



Original article

Mouse IDGenes: a reference database for genetic interactions in the developing mouse brain

Michaela Matthes^{1,2,†}, Martin Preusse^{3,4,†}, Jingzhong Zhang^{1,†},
Julia Schechter¹, Daniela Mayer¹, Bernd Lentens¹, Fabian Theis^{4,5},
Nilima Prakash^{1,*}, Wolfgang Wurst^{1,6,7,8,9,*}, Dietrich Trümbach^{1,6,*}

¹Institute of Developmental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, ²Technische Universität München-Weihenstephan, Lehrstuhl für Genetik, Emil-Ramannstr. 8, 85354 Freising, Germany, ³Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, ⁴Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, ⁵Technische Universität München, Zentrum Mathematik, Boltzmannstr. 3, 85747 Garching, Germany, ⁶Max-Planck-Institute of Psychiatry, Kraepelinstr. 2-10, 80804 München, Germany, ⁷Deutsches Zentrum für Neurodegenerative Erkrankungen e. V. (DZNE), Standort München, Schillerstr. 44, 80336 München, Germany, ⁸Technische Universität München-Weihenstephan, Lehrstuhl für Entwicklungsgenetik, c/o Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany and ⁹Munich Cluster for Systems Neurology (SyNergy), Adolf-Butenandt-Institut, Ludwig-Maximilians-Universität München, Schillerstr. 44, 80336 München, Germany

*Corresponding author: Tel: ++49-(0)89-3187-3341; Fax: ++49-(0)89-3187-3099;
Email: dietrich.truembach@helmholtz-muenchen.de

Correspondence may also be addressed to Wolfgang Wurst. Tel: ++49-(0)89-3187-4111; Fax: ++49-(0)89-3187-3099;
Email: wurst@helmholtz-muenchen.de and Nilima Prakash. Tel: ++49-(0)89-3187-2275; Fax: ++49-(0)89-3187-3099;
Email: nilima.prakash@helmholtz-muenchen.de

[†]These authors contributed equally to this work.

Citation details: Matthes, M., Preusse, M., Zhang, J., *et al.* Mouse IDGenes: a reference database for genetic interactions in the developing mouse brain. *Database* (2014) Vol. 2014: article ID bau083; doi:10.1093/database/bau083

Received 27 May 2014; Revised 18 July 2014; Accepted 23 July 2014

Abstract

The study of developmental processes in the mouse and other vertebrates includes the understanding of patterning along the anterior–posterior, dorsal–ventral and medial–lateral axis. Specifically, neural development is also of great clinical relevance because several human neuropsychiatric disorders such as schizophrenia, autism disorders or drug addiction and also brain malformations are thought to have neurodevelopmental origins, i.e. pathogenesis initiates during childhood and adolescence. Impacts during early neurodevelopment might also predispose to late-onset neurodegenerative

disorders, such as Parkinson's disease. The neural tube develops from its precursor tissue, the neural plate, in a patterning process that is determined by compartmentalization into morphogenetic units, the action of local signaling centers and a well-defined and locally restricted expression of genes and their interactions. While public databases provide gene expression data with spatio-temporal resolution, they usually neglect the genetic interactions that govern neural development. Here, we introduce Mouse IDGenes, a reference database for genetic interactions in the developing mouse brain. The database is highly curated and offers detailed information about gene expressions and the genetic interactions at the developing mid-/hindbrain boundary. To showcase the predictive power of interaction data, we infer new Wnt/ β -catenin target genes by machine learning and validate one of them experimentally. The database is updated regularly. Moreover, it can easily be extended by the research community. Mouse IDGenes will contribute as an important resource to the research on mouse brain development, not exclusively by offering data retrieval, but also by allowing data input.

Database URL: <http://mouseidgenes.helmholtz-muenchen.de>.

Introduction

Brain formation during vertebrate development is a complex process that has been studied for decades. The understanding of neuronal development is a prerequisite for the fight not only against neurodegenerative diseases, e.g. Parkinson's disease, but also toward neuropsychiatric disorders in particular schizophrenia, autism disorders and drug addiction.

The emergence of the neural tube from the neural plate and the patterning of these structures along their anterior–posterior, dorsal–ventral and medial–lateral axes are fundamental processes during vertebrate neural development. The formation of forebrain, midbrain, hindbrain and spinal cord is determined by well-defined and locally restricted expression of genes and their gene regulatory networks (1). Whereas the patterning of the dorso–ventral axis depends on the relative amounts of dorsalizing and ventralizing factors such as the bone morphogenetic protein (BMP) and Sonic hedgehog (Shh), respectively, the patterning along the anterior–posterior axis is usually accomplished by local signaling centers such as the isthmic organizer (IsO) (2). The IsO, which is necessary and sufficient for the development of mesencephalic and metencephalic structures, is located at the boundary between midbrain and hindbrain and is, therefore, also referred to as the mid-/hindbrain boundary (MHB). The IsO also controls the generation of clinically highly relevant cell populations such as the ventral midbrain dopaminergic neurons, which are involved in Parkinson's disease, schizophrenia and drug addiction, or the rostral hindbrain serotonergic neurons, which take part in mood disorders and depression. Therefore, the MHB or IsO is not only of developmental importance but also of high clinical relevance and thus subject of intense investigations (1–10). Up to now,

four stages are thought to be necessary for the development of the MHB: (i) positioning and establishment, (ii) induction, (iii) maintenance and (iv) morphogenesis (1, 2, 7).

Positioning of the future MHB is almost exclusively achieved by the cross-inhibitory interaction of orthodenticle homolog 2 (Otx2) and gastrulation brain homeobox 2 (Gbx2), two transcription factors initially expressed in the anterior and posterior part of the developing embryo, respectively. The inductive mechanism for these two and other factors of the IsO in the neural plate are still unknown. Wingless-type MMTV integration site family member 1 (Wnt1) and fibroblast growth factor 8 (Fgf8) are two factors secreted from the anterior and posterior region of the MHB, respectively. Wnt1 is required for the maintenance of the MHB, and Fgf8 is necessary for the patterning of the midbrain and rostral hindbrain. The engrailed genes *En1* and *En2* as well as the paired box transcription factors *Pax2* and *Pax5* act up- and downstream of Wnt1 and Fgf8, mediating their maintenance as well as patterning function at the MHB (1, 2).

Advances in understanding the signaling cascades that give rise to distinct neuronal populations open new prospects for clinical therapies, like stem cell-based treatments. On the other hand, it allows clinicians to classify malformations of the brain more precisely, as with the help of embryology and genetics the major categories of a classification are the causative genes and their pathways and not exclusively the clinical phenotype (8–10).

The gene expression in neural development has been subject to many large-scale studies, and the results were stored in publically available databases. The most important of these resources were recently reviewed (11) and in the following a few will be exemplified. The mouse gene expression database developed by Mouse Genome

Informatics (MGI) is a community resource for gene expression information from the laboratory mouse (12). It is designed as a database to collect and integrate raw expression data from a wide range of sources, such as RNA *in situ* hybridization, immunohistochemistry, western blots, northern blots and RT-PCR. Other databases focus on *in situ* hybridization data: The Allen Developing Mouse Brain Atlas [part of the Allen Brain Atlas (13)], GenePaint.org (14) and the e-Mouse Atlas of Gene Expression (15). Further, there is the Mouse Atlas of Gene Expression, which collects expression data based on serial analysis of gene expression (SAGE) (16). SAGE is more quantitative than *in situ* hybridization but lacks the high spatial resolution. While experimental methods like western and northern blots as well as RT-PCR are not suitable to derive information about exact spatial gene expression (e.g. within single cell populations), the main disadvantage of immunohistochemistry often represents the lack of a functional antibody. Notably, for many of the genes expressed at the MHB suitable antibodies are not available.

To facilitate, for example, dynamic modeling approaches, which necessitate a priori knowledge, i.e. highly curated data, a comprehensive collection of known genetic interactions containing spatial and temporal information is essential (17). These dynamic approaches are means to provide valuable insight into biological problems (17–20). Another interesting field for which integration of interaction and expression data was applied represents the prediction of new transcription factor binding sites (TFBSs) by using statistical models (21, 22). Although, plenty of gene expression data for the developing mouse brain are publicly accessible, interaction databases such as STRING (23), IntAct (24) or BioGRID (25) do not provide high spatial resolution on a developmental time scale. Thus, additional information about the interaction type [IEXP (inferred from experiment and/or expression pattern), ‘direct’, ‘direct signaling’, ‘indirect’, ‘indirect signaling’ or ‘maintenance’, which means to keep a gene active/inactive if it was already turned on/off] and mode (i.e. activation or repression) is not yet available for the specific genetic interaction network at the MHB and other brain regions.

We thus developed Mouse IDGenes, which represents a manually curated reference database for genetic interactions in the developing mouse brain focusing on the MHB, but with the possibility to add gene expression and interaction data of the central nervous system (CNS) with the help of a graphical user interface. The freely available database can be accessed via a Web interface through the URL <http://mouseidgenes.helmholtz-muenchen.de>. The Web interface offers detailed information about the expression of genes and their genetic interactions in the developing mid-/hindbrain region. Stored data were already

used in part to understand regulatory gene interactions on the systems level (26, 27). Therefore, the resource provides the possibility for the simulation of the processes occurring at the MHB, which is a unique feature of the presented Web page. The Mouse IDGenes project is conceived for a continuous expansion of stored gene expression and interaction data. Users can enter new data in the database via the Web interface. Currently, 89 spatio-temporally resolved *in vivo* gene expression data sets and 145 genetic interaction data sets from 154 original publications assigned to different anatomical regions at mouse embryonic developmental stages E8.5, E10.5 and E12.5 (Theiler Stages 13, 17 and 20) are available from the database.

