Database, Vol. 2012, Article ID bas029, doi:10.1093/database/bas029

Original article

CreZOO—the European virtual repository of Cre and other targeted conditional driver strains

Christina Chandras[1,*], Michael Zouberakis[1], Ekaterina Salimova[2], Damian Smedley[3], Nadia Rosenthal[2] and Vassilis Aidinis[1,*]

[1]Institute of Immunology, Biomedical Sciences Research Center Alexander Fleming, 34 Fleming Street, 16672 Athens, Greece, [2]EMBL, Mouse Biology Programme, Via Ramarini 32, 00015 Monterotondo-Scalo (RM), Italy and [3]European Bioinformatics Institute, Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK

***Corresponding author:** Tel: +30 210 9654382; Fax: +30 210 9654210; Email: V.Aidinis@Fleming.gr

Correspondence may also be addressed to Christina Chandras. Tel: +30 210 9656310 (ext. 159); Fax: +30 210 9653934; Email: chandras@fleming.gr

Submitted 3 June 2011; Revised 1 June 2012; Accepted 2 June 2012

The CreZOO (http://www.crezoo.org/) is the European virtual repository of Cre and other targeted conditional driver strains. These mice serve as tools for researchers to selectively 'switch off' gene expression in mouse models to examine gene function and disease pathology. CreZOO aims to capture and disseminate extant and new information on these Cre driver strains, such as genetic background and availability information, and details pertaining promoter, allele, inducibility and expression patterns, which are also presented. All transgenic strains carry detailed information according to MGI's official nomenclature, whereas their availability [e.g. live mice, cryopreserved embryos, sperm and embryonic stem (ES) cells] is clearly indicated with links to European and International databases and repositories (EMMA, MGI/IMSR, MMRRC, etc) and laboratories where the particular mouse strain is available together with the respective IDs. Each promoter/gene includes IDs and direct links to MGI, Entrez Gene, Ensembl, OMIM and RGD databases depending on their species origin, whereas allele information is presented with MGI IDs and active hyperlinks to redirect the user to the respective page in a new tab. The tissue/cell (special) and developmental (temporal) specificity expression patterns are clearly presented, whereas handling and genotyping details (in the form of documents or hyperlinks) together with all relevant publications are clearly presented with PMID(s) and direct PubMed links. CreZOO's design offers a user-friendly query interface and provides instant access to the list of conditional driver strains, promoters and inducibility details. Database access is free of charge and there are no registration requirements for data querying. CreZOO is being developed in the context of the CREATE consortium (http://www.creline.org/), a core of major European and international mouse database holders and research groups involved in conditional mutagenesis.

Database URL: http://www.crezoo.org/; alternative URL: http://www.e-mouse.org/

Introduction

The detailed knowledge gained through sequencing the mouse and human genomes has provided a comprehensive picture of gene arrangement and composition, revealing the startling genetic similarities between mouse and man (1). Together with recent advances in the manipulation of the mouse genome, this information has established the mouse as the premier organism for developing models of human disease and drug action (2). In recent years, highly productive research has been performed in Europe and worldwide on this important model organism to coordinate

Database, Vol. 2012, Article ID bas029, doi:10.1093/database/bas029

and mutate all protein-coding genes in the mouse, using a combination of gene trapping and gene targeting in C57BL/6 mouse embryonic stem (ES) cells, the majority of which are designed as conditional, activated only in specific tissue (special control) and/or life stage (temporal control), represented by the International Knockout Mouse Consortium (3,4). Conditional mutagenesis strategies are particularly relevant for the generation of mouse models of complex diseases, such as cancer, neurodegenerative diseases, atherosclerosis, obesity and diabetes, where the pathology affects multiple organs, thereby enabling the determination of the impact of a particular tissue or cell type on the pathogenesis.

Targeted conditional somatic mutagenesis defines an entirely new scientific field, already widely in use for the production of mutant mouse ES cell lines, each of which carries an altered or 'floxed' allele of a single gene. These mutant ES cells can be readily transformed into mice and the mutation activated by crossing the mouse bearing the floxed allele with a Cre recombinase driver strain to induce the mutation in spatially and temporally determined patterns (5). The full power of conditional mutant ES cell libraries and mice can therefore only be exploited with the availability of well-characterized mouse lines expressing Cre-recombinase in tissue, organ and cell type-specific patterns, to allow the creation of somatic mutations in defined genes (6).

