SPECIAL ARTICLE www.jasn.org

GUDMAP: The Genitourinary Developmental Molecular Anatomy Project

Andrew P. McMahon,* Bruce J. Aronow,† Duncan R. Davidson,‡ Jamie A. Davies,§ Kevin W. Gaido,‖ Sean Grimmond,¶ James L. Lessard,** Melissa H. Little,¶ S. Steven Potter,** Elizabeth L. Wilder,†† and Pumin Zhang,‡‡ for the GUDMAP project

*Department of Molecular and Cellular Biology and Harvard Stem Cell Institute, Harvard University, Cambridge, Massachusetts; †Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; ‡Medical Research Council Human Genetics Unit, Western General Hospital, and §Centre for Integrative Physiology, University of Edinburgh, Edinburgh, United Kingdom; ‖Hamner Institutes for Health Sciences, Research Triangle Park, North Carolina; ¶Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland, Australia; **Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; ††Renal and Urogenital Development, Kidney Injury and Repair, and Basic Science of Cystic Kidney Disease Research, National Institutes of Health Roadmap Interdisciplinary Research Working Group, Bethesda, Maryland; and ‡‡Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas

ABSTRACT

In late 2004, an International Consortium of research groups were charged with the task of producing a high-quality molecular anatomy of the developing mammalian urogenital tract (UGT). Given the importance of these organ systems for human health and reproduction, the need for a systematic molecular and cellular description of their developmental programs was deemed a high priority. The information obtained through this initiative is anticipated to enable the highest level of basic and clinical research grounded on a 21st-century view of the developing anatomy. There are three components to the Genitourinary Developmental Molecular Anatomy Project GUDMAP; all of these are intended to provide resources that support research on the kidney and UGT. The first provides ontology of the cell types during UGT development and the molecular hallmarks of those cells as discerned by a variety of procedures, including *in situ* hybridization, transcriptional profiling, and immunostaining. The second generates novel mouse strains. In these strains, cell types of particular interest within an organ are labeled through the introduction of a specific marker into the context of a gene that exhibits appropriate cell type or structure-specific expression. In addition, the targeting construct enables genetic manipulation within the cell of interest in many of the strains. Finally, the information is annotated, collated, and promptly released at regular intervals, before publication, through a database that is accessed through a Web portal. Presented here is a brief overview of the Genitourinary Developmental Molecular Anatomy Project effort.

J Am Soc Nephrol 19: 667–671, 2008. doi: 10.1681/ASN.2007101078

One in nine Americans has chronic kidney disease (CKD), and the same number is at increased risk for developing CKD. Approximately 12% of women in the United States have a problem conceiving or bringing a baby to term. Approximately 3% of all children are born with a recognized birth defect. Some of the most common are those that affect the genitourinary system, with cryptorchidism and hypospadias being the most common male birth defect. Furthermore, less obvious developmental deficiencies likely underlie or influence susceptibility to a range of diseases of the urogenital tract (UGT), as well as adversely affecting fertility. One example is the link between nephron number, determined at the outset by a fetal developmental program, and hypertension and renal disease later in life.[1]

This increased realization of a link between development and disease and the opportunities that might come from exploiting our developmental understanding toward the treatment of disease have thrown a spotlight on the normal developmental programs that orchestrate development of our organ systems. Which cells make up an organ? How are they generated? Which factors control these events in time and space, ensuring a co-herent, functional system? How are organs repaired? What controls the numbers of any given cell type and the overall balance of cell number in the organ? Addressing any of these pertinent and fascinating questions requires a solid foundation in developmental anatomy. The overarching goal of the Genitourinary Developmental Molecular Anatomy Project (GUDMAP) initiative is to establish this foundation for the UGT. The ensuing data generated by the initiative will enable, facilitate, and stimulate research into, and our understanding of, this critical organ network.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Andrew P. McMahon, Department of Molecular and Cellular Biology and Harvard Stem Cell Institute, Harvard University, 16, Divinity Avenue, Cambridge, MA 02138. Phone: 617-496-3757; Fax: 617-496-3763; E-mail: amcmahon@mcb.harvard.edu; consortium Web address: http://www.gudmap.org

E.L.W.'s current affiliation is Office of Portfolio Analysis and Strategic Initiatives, National Institutes of Health, Bethesda, Maryland.

