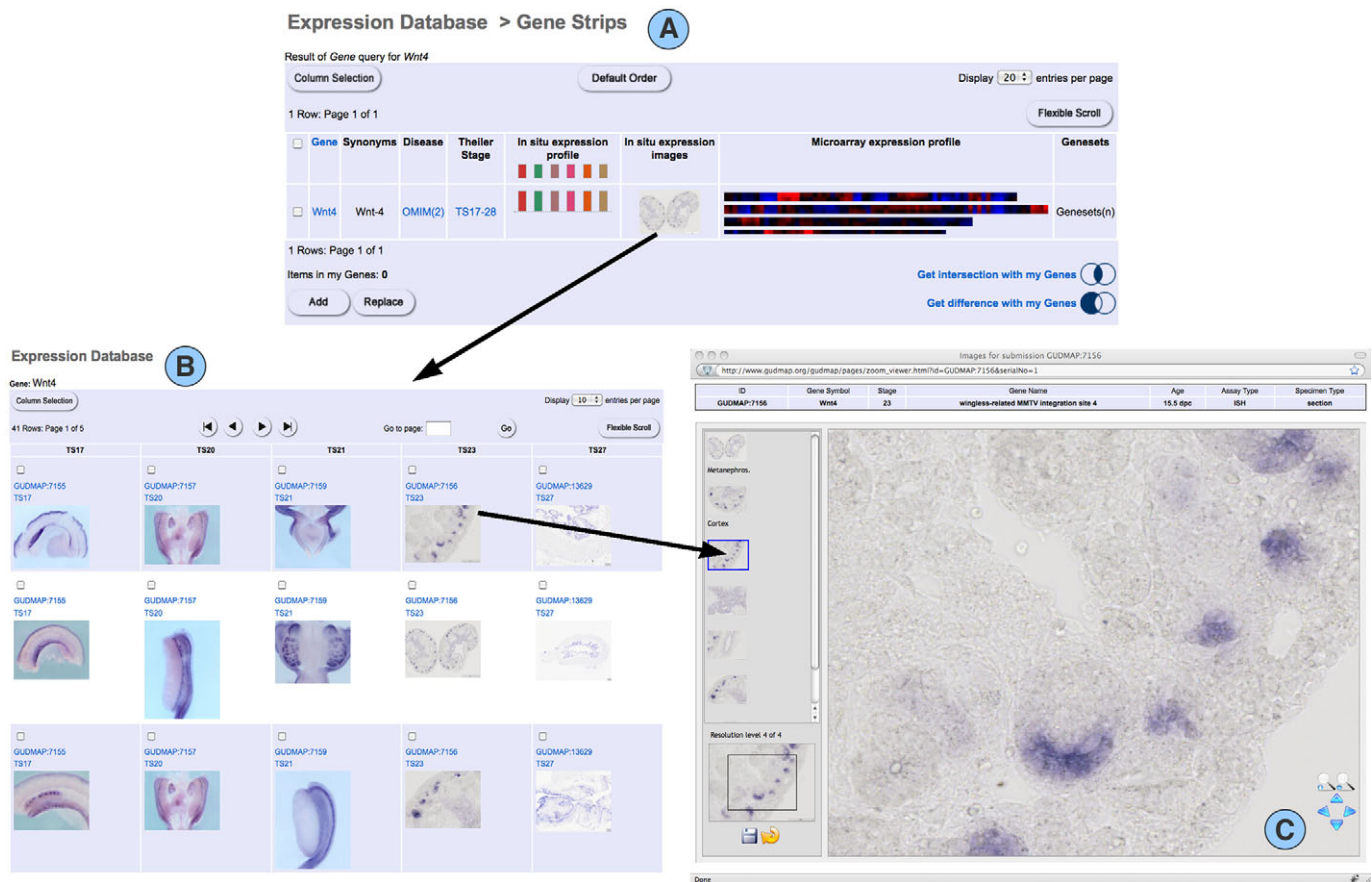


Fig. 5. GUDMAP image galleries and image viewer. (A) Gene expression summaries also link to an image gallery of all the in situ images from all entries for that gene. (B) Shown is a selection from the image gallery for *Wnt4*. Images are organised into columns by Theiler stage. Next to each image is a link to the individual entry from which the image came. (C) Whether accessing images from a GUDMAP entry page (see Fig. 4) or from the image gallery, all images on the GUDMAP web interface can be viewed in detail using the image viewer by clicking on the image thumbnail. The viewer enables users to zoom in on an image to pick up fine details and view other images from the same entry. Being able to view images easily and in such detail assists in understanding the annotations for a given entry and in understanding the expression of a gene in a given structural component(s).