Given the importance of this powerful Cre-recombinase mouse system and in the context of the CREATE (Coordination of resources for conditional expression of mutated mouse alleles) Coordination Action (http://www.creline.org/), a European CreZOO (http://www.crezoo.org/) has been created in an attempt to bring together existing and new Cre driver strains created and maintained in Europe, their complete characterization, and adequate coverage of cell and tissue types in which they are active. CreZOO currently hosts 174 publicly available mice and collects information from mice that are housed in EMMA (www.emmanet.org) (7), the ICS creER^{T2} zoo (www.ics-mci.fr/mousecre), and other scattered European institutes and laboratories that have donated their respective conditional systems to the MUGEN Mouse Database (www.mugen-noe.org/database). Its aim is to promote European collaboration toward unified mouse databases and deposition of existing Cre mouse strains into EMMA and to share and publicize information on mouse characteristics and availability. CreZOO's design is based on the users' convenience and is thus an easy-to-use database, offering a user-friendly query interface and providing instant access to the list of transgenic mice and alleles. CreZOO is hosted at the B.S.R.C. Alexander Fleming Institute (http://www.fleming.gr/).

Database design and implementation

CreZOO is a J2EE application built around a PostgreSQL database that consists of 66 tables. The database itself is fully normalized. The application follows the Model-View-Controller model and is based on a custom-built framework heavily dependant on XML descriptors. Three main tiers compose the CreZOO application:

- the Enterprise Java Bean (EJB) tier;
- session tier; and
- user interface (UI) tier

The EJB tier mainly maps database tables and provides the application with Java objects. These objects are combined and handled by the Session layer, the intermediate layer between the EJB tier and the UI tier, so as to provide the UI layer with the requested output. The UI tier, the layer that handles user interaction and delivers the content, consists mostly of Java Server Pages, custom Java Server Pages tag classes and JavaScript documents. The CreZOO application is hosted in a Glassfish 2.1 server and its source code is available under the GNU general public license at Githubs public repository (https://github.com/bioit/crezoo).

Database availability

CreZOO contents are freely accessible to all visitors and are not password protected. There is no need for visitors to login on entering CreZOO, they may browse the available data by clicking on the respective links (Mice and Alleles) on the homepage tool bar or via the search option, available in CreZOO at all times.

Data entry and curation

Data submission and modification is performed by CreZOO curators who have full access rights to all of the functionalities and contents of the database. The CreZOO curation team surveys journals and other online resources, collects all the data and processes the submission of each Cre driver strain individually before its deposition into the database. All data are reviewed and checked for accuracy and completeness of biological information, including genetic, strain and allelic features. In all cases and most importantly in those of unpublished Cre lines, CreZOO curators work closely with the responsible researcher, to bring the data into standardized formats, resolve issues pertinent to nomenclature and referential integrity and to ensure data accuracy and correctness. Transgenic driver strain data are currently under regular curation and are continuously revised, to maintain a constantly updated version of

CreZOO for the user to refer to. The date that the information for the respective mouse was last updated is also available for the user.

Data categorization

Data are categorized and can be retrieved in several ways. Detailed information is provided for each transgenic mouse presented to the user in three different tabs.

The first section presents in detail 'General' information on the particular transgenic mouse: CreZOO ID, inducibility information (YES/NO), details regarding the availability (European or International repository) and genetic background of the particular transgenic mouse are presented, thus providing the user with all the necessary information about its creation. Finally, users may also find the name of the donating and corresponding researcher(s) and a list of all former names that the particular transgenic mouse has had in the past.

The second tab named 'Gene & Allele Information' provides the user with useful information that include: Promoter name and symbol with links to related databases (MGI, Entrez Gene, Ensembl, RGD and OMIM), in addition to details about the expressed gene and allele information including MGI IDs where available. Most importantly, under this section, information with regard to the specificity of the transgene expression is provided to the user, through the use of anatomical terms according to MGIs Adult Mouse anatomy, with direct links to each term, as well as the developmental stage where expression is noted and respective links to PubMed journals where the information has been retrieved from if possible.