Brain regionalization model

We introduced a CNS regionalization model, which covers developmental stages E8.5, E10.5 and E12.5 (Theiler Stages 13, 17, and 20) representing three crucial stages in the development of the murine MHB and mid-/hindbrain region (Figure 1, Supplementary Table S1). The development of the MHB and of the mid-/hindbrain region initiates after gastrulation is finished. At E10.5, the establishment of the IsO at the MHB is completed, and its function at this stage is well characterized, meaning that most of the known interactions are taking place at this time point. The neural tube already exists at this stage, whereas specific neuronal populations have not developed yet. These neuronal populations, however, are first identifiable at around E12.5. Our model about the mouse anatomy is based on data reviewed from literature (28–32) and the MGI database (<http://www.informatics.jax.org>). To comply with the Edinburgh Mouse Atlas Project (EMAP), ontologies of mouse developmental anatomy, which provides a standard nomenclature for the description of normal and mutant mouse anatomy (33) and, therefore, to allow reusability of the data, we provide EMAP identifiers and descriptions for the defined brain regions (Supplementary Table S1). Because EMAP ontologies were recently updated to the EMAPA ontology (34), we mapped the presented brain regions also to these identifiers (Supplementary Table S1). The regionalization model was kept as general as possible, but as exact as necessary and covers initially the anterior–posterior compartmentalization of the neural tube into the brain vesicles and spinal cord. These brain vesicles correspond to the regions of the brain, which have already developed at a given developmental stage. Within these compartments, we further divided the regions on the anterior–posterior (i.e. longitudinal) axis into lateral and medial or dorsal and ventral parts depending on the corresponding developmental stage. Therefore, we developed a tripartition of a respective

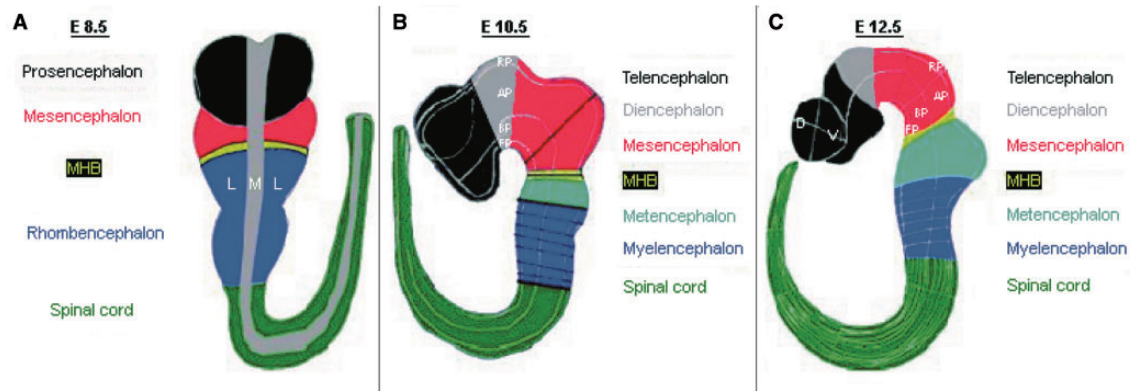


Figure 1. CNS regionalization of the mouse embryo at different developmental stages as used for the database structure. (A) Developmental stage E8.5: The whole embryo is divided into five regions along the anterior–posterior axis: prosencephalon, mesencephalon, MHB, rhombencephalon, spinal cord; along the medial–lateral axis the embryo is divided into medial and lateral, and the region in between is considered as mediolateral boundary. (B) Developmental stage E10.5: The embryo regionalization along the anterior–posterior and along the dorsal–ventral (previous medial–lateral) axis is as follows: telencephalon, diencephalon, mesencephalon, MHB, metencephalon (r1), myelencephalon (r2–r8) and spinal cord; for all CNS regions, the new regionalization along the dorsal–ventral axis is RP, AP, ABB, BP and FP. (C) Developmental stage E12.5: The mouse embryo is regionalized along the anterior–posterior axis as follows: telencephalon (anterior, posterior), diencephalon (anterior hypothalamus, posterior hypothalamus, prethalamus, thalamus), mesencephalon (anterior, posterior), MHB (anterior, posterior), metencephalon (r1), myelencephalon (r2–r8) and spinal cord (cervical, thoracic, lumbar, sacral, caudal); the telencephalon, diencephalon and MHB are subdivided along the dorsal–ventral axis into RP, AP, ABB, BP and FP; the mesencephalon, metencephalon and myelencephalon are subdivided into dorsal and ventral regions and/or neuronal populations along the dorsal–ventral axis; the spinal cord is subdivided into roof plate, dl1, dl2, dl3, dl4, dl5, dl6, v0, v1, v2, v3, v4, mn and floor plate, where dl1 to dl6 describe the dorsal interneurons, and v0 to v3 and mn denote the ventral interneurons (not shown).

CNS region, which has the general structure of vesicle→anterior–posterior localization→medial–lateral or dorsal–ventral localization (column ‘Vesicle’, ‘Structure AP’, ‘Structure ML/DV’ in [Supplementary Table S1](#), respectively). Whenever no further division was made, the description ‘all’ was used.

The CNS regions defined at developmental stage E8.5 are the prosencephalon, mesencephalon, MHB, rhombencephalon and spinal cord. Because the 2D neural plate has not yet folded up to give rise to the 3D neural tube, we only defined an anterior–posterior axis and a medial–lateral axis at this stage. At E10.5, the 3D neural tube undergoes further regionalizations on the anterior–posterior axis as well as on the dorso–ventral (previous medial–lateral) axis. For all CNS regions, the new regionalization on the dorso–ventral axis is roof plate (RP), alar plate (AP), alar basal boundary (ABB), basal plate (BP) and floor plate (FP). The prosencephalon separates along the anterior–posterior axis into the telencephalon and the diencephalon, the developing mesencephalon separates into anterior and posterior, the rhombencephalon splits into met- and myelencephalon and further into eight rhombomeres. The metencephalon was defined as consisting only of rhombomere one (r1).

Because these subdivisions are not defined as sharply as before and distinct neuronal populations have already arisen or are arising in the mid-/hindbrain region of the E12.5 mouse embryo, we additionally defined individual neuronal populations and new subdivisions of the regions

(compartments) tel-, di-, mesencephalon, MHB, met- and myelencephalon and also spinal cord. At this stage, the telencephalon as well as the diencephalon show additional subdivisions along the dorsal–ventral axis, and the mesencephalon is subdivided into regions or neuronal populations (or both) also along the dorsal–ventral axis. While at E12.5 the MHB is still subdivided along the anterior–posterior and dorsal–ventral axis, in the met- and myelencephalon as well as spinal cord, dorsal and ventral regions or neuronal populations (or both) are defined along the dorsal–ventral axis. Along the anterior–posterior axis, the diencephalon is now subdivided into anterior hypothalamus, posterior hypothalamus, prethalamus and thalamus. In addition, at E12.5, the developing spinal cord is subdivided into five anterior–posterior regions, namely cervical, thoracic, lumbar, sacral and caudal.

Database and Web page

Mouse IDGenes was implemented as a relational database using PostgreSQL (<http://www.postgresql.org>). Gene expression and interaction data were manually extracted from literature and stored in the database including references. Currently, the database contains 89 expression data sets and 145 genetic interactions. Genetic interactions as well as expression data sets are assigned to different anatomical regions at the mouse embryonic developmental stages E8.5, E10.5 and E12.5, as described before. To assess the quality of our data, we compared all interactions

with the STRING database (v 9.1), a comprehensive resource of protein–protein interactions (23). All interactions but one in the current Mouse IDGenes data set are present in STRING with a high-confidence score (>0.7). The database was made accessible online through a Java Web interface, which was implemented using the Java Servlet class and runs on an Apache Tomcat Server. The Web interface allows the user to browse the Mouse IDGenes database for expression data and genetic interactions. By subscribing to a mailing list, it facilitates communication with other users of the database, as well as with the developers of the database and the Web page. Data can be retrieved in a legible PDF file format and as tab-delimited flat files (by navigating to the ‘Download’ tab on the Web page). By allowing external user an easy input (using the ‘Input Data’ tab on the Web page), which will be curated and evaluated by the authors, Mouse IDGenes will be continuously updated and thereby stay an up-to-date research tool.

Detailed search

Navigation on the Mouse IDGenes Web page by the ‘Detailed Search’ tab allows users to search for gene expression and interaction data in specific regions of the embryonic mouse CNS at mouse embryonic developmental stages E8.5, E10.5 and E12.5 (radio button: ‘display genes according to the chosen region’). CNS regions and developmental stages (according to [Supplementary Table S1](#)) can be selected with the help of combo boxes. Having set a specific developmental stage and an anatomical structure, users are further able to analyze whether a specific gene of interest is expressed during that specific developmental stage in the selected anatomical structure.

Search for interactions

On the ‘Search for Interactions’ tab, the Mouse IDGenes Web page allows users to search for specific interactions between two genes of interest and for all interactions in which one specific gene of interest is involved (by the ‘Search for specific interactions’ button). Information about the region and at which stage the specific interaction takes place in the embryonic mouse CNS can be retrieved.

Users can also search whether two genes of interest display an overlapping gene expression (by the ‘Search for overlapping gene expressions’ button).

By selecting ‘all’ in either both or only one of the gene selection boxes, users can search for all stored interactions either of the whole database or in which a specific gene of interest is involved.