(Brunskill et al., 2008; Thiagarajan et al., 2011), the identification of previously unidentified subcompartments (Chiu et al., 2010; Georgas et al., 2009; Mugford et al., 2009) and the association between gene networks and key developmental processes (Brunskill et al., 2008; Chiu et al., 2010). It has also been referred to, or utilised by, a number of other studies investigating gene expression and development (Combes et al., 2009; Dallosso et al., 2009; Gerber et al., 2009; Parreira et al., 2009; Shah et al., 2010; Surendran et al., 2010).

Generation of an atlas of gene expression in the developing kidney

Brunskill et al. generated microarray gene expression data for 15 separate subcompartments of the developing kidney collected using either LCM of anatomical compartments based on lectin staining or FACS from fluorescently tagged transgenic mice (Brunskill et al., 2008). Analysis of these data enabled the mapping of changes in expression profile during successive stages of nephron development. Employing precise sampling of tissues for microarray expression analysis, the authors showed that developmentally related components display a high degree of correlative gene expression and that certain structures, for example the early proximal tubule, show highly restricted gene expression patterns. Furthermore, the bioinformatic identification of transcription factor binding sites in the proposed minimal

promoters of genes enriched in specific compartments of the developing kidney allowed the prediction of key genetic pathways crucial for the development of renal subcompartments. This included the identification of downstream targets of HNF1B in the early proximal tubule-enriched gene set and overlapping sets of Lcf/Lef targets in other compartments of the kidney. As such, this study represents the most comprehensive atlas of temporospatial gene expression for any developing organ. Microarray data were validated using ISH, the results of which are available from the GUDMAP database; for example, Entry ID GUDMAP:9112 (*Prnp*) (http://www.gudmap.org/gudmap/pages/ish_submission.html?id=GUDMAP%3A9112). The results of the analysis reported by Brunskill et al. (Brunskill et al., 2008) can be accessed on the GUDMAP website (http://www.gudmap.org/gudmap/pages/genelist_folder.html) and researchers can also download raw microarray data for their own analyses (see http://www.gudmap.org/Help/Download_Help.html).

Subcompartment-restricted anchor genes for the prioritisation of reporter mouse strains

The microarray data generated by Brunskill et al. (Brunskill et al., 2008) will allow continued bioinformatic interrogation for a number of purposes. Already, it has facilitated a key aim of the GUDMAP consortium – the identification of genes that specifically mark single

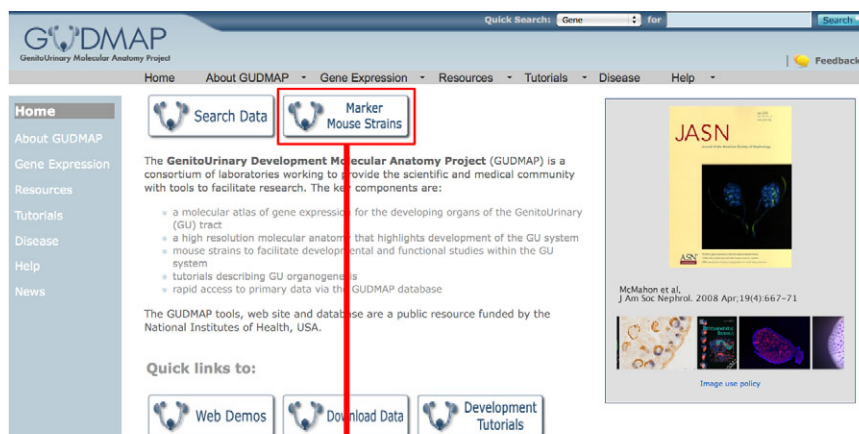


Fig. 6. GUDMAP transgenic mouse strains. Accessing the transgenic mouse strains generated by GUDMAP is simple. From the GUDMAP home page (www.gudmap.org), click on the 'Marker Mouse Strains' button. This takes you to a table of the novel transgenic mouse strains generated by the project. For each strain there are links to both the allele verification and characterisation information. There is also a link to JAX Mice & Services, from where the strains can be ordered.

