Xenopus Community White Paper 2014

Contents

Executive Summary

Xenopus is an essential vertebrate model system for biomedical research

Model animals are crucial to biomedical research. Among the commonly used model animals, the frog, *Xenopus,* has great impact because of its unique experimental advantages, cost effectiveness, and close evolutionary relationship with mammals. Over the past 50 years *Xenopus* has made many fundamental contributions to biomedicine and is a cornerstone of research in neurobiology, physiology, molecular biology, cell biology, and developmental biology. Notably, in 2012 Sir John Gurdon was awarded the Nobel Prize for his discovery in *Xenopus* that adult cells can be reprogrammed to become pluripotent, revolutionizing our understanding of cell differentiation and paving the way for methods to induce pluripotent stem cells in humans, transforming biomedical research. Looking forward, *Xenopus* continues to be a powerful system to study fundamental biological and disease mechanisms. Moreover, recent advances in high throughput DNA sequencing, genome editing, proteomics, and pharmacological screening are particularly well suited to *Xenopus* enabling rapid functional genomics and human disease modeling at a systems level.

This White Paper highlights recent contributions of *Xenopus* research to the NIH's mission and outlines current opportunities, including recommendations on how continued, focused NIH investment can maximize the impact of research with the *Xenopus* system.

NIH investment in research using *Xenopus*

The NIH's stated mission is "*to seek fundamental knowledge about the nature and behavior of living systems and to apply that knowledge to enhance health, lengthen life, and reduce illness and disability*". As one means to achieve these goals, the NIH has made a large and sustained investment in biomedical research using *Xenopus*. A search of the NIH RePORT database using the term "*Xenopus*" returned **344 active grants of over $112,000,000 in FY2013 and a total investment of more than $690 million in the last five years.** This includes ~$12 million awarded through a multi-institute "Genetic and Genomic Analysis of *Xenopus*" PAR-12-250/1. This has had broad and significant impact, supporting over 20 projects leading to many critical resources including draft *Xenopus* genome sequences, mutational resources, the *Xenopus* ORFeome and extensive epigenetic, transcriptomic and proteomic datasets. *Xenopus* research is benefiting from these resources developed in response to previous white papers. In addition the NICHD and the OD invested >$5 million to establish the National *Xenopus* Resource (NXR) in Woods Hole, MA and Xenbase, the *Xenopus* online bioinformatics database. The outstanding return on this investment is evident from the research contributions highlighted below.

Xenopus as a model system for human disease gene function

The pace of biological discovery using *Xenopus* is vigorous and has significantly improved our understanding of human disease genes and their mechanisms of action, justifying the NIH's investment. Below are a few example of the human health discoveries made in recent years:

Xenopus embryos are used for *in vivo* analysis of disease gene expression and function:

Congenital Heart Disease – *PNAS* 2011. 108: 2915-2920; Nature 2013. 504: 456-459
Craniofacial malformations – *Nature Genetics* 2012. 44: 709-713.
Bardet-Biedl and Meckel-Gruber Syndromes – *Science* 2010. 329: 1337-1340.
Mitochondrial Neuropathy – *Cell* 2012. 148: 752-764
Cancer – *Mol. Cell* 2013. 52: 1-13. *Dis. Model Mech* 2013. 6: 595-607
Regeneration – *Nature Cell Bio.* 2013. 15: 222-228; *Dev Cell* 2013. 24: 41-51
Wound healing – *PNAS* 2013. 110: 11029-11034
Nephronophthisis– *Nature Genetics* 2013. 45: 951-956
Bartsocas-Papas syndrome – *Science* 2013. 339: 1441-1445
Epilepsy – *Nature Neuroscience* 2011. 14: 505-512
Antiviral immunity - *PNAS* 2013. 110: 14342-14347

Xenopus egg extracts are used for *in vitro* biochemical studies:
Microtubule nucleation - *Cell* 2013. 152: 768-777
Fanconi Anemia – *Mol. Cell* 2009. 35: 704-715; *Science* 2011. 333: 84-87.
Cell cycle – *Nature* 2013. 500: 603-607.
Centromere and kinetochore assembly – *Nature Prot.* 2012. 7: 1847-1869

Xenopus oocytes are used to study gene expression and channel activity:
Pain sensation – *Nature* 2013. 490: 552-555.
Idiopathic basal ganglia calcification – *Nature Genetics* 2012 44: 254-256
Scott syndrome, blood coagulation – *Cell* 2012 151: 111-122
Renal Fanconi's Syndrome – *N Engl J Med.* 2010. 362: 1102-1109
Autosomal hereditary stomatocytosis – *Blood* 2013. 121: 3925-3935
Epileptic encephalopathy – *Nature Genetics* 2012. 44: 1255-1259
Hyperuricemia and gout – *PNAS* 2013. 110: 5223-5528

Xenopus as a model system for understanding fundamental biological processes

Xenopus also plays a crucial role in elucidating the basic cellular and biochemical mechanisms underlying the spectrum of human pathologies. A small fraction of the many recent discoveries are:

Xenopus continues to provide fundamental insight into nuclear reprograming and stem cells (*Science* 2013. 341: 1002-1005; *PNAS* 2011. 108: 17331-17336)
Xenopus contributes to our understanding of genome organization, regulation and epigenetics (*Nature* 2013. 502: 249-253; *Genome Research* 2013. 23: 201-16; *Cell* 2012. 151: 1200-1213)
Xenopus embryos and egg extracts reveal fundamental aspects of cell division (*Cell* 2013. 152: 768-777; *Nature* 2013. 500: 603-607; *Science* 2013. 341: 893-896)
Xenopus tadpoles reveal fundamental mechanisms of neuronal plasticity (*Cell* 2012. 151: 41-55; *PNAS* 2012. 109: 15924-15929)
Xenopus reveals new aspects of eukaryotic nuclear structure and function (*Cell* 2012. 150: 738-751; *Dev. Cell* 2013. 27: 1-13)
Xenopus embryos are used for studies of Wnt and TGF-β signal transduction (*Cell* 2013. 153: 1296-1311; *Science* 2013. 340, 867-870; *Nature* 2012. 485: 195-200)
Xenopus embryos are used for studying organ development (*Nature Genetics.* 2012. 44: 1382-1387; *Dev. Cell* 2012. 23: 292-304)
Xenopus contribute to our understanding of hematopoiesis and vascular development (*Dev. Cell* 2013. 26: 237-249; *Dev. Cell* 2013. 25: 132-143)
Xenopus contribute to our understanding of ontogeny and evolution of immunity (*PNAS* 2013. 110: 14342-14347; *Science* 2013. 343: 366-369; *Mucosal Immunol.* 2013. 6: 358-368)
Xenopus egg extracts provide key insights into DNA damage responses (*Mol. Cell* 2013. 50: 116-122; *Science* 2011 333: 84-87)
Xenopus embryos are used for studying morphogenesis (*Nature* 2013. 497: 374-377; *Dev. Cell* 2012 22: 775-787)
Xenopus are used for small molecule screens to develop therapeutics (*Nat Chem. Biol.* 2010. 6: 829-836; *Blood* 2009. 114: 1110-1122)

Recommendations for NIH investment in *Xenopus* Research

The discoveries listed above (a small subset of all *Xenopus* research) highlights the productivity and impact the *Xenopus* system can have on the NIH's goals to discover fundamental new knowledge and translate that knowledge to human health. *Xenopus* is a particularly compelling model system in the current fiscal climate because it provides excellent value for money allowing the study of conserved biomedical mechanisms at a fraction of the cost of mammalian models and at a much faster rate. To maximize the NIH's investment and realize the full potential of *Xenopus* research, ***we need to continue investment*** in this model. Representatives from the broad and diverse *Xenopus* community met in August 2013 at the Marine Biological Laboratory to establish priorities for continued investment that would have the largest impact on the NIH's mission. There was broad community-wide consensus that eight resources are currently needed, and these were prioritized into two categories: Immediate Needs and Essential Resources.

The Immediate Needs of the *Xenopus* research community:

The *Xenopus* community identified two needs that were the highest priority - renewed funding for the National *Xenopus* Resource and Xenbase. Both of these community resources expire in spring 2015, and their grant renewals are being submitted in May 2014.