INTRODUCTION TO GUDMAP

In September 2001, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) was advised by a working group of its National Advisory Council to form a strategic plan to consider how stem cells and developmental biology might be used to explore the repair and replacement of damaged organs and to determine how they might provide insights into the pathologic processes underlying developmental defects and disease. A separate working group of the National Advisory Council highlighted developmental biology as an understudied area of bladder research. These reports led the NIDDK to convene a panel of advisors in January 2003 to discuss approaches that would provide a fundamental description of the developing UGT.

The working group recommended several strategies to reach this goal: (*1*) A high-throughput *in situ* analysis of gene expression in the developing UGT, (*2*) a high-resolution analysis of a limited set of genes in time and space to define the emergence of anatomic and functional domains, and (*3*) the creation of a database to make these data readily and promptly available to the research community.

Requests for applications were issued in late 2003, and a funded consortium of seven laboratories was formed from the applicants late in 2004. The group consisted of six data-generating centers: Five in the United States (A.P.M., K.W.G., J.L.L., S.S.P., and P.Z.), one in Australia (S.G. and M.H.L.) and a database in the United Kingdom (D.R.D. and J.A.D.). Subsequently, consideration of the Strategic Plan for Pediatric Urology, developed by the NIDDK in 2006, and additional advice from an external panel of expert consultants led the NIDDK and the National Institute of Child Health and Human Development to add three collaborators to the group to maximize the effectiveness and coverage of the GUDMAP effort: Two in the United States (B.J.A. and Michelle Southard-Smith) and the other in Australia (Peter Koopman). GUDMAP is a 5-yr project jointly funded by the NIDDK and the National Institute of Child Health and Human Development.

AN ANATOMIC ONTOLOGY FOR THE ANNOTATION OF THE DEVELOPING UGT

Early discussions among the members of the consortium highlighted the need to develop a high-resolution anatomy-based ontology of the UGT. This ontology would not only provide a common descriptive language for cell types within the developing UGT, an essential prerequisite for any search algorithm and the logical construction of the critical GUDMAP database, but also draw these terms into a developmental series of hierarchical relationships among the cell populations of the UGT. M.H.L. spearheaded the important goal of a high-resolution anatomic ontology of the UGT in consultation with GUDMAP members, members of the EuReGene network (http://www.euregene.org), and other international experts in the field. The ontology was published in 2007.[2] This ontology provides a framework for annotating the expression data. Furthermore, the expression studies provide an ongoing test and challenge to the ontology.

The ontology will inevitably evolve as molecular approaches reveal molecularly distinct cell types that are not currently incorporated into histology-based anatomy. We urge that this ontology be adopted broadly within the research community. A common standard for annotating data will maximize the value of that data. For example, there will be a better understanding as to which cell type is being discussed if the terms are defined and accepted. Furthermore, this would more readily facilitate the incorporation of data generated outside of the GUDMAP consortium at some time in the future.

MOLECULAR ANATOMY OF THE DEVELOPING UGT

Whole-mount *in situ* hybridization (WISH) is being used as a low-resolution approach to map gene expression domains and identify patterns that are diagnostic of specific anatomic regions or individual cell types. WISH has been used generally to survey expression of specific classes of genes within the UGT. Transcriptional regulators, as a functional category, are valuable cell type–specific markers in several developmental studies.[3] Furthermore, their known regulatory actions can provide insights into possible actions within expressing cells, stimulating further experimental inquiry. Optimizing a WISH procedure for sensitivity, penetration, and specificity, a systematic, genome-scale screen was performed for the spatial expression of a majority of mouse transcriptional factors in at embryonic day 15.5 UGT (Yu *et al.* in preparation; http://www.gudmap.org/Research/Protocols/). The annotated data representing approximately 1500 mouse transcriptional factors can be queried through the GUDMAP database. In addition, an analysis of the expression pattern of genes classified as extracellular on the basis of gene ontology terms and previously described to be expressed during kidney development has been submitted.[4] A second comprehensive genome-scale screen of mammalian signals and their receptors is under way, together with prioritized gene sets based on spatial expression profiling of the developing UGT (see below). All data are released on a regular basis through the GUDMAP database. Together, these screens have identified a large number of genes that have spatially discrete patterns of expression and map to distinct regional domains within each organ system.