| Sort ID | Gene | Reporter Allele | Allele Verification (Zhang Lab) | Allele Characterisation (McMahon Lab) | Strain available from Jackson Laboratory | Organ* | Cell Type* |
|---------|-------------------|--------------------|---------------------------------|---------------------------------------|---|----------------------------------|--|
| 1 | Id3 | Id3-RFP | download | download | http://jaxmice.jax.org/strain/010983.html | Kidney | Broad expression including ureteric tree terminal branch excluding tip itself, renal cortical interstitium and renal capsule |
| 2 | Upk1b | Upk1b-RFP | download | download | http://jaxmice.jax.org/strain/010984.html | Lower urinary tract | Urothelium of ureter and epithelium of bladder |
| 3 | Klf3 | Klf3-GFP-Cre-ERT2 | download | download | http://jaxmice.jax.org/strain/010985.html | Kidney, reproductive system | Collecting duct, renal interstitium. Testis, ovary. |
| 4 | Osr2 | Osr2-RFP | download | download | http://jaxmice.jax.org/strain/010986.html | Kidney, male reproductive system | Cortical renal tubule, highest in proximal segment of s-shaped body. Urethra and testis. |
| 5 | Sox18 | Sox18-GFP-Cre-ERT2 | download | download | http://jaxmice.jax.org/strain/010987.html | Vasculature | Renal cortical and medullary vasculature |
| 6 | Cyp11a1 | Cyp11a1-GFP-Cre | download | download | http://jaxmice.jax.org/strain/010988.html | Male reproductive system | Leydig cell (and adrenal gland) |
| 7 | Ifitm3 (Fragilis) | Ifitm3-RFP | download | download | http://jaxmice.jax.org/strain/013139.html | Reproductive system | Reproductive system - germ cell |
| 8 | Akr1b7 | Akr1b7-TagRFP-T | download | download | http://jaxmice.jax.org/strain/013137.html | Kidney | Medullary vasculature |
| 9 | Upk3a | Upk3a-GCE | | download** | | | |
| 10 | Tmem100 | Tmem100-GCE | | download** | | | |
| 11 | S100b | S100b-GCE | | download** | | | |

Link out to JAX Mice & Services, where mice can be ordered.
<http://jaxmice.jax.org/index.html>

developmental compartments within the GU tract. Such genes will serve to prioritise the development of transgenic reporter mice of value to the research community. Using a stringent bioinformatic filter, Thiagarajan et al. (Thiagarajan et al., 2011) selected ~250 putative anchor genes, which are defined as genes with expression restricted to one of 11 subcompartments within the developing mouse kidney. Two hundred of these genes were validated using high-resolution SISH, thereby identifying 37 anchor genes across six compartments (early proximal tubule, medullary collecting duct, ureteric tip, renal vesicle, loop of Henle and renal corpuscle). Five anchor genes were identified for the medullary collecting duct (*Gsdmc4*, *Upk3a*, *Fam129a*, *Clmn* and *A1836003*), four in the renal corpuscle [*Gpsm3*, *Tard5*, RIKEN clone *C230096N06* (MGI 2416283) and *Vip*] and one each in the ureteric tip (*Slo4c1*), renal vesicle (*Npy*) and loop of Henle (*Umod*). Reflecting the initial observation of a strong cluster of proximal tubule genes, 25 of these anchor genes showed specific expression in the early proximal tubule.