1. **Renewal of the National *Xenopus* Resource (NXR):**
   -Will generate, maintain and distribute critical strains of *Xenopus laevis* and *tropicalis*.
   -Will serve as a training venue for husbandry and experimental techniques in *Xenopus*.
   -Will enhance the creation of *Xenopus* mutants using new genome editing technologies.
   -Will facilitate the use of *Xenopus* as a model for human disease.

2. **Renewal of Xenbase: the *Xenopus* model organism database:**
   -Will allow researchers to continue to access specialized highly integrated *Xenopus* data.
   -Will enable the continued dissemination of *Xenopus* data to the broader research community.
   -Will continue to provide essential data sharing infrastructure for NIH funded projects.
   -Will facilitate the use of *Xenopus* as a model for human disease.

Essential Resources for *Xenopus* research community:

In addition to these most-pressing needs, the community has identified six other Essential Resources that should be developed as soon as possible, so that *Xenopus* research can more effectively fulfill the missions of the NIH. The *Xenopus* community considers all of these additional resources to be essential, but understands that priorities must be set, and therefore ranks these as indicated below:

3. ***Xenopus* genome improvement**
4. **Improvement of *Xenopus* antibody resources**
5. **Improve Genome Editing Resources (TALEN/CRISPR) in *Xenopus***
6. **Enhancement of transgenic resources**
7. **Functional characterization of the genome – XvivoENCODE**
8. **Continued support for *Xenopus* training and meetings**

Overarching goal for *Xenopus* research:

Last, the community felt it important to highlight the fact that all of the above resources are essentially a means to achieve the overarching goal of optimizing the utility of *Xenopus* to study human disease and improve human health. *In the last section we have summarized how efforts should be made to* ***facilitate the use of Xenopus to model human disease****.*

Community Recommendations for Attaining Resources:

- Resource PARs should be released to enable renewed funding of the NXR and Xenbase.
- The *Xenopus* community also strongly advocates a renewal of the "Genetic and Genomic Analyses of *Xenopus"* PAR-12-250/1, or a similar mechanism. This is the most effective way to establish critical resources that accelerate research by the entire community. Such resources are essential to the NIH mission, but because they often are not hypothesis-driven, these proposals fare poorly in standard study sections, which typically lack the depth and breadth of expertise that is needed to properly evaluate them.
- The *Xenopus* Community also concluded that new RFAs/PARs to 1) promote the use of *Xenopus* for human disease modeling and 2) exploit ENCODE-type functional genomics in *Xenopus* would be timely and have great impact. New RFAs/PARs in these areas would leverage the established advantages of *Xenopus* with new technologies to facilitate systems-level analyses of disease *in vivo*. Such studies would have a significant impact on translation of big data to knowledge that would fit perfectly with the NIH BD2K initiative.
- Finally it is likely that some of these essential resources could be obtained through traditional R01 mechanisms and existing program announcements, with recognition in the review process that such projects will have a broad impact on the entire community.

Introduction

As outlined above in the Executive Summary, *Xenopus* continues to be a critical animal model for investigating the basic mechanisms underlying human health and disease. *Xenopus* not only provides a remarkably broad experimental platform, but as an amphibian it also bridges the gap between costly mammalian models and the evolutionarily more distant zebrafish model. Recent technological advances in DNA sequencing, high throughput screening and genome editing are particularly well suited to the frog. The community strongly believes that the full potential of the *Xenopus* model could be realized if additional resources were available. The goal of this White Paper is to outline these resources, justify their need, and provide a preliminary plan on how to obtain them.

To identify resources needed by the broad and diverse *Xenopus* community, *Xenopus* researchers met in August 2013 at the Marine Biological Laboratory in Woods Hole, MA. This group of roughly 55 PIs discussed the progress made since the 2011 *Xenopus* White Paper, identified emerging opportunities and agreed upon a prioritized list of resources that would maximize the NIH's investment and have the biggest impact on research using *Xenopus* to understand human health and disease.

The impact of NIH's investment in response to the previous 2009 and 2011 *Xenopus* white paper recommendations have been transformative including; the establishment of the National *Xenopus* Resource center (NXR), Xenbase the *Xenopus* bioinformatic database, draft *Xenopus* genome sequences, mutational resources, the *Xenopus* ORFeome and extensive epigenetic, transcriptomic and proteomic datasets. These enabling technologies, infrastructure and resources are having a major impact on most R01 funded *Xenopus* research. In balancing exciting new opportunities in the *Xenopus* system with the current fiscal climate, the *Xenopus* community identified two immediate needs along with six essential resources for which continued NIH support would protect previous investment and provide maximum impact on accelerating current research and training in *Xenopus*.

Immediate Needs identified by the *Xenopus* community are:

1. **Renewal of the National *Xenopus* Resource (NXR)**
2. **Renewal of Xenbase: the *Xenopus* bioinformatics database**

Essential Resource include:

3. ***Xenopus* genome improvement**
4. **Improvement of *Xenopus* antibody resources**
5. **Improve Genome Editing Resources (TALEN/CRISPR) in *Xenopus***
6. **Enhancement of transgenic resources**
7. **Functional characterization of the genome – XvivoENCODE**
8. **Continued support for *Xenopus* training and meetings**

Overarching goal for *Xenopus*:

Facilitate the use of Xenopus to model human disease

These resources represent the consensus view of the *Xenopus* research community. The priorities were established during the PI's meeting at the MBL and a draft document was the prepared by several members of the *Xenopus* community, including molecular, cell, and developmental biologists. (The authors are listed in Appendix 1.) The draft was posted on Xenbase for a period of 4 weeks and announcements were sent by email from Xine, the online *Xenopus* newsletter in order to solicit comments and input from the entire *Xenopus* community. Feedback was incorporated and the resulting White Paper represents a broad consensus of the community.

Renewal of the National *Xenopus* Resource (NXR)

1A. Impact of the National Xenopus Resource

The National *Xenopus* Resource (NXR) at the Marine Biological Laboratory maintains, breeds and distributes over 50 different inbred and transgenic lines of *Xenopus laevis* and *tropicalis*. Akin to JAX or ZIRC for mouse and fish, the US NXR is one of only three *Xenopus* stock centers in the world (the others are in the UK and Japan) that maintains these critical animal resources. Established in 2010 with strong community support and a P40 grant OD010997 (OD and NICHD, PIs Grainger and Horb) the NXR has had significant and growing impact on *Xenopus* research. It is the main repository to purchase the *X. laevis* inbred J strain frog, which was used for genome sequencing.

- In addition to providing investigator created transgenic and mutant lines to the community the NXR provides services to generate custom transgenic and mutant lines for the community.
- It serves as a centralized location where new technologies related to transgenics and genome editing can be optimized and propagated to the community.
- The NXR serves as the focal point for the community for advanced training. Two separate workshops (bioinformatics and imaging) have already been hosted and will be hosted again in 2014, and new workshops on different topics are being planned.
- The NXR also hosts the biennial PI meeting that serves as the focal gathering for Principal Investigators in the *Xenopus* community to discuss community
- The NXR also serves as a training venue for labs and universities to send individuals to learn new techniques and proper animal care.

The current NXR grant will expire in April 2015 and the *Xenopus* community has identified the **renewal of the NXR grant as a critical resource** that must be maintained and expanded.

1B. Why renewing the NXR is essential

The NXR, together with the European *Xenopus* Resource Center (EXRC) and the Japanese stock center, coordinates the global *Xenopus* effort through monthly video conferences. These meetings are used to coordinate efforts among the stock centers, avoid unnecessary duplication of work and to communicate region-specific news and updates. Within its first three years the NXR has become recognized throughout the world as the physical center point of the US *Xenopus* community. Listed below are the key reasons why the NXR grant should be renewed.