Furthermore, genes can be clustered into groups on the basis of these patterns. The three-dimensional view afforded by WISH analysis is particularly valuable in identifying specific patterns that might be harder to identify in section analysis. The low-resolution WISH data also provide an excellent prescreen for further high-resolution section *in situ* hybridization (SISH) analysis, increasing the likelihood that any gene will give useful information.

SISH and immunohistochemistry

(IHC) provide high-resolution approaches that enable gene activity to be mapped at single-cell or near-single-cell resolution within a tissue (http://www.gudmap.org/Research/Protocols/).
SISH also identifies synexpression clusters—genes that are "coexpressors"—prioritizing these for in-depth annotation of specific domains. Large-scale SISH screens have formed the basis for molecular anatomies of the developing mouse nervous system[3] and the adult brain,[5] enhancing research on the central nervous system. High-throughput SISH studies are under way within several groups within GUDMAP addressing each component part of the UGT. A complementary, comprehensive initiative for the kidney has been initiated within the EuReGene consortium (http://www.euregene.org).

The analysis of bulk transcriptional activity through microarray-based technology is providing unprecedented information into cell-, tissue-, and organ-type gene expression programs. Transcriptional analysis of the wild-type UGT at different stages of development[4,6–8] and within specific cellular or structural subcompartments of those organs[4,8–10] has provided temporal and spatial profiles of gene expression where *de novo* expression of a gene may potentially correlate with the emergence of new cell identities and distinct structures. In GUDMAP, microarray profiling has been used in several ways.

In the first instance, increased spatial resolution is being sought through the direct isolation of specific cell types or anatomic structures and their subsequent transcriptional profiling. This provides detailed insight into local gene expression programs. Cell populations of interest have been isolated by manual dissection, by laser capture microscopy,[11] or through FACS where markers are available (see below). After amplification of resulting cDNA, the gene expression profile is determined predominantly through the use of high-density Affymetrix arrays.

Currently, profiling data exist for most organs of the UGT. Within the kidney, 12 cellular compartments have been analyzed using these approaches. The primary transcriptional profiling data and the emerging expression studies based on these data are available on the GUDMAP Web site. Profiling is also being applied to mutant mouse strains in which the primary developmental defect is known but the underlying mechanism remains to be fully elucidated. For example, *Wnt4* mutants fail to generate epithelial renal vesicles, the precursor for the main body of the nephron.[12] Thus, *Wnt4* mutants form the basis of a screen that enriches for genes representative of the renal vesicle or its more differentiated nephron derivatives. Identifying genes whose expression is downregulated in *Wnt4* mutants identified several hundred candidates. Analysis of their expression showed that half are restricted to *Wnt4*-dependent structures.[13] Thus, the screen significantly enriches for genes that demarcate the target component of the kidney. In both instances, accompanying WISH and SISH of the resultant gene lists acts both to validate the data and to identify novel subcompartments of the structures initially profiled, thereby expanding our developmental ontology to include domains that are identifiable only *via* gene expression.

Figure 1 provides an overview of the workflow and data generation within GUDMAP. Although individual groups may have a more organ-specific interest, the consortium has attempted to maximize the effectiveness of the entire group's effort to define UGT development molecularly.