The fully annotated expression patterns, probe details and ISH images for this complete set of validated genes are available on the GUDMAP website. For example, *Spp2* in the early proximal

tubule can be found in Entry ID GUDMAP:9147 (http://www.gudmap.org/gudmap/pages/ish_submission.html?id=GUDMAP%3A9147) and *Umod* in the loop of Henle can be found in Entry ID GUDMAP:9104 (http://www.gudmap.org/gudmap/pages/ish_submission.html?id=GUDMAP%3A9104). Precise, curated descriptions of the molecular probes used for in situ assays are given on the GUDMAP site. This information can be used to determine which transcripts are being assayed, for example in the context of jointly analyzing in situ and microarray results. Microarray and in situ data for a given gene are best accessed from the gene expression summaries, where there are links to extensive information on the genes, genomic locations, phenotypes, functions and pathway information at the MGI, University of California at Santa Cruz (UCSC) Genome Browser and Kyoto Encyclopedia of Genes and Genomes (KEGG) sites. Researchers can use GUDMAP data, in the context of information from these resources, to manually explore relationships within gene lists produced by their own analyses or to explore the key results from bioinformatics analyses carried out in silico.

Use of GUDMAP data to interrogate a key developmental process

The availability of microarray data for individual anatomical compartments is a key strength of GUDMAP. Further analysis of genes enriched in a specific compartment based upon microarray data has assisted in the subdivision of these anatomical compartments into previously unidentified molecular subcompartments. A key example of this was reported by Georgas et al. (Georgas et al., 2009), who used GUDMAP microarray expression to identify genes with enriched expression in the renal vesicle and decreased expression in *Wnt4* mutants. Sixty-three genes were identified and subjected to high-resolution SISH, the result of which indicated a subdivision of this structure into distal renal vesicle (closest to the adjacent ureteric tip; e.g. *Papss2*, GUDMAP:8960, http://www.gudmap.org/gudmap/pages/ish_submission.html?id=GUDMAP%3A8960) and proximal renal vesicle (farthest from the adjacent ureteric tip; e.g. *Tmem100*, GUDMAP:8888, http://www.gudmap.org/gudmap/pages/ish_submission.html?id=GUDMAP%3A8888). From their SISH analysis, the authors went on to show that the distal part of the late renal vesicle, marked by *Lhx1* and *Bmp2*, fuses with the adjacent part of the collecting system, the ureteric tip, and that this process involves the insertion of cap mesenchyme-derived cells that express renal vesicle markers into the ureteric tip at the late renal vesicle stage. This is earlier than had previously been proposed.

Mugford et al. used WISH data contained in the GUDMAP database as the starting point for an investigation into the compartmentalisation of the nephron progenitor population (Mugford et al., 2009). Using the results of the genome-wide low-resolution WISH screen of transcription factors available on the GUDMAP site, the authors identified 45 genes expressed in the cap mesenchyme. These genes were selected for further analysis by high-resolution SISH and annotated using the GUDMAP anatomy ontology (<http://www.gudmap.org/Resources/Ontologies.html>). Based on the annotations of the in situ results, the authors revealed ten distinct categories of gene expression patterns, indicating a complexity of cell states that had not previously been appreciated. From this categorisation, Mugford et al. were able to begin to dissect the roles of transcription factors and signalling pathways in spatially and molecularly distinct cell populations in the cap mesenchyme during the earliest stages of nephron induction and differentiation.

A final example comes from the lower urinary tract, for which microarray analyses comparing its different regions at embryonic days 13 and 14 and subsequent WISH analyses have been undertaken. This revealed novel domains of gene expression on the dorsal genital tubercle. Bioinformatics interrogation of this gene cluster predicted a network of *Wnt7a*-associated genes involved in the epidermal development of the genital tubercle (Chiu et al., 2010).

Conclusions

The GUDMAP website provides free access to the most comprehensive gene expression dataset of the GU system in the developing mouse. Combining both in situ and microarray data it serves as a powerful resource for developmental biologists, clinicians and bioinformaticians alike. GUDMAP data have already provided insight into the progression of gene expression during nephrogenesis, the genetic regulatory mechanisms of kidney development and into gene expression patterning in the early nephron. The illustration of the resource that we provide here serves as a primer for researchers with an interest in the developing GU system.

Acknowledgements

We acknowledge help from members of the GUDMAP consortium whose data are represented on the GUDMAP website, including Blanche Capel, Samantha Jameson, Kevin Gaido, Sean Grimmond, Peter Koopman, Jim Lessard, Chad Vezina and Pumin Zhang and members of their groups. We especially thank M. Todd Valerius, Deborah Hoshizaki and Elizabeth Wilder for their contributions to the website. M.H.L. is a Principal Research Fellow of the National Health and Medical Research Council, Australia. This work is supported by the National Institutes of Health via DK070136 (M.H.L.), DK070200 (D.D.) and DK070181 (A.P.M.).