Transgenic and mutant lines

- Many researchers are unable to breed and distribute the transgenic lines they created with NIH funding. As a result valuable animals are lost or need to be independently generated in different laboratories. By maintaining and distributing these lines the NXR maximizes the NIH investment and allows for more efficient use by the community.
- The NXR has successfully created new germ line transgenic lines for the community, and it continues to create new lines based on demands from the community. It acts as the only US location for researchers to obtain specific *Xenopus* lines.
- With the recent advent of TALEN and CRISPR knockout technologies it has become feasible to create specific mutant *Xenopus* lines. The NXR is at the forefront in implementing wider use of these technologies for the community. These are new techniques and the NXR serves to develop these new technologies for wider use by the community.
- The NXR also works with individual scientists to create and breed their individual mutant of interest, as well as helping researchers navigate the complexities of creating TALENs and sgRNAs targeting specific loci in the genome.

Training and dissemination

- In the first three years the NXR has hosted workshops for advanced training in bioinformatics and imaging. These workshops train *Xenopus* researchers how best to analyze and interpret their data, advancing data acquisition and publication to the wider research community. These workshops serve as the springboard for dissemination of new emerging technologies to the wider research community, as well as helping translate big data to knowledge in individual labs.
- The NXR acts as the physical location for the *Xenopus* community to congregate. The *Xenopus* PI meeting has been held twice at the NXR in 2011 and 2013, bringing together *Xenopus* researchers. These meetings allow the researchers to cohesively determine the best way forward. These meetings will continue to be held every two years at the NXR, opposite the larger international *Xenopus* meeting.
- The NXR is a key component of the larger strategic plan outlined in this white paper to accelerate the use of *Xenopus* as model for human disease models, system-level proteomics and functional genomics. As new and upcoming resources are generated the NXR is essential to the development and dispersal to the wider community.
- The NXR provides a unique location for researchers (including those not familiar with using *Xenopus*) to perform experiments using NXR resources, known as a research hotel service.
- The NXR has become the place where new emerging technologies are optimized, standardized and disseminated.
- The NXR provides a unique location to help researchers overcome any stumbling blocks that may inhibit the advancement of their research. The NXR provides training in *X. tropicalis* husbandry and experimentation; helping *X. laevis* researchers integrate the newer diploid genetic model into their research.

1C. How we should proceed

At the most recent *Xenopus* PI meeting (August 23, 2013) there was unanimous strong support from the community for the renewal of the NXR grant, which will be submitted in May 2014 (PIs: Rob Grainger and Marko Horb). The main goals of the NXR for the next five years will be:

- *Continue to obtain transgenic lines from the community.* The NXR has acquired over 50 transgenic lines from individual labs and has bred them for distribution to the community. The demand for these transgenic lines has grown each year. However, there still many lines in the community that need to be acquired and bred into the NXR.
- *Generate new transgenic lines.* The NXR has been successful in generating several new transgenic lines, both through individual requests and community initiated requests. Demand for new lines continues to increase each year and the NXR must continue to meet these demands.
- *Generate new mutant lines.* The NXR has integrated TALEN and CRISPR/Cas mutagenesis into its portfolio as a service-based fee. Community-requested mutant lines can be generated based on demand, ideally as a supplement to the NXR grant. In the coming years the NXR will standardize these techniques and generate new mutant lines for the community.
- *Maintain and enhance advanced training for the community.* The NXR has already hosted two different workshops (twice for Bioinformatics and once for Imaging) for advanced training. In the next phase the NXR will continue to offer these workshops each year and expand them to include new topics.

Renewal of Xenbase: the *Xenopus* bioinformatic database

2A. Impact of Xenbase

Xenbase, the *Xenopus* model organism bioinformatic database, is an essential web accessible resource that integrates the diverse genomic, expression and functional data available from *Xenopus* research. Comparative functional genomics between humans and model organisms has led to a wealth of discoveries and in the post genomic world with tens of thousands of research publications annually and accelerating large-scale data sets being produced, databases like Xenbase are essential to translate this vast body of data into a meaningful biological synthesis. Xenbase enhances the value of *Xenopus* data through high quality curation, data integration, providing bioinformatics tools optimized for *Xenopus* experiments, and linking *Xenopus* data to human genes. Xenbase also plays an indispensable role in making *Xenopus* data accessible to the broader biomedical community through data sharing with NCBI, UniProtK and Ensembl. Thus Xenbase enables discovery, allowing investigators to make novel connections between molecular pathway in *Xenopus* and models of human disease. This is critical to the NIH's new Big Data to Knowledge (BD2K) initiative, which seeks to enable biomedical scientists to capitalize more fully on the avalanche of diverse data being generated.

Recognizing the essential and growing role of Xenbase, and with unanimous community support from the 2009 *Xenopus* White paper, the NICHD funded an expansion of Xenbase (P41 HD064556, PIs Zorn and Vize, 06/2010-05/2015). As a result, Xenbase content, tools and usage have grown tremendously. These include a new genome browser, new BLAST server, powerful gene expression and text mining tools, antibody support, and ongoing annotation of the scientific literature. Xenbase now has over 1000 unique visitors a day. It hosts over 15,500 gene pages, supports both the *Xenopus tropicalis* and *laevis* genomes, contains >44,000 *Xenopus* papers from the literature, >50,000 gene expression images and provides a number of unique large-scale datasets not available at other databases. **In order to maintain these essential services renewed funding for Xenbase is an immediate top priority of the *Xenopus* research community**.

2B. Why renewing Xenbase is essential

Maintenance of Xenbase is critical to protect and maximize the NIH's >$100 million annual investment in *Xenopus* research. Xenbase plays a critical role in making data from this research accessible in a meaningful way. Without Xenbase much of the *Xenopus* data generated from R01 funded investigations would remain largely buried in the scientific literature.

It is essential to maintain and expand the functionality of Xenbase for a number of reasons. Firstly, Xenbase has become the single most important clearinghouse for *Xenopus* data, serving both *Xenopus* investigators and the broader biomedical research community. Xenbase adds value by providing high quality annotation, tools specific for *Xenopus* research and the integration of diverse data types in a way not available at any other single database. Xenbase inter-relates *Xenopus* genomic, epigenetic, mRNA and protein sequence, with gene expression and gene function as well as physical reagents such as antibodies, morpholinos and transgenic lines from the literature. Xenbase links all this *Xenopus* data to other model organism and human disease databases such as OMIM. By administering the HGNC approved *Xenopus* gene nomenclature Xenbase associates *Xenopus* gene centric data to the correct human genes and then through automated weekly data sharing Xenbase provides this curated *Xenopus* data to many external resources including; NCBI, Entrez Gene, UniProtK and Ensembl.

Xenbase has become a virtual focal point for the *Xenopus* community with over 1200 registered users and hosts a website of community announcements, protocols, educational material and community resources. Xenbase provides critical data sharing infrastructure for the *Xenopus* Stock centers and many other NIH funded initiatives including, pre-release *Xenopus* genomes (Rokhsar & Harland; HD065705); epigenomics data (Veenstra; HD069344, Baker HD076839); the ORFeome (Stukenberg et al; HD069352); and large-scale expression screens (Amaya, Kirschner, Pollet, Perone, Ueno, Wheeler & Zorn labs). It also has become clear that in order to keep up with technological advances and the

increasing pace of *Xenopus* research the community needs Xenbase to develop new infrastructure, data pipelines, bioinformatics tools and annotation teams to support the exponential growth in NextGen sequence data and the increasing use of *Xenopus* phenotypes as a tool to model human diseases. Immediate plans include support for the *Xenopus* transcriptome (Khokha et al; GM099149), *Xenopus* proteome (Kirschner; GM103785) and XenMine bioinformatics tools (Baker; GM095346).

Without Xenbase the data from these important projects would be much less accessible. Thus Xenbase has a broad and significant impact on *Xenopus* researchers, the broader biomedical community and it is essential to the NIH's mission. In particular, the role of Xenbase is well aligned with the NIH's new Big Data to Knowledge (BD2K) initiative, which seeks to enable the integration and translation of large, complex and diverse data sets into meaningful knowledge that can improve human health.