The displayed interactions follow an overall scheme consisting of the attributes ‘effect’, ‘type’ and ‘name’ for a

genetic interaction. The attribute ‘effect’ can be either activation, i.e. turning on gene expression, or repression, which is defined as shutting down gene expression. The attribute ‘type’ of an interaction is defined by the following six values:

- direct: We define a direct interaction in case interaction partner 1 binds directly to the promoter of interaction partner 2.
- direct signaling: In case of a ligand that activates a signaling cascade or any other component of a signaling cascade that does not directly interact with (or binds to) the promoter of a target gene of this signaling pathway, a direct signaling interaction refers to the activation/repression of a direct target gene of this signaling pathway.
- indirect: Interaction partner 1 does not bind directly to the promoter of interaction partner 2, and signaling or genetic interaction cascades have to occur between the two interaction partners.
- indirect signaling: In case of a ligand that activates a signaling cascade or any other component of a signaling cascade that does not directly interact with (or binds to) the promoter of a target gene of this signaling pathway, an indirect signaling interaction refers to the activation/repression of an indirect target gene by a direct target gene of this signaling pathway.
- maintenance: Interaction partner 1 is not required to activate (i.e. turn on) or to repress (i.e. turn off) the promoter (or expression) of interaction partner 2, but to keep this promoter (or expression) activated (‘on’) or repressed (‘off’) over (a longer period of) time.
- IEXP: Inferred from experiment (e.g. loss of function/gain of function) and/or expression pattern

Currently, the database contains these general interaction schemes: direct activation, direct signaling activation, IEXP activation, maintenance activation, indirect activation, indirect signaling activation, direct repression, direct signaling repression, maintenance repression, IEXP repression, indirect repression and indirect signaling repression. The attribute ‘name’ of an interaction, which is displayed on the Mouse IDGenes Web page, is composed of the official gene symbol according to MGI of interaction partner 1, followed by an arrow symbol from the third column of [Table 1](#) and finally the gene symbol of interaction partner 2.

As pointed out before, the maintenance activation occurs over a longer period to cause a downstream effect. Time-wise, such an interaction can occur over several developmental stages as, for example, is the case of the development of midbrain dopaminergic (mDA) neurons. There, initially a Wnt1-regulated network together with a Shh-controlled genetic cascade establishes the mDA

Table 1. General scheme for interactions used at the Mouse IDGenes Web page is listed; interactions are divided into a type, which can be 'direct', 'direct signaling', 'indirect', 'indirect signaling', 'IEXP' and 'maintenance', as well as an effect, namely 'activation' or 'repression' of gene expression

Interaction type	Interaction effect	Symbol
Direct	Activation	- >
Direct signalling	Activation	- >
Indirect	Activation	-- >
Indirect signalling	Activation	-- >
IEXP	Activation	- >
Maintenance	Activation	- >
Direct	Repression	-
Direct signalling	Repression	-
Indirect	Repression	--
Indirect signalling	Repression	--
IEXP	Repression	-
Maintenance	Repression	-

Direct: a transcription factor (interaction partner 1) directly binds to the promoter of a gene (interaction partner 2); direct signaling: an activation/repression of a direct target gene of a specific pathway initiated by a ligand that activates the signaling cascade of this pathway; indirect: interaction partner 1 does not bind directly to the promoter of interaction partner 2, and signaling or genetic interaction cascades have to occur between the two interaction partners; indirect signaling: an activation/repression of an indirect target gene through a direct target gene of a specific pathway initiated by a ligand that activates the signaling cascade of this pathway; IEXP: inferred from experiment (e.g. loss of function/gain of function) and/or expression pattern; maintenance: interaction partner 1 is not required to activate (i.e. turn on) or to repress (i.e. turn off) the promoter (or expression) of interaction partner 2 but to keep this promoter (or expression) activated ('on') or repressed ('off') over (a longer) time.

progenitor domain, by maintaining *Otx2* expression in the ventral midbrain (35). Another example is *Lmx1b*, which is known to be necessary for the initiation of *Fgf8* expression and for the maintenance of several other genes including *Engrailed 1* (*En1*), *En2* and *Wnt1* (3). *Lmx1b*, therefore, falls into the interaction scheme 'maintenance activation' for *En1*, *En2* and *Wnt1*. On our Web page, these interactions are depicted in the following way: *Lmx1b* → *En1*, *Lmx1b* → *En2* and *Lmx1b* → *Wnt1*, meaning that *Lmx1b* keeps the expression of *En1*, *En2* and *Wnt1*, respectively, on overtime.

Furthermore, cooperative interactions, i.e. interactions with more than two interaction partners, can be stored in the database, for example, if two transcription factors or one transcription factor and another cofactor will bind on the promoter of a target gene, i.e. interaction partner 2.

Data input

One of the core functions of Mouse IDGenes is the possibility for data input by the user. This feature allows the permanent update of data by the respective experts in the field

of developmental neurosciences. To input new gene expression and interaction data into the database, the user is asked to maintain the overall CNS regionalization scheme (according to [Supplementary Table S1](#)), by choosing and subsequently storing specifications given by combo boxes, which follow our CNS model ([Figure 2A](#)). The database so far has stored data exclusively about the developmental stages E8.5 to E12.5, as these are the crucial stages in the establishment of the MHB and development of the mid-/hindbrain region, which is the authors' main research interest. To maintain the high degree of curation in the current database, it is necessary to also introduce at least one relevant publication as well as the corresponding PubMed ID.

Users can input gene expression data as well as interaction data into the database by selecting the respective type of data on the Web page. All fields that are mandatory for data input as well as storage in the database are labeled by an asterisk ([Figure 2A](#) and [B](#)), which ensures the completeness of expression and interaction data sets. In addition, it is automatically controlled from the PubMed abstract by using the link to PubMed whether the given year of the publication and the entered PubMed ID are consistent (compare with the fields 'Author and Year' and 'PubmedID' in [Figure 2A](#) and [B](#)); otherwise, a warning message is displayed, and storage of the data is prevented. Further, gene symbols from MGI are auto-completed after typing some letters in the field 'Gene' ([Figure 2A](#)) or 'First Factor' or 'Target Gene' ([Figure 2B](#)), and they are internally stored via MGI identifiers. The use of predefined lists by combo boxes for the selection of e.g. a specific brain region and/or the interaction type helps to comply with our brain regionalization model as well as the model for genetic interactions and therefore ensures data consistency. In case of extending the database for an interaction, it is possible to input also cofactors and more targets of an interaction ([Figure 2B](#)). After completion of the data set, the user is asked to review the input before final submission to the database. Constraints in the PostgreSQL database prevent data from being duplicated when stored.

To control for incorrect input, the data will be regularly curated by the authors. Before curation, new data will be distinguishable on the Web page from already curated data by labeling the not yet validated gene expression or interaction data on the Web page ([Figure 2C](#)).

Confirmation of new data is performed by

- reading the given publications,
- comparing the indicated gene expression and/or genetic interaction,
- comparing the developmental stage as well as the brain regions with the data entered into the Mouse IDGenes database.

A Expression Interaction

Input expression data

already stored in database, * required

Gene*

Stage* 10.5

Region* Mesencephalon

Anterior-Posterior* all

Dorsal-Ventral* choose

Please select a region in the medial-lateral/dorsal-ventral axis or all

Literature

Author and Year* PubmedID*

add a publication

B Expression Interaction

Input interaction data

already stored in database, * required

add a cofactor remove a cofactor

First Factor*

add a target remove a target

Target Gene*

Type* direct

Effect* activation

Stage* choose

Region* please choose a stage

Anterior-Posterior* please choose a region

Medial-Lateral* please choose a region

Literature

Author and Year* PubmedID*

add a publication

C

Nkx6-1 -> Gata2: IEXP,activation 77
Data not validated yet!
77 Craven et al. 2004

Nkx6-1 -> Gata3: IEXP,activation 77
Data not validated yet!
77 Craven et al. 2004

Figure 2. Dialogue of the data input and output of not yet validated data as found on the Mouse IDGenes Web interface. Input dialogue for (A) expression data and (B) interaction data. (C) The output of not yet validated data.

A mailing list has been set up for users to discuss their plan to submit new data.

After verification of new data, the flag ‘Data not validated yet!’ is removed from the database and the Web page; otherwise, if the data could not be verified, they are deleted from the database.

Output

The output of a specific search either in one of the gene expression or the interaction menu items appears on the same page underneath the user selection (Figure 3B), organized in a table. In case of the gene expression data (radio button: ‘display genes according to the chosen region’ (Figure 3A) in the ‘Detailed Search’ menu item), the output table indicates the developmental stage the user is looking at. The output further shows the specific subdivision of the chosen anatomical area and which genes are expressed there in the column ‘Region’. Additionally, links to the public databases NCBI, MGI, UCSC and Ensembl are given in the column ‘Expression Data’, where further general information about a displayed gene can be retrieved. Most importantly, links to literature references are indicated as evidence of gene expression information. The last column ‘Interactions’ of the output shows the

interaction data, which can be retrieved for the actual brain region at the actual developmental stage with the corresponding literature references.

Application of the database

Prediction of a new Wnt1 target based on Mouse IDGenes

To demonstrate the usefulness of the Mouse IDGenes database for other research applications, we chose an example from our own scientific interests focused on the role of the Wnt signaling pathway in the development of the mid-/hindbrain region and of neuronal populations located in this region, such as the ventral midbrain dopaminergic neurons. In this context, the Wnt signaling pathway plays a crucial role because it participates in the regulation of regional patterning, cell cycle, cell fate specification, cell differentiation and cell survival. It is also involved in various human diseases (36).

The manually curated, and thus, highly reliable data set of Mouse IDGenes provides an ideal basis to further analyze, for example, the complex gene regulatory network at the MHB and in the ventral midbrain in which Wnt1/ β -catenin signaling has so far been implicated (37, 38).