Competing interests statement

The authors declare no competing financial interests.

References

- Amberger, J., Bocchini, C. A., Scott, A. F. and Hamosh, A. (2009). McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.* **37**, D793-D796.
- Baldock, R. A., Bard, J. B., Burger, A., Burton, N., Christiansen, J., Feng, G., Hill, B., Houghton, D., Kaufman, M., Rao, J. et al. (2003). EMAP and EMAGE: a framework for understanding spatially organized data. *Neuroinformatics* **1**, 309-325.
- Bard, J. (2007). Systems developmental biology: the use of ontologies in annotating models and in identifying gene function within and across species. *Mamm. Genome* **18**, 402-411.
- Bard, J. L., Kaufman, M. H., Dubreuil, C., Brune, R. M., Burger, A., Baldock, R. A. and Davidson, D. R. (1998). An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. *Mech. Dev.* **74**, 111-120.
- Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I. F., Soboleva, A., Tomashevsky, M., Marshall, K. A. et al. (2009). NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* **37**, D885-D890.
- Brunskill, E. W., Aronow, B. J., Georgas, K., Rumballe, B., Valerius, M. T., Aronow, J., Kaimal, V., Jegga, A. G., Yu, J., Grimmond, S. et al. (2008). Atlas of gene expression in the developing kidney at microanatomic resolution. *Dev. Cell* **15**, 781-791.
- Chiu, H. S., Szucsik, J. C., Georgas, K. M., Jones, J. L., Rumballe, B. A., Tang, D., Grimmond, S. M., Lewis, A. G., Aronow, B. J., Lessard, J. L. et al. (2010). Comparative gene expression analysis of genital tubercle development reveals a putative appendicular Wnt7 network for the epidermal differentiation. *Dev. Biol.* **344**, 1071-1087.
- Combes, A. N., Lesieur, E., Harley, V. R., Sinclair, A. H., Little, M. H., Wilhelm, D. and Koopman, P. (2009). Three-dimensional visualization of testis cord morphogenesis, a novel tubulogenic mechanism in development. *Dev. Dyn.* **238**, 1033-1041.
- Dallosso, A. R., Hancock, A. L., Szemes, M., Moorwood, K., Chilukamari, L., Tsai, H. H., Sarkar, A., Barasch, J., Vuononvirta, R., Jones, C. et al. (2009). Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 5q31 in Wilms' tumor. *PLoS Genet.* **5**, e1000745.
- Davidson, D. and Baldock, R. (2001). Bioinformatics beyond sequence: mapping gene function in the embryo. *Nat. Rev. Genet.* **2**, 409-417.
- Davidson, E. H. (2009). Developmental biology at the systems level. *Biochim. Biophys. Acta* **1789**, 248-249.
- Davies, J. A. (1999). The kidney development database. *Dev. Genet.* **24**, 194-198.
- Georgas, K., Rumballe, B., Wilkinson, L., Chiu, H. S., Lesieur, E., Gilbert, T. and Little, M. H. (2008). Use of dual section mRNA in situ hybridisation/immunohistochemistry to clarify gene expression patterns during the early stages of nephron development in the embryo and in the mature nephron of the adult mouse kidney. *Histochem. Cell Biol.* **130**, 927-942.
- Georgas, K., Rumballe, B., Valerius, M. T., Chiu, H. S., Thiagarajan, R. D., Lesieur, E., Aronow, B. J., Brunskill, E. W., Combes, A. N., Tang, D. et al. (2009). Analysis of early nephron patterning reveals a role for distal RV proliferation in fusion to the ureteric tip via a cap mesenchyme-derived connecting segment. *Dev. Biol.* **332**, 273-286.
- Gerber, S. D., Steinberg, F., Beyeler, M., Villiger, P. M. and Trueb, B. (2009). The murine Fgfr1 receptor is essential for the development of the metanephric kidney. *Dev. Biol.* **335**, 106-119.
- Hoy, W. E., Hughson, M. D., Bertram, J. F., Douglas-Denton, R. and Amann, K. (2005). Nephron number, hypertension, renal disease, and renal failure. *J. Am. Soc. Nephrol.* **16**, 2557-2564.
- Little, M. H., Brennan, J., Georgas, K., Davies, J. A., Davidson, D. R., Baldock, R. A., Beverdam, A., Bertram, J. F., Capel, B., Chiu, H. S. et al. (2007). A high-resolution anatomical ontology of the developing murine genitourinary tract. *Gene Expr. Patterns* **7**, 680-699.
- McMahon, A. P., Aronow, B. J., Davidson, D. R., Davies, J. A., Gaido, K. W., Grimmond, S., Lessard, J. L., Little, M. H., Potter, S. S., Wilder, E. L. et al. (2008). GUDMAP: the genitourinary developmental molecular anatomy project. *J. Am. Soc. Nephrol.* **19**, 667-671.