2C. How we should proceed

With unanimous support from the community, Drs. Vize and Zorn are submitting a competitive renewal for Xenbase in early 2014. In order for this to be received effectively it is critical that the NIH re-issue a PAR such as the PA-08-180 "Resource Program Grants in Bioinformatics" that funded the previous Xenbase P41 grant and which is specifically addresses model organism databases. This is essential because Resource initiatives such as Xenbase cannot be properly reviewed in study sections in comparison to hypothesis driven R01 applications. The current plans for renewal of Xenbase, which will maintain operations and allow the addition of new data types and functionality are:

- Aim1 Maintain Xenbase and continue ongoing curation of *Xenopus* research data.
- Aim2 Enhance Xenbase support for large-scale NextGen sequencing data.
- Aim3 Develop support to enhance *Xenopus* as a model of human diseases.
- Aim4 Expand data sharing infrastructure and bioinformatic tools.

Importantly, these proposed improvements are guided by user feedback from the PIs meeting, the Xenbase external advisory board and recent surveys. The community identified three areas in which their research requires new Xenbase support. First, there is an urgent need for better support of the avalanche of NextGen sequence data being generated (Aim 2). Second, the community strongly supports the development of mutant (CRISPR/TALEN) and phenotype tools that use cross-species ontologies enabling the comparison of *Xenopus* gene-phenotypes to identify and characterize human disease models (Aim 3). Lastly, as the community increasingly exploits *Xenopus* for systems level analyses they require more sophisticated bioinformatic tools to mine large data sets (Aim 4).

Proposed Funding: Xenbase should be funded via the reissue of a Resource Program PAR.

Xenopus Genome Improvement

3A. Impact of the Xenopus genome

Improving the *Xenopus* genome sequence is a top priority of the research community. We are in the midst of a transformative time in biomedical research, with advances in quantitative DNA sequencing technology and a revolution in genome editing. These are enabling an unprecedented analysis of cell function, development, homeostasis and disease at both a genome-wide level and high-resolution mechanistic depth. The experimental advantages of *Xenopus* make it ideally suited for this type of systems-level analysis. In order to efficiently realize this potential a complete highly annotated *Xenopus* genome sequence is required.

Recognizing the importance of *Xenopus* the NIH and the Department of Energy (DoE/JGI) have made a substantial investment in sequencing the *Xenopus* genomes resulting in the first X. *tropicalis* genome assembly published in 2010. With strong community support and funding from R01s GM086321 and HD065705, the Rokhsar and Harland groups have generated an improved *X. tropicalis* genome version 8 assembly (Xt.v8) as well as a draft *X. laevis* genome sequence (Xl.v8) both of which are due to be released in 2014. Pre-release versions of the *tropicalis* (v7) and *laevis* (v7) genomes are available on Xenbase; their high value is reflected in their substantial usage by the *Xenopus* community. The current *Xenopus* genomes are outstanding resources that continue to have a tremendous impact on all areas of *Xenopus* research. However, consistent with experience from human, mouse, and zebrafish genomes, more work is needed to complete a fully annotated a reference sequence. Continued investment in two key areas is required to realize the full potential of the *Xenopus* genomes.

- Improve the *Xenopus* genomes – finishing reference sequence, gap closure
- Curation and Annotation of the Genome

3B. Why improve the Xenopus genome?

A complete genome sequence with well-annotated gene models is essential for virtually all studies of gene expression and function. Identification of the transcription/translation start sites and intron-exon boundaries are critical for designing gene disruption experiments using morpholinos, shRNAs, TALENs and CRISPRs. Well-annotated genomes are critical for identifying cis-regulatory elements and essential for systems-level functional genomics. Complete gene models are also essential for proteomics and thus are critical to the large and very productive *Xenopus* cell biology community that relies on mass-spectroscopy analysis. Both the *tropicalis* and *laevis* genomes need improvement. In addition to being essential for experiments in both species comparative genomics between *laevis* and *tropicalis* has proven to be a powerful tool for characterizing conserved non-coding regulatory elements. Moreover, because *laevis* underwent an ancient whole genome duplication, the two *Xenopus* genomes offer a unique model of vertebrate gene diversification and genomic evolution.

- Current status of the genomes: The *tropicalis* genome is ~1.5 Gbp with haploid chromosome number of N=10. The allotetraploid *laevis* genome is twice as large at ~3 Gbp with disomic inheritance (two segregating alleles at each locus) and a haploid chromosome number of N=18. The newest *tropicalis* Xt.v8 assembly was generated using improved computational methods, incorporating additional BAC-end data, and by integrating a dense genetic map, resulting in a more contiguous assembly mapped to chromosomes with 50% of the contigs > 72 kb in length. The X. *laevis* Xl.v8 (contig N50 length ~20 kb) benefited from collaboration with the Japanese *Xenopus* Genome Project (Masanori Taira and Asao Fujiyama, et al., U. Tokyo and NIG), which provided BAC and fosmid end sequences along with FISH to localize sequence to chromosomes. Both Xt.v8 and Xl.v8 draft genomes have been annotated with standard automated algorithms resulting in ~25,000 gene “models” (exon-intron structure predictions) in X. *tropicalis* and ~40,000 in X. *laevis*, including almost all known RefSeq cDNAs.
- Limitations of current genomes: Although these genome assemblies are high quality and a tremendous resource, there are still considerable limitations including a significant number of gaps,

many incomplete gene models, and incomplete annotation of non-coding elements. In the current *tropicalis* assembly there are still 31,431 within-scaffold gaps spanning an estimated 25.2 Mb of sequence. These gaps hamper automated prediction algorithms and can produce erroneous gene models. In addition, alternative splicing has in many instances resulted in multiple gene models predicted per locus, but the confidence of the various models is not clear as exon combinations and frequencies have not been validated. Indeed independent assessments of the *tropicalis* genome by the Gilchrist and Veenstra groups suggests that only 50% of the gene models are high confidence with ~20% of gene models missing 100 bp or more of the 5' and 3' mRNAs, and up to 13% of the models lacking a complete ORF. The machine generated gene models have not been refined by proteomic or RNA-seq data that has recently become available, nor have ENCODE approaches to define the transcription start sites such as TSS-seq or ChIP-seq of chromatin modifications been employed. Thus there is a pressing need to improve the genome sequence and its annotation but there are currently limited funds dedicated to this task.

3C. How we should proceed

A *Xenopus* Genome Steering Committee was established in 2009, which includes a diverse group of PIs and a panel of bioinformaticians. This committee, together with the broad *Xenopus* community prioritized two key areas of genome improvement; 1) closing gaps in the assembly and finishing the sequence; and 2) improving gene model annotations. *Xenopus* can leverage lessons learned from the genome projects of human and other species and incorporate many different data sources to improve the *Xenopus* genome sequence and annotation.

- Leverage data from currently funded ongoing projects including proteomics, ORFeome, RNA-seq transcription profiling, TSS-capture methods and ChIP-seq epigenetic analysis.
- It will be necessary to support new projects to generate new data such as targeted amplification of all putative missing mRNAs, sequencing of additional BACs and fosmids for gap closure and additional genetic mapping.
- Continued bioinformatics support is critical to maximize the value of *Xenopus* as a model for human disease. New bioinformatics tools and analytical approaches will need to be developed to integrate the diverse data types described above. Currently there are no funded efforts to coherently integrate these different data types into a quality set of reference gene annotations. We anticipate that a certain amount of manual curation will also be needed, as this has remained the gold standard from the human genome and multiple models including mouse, fly and worm. The community strongly recommended that simultaneous efforts to improving both *tropicalis* and *laevis* genomes be pursued. One compelling reason is that comparisons between the two species can help accelerate annotation as gene models validated in one can be used as templates models in the other species. In addition to accelerating genome annotation this comparative approach has proven very powerful for characterizing conserved regulatory elements (e.g., comparisons across mammals and across various *Drosophila* species). Analysis of the conserved synteny between *Xenopus*, human, and other vertebrate model systems, as well as between the two *Xenopus* species, allows precise assignment of orthology across species and will facilitate the use of *Xenopus* for human disease modeling. Xenbase will serve as a centralized repository for data integration and facilitate dissemination to public databases (NCBI, UCSC and Ensembl). To this end members of the *Xenopus* Genome steering committee have been in communication with NCBI leadership to improve data pipelines and increase the visibility of *Xenopus* data.