A Developmental stage: display genes according to the chosen region

Region: select gene name to search for correlated anatomical structure

Anterior-Posterior:

Dorsal-Ventral:

B

Stage: E12.5

Region	Expression Data	Interactions
Mesencephalon		
anterior,roof plate	<p>Fgf8 NCBI Ensembl ucsc MGI id 50 31 38 49 55 56 57</p> <p>Gad1 Ensembl MGI id ucsc NCBI 95 65</p> <p>Gad2 ucsc MGI id Ensembl NCBI 99</p> <p>Lhx1 Ensembl NCBI MGI id ucsc 65 94</p> <p>Lhx5 MGI id Ensembl ucsc NCBI 94</p> <p>Lmx1a ucsc MGI id Ensembl NCBI 53</p> <p>Lmx1b ucsc MGI id Ensembl NCBI 54 53 55</p> <p>Msx1 ucsc NCBI Ensembl MGI id 18</p> <p>Otx2 Ensembl MGI id NCBI ucsc 36 8 35</p> <p>Pou4f1 Ensembl MGI id ucsc NCBI 85 65</p> <p>Sef Ensembl ucsc MGI id NCBI 48 49</p> <p>Six3 NCBI MGI id Ensembl ucsc 81</p> <p>Spry2 Ensembl NCBI ucsc MGI id 46 49 47</p> <p>Spry4 MGI id ucsc Ensembl NCBI 47 46</p> <p>Tapp1 MGI id ucsc NCBI Ensembl 48</p> <p>Wnt1 ucsc MGI id NCBI Ensembl 34 38 39 55 40</p> <p>8 Prakash et al. 2006, 18 Andersson et al. 2006, 31 Davis et al. 1988, 34 Rowitch and McMahon 1995, 35 Simeone et al. 1992, 36 Simeone et al. 1993, 38 Davis and Joyner 1988, 39 Wilkinson et al. 1987, 40 Parr et al.1993, 46 Zhang et al. 2001, 47 Minowada et al. 1999, 48 Bouchard et al. 2005, 49 Lin et al. 2002, 50 Echevarria et al. 2004, 53 Mishima et al. 2009, 54 Asbreuk et al. 2002, 55 Guo et al. 2007, 56 Lin et al. 2003, 57 Marnoka et al. 1998, 65 Kala et al. 2009, 81 Oliver et al. 1995, 85 Fedtsova and Turner 1995, 94 Sheng et al. 1997, 95 Guimera et al. 2006, 99 Katarova et al. 2000</p>	<p>Lhx1-> Gad1 IEXP activation 74</p> <p>Lhx1-> Gad2 IEXP activation 74</p> <p>Lhx5 -> Gad1 IEXP activation 74</p> <p>Lhx5 -> Gad2 IEXP activation 74</p> <p>Wnt1 -> Lmx1b IEXP activation 55</p> <p>55 Guo et al. 2007, 74 Nakatani et al. 2007</p>

Figure 3. Output window of expression data and interaction data by the use of the ‘Detailed Search’ option on the Mouse IDGenes Web page. (A) Search dialogue on the menu item ‘Detailed Search’. In this example, expression and interaction data for the roof plate of the anterior mesencephalon at embryonic day 12.5 are requested. (B) Output for gene expression data as well as interaction data for the request according to (A).

We used the Mouse IDGenes database to predict novel direct or indirect targets of the Wnt1/ β -catenin signaling pathway that might be involved in the development of the mid-/hindbrain region and of associated neuronal populations and validated our results experimentally.

We obtained known targets with the interaction type ‘direct signaling’ and ‘indirect signaling’ of the Wnt1/ β -catenin pathway from the Mouse IDGenes Web page by choosing Wnt1 as search term in the ‘Search for Interactions’ menu as well as from literature searches (Figure 4A) (35, 39–52). One hallmark of the Wnt1/ β -catenin signaling pathway is the stabilization and nuclear translocation of cytoplasmic β -catenin (Figure 5). In the absence of a Wnt1 signal, the Lef1/Tcf transcription factors are bound to the promoter regions of the direct Wnt1 target genes together with other co-repressors, thereby

inhibiting the activation of these genes. In the presence of a Wnt1 signal, the replacement of these co-repressors and binding of β -catenin to the Lef1/Tcf transcription factors activates the transcription of the direct Wnt1 target genes. The frequent presence of, in particular, evolutionary conserved Lef1/Tcf TFBSs in the promoter region of a gene is therefore indicative that this gene might be a direct target gene of the Wnt1/ β -catenin signaling pathway. An indirect target gene of the Wnt1/ β -catenin pathway was defined as a gene that is upregulated on Wnt1/ β -catenin signaling activity but is not directly bound by β -catenin and Lef1/Tcf transcription factors and thus requires another mediator, i.e. an intermediate gene regulatory step.

We performed an *in silico* promoter analysis for Wnt1 target genes of the interaction type ‘direct signaling’ and ‘indirect signaling’ in the training data set (Figure 4A) as

A training set for the SVM classification

Gene Symbol	Number of conserved binding sites	Average matrix similarity	Class	Ref
Otx2	2	0.901	1	35;39;40
Lmx1a	1	0.911	1	39;40
En1	0	0.844	1	41-43
Ccnd1	1	0.831	1	44;45;40
Fgf20	7	0.914	1	46
Dkk1	4	0.923	1	46
Sp5	8	0.934	1	47
Msx1	0	0.96	2	39;40
Pitx3	0	0.827	2	39
Gbx2	0	0.8	2	48;49
Shh	1	0.948	2	50
Pou4f1	0	0.903	2	51
Six3	1	0.843	2	52

B test set and predicted classes

Gene symbol	Number of conserved binding sites	Average matrix similarity	Predicted class	Ref
Fgf8	0	0.763	2	53;54
Lef1	2	0.933	1	55;56
Islet1	0	0.82	2	57-59
Dkk3	2	0.914	1	*

* Zhang, J. and Prakash, N. et al., unpublished in vivo data

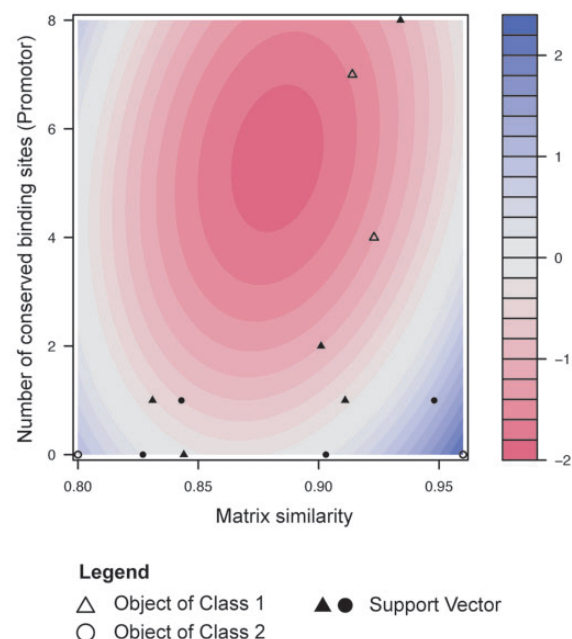
C SVM classification plot

Figure 4. Training and test data for SVM classification. (A) Training data containing experimentally validated Wnt1 target genes. Genes were extracted from Mouse IDGenes, and number of conserved binding sites and average matrix similarity were computed with Genomatix MatInspector. Class 1 contains Wnt1 targets of the type ‘direct signaling’ (i.e. direct Lef1/Tcf target genes), whereas class 2 includes Wnt1 targets of the type ‘indirect signaling’. (B) Result of classification of selected genes from the test set. Direct Lef1/Tcf targets have more conserved binding sites and a higher average matrix similarity. (C) SVM classification (contour) plot showing training data. Filled objects indicate support vectors, blank objects remaining data points. Red color indicates decision values of class 1 (i.e. direct Lef1/Tcf binding), while blue color indicates decision values of class 2 (indirect Lef1/Tcf binding).

well as in the test set (Figure 4B) consisting of interesting target genes for which the interaction type is not yet known in the CNS (53–59). The MatInspector (Genomatix) program was used to identify Lef1/Tcf binding sites in the promoter sequences of these target genes with help of predefined position weight matrices (PWMs) (60). For classification of the interaction type, we assigned ‘direct signaling’ target genes of Wnt1/ β -catenin pathway to class 1 and ‘indirect signaling’ targets to class 2. We computed two parameters for each gene: (i) the number of evolutionary conserved Lef1/Tcf TFBSs (V\$LEFF). Orthologous promoter sequences from human, chimp, mouse, rat, dog, horse, cow and opossum were taken into account, and only binding sites present in at least two species were considered. (ii) The average matrix similarity of all Lef1/Tcf binding sites in the promoter, a score indicating the similarity of a predicted binding site with the consensus matrix of the TF (60, 61). It lies in [0, 1] and reaches 1 only if the predicted sequence corresponds to the most conserved nucleotide at each

position. While the matrix similarity allows the assessment of the structural quality of a binding site, the number of conserved binding sites supports the possibility that these binding sites might be functionally relevant (62). It is assumed that in case of direct interactions the matrix similarity is higher and conserved binding sites are more frequent than in case of indirect interactions.