MOUSE STRAINS FOR THE ANALYSIS OF UGT DEVELOPMENT

Transgenic approaches enable the generation of novel mouse strains in which key structures are labeled in an organ of interest.[14] The generation of novel mouse strains is an important part of the GUDMAP initiative. The consortium goal has been to identify cell-type or anatomic regions of interest within the UGT by the expression of a specific gene. Gene targeting is then used to introduce, under the control of the cell/tissue-type marker gene, expression constructs that enable these cells to be identified, isolated, and

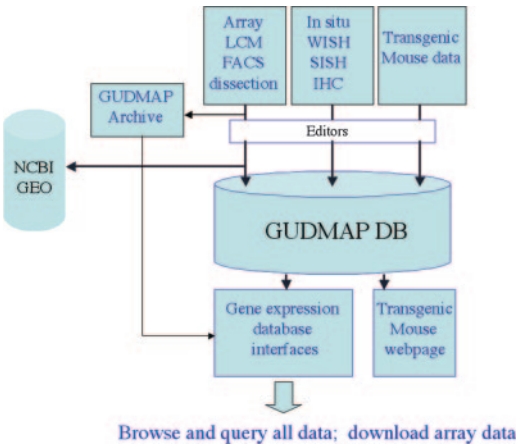

Figure 1. Data flow in the GUDMAP. Gene expression results from microarray and *in situ* assays are submitted by consortium laboratories to a central editorial office. Microarray data includes .cel, .chp, .rpt, and .txt files. *In situ* data include images, definitive descriptions of probes, and experimental details; gene expression patterns are annotated with a standard anatomic ontology by the data-producing laboratories. *In situ* and microarray submissions are curated by editors and entered into the database within a few days of receipt. For rapid data release, uncurated microarray files are made available for download from an archive Web page on the day of submission. A suite of interfaces to the GUDMAP database make the data available to researchers. In addition, other bioinformatics resources will have direct programmatic access to the GUDMAP database. Curated information about the consortium mouse strains is also available from the Web site.

genetically manipulated. In one example, a GFP:Cre fusion protein encoding gene[15] is introduced through embryonic stem cell–mediated gene targeting. In mice resulting from germline transmission of the targeted embryonic stem cell genome, the cell type of interest can be readily identified by GFP fluorescence in the native tissue. In addition, cells can be FACS sorted and subjected to high-resolution transcriptional profiling, or the fate of the cell type of interest determined whereby CRE-mediated recombination is used to activate any of a number of Cre reporter mouse strains.

The consortium goal is to generate 30 to 40 new mouse lines. Each of these strains will undergo an initial analysis within the consortium to verify that the reporter expression matches the expression of the endogenous gene to which it is targeted and to confirm the activity of any additional regulatory components (*e.g.*, Cre-based DNA excision). The choice of gene/cell type to be targeted to date has been directed either by an existing objective to characterize a specific cell population or on the advice of the research community with respect to cells types for which current tools were inadequate.

Details of the selected genes, targeting events, and progress can be viewed at the GUDMAP Web site (http://www.gudmap.org/Resources/MouseStrains/). All strains will be made publicly available through a third-party distribution center. Selection of the center is under way. Community input is welcomed in the selection of cell types for marking, candidate genes for targeting, and vector selection for the generation of new mouse strains (http://www.gudmap.org/Contact/Mice.html).

GUDMAP DATABASE

The Web-based GUDMAP database (accessible through http://www.gudmap.org) houses all of the data generated by the consortium (Figure 2). This includes primary WISH, SISH, IHC, microarray-based transcriptional profiling, mouse strain characterization data together with subsequent follow-up analyses of these data sets, and information regarding methods, research tools, and community resources relevant to the UGT.