Proposed funding mechanisms: We anticipate that number R01 level projects for both data generation as well as data integration will be required to achieve these goals. These could be funded through the existing *Xenopus* Genetics and Genomics PAR (PAR-12-250), which we strongly recommend be renewed.

Improvement of *Xenopus* antibody resources

4A. Impact of improving antibody resources

Current cell and developmental biology research is moving towards systems level analyses performed in the context of the whole genome and proteome. Despite the advance in *Xenopus* genomic and genetic resources, the availability of proteomic tools including specific antibodies specific to *Xenopus* proteins remains limited. The generation of a wide range of antibodies would facilitate all aspects of research in *Xenopus*, including neurobiology, immunology, developmental, cellular, and molecular biology. There is an acute need for a community-based resource containing a comprehensive collection of tools that allows one to specifically detect, purify and analyze the *Xenopus* proteome. Antibodies are the most useful category among such tools, given their high affinity interaction with the antigen, and a large number of developed immunoassays. Systematic generation of specific antibodies and recombinant antibody libraries is a priority for the *Xenopus* community.

4B. Why develop antibody resources for Xenopus?

Xenopus has been at the forefront of cellular, molecular, and developmental biology for many decades. However, the broader impact of the *Xenopus* model system for *in vivo* cell, biological and biochemical approaches has been compromised by the lack of high affinity, high specificity antibodies. Outside of the fields of cell cycle control, DNA repair and cytoskeletal regulation, relatively few *Xenopus*-specific antibodies have been generated, and while there are some notable exceptions, there is relatively low cross-reactivity with existing or commercial antibodies. Therefore, there is a widespread and acutely perceived need to develop antibodies for the *Xenopus* community, and in parallel, to have an open access database of antibodies tested by the *Xenopus* community.

To address this need, Dr. Anna Philpott carried out a broad survey at the most recent *Xenopus* PI Meeting at the MBL, Woods Hole in August 2013, to prioritize specific classes of proteins, for which antibodies are needed, and determine how to develop better antibody resources and databases. The majority of the attendees voiced strong support to the establishment of an antibody resource, as one of the main priorities for the community. Purchasing and testing commercial antibodies frequently waste of time and financial resources, as many antibodies are poorly characterized or made towards non-conserved epitopes that are not present in *Xenopus* proteins.

4C. How we should proceed

This White Paper therefore supports the generation of specific high-affinity antibodies against *Xenopus* proteins at a large scale, with emphasis on high throughput approaches. This resource will be essential for exploiting the advantages of the *Xenopus* system to its full impact as a biomedical model. Short-term and long-term solutions are proposed.

- An immediate need is to establish individual funding to create a library of antibodies. Funding of investigator-initiated grants through the *Xenopus* PAR that will help develop the generation of new antibodies, such as has been proposed using in a pilot llama immunization project by Sergei Sokol.
- In order to obtain information on existing antibodies, Drs. Anna Philpott and Chenbei Chang, and Xenbase curators surveyed the community to collect and organize data on antibodies generated by community members as well as commercially available antibodies raised against other species that have been tested for cross-reactivity to *Xenopus*. These data were submitted to Xenbase.
- Xenbase developed and is actively curating a *Xenopus* antibody database containing information on all of the community-generated, commercial and published antibodies that have been validated in *Xenopus*.
- Annual surveys of *Xenopus* PIs will help keep this database up-to-date. Moreover, the antibody database will be cross-referenced with Xenbase gene-specific pages to seamlessly indicate antibody resources available to proteins of interest. Additional input will be solicited at the annual meetings (International *Xenopus* meeting and PI meeting).

- In the longer term, the *Xenopus* community must generate and maintain a collection of antibody tools for *Xenopus* antigens that are made widely available. Most useful antibody tools would be those that can be easily maintained and be readily accessible to the community, such as defined (sequenced) monoclonal antibodies or recombinant antibody libraries. Conventional monoclonal antibodies should be deposited in the Developmental Studies Hybridoma Bank.
- Special attention should be given to the development of recombinant antibody libraries, including those generated from single domain antibodies. Compared to the cost involved in the production and maintenance of conventional antibodies, recombinant antibody technologies are more economical and will serve better as a large proteomic tool resource. Whereas single recombinant IgG heavy chains can be used in immunochemical studies, they have modest affinities and protein stability. By contrast, naturally occurring single domain antibodies (nanobodies) from llamas and camels have excellent stability and comparably high affinity. Finally, nanobody cDNA libraries can be easily expanded and maintained in bacterial and eukaryotic systems.

4D. Anticipated outcomes

After the PI meeting at MBL, Woods Hole, in August 2013, a joint effort was initiated by Sergei Sokol, Anna Philpott, Matt Guille and several other *Xenopus* PIs to prepare a number of purified *Xenopus* antigens for a pilot llama immunization project aimed at assessing the utility of *Xenopus* antigen-specific nanobody libraries for the community. If this pilot project is successful, it is anticipated that large-scale antibody libraries will be made. Individual research groups would next take advantage of these libraries for the selection of high-affinity antigen binders of appropriate specificity using phage display protocols, and optimize their usage based on their individual projects. Both these generated libraries, and individual antigen-specific nanobodies will be maintained by the European *Xenopus* Stock Centre or the National *Xenopus* Resource at the MBL and will be freely distributed to *Xenopus* investigators.

We anticipate that acting upon this priority would result in the generation of a large number of antibodies that will be invaluable for a wide range of applications. We envision that both conventional and recombinant antibodies will be utilized for the analysis of protein-protein interactions and gene regulatory networks (via chromatin immunoprecipitation), in structural biology, for functional interference and signal transduction studies. Subsequently, panels of antibody tools will be developed with specificities for particular cellular components, chromatin modifications or signaling pathway components that could be used for antigen recognition *in vitro*, in fixed cells and in vivo, during live imaging.

Proposed funding mechanisms: We anticipate that number R01/R21 level projects will be required to achieve these goals. These could be funded through the existing *Xenopus* Genetics and Genomics PAR (PAR-12-250), which we strongly recommend be renewed.

Improve Genome Editing Resources (TALEN/CRISPR) in *Xenopus*

5A. Impact

Xenopus has been a powerful system for rapid functional genomics using both mRNA-mediated over expression and morpholino knockdown. Recent genetic modification techniques (TALENs and CRISPR/Cas) have been shown to work exceptionally well in *Xenopus,* making *Xenopus* amenable to gene knockout studies that will help expand the repertoire of experimental possibilities. To facilitate the use and impact of *Xenopus* as a model for human health and disease, these new loss-of-function technologies must continue to be optimized, standardized and disseminated. Greater flexibility for knockout and knockdown technology will allow *Xenopus* investigators to fully exploit loss-of-function in *Xenopus* oocytes, cell free extracts, developing embryos and adults. These approaches are essential for fully exploiting the system to model diseases.

Enhancement of this new genome editing, enabling the mutation and genetic modification of any sequence in the genome, will be transformative, especially when coupled to the established experimental advantages of *Xenopus*. Moreover establishing mutant frog lines with this technology in key genes that function in many aspects of development and disease (such as oncogenes) is critical and will expand functional genomic analysis to later development stages, which will be essential for human disease modeling and regenerative biology studies.

These technologies are easily adapted to the *Xenopus* system as they only require microinjection into early embryos and can be used in either *X. laevis* or *X. tropicalis*. Although gene function can be assessed with bi-allelic gene targeting in F0 embryos and tadpoles, due to the mosaic nature of the mutations heterozygous animals must be bred and maintained to create true null homozygous mutant offspring. Another important consideration is that most *Xenopus* laboratories are not experienced in breeding and maintaining mutant lines, particularly with the diploid *X. tropicalis* that is ideal for genetic studies. To facilitate the use of this technology the NXR now offers the generation of mutant lines as a fee for service. Although *X. laevis* is allotetraploid, which complicates its use for genetic knockout studies, the gene editing feature of TALENs/CRISPRs, such as the introduction of epitope tags into any protein coding gene in the genome, is likely to be extremely powerful in combination with the excellent cell biological and biochemical approaches that are standard in *X. laevis*. In addition, by utilizing the *X. laevis* and *X. tropicalis* inbred lines deficient cell or tissues cold be transferred in normal recipients for further characterization of the process of interest, and the inbred genetic background would also minimize effect from gene polymorphism. In order to disseminate this technology to the community as quickly as possible and to promote wider use of *X. tropicalis* requires integration with the NXR.