The third and final section for each transgenic mouse 'Documents' provides users with handling and genotyping instructions in the form of hyperlinks or documents, the primary reference with a PubMed link if the mouse has been published, as well as additional references where the particular mouse has also been referenced, used and/or characterized and may contain additional related information.

Detailed allele data may also be retrieved via the allele index page. On finding the allele of choice, users are presented with the allele name, symbol and MGI ID if applicable together with a list of mice that carry the particular allele and the promoter(s) involved in driving gene expression.

Overall, gene/promoter and allelic nomenclature is assigned according to rules and guidelines for mouse genes and strains given by MGI [(8) http://www.informatics.jax.org/mgihome/nomen/strains.shtml] and based on Gene Ontology (GO) [(9) http://www.geneontology.org] and international nomenclature standards.

The CreZOO currently holds (174) listed transgenic mice and (192) related alleles.

Querying the database

CreZOO provides a browsing/filtering interface to the underlying data, allowing formulation of queries, to narrow down the list of potential results and to coordinate screening of certain mutant strains. Automated drop-down menus that include Promoters (official gene symbols), Inducibility and Site of Expression (assigned MA terms) or even a combination of the aforementioned selections allow further refinement of the transgenic mouse list, according to special characteristics defined by the user. Another way in which stored data can be queried is via the function formulating ad hoc (string) queries. At all times, the user can enter the desired query word(s) into the available text box and seek for the particular word(s) at all levels of the database. In all cases, a chart is generated automatically with the corresponding transgenic mice, alleles, promoters or expressed genes that in turn can be accessed by simply clicking on the respective name from the list that appears.

A detailed help section is also available in CreZOO, on the top right-hand corner at all times. In-depth directions on how to view different parts of CreZOO are provided. These are categorized in three different sections: (i) how to find a transgenic mouse; (ii) how to find an Allele of interest and all related Transgenic Mice; and (iii) how to use the CreZOO search engine. Under each of the offered options, step-by-step information is presented, aiming to provide directions for the user to retrieve more efficiently and easily the transgenic tool of choice and providing explanatory information on how to navigate through CreZOO.

Other database characteristics

On the top right-hand side of CreZOO, users are provided at all times with an 'About' page, which not only presents an overview of CreZOO, but because user feedback is fundamental for improving overall database usability and data accuracy, we provide email contact addresses for visitors to email the CreZOO developing team with any comments, on either the website in general or a particular transgenic strain. CreZOO user opinions are extremely useful to us, as we can improve our database, data presentation and update them according to the feedback obtained.

CreZOO, striving to stay up to date with technological advances, has developed Web Services for all transgenic mice and is contributing its data to the Cre8 portal (http://www.creline.org/search_cre_mice). The CREATE consortium representing a core of major European and international mouse database holders and research groups involved in conditional mutagenesis has created a unified portal for worldwide access to these critical resources (6). All databases contribute seven common data-fields from their resources, which include: Transgene or Knock-in Name, MGI

Database, Vol. 2012, Article ID bas029, doi:10.1093/database/bas029

ID, Promoter/Locus, Site of expression, PubMed ID, IMSR status and Inducibility, and can be collectively viewed and searched through the Cre8 portal.