The Web-based GUDMAP database has been publicly accessible since April 2006. Consortium members submit data on a monthly basis to the GUDMAP editorial office, where they are curated and made available within the database, normally within a few days of receipt. Importantly, the database provides an effective way to query and view data generated by the consortium. The initial priority was to provide immediate access to data from a variety of simple interfaces. Accordingly, typing any text term in the quick search slot enables the querying of the entire database. Alternatively, entries in the database can be browsed in table form; columns in the table can be re-ordered to segregate WISH, SISH, IHC, and array data or to sort data by developmental age/stage, by gene symbol, and so forth. Standard query interfaces allow researchers to list genes expressed in any particular anatomic structure or to find the expression pattern of any particular gene; an advanced interface lets users build custom queries to access every field in the database. Database users can choose to focus on particular parts of the UGT, for example metanephros or female reproductive system. Using any of these functions, researchers can then make collections of database entries and explore their intersection with the results of further queries. In this way, for example, one can find genes expressed in the nephrogenic zone in the newborn kidney but not in the nephrogenic zone at embryonic day 15.5.

The database is an evolving resource. We anticipate that the database will provide sophisticated views of data to enable researchers to build new hypotheses about gene function in the UGT. For example, a researcher will be able enter a list of genes, perhaps from his or her own work, and view all related data in the database, enabling him or her to explore pathways and gene interactions on the basis of a detailed picture of gene expression in the genitourinary system. Pages linked to the database will list congenital genitourinary diseases and associated

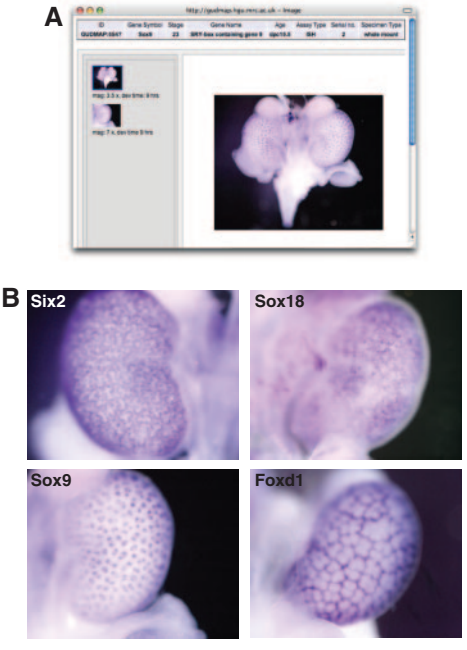

Figure 2. Gene expression patterns in the GUDMAP database. (A) Web shot of a WISH screen page in the GUDMAP database. (B) WISH analysis of mammalian transcriptional regulators in embryonic day 15.5 UGT. Four images from the database reflect distinct regional domains in the expression of the indicated regulators (Jing Yu, unpublished data).

genes with expression data from the project. Other pages will list details of transgenic mice generated by the project and their preliminary phenotypic characterization.

In addition to these functions, the database will provide the outlet for bioinformatic studies that will maximize the information and knowledge gained from available GUDMAP data. The key bioinformatic challenge for the consortium will be to interrogate, integrate, and mine all data sets to identify the best unique molecular descriptor for every known anatomic component within the UGT. We also expect to define new molecular subcompartments in the process. This will in turn inform the biology and direct decision making with regard to development of further animal tools in this field. We anticipate that the availability of the data to the community will result in the identification of synexpression groups, whereby coexpression of genes will assist the elucidation of transcriptional regulatory networks.

The consortium data will also enable meta-analyses, for example, comparison of UGT development with development of other organ systems to reveal shared regulatory principles, with mouse models of disease and dysgenesis of the UGT to explore mechanism and pathologies, and eventually intersecting murine and human data to inform better on patient material. Finally, the protocols, tutorials, links, and resources on the GUDMAP Web site should make this information base an important "one-stop shop" for researchers exploring the UGT. In keeping with this idea, efforts are under way to expand the coverage of this site, for example, synthesizing information on human genetic diseases of the UGT and highlighting relevant mouse models.

ACKNOWLEDGMENTS

We acknowledge the tremendous group effort of many individual researchers in their own laboratories toward the GUDMAP goals. A.P.M. is particularly grateful to Todd Valerius and Jing Yu for the preparation of figures in this article.

DISCLOSURES

A.P.M. is a consultant for Merck.