We trained a support vector machine (SVM) with ‘direct signaling’ targets (class 1) and ‘indirect signaling’ targets (class 2) using both parameters (Figure 4C). Classification via SVM has been successfully used not only for feature selection of microarray data (63–65) but also to integrate expression as well as genomic data, e.g. evolutionary conservation or binding site clusters for the improvement of TFBS prediction (21, 22). The statistical model was calculated with help of the ksvm function from the kernlab package in R statistical software by using a polynomial kernel matrix similarity (degree = 2), a cost parameter $C=1$ and a 13-fold cross validation

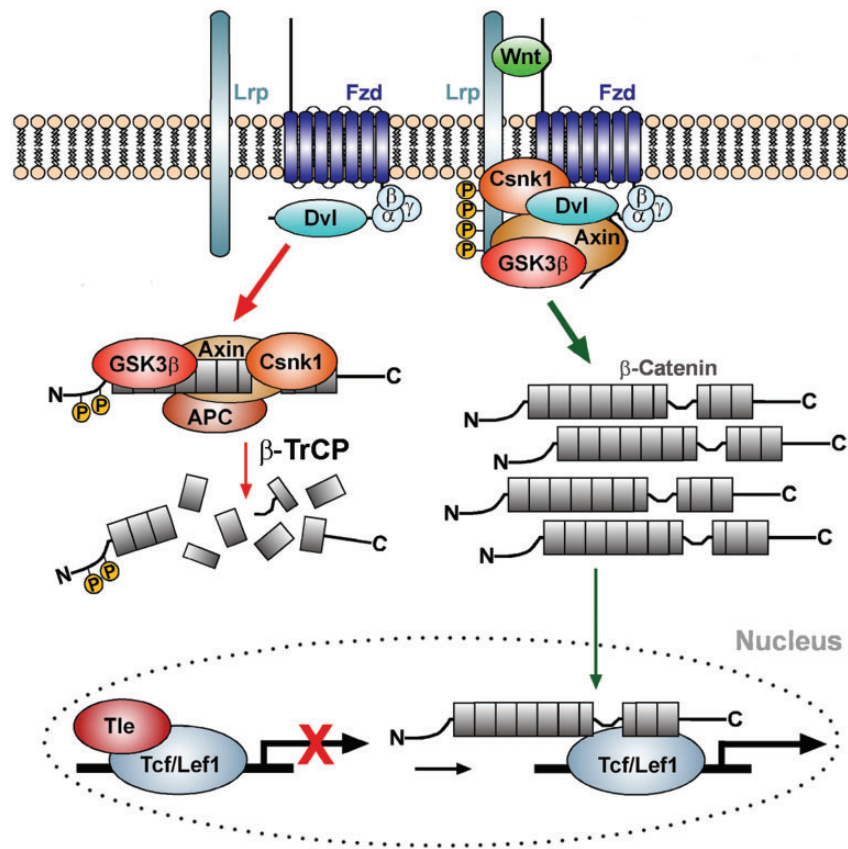


Figure 5. The Wnt/ β -catenin signaling pathway. Left (red arrows): In the absence of Wnt ligand, β -catenin is bound by the destruction complex consisting of the scaffolding proteins Axin and Adenomatosis polyposis coli (APC), and the protein kinases Glycogen synthase kinase 3 beta (GSK3 β) and Casein kinase I (Cskn1), and sequentially phosphorylated by these kinases. Phosphorylated β -catenin binds to and is ubiquitinated by the E3 ubiquitin ligase β -TrCP, thereby targeting it for proteasomal degradation. In the absence of Wnt ligand, lymphoid enhancer binding factor 1 (Lef1) or T cell-specific (TCF) transcription factors are bound to the promoters of Wnt target genes in the cell nucleus together with co-repressors of the Groucho/transducin-like enhancer of split (Tle) family proteins, thereby repressing their expression. Right (green arrows): On binding of Wnt ligand to the Frizzled (Fzd) receptor and low-density lipoprotein receptor-related protein (Lrp) co-receptor complex, Axin and GSK3 β are recruited to the cell membrane via Dishevelled (Dvl) and the destruction complex falls apart. Unphosphorylated β -catenin accumulates in the cytosol and translocates into the nucleus, where it binds to the Lef1/TCF transcription factors and activates Wnt target genes by displacing the co-repressors and recruiting co-activators to this complex. Properties of Lef1/Tcf binding sites in the promoters of known Wnt target genes, i.e. the number of conserved Lef1/Tcf binding sites as well as the averaged matrix similarity, were used to train a classifier and to predict direct or indirect interactions of potentially new target genes and Lef1/Tcf transcription factors in the Wnt/ β -catenin signaling pathway.

(leave-one-out cross validation). Class 1 consists of *Otx2*, *Lmx1a*, *En1*, *Cnd1*, *Dkk1* and *Sp5* and class 2 consists of *Msx1*, *Pitx3*, *Gbx2*, *Shh*, *Pou4f1* (*Brn3a*) and *Six3* (Figure 4A). Additionally, *Fgf20* was included in class 1. Interaction data for this gene cannot be retrieved from Mouse IDGenes, as expression of *Fgf20* in the mouse neural tube starts only after E12.5 (66). We assigned *En1* to class 1 because it was shown that *En1* expression under the control of the *Wnt1* enhancer in mice rescues the *Wnt1*-mid-/hindbrain phenotype (41, 42), thus indicating that *En1* is the downstream target of *Wnt1* signaling in mid-/hindbrain development, and because a direct interaction of the promoter of the homologous gene *engrailed-2* (*En2*) with the Lef1/Tcf transcription factor was observed in the frog (43). For our statistical model, we obtained a training error of 7.69% and a cross-validation error of 46.15%.

To further elucidate the *Wnt1*-controlled gene regulatory network at the MHB, we applied SVM prediction by using the number of conserved Lef1/Tcf binding sites and the average matrix similarity in the promoters of four interesting genes, *Fgf8*, *Lef1*, *Islet1* and *Dkk3*, representing the test set (from Figure 4B) to predict whether they are directly or indirectly activated by the *Wnt1*/ β -catenin pathway. For these four genes, it is not known whether they are direct or indirect targets of *Wnt1* in neural tissues, but it was observed that *Lef1* and *Dkk3* are co-expressed with *Wnt1* in the midbrain (Götz, S. et al., unpublished data), whereas *Fgf8* and *Islet1* are not co-expressed with *Wnt1* but depend on *Wnt1* expression in the mid-/hindbrain region (53, 54, 58, 59). Using the SVM on the test set (Figure 4B), our analysis predicts that *Lef1* and *Dkk3* are direct targets of Lef1/Tcf-mediated *Wnt1*/ β -catenin

signaling, whereas *Fgf8* and *Islet1* are not direct target genes of this signaling pathway and interact indirectly with the Wnt1 signaling cascade. *Fgf8* and *Islet1* therefore would represent Wnt1/ β -catenin target genes that are most likely activated by other genes that in turn are activated by Wnt1/ β -catenin signaling. Direct binding of Lef1/Tcf transcription factors to the Lef1 promoter was shown in colon cancer/lymphocytes (55) as well as in HEK 293 cells (56), which is in accordance with our prediction. A direct interaction of Lef1/Tcf transcription factors with a larger promoter region of the *Islet1* gene was shown in embryonic heart tissue but not in neural tissues (57). However, direct activation of *Fgf8* and *Islet1* in the CNS by Lef1/Tcf binding sites was up to now never observed although it was inferred from mutant mouse embryo analyses (53, 54, 58, 59).

Experimental validation of the predicted Wnt1 target gene *Dkk3*

To show the predictive power of our approach, we experimentally validated the so far unknown Wnt1 target gene *Dkk3* as a direct target gene of the Lef1/Tcf-mediated Wnt1/ β -catenin signaling cascade *in vitro*.

To identify conserved Lef1/Tcf binding sites in ~700 bp extended promoter region (as used for the SVM prediction) of the *Dkk3* gene of five different mammalian species (mouse, rat, cow, pig and opossum), we applied the DiAlign TF program in the Genomatix software suite GEMS Launcher to evaluate the overall promoter similarity and to identify conserved Lef1/Tcf binding sites in these regions. For the alignment, we chose the five most conserved *Dkk3* promoter sequences among 14 organisms from Genomatix homology group Hg3927. Four Lef1/Tcf binding sites were predicted in the putative mouse *Dkk3* promoter, of which one [binding site 'c', the most proximal Lef1/Tcf binding site to the transcription start site (TSS)] was highly conserved among all five species (Figure 6A). To determine whether these predicted Lef1/Tcf binding sites control the Wnt1/ β -catenin and Lef1/Tcf-mediated activation of the murine *Dkk3* promoter, we cloned a 744-bp-long fragment of this promoter containing three of the four predicted Lef1/Tcf binding sites into a promoter-less luciferase reporter vector (Figure 6E). Co-transfection of increasing amounts of rat Lef1 complementary DNA (67) or a constitutively active β -catenin [Δ N- β -catenin, which mimics the activation of Wnt1 signaling, (68)] into 'Wnt-responsive' HEK293T cells [exhibiting a basal level of Wnt1/ β -catenin signaling activity, Prakash, N. et al., unpublished data, (69)] led to a dose-dependent activation of luciferase expression mediated by this mouse *Dkk3* (mDkk3) promoter fragment (Figure 6F and G), indicating that the promoter of the mDkk3 gene is a direct target of

Lef1-mediated Wnt1/ β -catenin signaling in this *in vitro* context. Additional *in vivo* evidence indicates that the murine *Dkk3* gene is also a direct target of Lef1-mediated Wnt1/ β -catenin signaling in the mouse ventral midbrain (Zhang, J. and Prakash, N., unpublished data).

To evaluate whether the predicted Lef1/Tcf binding sites in the murine *Dkk3* promoter are functional, we mutagenized each of these binding sites either individually (mutation of a single Lef1/Tcf binding site) or altogether (mutation of all three Lef1/Tcf binding sites) within the mDkk3 promoter fragment such that they cannot be recognized by Lef1/Tcf transcription factors anymore (Figure 6B–E) (46). Site-directed mutagenesis of single or all three Lef1/Tcf binding sites in the mDkk3 promoter/reporter constructs revealed that

- luciferase activity was significantly decreased relative to the wild-type mDkk3 promoter after co-transfection of Lef1 cDNA (Figure 6H),
- the activation of the mDkk3 promoter/reporter construct carrying a mutagenized Lef1/Tcf binding site 'c' by Lef1 was completely abolished relative to the pcDNA3.1 control, in contrast to a still significant activation of the mDkk3 promoter/reporter constructs carrying a mutagenized Lef1/Tcf binding site 'a' or Lef1/Tcf binding site 'b' (Figure 6I).

This result strongly suggests that the most conserved (by position and sequence similarity among five mammalian species) and proximal (relative to the TSS) Lef1/Tcf binding site 'c' is the functionally most important of the three Lef1/Tcf binding sites for Lef1-mediated activation of the murine *Dkk3* gene by Wnt1/ β -catenin signaling.

