Proposal for the Generation of Novel BAC Libraries from Commonly Used Inbred Strains of Mice

Submitted by David R. Beier, M.D., Ph.D.
Genetics Division
Brigham and Women's Hospital/Harvard Medical School
20 Shattuck St
Boston, MA 02115
beier@rascal.med.harvard.edu
ph: 617 732 7796
fax: 617 264 5292

On behalf of members of the International Mammalian Genome Society and interested investigators.

Importance of the organism: The importance of the mouse for biomedical research cannot be overstated. Genetic investigations of mouse mutations were initiated at the turn of the century, shortly after the rediscovery of Mendel's laws. Early efforts were notable for the development of inbred strains as well as the study of spontaneous murine mutations and strain-specific phenotypes. More recently, the development of technologies to manipulate the mouse germ line by transgenesis or homologous recombination has made the mouse the definitive system for the study of mammalian gene function. With the rapid progress in the characterization of the mouse genome, and the application of efficient methods of mutagenesis such as ENU treatment or gene-trapping, potential for further progress is unparalleled.

Use of the BAC Libraries: An area that holds considerable challenge in mouse genetic investigation is the characterization of genes that contribute to complex traits and strain-specific phenotypes. While excellent analytical strategies and computational tools permit the localization of causal loci to small genetic intervals, validating the effects of candidate genes is problematic. BAC transgenesis represents a means to this end. This approach allows the investigators to introduce many genes efficiently in a context that permits their "normal" expression. This strategy may, in fact, be the only means to retain strain-specific expression with fidelity. This is crucial for analysis of complex traits, as it may be that modest effects in expression patterns account for the phenotype of interest.

It is also appropriate to note that protocols for efficient generation of transgenic mice using BACs are well described [1]. In addition, extremely elegant strategies for modification of BACs by homologous recombination in bacteria have been recently developed [2]. Thus the means have already been established for the routine utilization of BAC transgenesis to test strain-specific gene function and to modify these, e.g. by site-directed mutagenesis, in order to test hypotheses.

Since BAC libraries of the mouse already exist the question arises as to why additional libraries are required. BAC libraries derived from the 129/Sv substrains are useful as sources of isogenic DNA for homologous recombination into ES cells, and libraries from C57BL/6J were generated specifically for the public domain sequencing effort. However, genetic contributions to complex traits are most often uncovered by analysis of crosses between a variety of inbred strains, as are single gene modifier mutations whose identification promise to give unique insight into the function of the respective main effect gene. These strains include a number of "laboratory" *Mus musculus* (e.g., A, DBA/2, BALB/c, C3H) strains as well as "feral" strains derived from *Mus musculus castaneus and Mus spretus*. Thus, libraries derived from the strains commonly used for these studies will be necessary for identifying the causal loci

Finally, it should be noted that the need for generating additional BAC libraries has been previously discussed. In 1998, then NIH Director Dr. Harold Varmus convened a distinguished group of 60 national and international scientists for the purpose of defining and establishing priorities for the production of mouse genomics and genetics resources [3]. Item 1 of their report included the following recommendation: "Two BAC libraries of 10X coverage should be constructed using different restriction enzymes; the strains chosen should be those that have been designed as reference strains. Five to ten additional libraries of lower coverage should also be constructed from a variety of commonly used mouse strains." A follow-up meeting held later that year further supported this concept. Thus the proposal to generate BAC libraries from

commonly used strains has had high-level scientific review and has been consistently endorsed, but has yet to be realized.

Research community: At the recent 15th International Mouse Genome Conference, investigators met to review the recent solicitation for recommendations regarding the generation of novel BAC libraries; this proposal is a consequence of that discussion. The larger community was also queried using the Jackson Lab Mouse Genome Informatics List Server. The strains recommended by this group are discussed below.

In addition, it is important to note that these resources will serve a much larger community that has only recently appreciated the importance of strain-specific effects in the study of mammalian gene function. This is occurring mostly in the context of the analysis of mutant phenotypes generated by homologous recombination. It is rare these days that an investigator does not report background effects modifying the effects of these engineered mutations. Given that many of these were generated as models of human disease, and the modifications of these phenotypes have significant clinical implications, the further investigation of these effects is compelling. Again, strain-specific BAC libraries will be an important means to identify the causal loci.