- Mugford, J. W., Yu, J., Kobayashi, A. and McMahon, A. P. (2009). High-resolution gene expression analysis of the developing mouse kidney defines novel cellular compartments within the nephron progenitor population. *Dev Biol.* **333**, 312-323.
- Parkinson, H., Kapushesky, M., Kolesnikov, N., Rustici, G., Shojatalab, M., Abeygunawardena, N., Berube, H., Dylag, M., Emam, I., Farne, A. et al. (2009). ArrayExpress update-from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.* **37**, D868-D872.
- Parreira, K. S., Debaix, H., Cnops, Y., Geffers, L. and Devuyst, O. (2009). Expression patterns of the aquaporin gene family during renal development: influence of genetic variability. *Pflugers Arch.* **458**, 745-759.
- Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y. Y., Burton, N., Rao, J. G., Fisher, M., Baldock, R. A., Davidson, D. R. and Christiansen, J. H. (2010). EMAGE mouse embryo spatial gene expression database: 2010 update. *Nucleic Acids Res.* **38**, D703-D709.
- Rumballe, B., Georgas, K., Wilkinson, L. and Little, M. (2010). Molecular anatomy of the kidney: what have we learned from gene expression and functional genomics? *Pediatr. Nephrol.* **25**, 1005-1016.
- Schedl, A. (2007). Renal abnormalities and their developmental origin. *Nat. Rev. Genet.* **8**, 791-802.
- Shah, M. M., Sakurai, H., Sweeney, D. E., Gallegos, T. F., Bush, K. T., Esko, J. D. and Nigam, S. K. (2010). Hs2st mediated kidney mesenchyme induction regulates early ureteric bud branching. *Dev Biol.* **339**, 354-365.
- Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. L. and Rosse, C. (2005). Relations in biomedical ontologies. *Genome Biol.* **6**, R46.
- Smith, C. M., Finger, J. H., Hayamizu, T. F., McCright, I. J., Eppig, J. T., Kadin, J. A., Richardson, J. E. and Ringwald, M. (2007). The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Res.* **35**, D618-D623.
- Surendran, K., Boyle, S., Barak, H., Kim, M., Stomberski, C., McCright, B. and Kopan, R. (2010). The contribution of Notch1 to nephron segmentation in the developing kidney is revealed in a sensitized Notch2 background and can be augmented by reducing Mint dosage. *Dev Biol.* **337**, 386-395.
- Theiler, K. (1989). *The House Mouse: Atlas of Embryonic Development*. New York: Springer.
- Thiagarajan, R. D., Georgas, K. M., Rumballe, B., Lesieur, E., Chiu, H. S., Taylor, D., Tang, D. T. P., Grimmond, S. and Little, M. H. (2011). Identification of anchor genes during kidney development defines ontological relationships, molecular subcompartments and regulatory pathways. *PLoS ONE* **6**, e17286.
- Unbekandt, M. and Davies, J. (2009). Control of organogenesis: towards effective tissue engineering. In *Fundamentals of Tissue Engineering and Regenerative Medicine* (ed. U. Meyer, T. Meyer, J. Handschel and H. P. Wiesmann), pp. 61-70. Berlin, Germany: Springer.