Lef1/Tcf transcription factors were predicted to directly activate the mouse *Dkk3* gene in the context of Wnt1/ β -catenin signaling by our SVM analyses. Therefore, the experimental validation of Lef1/Tcf binding site 'a' and Lef1/Tcf binding site 'c' with a matrix similarity of 0.867 and 0.961, respectively, confirm the result of our SVM classification, indicating that direct Lef1/Tcf targets have in general more conserved binding sites than indirect targets (Figure 4B). Altogether, our experimental results indicated that the activation of the mouse *Dkk3* gene is mediated at least in part by the predicted Lef1/Tcf binding sites in its promoter region and therefore highlight the importance and predictive power of a database combining both expression and interaction data.

Future directions

So far, the Mouse IDGenes database offers spatially resolved and manually registered data about the developmental stages E8.5, E10.5 and E12.5 of the mouse mainly

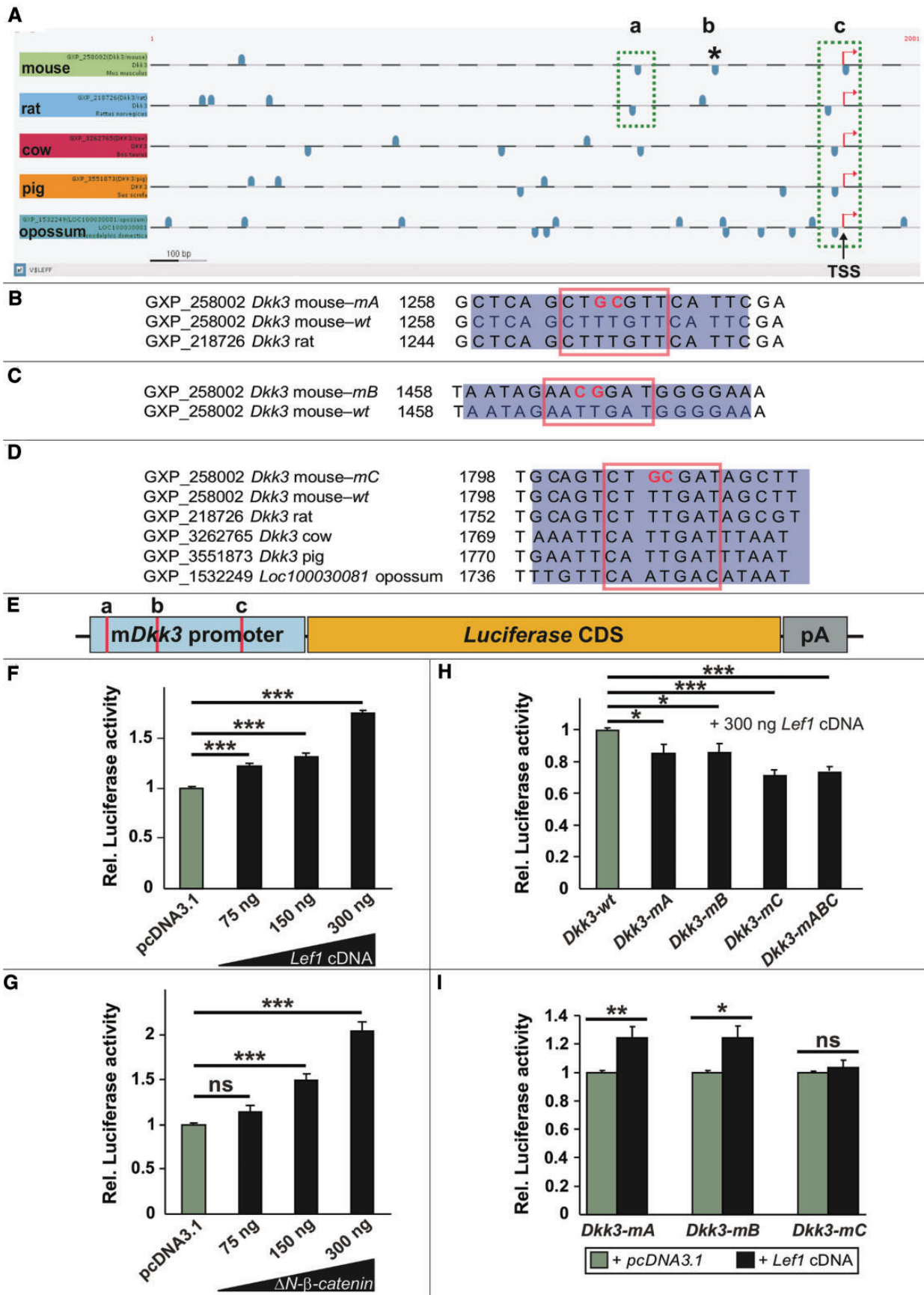


Figure 6. Mouse *Dkk3* is a direct target gene of Lef1-mediated Wnt/ β -catenin signaling. (A) Representation of the putative *Dkk3* promoter (5' proximal) regions from mouse, rat, cow, pig and opossum and of the predicted Lef1/Tcf binding sites on the sense (upper blue boxes) and antisense

(continued)

with gene expressions and interactions important for the development of the mid-/hindbrain region because it harbors important neuronal populations that are implicated in several neurodevelopmental human diseases (1, 2, 7). We aim to enlarge the data sets to more entries that would provide a good representation of the known gene expression patterns and interactions in the developing mouse CNS. This aim is already facilitated by allowing users to input data to the database. With this, we intend to attract experts in different fields of developmental neurosciences to update the existing platform so that Mouse IDGenes becomes the data source of choice, for experimental and *in silico* analyses related to gene expression and interaction data in the developing murine CNS. Additionally, a broader data set will lead to improvement of dynamic modeling projects and to more precise prediction methods (26, 27). As demonstrated, the genetic interaction data stored in the Mouse IDGenes database can be used for the prediction of Wnt target genes, but with this database in combination with publicly available PWMs (70), it is also possible to predict new targets of other signaling pathways (e.g. Shh, Fgf8 or BMPs), which are equally important for CNS development. With our mailing list, we seek to develop an open platform, which eases the communication between neuroscientists.

Materials and methods

Bioinformatics prediction of Lef1/Tcf binding sites in the promoter regions of Wnt/ β -catenin target genes

To discriminate interactions of the type ‘direct signaling’, i.e. genes bound by Lef1/Tcf transcription factors activated

by Wnt1 signaling from correlations of the type ‘indirect signaling’ and to predict the interaction type of new target genes, a support vector machine (SVM) was applied. The data matrix of the SVM is composed of two variables (columns) for each object (row), i.e. the frequency as well as the averaged matrix similarity of conserved TFBSs for each promoter sequence of a specific gene. The classification of the interaction type in training data set for each object (gene), meaning whether Lef1/Tcf transcription factor binds directly or indirectly to the promoter sequence, was derived from the Mouse IDGenes database. Promoter sequences for Wnt/ β -catenin target genes were derived from the ELDorado genome database (Genomatix/Germany), versions 12-2010 and 08-2011. Orthologous promoter sequences from different mammalian species of the Genomatix homology group (human, chimp, mouse, rat, dog, cow, pig and opossum) were analyzed using the MatInspector program (with the Matrix Family Library Version 8.4) from Genomatix to identify potential Lef1/Tcf binding sites and to extract matrix similarities. Conserved binding sites were determined by using the DiAlign TF program (Genomatix) and by searching for common Lef1/Tcf sites occurring at the same position in (aligned) orthologous promoter sequences of each Wnt/ β -catenin target gene. The length of the promoter regions used for the detection of Lef1/Tcf sites to derive the total number and the average matrix similarity of binding sites both included in the SVM algorithm were generally in the range of 600–1400 bp. In addition, a longer promoter region of mouse Dkk3 mRNA (with the Genbank identifier AK013054, firstly detected in whole body of 10- and 11-day-old mouse embryos) was defined as 1800 bp upstream, including the proximal region, and 200 bp downstream of the TSS. Dkk3 promoter sequences of 2000 bp length from

(lower blue boxes) strands within these Dkk3 promoter regions. The most conserved (by sequence similarity and position) and proximal (relative to the TSS) predicted Lef1/Tcf binding sites were designated as ‘a’ and ‘c’ (green dotted boxes). The Lef1/Tcf binding site ‘b’ is only predicted in the mouse Dkk3 promoter (asterisk). (B–D) Sequence alignments of the mutated (m) and wild-type (wt) Lef1/Tcf binding site ‘a’ in the mouse and rat Dkk3 promoter regions (B), ‘b’ in the mouse Dkk3 promoter region only (C) and ‘c’ in the mouse, rat, cow, pig and opossum Dkk3 promoter regions (D). The blue rectangles delimit the sequence, and the red boxes frame the core sequence of the corresponding Lef1/Tcf binding site. The red bold letters indicate the mutagenized nucleotides in the corresponding Lef1/Tcf binding site core sequence. (E) Schematic drawing of the murine Dkk3 (mDkk3) promoter/luciferase reporter construct used in the following experiments, and of the approximate position of the three predicted and partly conserved proximal Lef1/Tcf binding sites within this promoter fragment (red bars). CDS, coding sequence; pA, polyadenylation signal. (F–I) Luciferase reporter assays in HEK293T cells using the wild-type and mutated mDkk3 promoter/reporter construct depicted in (E). (F) Co-transfection of increasing amounts of Lef1 cDNA led to a dose-dependent activation of the wild-type mDkk3 promoter relative to the ‘empty’ (pcDNA3.1) vector control. (Rel. luciferase activities: pcDNA3.1, 1.0 ± 0.01 ; 75 ng Lef1 cDNA, 1.21 ± 0.03 ; 150 ng Lef1 cDNA, 1.32 ± 0.035 ; 300 ng Lef1 cDNA, 1.74 ± 0.04). (G) Co-transfection of increasing amounts of a constitutively active β -catenin (ΔN - β -catenin) led to a dose-dependent activation of the wild-type mDkk3 promoter relative to the ‘empty’ (pcDNA3.1) vector control. (Rel. luciferase activities: pcDNA3.1, 1.0 ± 0.01 ; 75 ng ΔN - β -catenin, 1.14 ± 0.07 ; 150 ng ΔN - β -catenin, 1.49 ± 0.06 ; 300 ng ΔN - β -catenin, 2.04 ± 0.11). (H) Site-directed mutagenesis of single and of all three predicted Lef1/Tcf binding sites (‘a’, ‘b’, ‘c’, ‘abc’) within the mDkk3 promoter fragment (Dkk3-mA, Dkk3-mB, Dkk3-mC, Dkk3-mABC) resulted in a site-specific and significant decrease of luciferase activity relative to the wild-type mDkk3 promoter (Dkk3-wt) after co-transfection of 300 ng Lef1 cDNA. (Rel. luciferase activities: Dkk3-wt, 1.0 ± 0.01 ; Dkk3-mA, 0.85 ± 0.05 ; Dkk3-mB, 0.86 ± 0.06 ; Dkk3-mC, 0.71 ± 0.04 ; Dkk3-mABC, 0.73 ± 0.03). (I) Site-directed mutagenesis of the most conserved (across species) and proximal (relative to the TSS) Lef1/Tcf binding site ‘c’ in the mDkk3 promoter completely abolished the activation of this promoter after co-transfection of 300 ng Lef1 cDNA relative to the ‘empty’ vector control (pcDNA3.1). (Rel. luciferase activities: Dkk3-mA: pcDNA3.1, 1.0 ± 0.01 ; Lef1 cDNA, 1.24 ± 0.08 ; Dkk3-mB: pcDNA3.1, 1.0 ± 0.01 ; Lef1 cDNA, 1.24 ± 0.08 ; Dkk3-mC: pcDNA3.1, 1.0 ± 0.01 ; Lef1 cDNA, 1.04 ± 0.05). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; ns, not significant.