Finally, it should be appreciated that the investigation of complex traits in murine systems is robust. A search of the Pubmed database using the keywords "complex trait OR QTL AND mice" identified 73 papers published in the year 2001 alone. These reports included investigations into the genetic contribution to airway hyperresponsiveness, alcohol and barbiturate withdrawal, anxiety, atherosclerosis, body weight, bone density, brain structure, cholesterol gallstone formation, circadian behavior, heart regeneration, immune function, infectious disease susceptibility, locomotor activating effects of cocaine and ethanol, milk production, morphine antinociception, non-insulin-dependent diabetes, obesity, plasmacytoma susceptibility, seizure susceptibility, systemic lupus erythematosus, and transgene methylation.

Status of genomic sequencing: The public domain effort for sequencing of the mouse genome is using a C57BL/6J BAC library generated by Dr. Pieter de Jong and is making excellent progress. A BAC map has been completed and 2-3X shotgun sequence coverage has been deposited. In addition, a proprietary annotated genome sequence of the mouse has been developed by Celera using a plasmid shotgun sequencing strategy.

Other genomic resources: A wide variety of genomic resources exist for genetic analysis in the mouse. Dense microsatellite and radiation hybrid maps have been generated, largely as a consequence of efforts at the Whitehead Institute [4,5]. A large amount of EST sequence has been generated (primarily by Washington University and RIKEN) and recently a large collection of full-length cDNAs have been reported and annotated by RIKEN [6]. Preliminary studies to identify SNPs in the mouse have demonstrated that they are common and readily detected (due to the homozygous nature of inbred strains) [7], and it is an aim of the public domain sequencing effort to generate a dense SNP map over the course of the next year (Lindblad-Toh, personal communication). In addition to these tools there are biological resources, such as the mutant mice that are being generated from the NIH-supported mutagenesis centers at Baylor, The Jackson Laboratory, and Oak Ridge; as well as the recently funded effort to develop public domain STS-tagged gene-trapped ES cell libraries.

Also worthy of mention are computational tools that are useful for genetic analysis in the mouse. With respect to this specific proposal, one should recognize the ongoing development of statistical analysis software for QTL mapping such as MapManager (http://mcbio.med.buffalo.edu/mapmgr.html) and MapMaker/QTL (http://www-genome.wi.mit.edu/ftp/distribution/software/). Also of note are the tools that allow comparative analysis of genome sequence, such as the program VISTA (http://www-gsd.lbl.gov/vista/). One version of this facilitates the identification of differences between closely related species; this has obvious application to the ultimate elucidation of strain- and species-specific effects of mouse phenotypes.

Strains: A list of investigators' strains preferences and research projects are contained in Tables 1 and 2. There is considerable interest in a library from the CAST/Ei strain, the most commonly used *Mus castaneus*-derived inbred strain. This is likely due to the fact that its common use in genetic mapping studies (as a consequence of its abundant polymorphism with inbred lines) has led to the appreciation that it has frequent strain-specific effects on the traits being investigated. SPRET/Ei, derived from *Mus spretus*, is of interest for similar reasons. However, it is used less commonly due to limited availability and difficulty in breeding; Dr. Jean-Louis Guénet has proposed that the more robust breeding line SEG/Pas be imported to the U.S. for distribution and a BAC library generated from that. With respect to common laboratory strains, BALB/cByJ is of interest due to its frequent use in immunological studies; A/J and DBA/2J are of interest due to their polymorphism with respect to a variety of clinically-relevant traits and their representation in RI and consomic substitution panels; and other strains are recommended as noted in tables 1 and 2.

Ultimately, it will be important to coordinate the selection of strains (and substrains) for BAC library generation with ongoing efforts in mouse genome analysis. This could perhaps be done best by enlisting the participation of members of the recently established Mouse Sequencing Liaison Group (http://www.ncbi.nlm.nih.gov/genome/guide/M_musculus.html), chaired by Dr. Wayne Frankel, in the review process.

Genome size: $3x10^9$ base pairs.

Availability of a DNA source: The inbred strains proposed are commonly available from sources such as the Jackson Laboratory.

Specifications: The depth of the library should be sufficient such that the entire genome is likely to be represented and of sufficient depth that there will be multiple clones covering a single locus. This will be desirable to insure that most genes will be fully contained within a BAC and not reside only at a breakpoint. It need not be as deep as required for genome sequencing, and 5X coverage will be useful. In addition, it would be desirable for the BACs to be as large as feasible (e.g. 200 Kb) because this would reduce the effort and cost required for generating BAC transgenic lines and minimize the loci that will be disrupted at breakpoints. No unusual vectors are required.