5B. What is needed to enhance genome editing technologies in Xenopus

- The new genome targeting approaches will permit rapid development of numerous mutant *Xenopus* lines. This should be optimized and standardized for both *X. tropicalis* and *X. laevis*.
- Generate community requested mutant lines, such as key disease related genes; this would overlap well with development of *Xenopus* as a non-mammalian model system for human disease research.
- To facilitate creating knockouts will require a more complete annotation of both genomes, as outlined in the genome improvement portion of this white paper (section 3).
- Generate advanced tools for establishing knock-in strategies in *Xenopus*, including gene replacement, gene editing and targeted insertion.
- Create new transgenic lines that ubiquitously express humanized Cas9. This will allow for easier creation of single and double/triple mutant lines by simple microinjection of sgRNA(s). This overlaps well with the transgenesis portion of this white paper (section 6).
- Combine targeted insertion with large-scale proteomics, which can be achieved with the thousands of embryos that can be obtained from a single fertilization in *Xenopus*.

- Standardize and enhance use of different TALEN and CRISPR systems, such as dominant active/ repressor and nickase versions of Cas9.
- Optimize a strategy to target TALENs to germ cells only, thus bypassing somatic mutations in the F0 generation.
- Bioinformatic support to identify CRISPR/TALEN sites in the *Xenopus* genomes as well as data on the resulting mutant animals and phenotypes needs to be collated.

5C. How we should proceed

This genome editing technology will transform the use of *Xenopus* in the biomedical research community and the NIH should encourage and enhance its use by the community. Since one of the difficulties (cost and space) within the *Xenopus* community is the raising and breeding of lines of *Xenopus* the first step is to encourage labs to interact with the NXR to create specific mutants. As this is easiest to achieve in *Xenopus tropicalis*, this would be the logical step forward. However, since many labs work mainly with *laevis* this would require labs to buy new housing for tropicalis, and at the moment this is not an achievable immediate goal. Administrative supplements to existing *laevis* grant should alleviate this bottleneck and accelerate the use of this technology in *Xenopus*.

Additional R01 grants focused on enhancing TALENs and CRISPR/Cas and developing knock-in strategies in *Xenopus* should be funded via the PAR: Genetics and Genomic Analysis in *Xenopus*. As there is only one year left on this PAR, extending this PAR would benefit the entire biomedical research community. Additional grants should be funded through an R24 mechanism targeted for resource related projects on animal models.

The best way to maximize the utility of this technology would be a coordinated effort within the community centered at the NXR to generate mutations in several hundred key genes that are studied by many investigators. The long-term goal would be to mutate all genes in the genome and store these as froze sperm that could be easily distributed to the community. This would generate a resource similar to JAX for mice and ZIRC for fish, which have proven to be very successful. Further, it would provide an overall cost savings by reducing duplication of effort by individual investigators. With this in mind the NXR is soliciting a list of genes from the *Xenopus* community to prioritize for mutation. Importantly this initiative would require a substantial investment and the community strongly urges the NIH to issue a resource PAR enabling a consortium of investigators, led by the NXR. In addition it will be important for Xenbase to build tools to support data related to CRISPR/TALENS and the resulting mutant phenotypes, which are part of the planed Xenbase renewal. Together the proposed generation, distribution and data collection would be a modest, but highly productive, investment relative to the already substantial NIH investment.

Enhancement of transgenic resources

6A. Summary

There is a strong and pressing need for transgenic *Xenopus* resources. Significant results in cell and developmental biology have come from using transgenic systems in frogs, but few transgenic *Xenopus* lines are currently available. Characterized transgenic lines and production of new lines are critical for advances in basic cell and developmental biology, pluripotent stem cell development, and understanding the developmental origins of adult disease. Transformative advances will come from combining transgenic resources with established advantages of the *Xenopus* model. Although there are several highly efficient methods to generate transgenic *Xenopus*, individual researchers often lack the space and time and technical expertise required to make and characterize them. *Xenopus* stock centers are geared to generate or facilitate production of transgenic lines as well as accept lines for distribution. Community input and working groups will be coordinated by NXR to organize production of the most commonly desired transgenic lines.

6B. Transgenic Resources: Current status and future promise

Transgenic resources in *Xenopus* have revealed significant biological insights and provide paths for expanded development of the resources.

- **Reporter lines:** These lines allow visualization of a subcellular component, tissue, or whole body via expression of a fluorescent protein. The great variety of uses of such lines include live imaging, lineage tracing, developmental signaling pathways, enhancer trap screens, analysis of promoter/enhancer activity of non-coding sequences and screening endocrine disrupting chemicals. These uses are particularly advantageous in *Xenopus* due to the large number of offspring, transparent larvae, and ease of embryonic gene expression and tissue manipulation. A few tissue reporter lines are currently available, including some cellular components, rod cells, heart myocardium, and exocrine pancreas, but expanding this set to other tissues such as neural crest, specific brain regions, liver, lung, kidney is necessary to realize the full potential of the use of *Xenopus* to model developmental and disease processes.
- **Inducible lines:** These lines allow ubiquitous inducible expression of transgenes, utilizing heat shock or small molecule ligands as inducing agents. Numerous early and late developmental events have been studied using these technologies. Inducible lines are especially useful to avoid undesired early embryonic effects and are convenient to use because a single line is sufficient to realize the experiments. However, the value of these lines is limited to a relatively narrow range of experiments compared to binary expression systems (see next).
- **Binary expression systems:** These versatile systems enable precise temporal and spatial control of transgene expression and are comprised of two parts: 1) a promoter line controlling tissue-specific expression of a ligand-activated transgenic transcription factor and 2) a transgene line activated upon ligand addition in a specific tissue determined by the promoter line. Systems shown to work in frogs include the GAL4/UAS, Tet-On, and Cre/lox systems. Mechanisms of regeneration and development of muscle, liver, and pancreas development have been elucidated using these systems. A significant advantage of binary expression systems comes from the combinatorial possibilities of having separate promoter and transgene lines, which multiplies the combinations of tissue specificity and transgenes of interest readily available to researchers. Despite this great potential, few characterized transgenic lines suitable for future study of disease processes are currently available.

Promising opportunities exist to merge transgenic resources with the advantages of the *Xenopus* model and with resource-building efforts (genome sequences, ORFeome, XvivoENCODE). We envision the first high-throughput *in-vivo* functional-genomic analyses in specific tissues and developmental time points, which are not feasible in any other model system. It is clear that transgenic lines are needed both for both *X. laevis* and *X. tropicalis*. Although *X. laevis* has a longer generation time its large size make it

more amenable to cell biology and biochemistry, which would benefit from transgenic lines. The shorter generation time and diploid nature of *X. tropicalis* make it ideally suited for binary transgenic systems and more classical genetics.

6C. How we should proceed

Three main steps are required to realize additional transgenic lines.

- **Consolidation of existing transgenic constructs** from the research community. A number of excellent transgenic constructs have already been generated such as the modular gateway compatible pTrangenesis plasmids and cell-component GFP-fusion reporters. In most cases however these have only been used in transient assays and transgenic animals are not available. The first step will be to consolidate these constructs and prioritize those with the most wide spread utility to generate lines. This is already underway as the European and Japanese *Xenopus* stock centers distribute these plasmids to the community and the pTransgenesis system is used at the NXR in transgenic line production. The NXR and EXRC (and in some instances with help from the Japanese stock center) have already established a network to exchange lines.
- **Identify tissue-specific promoter/enhancers** Promoters from other species often work in frogs and novel candidate promoters will be identified from the published literature as well as from ENCODE-type prediction of cis-regulatory modules in the *Xenopus* genome. In both cases, promoters must be validated in frogs for appropriate expression characteristics for use as a reporter or for tissue-specific transgene expression. Additional transgenic vectors will also need to be generated with the emphasis on Gateway compatible modules that can be reused for many purposes.
- **Production and characterization of lines** After promoters and constructs have been validated with transient assays, transgenic founders need to be made, reared, and tested for germ-line transmission as well as continued appropriate transgene expression. Transgenic lines will be produced in the inbred J strain *X. laevis*, which will permit adoptive cell, tissues or organ transfers.