REFERENCES

1. Hoy WE, Hughson MD, Bertram JF, Douglas-Denton R, Amann K: Nephron number, hypertension, renal disease, and renal failure. *J Am Soc Nephrol* 16: 2557–2564, 2005
2. Little MH, Brennan J, Georgas K, Davies JA, Davidson DR, Baldock RA, Beverdam A, Bertram JF, Capel B, Chiu HS, Clements D, Cullen-McEwen L, Fleming J, Gilbert T, Herzlinger D, Houghton D, Kaufman MH, Kleymenova E, Koopman PA, Lewis AG, McMahon AP, Mendelsohn CL, Mitchell EK, Rumballe BA, Sweeney DE, Valerius MT, Yamada G, Yang Y, Yu J: A high-resolution anatomical ontology of the developing murine genitourinary tract. *Gene Expr Patterns* 7: 680–699, 2007
3. Gray PA, Fu H, Luo P, Zhao Q, Yu J, Ferrari A, Tenzen T, Yuk DI, Tsung EF, Cai Z, Alberta JA, Cheng LP, Liu Y, Stenman JM, Valerius MT, Billings N, Kim HA, Greenberg ME, McMahon AP, Rowitch DH, Stiles CD, Ma Q: Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* 306: 2255–2257, 2004
4. Challen G, Gardiner B, Caruana G, Kostoulias X, Martinez G, Crowe M, Taylor DF, Bertram J, Little M, Grimmond SM: Temporal and spatial transcriptional programs in murine kidney development. *Physiol Genomics* 23: 159–171, 2005
5. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR, Fliss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR, Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawal R, Kidney JM, Knapik RH, Kuan CL, Lake JH, Laramee AR, Larsen KD, Lau C, Lemon TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE, Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA, Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova NV, Sivisay T, Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN, Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY, Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B, Zwingman TA, Jones AR: Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445: 168–176, 2007
6. Challen GA, Martinez G, Davis MJ, Taylor DF, Crowe M, Teasdale RD, Grimmond SM, Little MH: Identifying the molecular phenotype of renal progenitor cells. *J Am Soc Nephrol* 15: 2344–2357, 2004
7. Martinez G, Georgas K, Challen GA, Rumballe B, Davis MJ, Taylor D, Teasdale RD, Grimmond SM, Little MH: Definition and spatial annotation of the dynamic secretome during early kidney development. *Dev Dyn* 235: 1709–1719, 2006
8. Schwab K, Patterson LT, Aronow BJ, Luckas R, Liang HC, Potter SS: A catalogue of gene expression in the developing kidney. *Kidney Int* 64: 1588–1604, 2003
9. Caruana G, Cullen-McEwen L, Nelson AL, Kostoulias X, Woods K, Gardiner B, Davis MJ, Taylor DF, Teasdale RD, Grimmond SM, Little MH, Bertram JF: Spatial gene expression in the T-stage mouse metanephros. *Gene Expr Patterns* 6: 807–825, 2006
10. Mitchell EK, Taylor DF, Woods K, Davis MJ, Nelson AL, Teasdale RD, Grimmond SM, Little MH, Bertram JF, Caruana G: Differential gene expression in the developing mouse ureter. *Gene Expr Patterns* 6: 519–538, 2006
11. Potter SS, Hartman HA, Kwan KM, Behringer RR, Patterson LT: Laser capture-microarray analysis of Lim1 mutant kidney development. *Genesis* 45: 432–439, 2007
12. Stark K, Vainio S, Vassileva G, McMahon AP: Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by Wnt-4. *Nature* 372: 679–683, 1994
13. Valerius MT, McMahon AP: Transcriptional profiling of Wnt4 mutant mouse kidneys identifies genes expressed during nephron formation. *Gene Expr Patterns* 2008, in press
14. Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, Heintz N: A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* 425: 917–925, 2003
15. Gagneten S, Le Y, Miller J, Sauer B: Brief expression of a GFP cre fusion gene in embryonic stem cells allows rapid retrieval of site-specific genomic deletions. *Nucleic Acids Res* 25: 3326–3331, 1997