Conclusion/Discussion and future prospects

CreZOO is a comprehensive and user-friendly source of information for retrieving mouse transgenic tools. Core data include inducibility, genetic background, strain, promoter and allelic information, expression patterns and details on the handling and genotyping of mice, together with links to related published articles. Data are integrated through a combination of expert human curation, and use a variety of controlled/structured vocabularies (ontologies). CreZOO is constantly being updated at the curation level to enrich the existing data collection and to upgrade the CreZOO software. Our ultimate goal is to provide as much information as possible from the mouse genome to the mouse phenome. The MA ontology and annotation schema are currently being used to support Cre or other targeted conditional expression annotations and querying capabilities for transgenic mouse data. CreZOO, staying up to date with the use of structured vocabularies for transgenic mouse descriptions, is planning to collaborate with EMAP (http://www.emouseatlas.org/emap/home.html) (8) to best characterize mouse developmental stages and temporal Cre or other targeted conditional expression, thus enabling its users to detect the desired developmental stage via simple keyword(s) search and to eventually make developmental stage assignment as simple as possible. Database content management and data accuracy are some of the key elements for successful database projects. Moreover, sustainability and effectiveness of the database and informatics infrastructure, as measured by the willingness of scientists to use them and deposit their data in such databases for the scientific benefit of the whole community, are of considerable value for a successful database. CreZOO currently brings together Cre and other targeted conditional driver strain information from several European institutes and laboratories, which present their respective conditional systems through the MUGEN Mouse Database, the ICS creER^{T2} zoo and EMMA. CreZOO has made a first attempt to bring together existing and new Cre driver strains created and maintained in Europe for researchers to retrieve more easily and be presented with well-characterized mouse lines expressing Cre-recombinase in tissue, organ and cell type-specific patterns. To provide incentives to potential contributors and researchers to submit their targeted conditional driver strains to CreZOO, our aim is to continue acting as the European specialized database providing a high standard repository, exclusively for Cre and other targeted conditional driver strains. CreZOO taking into account the necessity for database interoperability and combinational data from different resources (9) has already developed web services and is contributing its data to the Cre8 portal (http://www.creline.org/search_cre_mice). The Cre8 portal, created in the context of the CREATE consortium, aims to integrate information from European and International Cre driver strains, through a virtual international application of genetically modified mice expressing Cre recombinase (6). Participating databases (CreZOO/EU [http://www.crezoo.org/], MGI's Recombinase/USA [http://www.informatics.jax.org/recombinase.shtml] and Cre-X-Mice/Canada [http://nagy.mshri.on.ca/cre_new/index.php]) are exchanging/contributing seven agreed datafields to the cre8 portal, thus fulfilling the urgent need on the production and expansion of an international creZOO to complement the conditional mouse alleles currently being generated worldwide. The improved ability to combine relevant information from different online resources will provide additional value to each of the individual interoperable databases, but most importantly to the whole community using mouse as a model organism.

Acknowledgements

We would like to thank CREATE partners for their valuable input and constructive criticism.

Funding

Seventh Framework Programme of the European Commission, Project CREATE [Grant FP7-HEALTH-2007-223487-CREATE]. Funding for open access charge: FP7-HEALTH-2007-223592-IDCC.

Conflict of interest. None declared.

References

1. Brown,S.D., Hancock,J.M. and Gates,H. (2006) Understanding mammalian genetic systems: the challenge of phenotyping in the mouse. *PLoS Genet.*, **2**, e118.
2. Rosenthal,N. and Brown,S. (2007) The mouse ascending: perspectives for human-disease models. *Nat. Cell Biol.*, **9**, 993–999.
3. Collins,F.S., Rossant,J. and Wurst,W. (2007) International Mouse Knockout Consortium; A mouse for all reasons. *Cell*, **128**, 9–13.
4. Ringwald,M., Iyer,V., Mason,J.C. *et al*. (2011) The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium. *Nucleic Acids Res.*, **39**, D849–D855.
5. Branda,C.S. and Dymecki,S.M. (2004) Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell.*, **6**, 7–28.

6. Smedley,D., Salimova,E. and Rosenthal,N. (2011) Cre recombinase resources for conditional mouse mutagenesis. *Methods*, **53**, 411–416.
7. Wilkinson,P., Sengerova,J., Matteoni,R. *et al*. (2010) EMMA–mouse mutant resources for the international scientific community. *Nucleic Acids Res.*, **38**, D570–D576.
8. Baldock,R.A., Bard,J.B., Burger,A. *et al*. (2003) EMAP and EMAGE: a framework for understanding spatially organized data. *Neuroinformatics*, **1**, 309–325.
9. Smedley,D., Schofield,P., Chen,C.K. *et al*. (2010) Finding and sharing: new approaches to registries of databases and services for the biomedical sciences. *Database*, **2010**, baq014.