A larger initial investment is recommended to jump-start the widespread use of binary expression systems. A strategic effort with community input to produce *a priori* a critical mass of 10 to 15 promoter and transgene lines that could be used combinatorially would enable a large number of projects "off the shelf". Production of such a select set of transgenic lines could be done at the NXR or in a large lab or group of labs funded by the PAR - Genetic and Genomic Analyses of *Xenopus*. With promoter and transgene lines in hand, it would then be cost-effective for individual R01 recipients to simply use these lines or custom make a single promoter or transgene line.

Functional characterization of the genome – XvivoEncode

7A. The Need and Opportunity

It is now clear that most of the non-coding genome plays critical roles in gene regulation during development, homeostasis, disease and repair. Although we are beginning to appreciate the importance of genome regulation we have only scraped the surface of our mechanistic understanding, and *Xenopus* is poised to be an instrumental model in these studies. The ENCODE consortium has provided a critical toolkit in the overall understanding of genomic regulation. Multiple insights have been gained and the interdisciplinary nature of the project has led to an emergence of new technologies and concepts governing genomic architecture. While ENCODE has advanced our understanding and provided new molecular and biochemical tools, the effort has primarily focused on transformed human cell lines, which are clearly not indicative of in vivo tissue. The next critical step is to apply ENCODE-type analysis to normal development and in disease progression in vivo, but multiple factors complicate these efforts:

1. The lack of homogeneous cellular population of untransformed primary cells.
2. The inability to amass the numbers of cells required for examination.
3. The difficulty in validating the genomic findings *in vivo*.
4. The inability to functionally test genomic regions *in vivo.*
5. The inability to trace primary cells to their earliest developmental origins in the embryo.

We propose that an in vivo ENCODE project using both *X. tropicalis* and *X. laevis* be launched. The rationale is that human and other genomes can only be exploited to the extent its regulation and the functional elements are understood during conditions of development and disease. We strongly propose that *Xenopus* fill an important niche in the ENCODE endeavor as use of this system alleviates many of the above difficulties that are inherent in human as well as most other model systems. Further, it provides exciting opportunities not found in other systems.

- Frogs are inexpensive, their embryos are plentiful (thousands of synchronously developing embryos), and highly accessible at all stages of development. As a powerful tetrapod vertebrate model system, *Xenopus* is well positioned between distant metazoans (such as *Drosophila* and *C. elegans*) and fish and less accessible mammalian models. As such, the *Xenopus* system provides a powerful platform to investigate normal vertebrate development and its deregulation in disease.
- Simple dissections and sorting flow cytometry methods can yield homogeneous populations of primary “ex vivo” cells throughout the life cycle of *Xenopus* from the earliest fertilized egg and embryos to the adult. Millions of total cells can be obtained; even tissues or cell types with small amounts of starting material can be purified from a transgenic animal line or by using dissections. Recent progress in sample library preparation methods allow the production of high quality ChIP profiles from relatively low numbers of cells (two or three replicates of 40,000 cells). This offers the tantalizing prospect of true ex vivo profiling of specific embryonic regions and cell types during early vertebrate development.
- The availability of both *X. laevis* and *X. tropicalis* genomes offers tantalizing prospects that go beyond what a single genome can offer. Comparative genomics approaches are very powerful due to high degrees of gene orthology, high conservation both within genes and in non-coding elements such as long non-coding RNA and regulatory elements. Moreover, relative to *X. tropicalis* a whole genome duplication event has occurred in *X. laevis*. Genome duplication events have happened at the root of vertebrate evolution and are thought to contribute to diversification due to selective gene loss and neo-functionalization. Segmental and whole genome duplications as well as sequence loss (aneuploidy) also play a role in cancer. Therefore the two *Xenopus* genomes significantly boost the utility of either genomes and offers an important and unique new vertebrate model system for genome duplication and evolution.

- The two species can be used to make chimeric explant or hybrids that allow genomic separation technologies to effectively investigate how chromatin states change in response to signaling from different cell types.
- As John Gurdon's nobel prize illustrates *Xenopus* are and excellent model system to investigate the in vivo mechanism of chromatin modulation during development and cellular reprogramming.
- In the 20th century, *Xenopus* was a key model that led the way in functional analysis of vertebrate genes using classical molecular and embryological analyses. In the 21st century, *Xenopus* will play a critical role in testing predictions that emerge from post-genomic big data projects.

7B. How we should proceed

To provide exciting new inroads into the genomics of birth defects, cancer, aging, regeneration and cellular reprogramming and to exploit *Xenopus* genomics, biochemistry, embryology and molecular biology, it is necessary to assemble and implement a new ENCODE type initiative, referred to as XvivoENCODE. A three-pronged approach is proposed for maximal efficiency and utility.

- **First,** we will piggyback on the newest technologies developed over the past 10 years through the NHGRI funded ENCODE project. Therefore the XvivoENCODE project will be highly cost-effective and make rapid progress in 'catching up' with the encyclopedia of functional elements in the *Xenopus* genome.
- **Second,** we will form an interdisciplinary consortium consisting of computer scientists, engineers, biologists, embryologists, ecologists and clinicians to optimally utilize expertise between disparate fields.
- **Third,** we will develop new bioinformatics tools and analytical approaches to integrate these diverse data types: Development of an xENCODE track hub portal (visualization of all genomic data, raw and low level data such as peak sets) that works well with both with external resources (for example the UCSC browser) and the two main currently funded community resources, Xenbase (curated information on genes and gene expression) and XenMine (data integration, database queries, integration with Galaxy).

As the *Xenopus* community moves past genome sequencing, it is critical to be able to examine genes with a modern view. The human ENCODE project developed a new means to define a gene – the union of genomic sequences encoding a coherent set of potentially overlapping functional products. This includes not simply the ORF, which is a commonly accepted definition, but also the TSS, regulatory elements and the chromatin state. While the *Xenopus* community completes genomic sequencing and is able to accurately define simple gene models, it is imperative that we move toward a modern understanding of genes in the context of development and disease, which includes identification of all relevant promoters, enhancers, lncRNAs, miRNAs and their topological genomic interactions that define these regions and allow for accurate function. Given the experimental repertoire possible in *Xenopus* and its cost effectiveness, *Xenopus* is ideally suited for ENCODE-type functional genome studies and has the potential to provide a level of analysis not yet been achieved in other systems. Further, *Xenopus*, effectively could be used both to develop a clear encyclopedia of gene regulation, even in individual in vivo cells, but more importantly can provide a platform to study the function of these genes.

7C. Funding Mechanism

Because XvivoENCODE is envisioned to operate in a collaborative, community-centric fashion, a U-type funding mechanism would best suit the needs best. The *Xenopus* Genome Consortium is assembling ideas and data to write such collaborative grants.

Continued support for *Xenopus* Training and Meetings

8A. Impact

Training the next generation of scientists is an important component of the NIH's mission and a top priority of the *Xenopus* community. Training in *Xenopus* research is robust with the continued entry of new students, postdocs and junior faculty to the field and formal training through courses and meetings. These help enhance the NIH investment by providing scientists in the *Xenopus* field with access to the latest technology and opportunities to exchange research ideas. Historically, there has been one main course, the Cold Spring Harbor Laboratory (CSHL) course on Cell and Developmental Biology of *Xenopus*, and one international *Xenopus* meeting held every other year. More recently, advanced training workshops have been developed by the NXR, which provide more in-depth teaching focused emerging cutting edge technologies. Continued support for these and similar training initiatives is essential to maintain and propagate the use of *Xenopus* in biomedical research.