five different mammalian species (mouse, rat, cow, pig and opossum) were analyzed with the MatInspector to predict Lef1/Tcf binding sites.

Cloning of a mouse Dkk3 (mDkk3) promoter/reporter vector

A 744-bp fragment of the putative mDkk3 promoter (Entrez Gene ID: 50781; chromosome 7, strand: –, position: 112 158 266 to 112 159 009 bp, NCBI build 38) was amplified from C57BL/6 mouse genomic DNA by PCR using the forward primer 5'-*ctcgag*TGACCAGATCCAGC TTGCA-3' and reverse primer 5'-*aagctt*CCTCCTGAGG GTAGTTGAGA-3' that included an *Xho*I and *Hind*III restriction site (underlined sequences in italics), respectively. The amplified fragment was cloned into the pCR®II TOPO TA vector (TOPO® TA Cloning® Kit, Life Technologies/Germany) and sequenced throughout its entire length (Sequiseive/Germany). The mDkk3 promoter fragment was excised from the pCR®II TOPO TA vector by *Xho*I/*Hind*III digestion and subcloned into an *Xho*I/*Hind*III-digested pGL3-Basic Vector (Promega/USA).

Site-directed mutagenesis of the mDkk3 promoter fragment

Site-directed mutagenesis of the most conserved and proximal (relative to the TSS) Lef1/Tcf binding sites predicted in the 744-bp mDkk3 promoter fragment was done using the QuickChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies/USA) according to the manufacturer's instructions. Mutagenic primers were the following: Dkk3-mA: 5'-ccagcttcagctcagct**gcgttc**attcgaa ttgggtg-3'; Dkk3-mB: 5'-gtccaagagatcccgataataga**acggatgg** ggaatagtaaaggaa-3'; Dkk3-mC: 5'-ggtgctctcagct**tcgca** tagctttccgggac-3' (mutagenized nucleotides in bold; core sequence of the corresponding Lef1/Tcf binding site in italics). Mutated promoter fragments were confirmed by sequencing (Sequiseive).

Cell culture, transfections and luciferase reporter assays

HEK-293T cells were kept at 37°C and 5% CO₂ in DMEM medium + 10% fetal calf serum/glutamine (Life Technologies). HEK-293T cells (1.25 × 10⁵ cells/well of a 24-well plate) were co-transfected with 300 ng/well pGL3-mDkk3 promoter/reporter vectors (wild-type and mutagenized sequences), 30 ng/well pRL-SV40 (as internal transfection control, Promega) and 150–225 ng/well pcDNA3.1 (Life Technologies) 'empty' vector, alone or together with 75 ng/well, 150 ng/well or 300 ng/well

constitutively active Δ N- β -catenin (68) or rat Lef1 cDNA (67) using Lipofectamine 2000 (Life Technologies). The total amount of plasmid DNA transfected in each well was 630 ng. Cells were lysed in Passive Lysis Buffer (Promega) after 24 h, and Firefly and Renilla Luciferase luminescence were measured in a Centro LB 960 luminometer (Berthold Technologies/Germany) using the Dual-Luciferase® Reporter Assay System (Promega) according to the manufacturer's instructions. Firefly luminescence was normalized against Renilla luminescence for each well. Assays were performed in triplicates, and data are derived from three independent experiments.

Statistical analyses

All values shown are mean ± SEM. Statistical significance between groups was assessed by two-tailed independent-samples *t* tests using the SPSS 18.0 software (SPSS Inc./USA). A value of $P < 0.05$ was considered significant.

Supplementary Data

Supplementary data are available at Database Online.

Acknowledgments

The authors thank Sebastian Götz for extensive testing of the current version of the database. Their genuine thanks go further to Dr Regina Peis for her help regarding the *in silico* promoter analysis.

Funding

The 'Deutsche Forschungsgemeinschaft' ('German Research Foundation') within the framework of the Munich Cluster for Systems Neurology [EXC 1010 SyNergy]; the EU grant 'Systems Biology of Stem Cells and Reprogramming' [FP7-Health-F4-2010-242129]; the 'Bundesministerium für Bildung und Forschung' in the consortium 'From Disease Genes to Protein Pathways' [FKZ01GS0858]. Funding for open access charge: Helmholtz Zentrum München – Deutsches Forschungszentrum für Gesundheit und Umwelt (HMGU).

Conflict of interest. None declared.

References

1. Prakash,N. and Wurst,W. (2004) Specification of midbrain territory. *Cell Tissue Res.*, 318, 5–14.
2. Wurst,W. and Bally-Cuif,L. (2001) Neural plate patterning: upstream and downstream of the isthmic organizer. *Nat. Rev. Neurosci.*, 2, 99–108.
3. Hoekstra,E.J., Mesman,S., de Munnik,W.A. *et al.* (2013) LMX1B Is Part of a transcriptional complex with PSPC1 and PSF. *PLoS One*, 8, e53122.
4. Crespo-Enriquez,I., Partanen,J., Martinez,S. *et al.* (2012) Fgf8-related secondary organizers exert different polarizing planar instructions along the mouse anterior neural tube. *PLoS One*, 7, e39977.

5. Sato, T. and Joyner, A.L. (2009) The duration of Fgf8 isthmic organizer expression is key to patterning different tectal-isthmo-cerebellum structures. *Development*, 136, 3617–3626.
6. Guo, C., Qiu, H.-Y., Huang, Y. *et al.* (2007) Lmx1b is essential for Fgf8 and Wnt1 expression in the isthmic organizer during tectum and cerebellum development in mice. *Development*, 134, 317–325.
7. Dworkin, S. and Jane, S.M. (2013) Novel mechanisms that pattern and shape the midbrain-hindbrain boundary. *Cell. Mol. Life Sci.*, 70, 3365–33674.
8. Merlini, L., Fluss, J., Dhoub, A. *et al.* (2013) Mid-hindbrain malformations due to drugs taken during pregnancy. *J. Child Neurol.*, 29, 538–544.
9. Barkovich, A.J., Millen, K.J. and Dobyns, W.B. (2009) A developmental and genetic classification for midbrain-hindbrain malformations. *Brain*, 132, 3199–3230.
10. Doherty, D., Millen, K.J. and Barkovich, A.J. (2013) Midbrain and hindbrain malformations: advances in clinical diagnosis, imaging, and genetics. *Lancet Neurol.*, 12, 381–393.
11. Pollock, J.D., Wu, D.-Y. and Satterlee, J.S. (2014) Molecular neuroanatomy: a generation of progress. *Trends Neurosci.*, 37, 106–123.
12. Finger, J.H., Smith, C.M., Hayamizu, T.F. *et al.* (2011) The mouse gene expression database (GXD): 2011 update. *Nucleic Acids Res.*, 39, D835–D841.
13. Lein, E.S., Hawrylycz, M.J., Ao, N. *et al.* (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445, 168–176.
14. Visel, A., Thaller, C. and Eichele, G. (2004) GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.*, 32, D552–D556.
15. Christiansen, J.H., Yang, Y., Venkataraman, S. *et al.* (2006) EMAGE: a spatial database of gene expression patterns during mouse embryo development. *Nucleic Acids Res.*, 34, D637–D641.
16. Siddiqui, A.S., Khattri, J., Delaney, A.D. *et al.* (2005) A mouse atlas of gene expression: Large-scale digital gene-expression profiles from precisely defined developing C57BL/6J mouse tissues and cells. *Proc. Natl Acad. Sci. UUSA*, 102, 18485–18490.
17. Aswani, A., Keranen, S., Brown, J. *et al.* (2010) Nonparametric identification of regulatory interactions from spatial and temporal gene expression data. *BMC Bioinformatics*, 11, 413.
18. von Dassow, G., Meir, E., Munro, E.M. *et al.* (2000) The segment polarity network is a robust developmental module. *Nature*, 406, 188–192.
19. Ma, W., Lai, L., Ouyang, Q. *et al.* (2006) Robustness and modular design of the *Drosophila* segment polarity network. *Mol. Syst. Biol.*, 2, 70.
20. Turning, A.M. (1952) The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B*, 237, 37–72.
21. Holloway, D.T., Kon, M. and DeLisi, C. (2005) Integrating genomic data to predict transcription factor binding. *Genome Inform.*, 16, 83–94.
22. Nykter, M., Lähdesmäki, H., Rust, A. *et al.* (2009) A data integration framework for prediction of transcription factor targets. *Ann. N. Y. Acad. Sci.*, 1158, 205–214.
23. Franceschini, A., Szklarczyk, D., Frankild, S. *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, 41, D808–D815.
24. Kerrien, S., Aranda, B., Breuza, L. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.*, 40, D841–D846.
25. Chatr-aryamontri, A., Breitkreutz, B.-J., Heinicke, S. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.*, 41, D816–D823.
26. Wittmann, D.M., Blöchl, F., Trümbach, D. *et al.* (2009) Spatial analysis of expression patterns predicts genetic interactions at the mid-hindbrain boundary. *PLoS Comput. Biol.*, 5, e1000569.
27. Hock, S., Ng, Y.-K., Hasenauer, J. *et al.* (2013) Sharpening of expression domains induced by transcription and microRNA regulation within a spatio-temporal model of mid-hindbrain boundary formation. *BMC Syst. Biol.*, 7, 48.
28. Puelles, L., Harrison, M., Paxinos, G. *et al.* (2013) A developmental ontology for the mammalian brain based on the prosomeric model. *Trends Neurosci.*, 36, 570–578.
29. Vieira, C., Pombero, A., García-Lopez, R. *et al.* (2010) Molecular mechanisms controlling brain development: an overview of neuroepithelial secondary organizers. *Int. J. Dev. Biol.*, 54, 7–20.
30. His, W. (1892) Zur allgemeinen morphologie des gehirns. *Arch. Anat. EntwGesch.*, 1892, 346–383.
31. Jessell, T.M. (2000) Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat. Rev. Genet.*, 1, 20–29.
32. Caspary, T. and Anderson, K.V. (2003) Patterning cell types in the dorsal spinal cord: what the mouse mutants say. *Nat. Rev. Neurosci.*, 4, 289–297.
33. Bard, J.B.L., Kaufman, M.H., Dubreuil, C. *et al.* (1998) An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. *Mech. Dev.*, 74, 111–120.
34. Hayamizu, T., Wicks, M., Davidson, D. *et al.* (2013) EMAP/EMAPA ontology of mouse developmental anatomy: 2013 update. *J. Biomed. Semantics*, 4, 15.
35. Prakash, N. and Wurst, W. (2006) Genetic networks controlling the development of midbrain dopaminergic neurons. *J. Physiol.*, 575, 403–410.
36. Clevers, H. and Nusse, R. (2012) Wnt/ β -catenin signaling and disease. *Cell*, 149, 1192–1205.
37. van Amerongen, R. and Nusse, R. (2009) Towards an integrated view of Wnt signaling in development. *Development*, 136, 3205–3214.
38. Prakash, N. and Wurst, W. (2007) A Wnt signal regulates stem cell fate and differentiation *in vivo*. *Neurodegener. Dis.*, 4, 333–338.
39. Chung, S., Leung, A., Han, B.-S. *et al.* (2009) Wnt1-lmx1a forms a novel autoregulatory loop and controls midbrain dopaminergic differentiation synergistically with the SHH-FoxA2 pathway. *Cell Stem Cell*, 5, 646–658.
40. Omodei, D., Acampora, D., Mancuso, P. *et al.* (2008) Anterior-posterior graded response to Otx2 controls proliferation and differentiation of dopaminergic progenitors in the ventral mesencephalon. *Development*, 135, 3459–3470.
41. McMahon, A.P., Joyner, A.L., Bradley, A. *et al.* (1992) The mid-brain-hindbrain phenotype of Wnt-1–Wnt-1– mice results from

- stepwise deletion of engrailed-expressing cells by 9.5 days post-coitum. *Cell*, 69, 581–595.
42. Danielian, P.S. and McMahon, A.P. (1996) Engrailed-1 as a target of the Wnt-1 signalling pathway in vertebrate midbrain development. *Nature*, 383, 332–334.
 43. McGrew, L.L., Takemaru, K.-I., Bates, R. *et al.* (1999) Direct regulation of the *Xenopus* engrailed-2 promoter by the Wnt signaling pathway, and a molecular screen for Wnt-responsive genes, confirm a role for Wnt signaling during neural patterning in *Xenopus*. *Mech. Dev.*, 87, 21–32.
 44. Shtutman, M., Zhurinsky, J., Simcha, I. *et al.* (1999) The cyclin D1 gene is a target of the β -catenin/LEF-1 pathway. *Proc. Natl Acad. Sci. USA*, 96, 5522–5527.
 45. Tetsu, O. and McCormick, F. (1999) β -Catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature*, 398, 422–426.
 46. Chamorro, M.N., Schwartz, D.R., Vonica, A. *et al.* (2005) FGF-20 and DKK1 are transcriptional targets of β -catenin and FGF-20 is implicated in cancer and development. *EMBO J.*, 24, 73–84.
 47. Fujimura, N., Vacik, T., Machon, O. *et al.* (2007) Wnt-mediated down-regulation of Sp1 target genes by a transcriptional repressor Sp5. *J. Biol. Chem.*, 282, 1225–1237.
 48. Ellisor, D., Rieser, C., Voelcker, B. *et al.* (2012) Genetic dissection of midbrain dopamine neuron development *in vivo*. *Dev. Biol.*, 372, 249–262.
 49. Li, B., Kuriyama, S., Moreno, M. *et al.* (2009) The posteriorizing gene *Gbx2* is a direct target of Wnt signalling and the earliest factor in neural crest induction. *Development*, 136, 3267–3278.
 50. Joksimovic, M., Yun, B.A., Kittappa, R. *et al.* (2009) Wnt antagonism of Shh facilitates midbrain floor plate neurogenesis. *Nat. Neurosci.*, 12, 125–131.
 51. Lee, H.-Y., Kléber, M., Hari, L. *et al.* (2004) Instructive role of Wnt/ β -catenin in sensory fate specification in neural crest stem cells. *Science*, 303, 1020–1023.
 52. Braun, M.M., Etheridge, A., Bernard, A. *et al.* (2003) Wnt signaling is required at distinct stages of development for the induction of the posterior forebrain. *Development*, 130, 5579–5587.
 53. Lee, S.M., Danielian, P.S., Fritzsche, B. *et al.* (1997) Evidence that FGF8 signalling from the midbrain-hindbrain junction regulates growth and polarity in the developing midbrain. *Development*, 124, 959–969.
 54. Yang, J., Brown, A., Ellisor, D. *et al.* (2013) Dynamic temporal requirement of Wnt1 in midbrain dopamine neuron development. *Development*, 140, 1342–1352.
 55. Hovanes, K., Li, T.W.H., Munguia, J.E. *et al.* (2001) β -catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nat. Genet.*, 28, 53–57.
 56. Filali, M., Cheng, N., Abbott, D. *et al.* (2002) Wnt-3A/ β -catenin signaling induces transcription from the LEF-1 promoter. *J. Biol. Chem.*, 277, 33398–33410.
 57. Lin, L., Cui, L., Zhou, W. *et al.* (2007) β -Catenin directly regulates *Islet1* expression in cardiovascular progenitors and is required for multiple aspects of cardiogenesis. *Proc. Natl Acad. Sci. USA*, 104, 9313–9318.
 58. Mastik, G.S., Fan, C.-M., Tessier-Lavigne, M. *et al.* (1996) Early deletion of neuromeres in Wnt-1/- mutant mice: evaluation by morphological and molecular markers. *J. Comp. Neurol.*, 374, 246–258.
 59. Porter, J.D. and Baker, R.S. (1997) Absence of oculomotor and trochlear motoneurons leads to altered extraocular muscle development in the Wnt-1 null mutant mouse. *Dev. Brain Res.*, 100, 121–126.
 60. Cartharius, K., Frech, K., Grote, K. *et al.* (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21, 2933–2942.
 61. Quandt, K., Frech, K., Karas, H. *et al.* (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.*, 23, 4878–4884.
 62. Cohen, C.D., Klingenhoff, A., Boucherot, A. *et al.* (2006) Comparative promoter analysis allows de novo identification of specialized cell junction-associated proteins. *Proc. Natl Acad. Sci. USA*, 103, 5682–5687.
 63. Zhou, X. and Tuck, D.P. (2007) MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data. *Bioinformatics*, 23, 1106–1114.
 64. Augustin, R., Lichtenthaler, S.F., Greeff, M. *et al.* (2011) Bioinformatics identification of modules of transcription factor binding sites in Alzheimer's disease-related genes by in silico promoter analysis and microarrays. *Int. J. Alzheimers Dis.*, 2011, 154325.
 65. Trümbach, D., Graf, C., Pütz, B. *et al.* (2010) Deducing corticotropin-releasing hormone receptor type 1 signaling networks from gene expression data by usage of genetic algorithms and graphical Gaussian models. *BMC Syst. Biol.*, 4, 159.
 66. Hajihosseini, M.K. and Heath, J.K. (2002) Expression patterns of fibroblast growth factors-18 and -20 in mouse embryos is suggestive of novel roles in calvarial and limb development. *Mech. Dev.*, 113, 79–83.
 67. Kobiela, K., Kobiela, A. and Trzeciak, W.H. (1999) Cloning of the lymphoid enhancer binding factor-1 (Lef-1) cDNA from rat kidney: homology to the mouse sequence. *Acta Biochim. Pol.*, 46, 885–888.
 68. Chenn, A. and Walsh, C.A. (2002) Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science*, 297, 365–369.
 69. Nakamura, R., Hunter, D., Yi, H. *et al.* (2007) Identification of two novel activities of the Wnt signaling regulator Dickkopf 3 and characterization of its expression in the mouse retina. *BMC Cell Biol.*, 8, 52.
 70. Mathelier, A., Zhao, X., Zhang, A.W. *et al.* (2013) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, 42, D142–D147.