8B. How we should proceed

Courses, meetings, workshops and training grants help develop, maintain, and renew human and scientific resources for *Xenopus* researchers and contribute to the training mission of the NIH. We strongly encourage the NIH to provide continues support for training in the following areas:

- **Renewal of the CSHL grant for the course on Cell and Developmental Biology of *Xenopus*.** The CSHL *Xenopus* course has been offered each year since 1992, training over 300 students in basic practices, along with a range of advanced experimental approaches. Nearly 30 students who have taken the course are currently PIs working in the field, and one individual who was a student in the course in 2005 is now sending his own graduate student this year. This course is critical for disseminating evolving techniques throughout the *Xenopus* community in a timely fashion, and is taught by faculty who pioneered or refined the methodologies. The CSHL course is thus a vital resource for the *Xenopus* community in two ways: it offers an opportunity for postdocs and graduate students new to this system to acquire a foundation, while providing a means for more experienced participants to gain some familiarity with advanced approaches. It has also been a testing ground for new techniques. Finally, it provides a means for participants (students and teachers) to establish networks across multiple subdisciplines and extend from the US to Europe, Asia, and Latin America. Continued support for the CSHL *Xenopus* course will help maintain the vibrancy and usefulness of the *Xenopus* system.
- **Support for the biennial International *Xenopus* meeting.** In 2014 the 15th International *Xenopus* meeting will be held in California at the Asilomar Conference Center. These international meetings have been held every other year since 1984, with meetings held in the USA every fourth year. These are very well attended meetings, and unlike other meetings speakers are required to pay for themselves. The NIH has, in the past, supported these meetings, but increased support would help improve the number of students and young faculty attending.
- **Support for financial scholarships for the new advanced training workshops.** NXR has begun offering advanced training workshops. These are small (20-25 students) weeklong intensive workshops created and taught by experts in the field. The goal is to offer 2-3 workshops each year in important areas that impact research in *Xenopus*. To date these have included Bioinformatics and Imaging, with plans for future workshops in genome editing, egg extracts and husbandry. There is high-level enthusiasm for these workshops, but many have requested financial aid, which currently is not available. Supplemental funding to the NXR to allow it to provide tuition assistance will benefit the entire community.
- **Support for T32 training grants and individual fellowships** is an important mechanism by which the NIH supports graduate students and postdoctoral researchers using *Xenopus*. The *Xenopus* community encourages the NIH to continue this and to support institutional T32 training grants that include *Xenopus* training as a component of the program.

Facilitate the Use of *Xenopus* to Model Human Disease

A. Impact

A major opportunity identified by the research community is to accelerate the use of *Xenopus* as a model for human diseases. All of the essential resources in this white paper, **will help achieve the overarching goal of optimizing the utility of *Xenopus* to study human disease and improve human health**. By capitalizing on its established experimental advantages coupled with recent advances in genome editing and high throughput analysis *Xenopus* can provide a cost effective platform to rapidly identify, validate and characterize genes involved in human disease as well as provide mechanistic insight into potential therapeutic design. Beginning with John Gurdon's initial identification of nuclear reprogramming and Tim Hunt's identification of cyclins, over the years *Xenopus* research has made many seminal discoveries that have helped elucidate the cellular mechanisms underlying several human diseases. By revealing fundamental biological processes *Xenopus* research informs our understanding of how gene dysregulation can lead to disease and has provided critical insight into how these pathways might be manipulated for regeneration and repair.

Xenopus has a number of features that make it ideally suited to model human disease. The recent completion of sequencing of both *Xenopus* genomes has revealed a high level of gene colinearity with humans. Combined with the fact that organogenesis in *Xenopus* is similar to humans both genetically and anatomically makes this model system particularly attractive for investigating the molecular mechanics of human disease. Indeed for some organ systems such as lungs or limbs comparable studies in fish are not possible. The rapid functional genomics possible in *Xenopus*, with both gain-of-function and loss-of-function allows the analysis of hundreds of candidate disease genes and characterization of their interacting pathways in just a few months at a fraction of a cost compared to mice. Numerous studies have already demonstrated the utility of *Xenopus* in studying various human diseases, and as an excellent model to functionally test the role of disease-related mutations. For example, *Xenopus* has proved useful in investigating congenital heart disease, heterotaxy, diabetes, ciliogenic-related disorders, kidney disease as well as brain disorders.

With the proliferation of Genome-wide association studies (GWAS), an increased understanding of gene-environment interactions and the increasing use of patient specific genome sequencing for personalized medicine the amount of gene variant information related to human disease is increasing exponentially. The major challenge will be deciphering, validating and functionally characterizing specific gene variants/mutations and the impact they have on disease. This requires additional experimentation that in many cases cannot be accomplished in tissue culture. For example, many of the variants identified in GWAS studies do not cause protein truncations or even changes in amino acid, but rather create nucleotide changes in introns or intergenic regulatory regions that may only be active in the *in vivo* setting. With genome editing technologies and transgenics *Xenopus* is an excellent model system to interrogate candidate gene variants and to characterize pathways of human disease.

B. How can Xenopus be used for modeling and studying human disease?

- Generate tools to induce and model human diseases in developing *Xenopus* embryos. Enhanced development of mutant alleles in *Xenopus* using TALENs and CRISPR/Cas should be encouraged.
- In addition to providing an effective system for *in vivo* screening of candidate GWAS genetic variants in human disease, unbiased studies of organ formation and function can reveal *Xenopus* phenotypes similar to human diseases and thus implicate novel disease candidates.
- Functional epistasis experiments in *Xenopus* can elucidate how different genes implicated in the same disease can interact in regulatory pathway, revealing alternative drug targets.
- Create new transgenic lines to mark specific lineages associated with human disease that can be used by *Xenopus* researchers.

- Identify methods to rapidly introduce disease-specific alleles to study their altered function in the developing embryo. For example, human mutations in the neonatal diabetes candidate gene, Rfx6, were shown to produce nonfunctional proteins in developing *Xenopus* embryos.
- Many adult human diseases arise as a result of developmental defects. *Xenopus* is an excellent model to study the origins of such adult diseases.
- There is an increasing awareness that gene-environment interactions impact disease susceptibility at both the genetic and epigenetic level. *Xenopus* has long been used for teratogenic assays to test the exposure of tadpoles to toxins on the development of specific cell types.
- *Xenopus* is an excellent system for small molecule screens to assess the effects of newly described drugs. Screens can include in vivo analysis of organ development and function as well as biochemical screens in cell free extract focused on specific drug targets.
- *Xenopus* is very useful to study regeneration and methods to improve regeneration in mammals. The tail, limb and eye regenerate in *Xenopus* and studies investigating why regeneration occurs in early stages of *Xenopus* development and not later stages will be useful in identifying the key components required to promote regeneration.
- Robust cell biology, biochemistry and cell-free extracts as well as the compatibility with small molecule screens, enables structure-function analyses of potential therapeutic-target interactions.
- How do disease-causing mutations influence proteomics? Transgenic *Xenopus* can be created that carry specific mutations to study how these mutations affect protein function.

C. How we should proceed

Funding of the essential resources listed in this white paper above is the most immediate way to achieve the overarching goals of facilitating the use of *Xenopus* for human disease. In addition the community recognized that focused initiatives with specific funding opportunities (FOAs) to encourage the use *Xenopus* as a tool to study human disease mechanisms would have a lot of impact. Indeed many NIH institutes have issued PARs calling for novel animal models for human disease modeling. We encourage the NIH to issue PARs specifically for the development of non-mammalian model systems in disease modeling. We identified the following opportunities, which could be leveraged by new FOAs.

- Establish a cooperative network in the *Xenopus* community to identify key target genes for mutagenesis. Such a network could develop pipelines to systemically screen human gene variants in *Xenopus*. This can be achieved in collaboration with the NXR.
- Develop a chemical screening center that can be used to test various compounds and their effects on the development of specific organ systems and for understanding the molecular basis of how various chemicals function.
- Enhance bioinformatics resources to correlate phenotypes from *Xenopus* gene function studies with human diseases and mouse models.
- Incorporate the *Xenopus* model as a component of larger multi-system consortia studying human disease. For example, screening in *Xenopus* should be combined with the current approaches of patient phenotyping, human genetics, and screening candidate variants in silico and in tissue culture.

Appendix 1 – Authors of the 2014 *Xenopus* Community White Paper

Marko Horb
Aaron Zorn

Contributors
Julie Baker
Daniel Buchholz
Sally Moody
Dan Rokhsar
Sergi Sokol
Gert Veenstra
Mustafa Khokha