Time frame: The enthusiasm of the research community for these reagents indicates that they will be put to use as soon as they can be generated.

Other support: BAC libraries corresponding to 2X genome coverage have been generated for the A/J and SJL/J mouse strains by Dr. Cory Teuscher (see table 2). I am unaware of any other proposals for the generation of strain-specific mouse BAC libraries.

Literature cited:

1. Heintz, N. Analysis of mammalian central nervous system gene expression and function using bacterial artificial chromosome-mediated transgenesis. *Hum Mol Genet* **9**, 937-943. (2000).

2. Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A. & Copeland, N.G. A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65. (2001).

3. Battey, J., Jordan, E., Cox, D. & Dove, W. An action plan for mouse genomics. *Nat Genet* **21**, 73-75. (1999).

4. Dietrich, W., Katz, H., Lincoln, S.E., Shin, H.S., Friedman, J., Dracopoli, N.C. & Lander, E.S. A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**, 423-447. (1992).

5. Hudson, T.J. *et al.* A radiation hybrid map of mouse genes. *Nat Genet* **29**, 201-205. (2001).

6. Kawai, J. *et al.* Functional annotation of a full-length mouse cDNA collection. *Nature* **409**, 685-690. (2001).

7. Lindblad-Toh, K. *et al.* Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* **24**, 381-386. (2000).

Table 1: Investigators interested in BAC libraries: strain preferences.

			CAST/Ei	SPRET/Ei	Balb/C	A/J	DBA/2	C3H	SJL	PWD	NOD	FVB
Ackerman, Sue	sla@jax.org	The Jackon Laboratory, Bar Harbor, ME	v									
Avner, Philip	pavner@pasteur.fr	Unité Génétique Moléculaire Murine Institut Pasteur Paris France						v				
Balling, Rudi	balling@gbf.de			v								
Beier, David	beier@rascal.med.harvard.edu	Genetics Division, Brigham and Women's Hospital/Harvard Medical School, Boston, MA				v	v					
Bilger, Andrea	bilger@oncology.wisc.edu							v				
Bix , Mark		University of Washington	v		v							
Carlson, George	gac@mri8.mri.montana.edu		v						v			
Cascalho, Marilia	cascalho.marilia@mayo.edu	Transplantation Biology Medical Sciences 2-113 Department of Surgery and Immunology Mayo Medical School Rochester, MN			v							
Denny , Paul	paul@har.mrc.ac.uk	Genome Group Leader MRC UK Mouse Genome Centre & Mammalian Genetics Unit Harwell, Oxfordshire, UK.			v							
D'Eustachio , Peter	deustp01@mcrcr0.med.nyu.edu											
Disteche, C.	cdistech@u.washington.edu											
Everett, Eric T.	eeverett@iupui.edu	Departments of Oral Facial Development and Dermatology Indiana University Schools of				v						

		Dentistry and Medicine Indianapolis, IN										
Fisher, Elizabeth M C	e.fisher@ic.ac.uk		v									
Foley, Kevin	kfoley@mpi.com	Group Leader Mouse Models Department Millennium Pharmaceuticals, Inc. Cambridge, MA			v							
Forejt , Jiri	forejt@biomed.cas.cz									v		
Gibson, John	j.gibson@cgiar.org					v						
Guenet, Jean-Louis	guenet@pasteur.fr	Institut Pasteur 25, PARIS,	v	v*								
Herault , Yann	herault@cnrs-orleans.fr				v			v				
Iraqi, Fuad	F.IRAQI@CGIAR.ORG	Genetics of Disease Resistance (GDR) International Livestock Research Institute (ILRI) Nairobi, Kenya										
Lammert, Frank	flammert@post.klinikum.rwth-aachen.de	Department of Medicine III Klinikum RWTH Aachen Germany	v									
Legge, Mile	Mike.legge@stanebow.atago.ac.nz						v				v	
Letts, Verity	val@jax.org	The Jackon Laboratory, Bar Harbor, ME	v									
Libert, Claude	CLAUDE.LIBERT@lmb001.rug.ac.be		v									
Malo, Danielle	danielle.malo@mcgill.ca											
Mao, Jian-Hua	jmao@cc.ucsf.edu	UCSF Cancer Center		v								
Mcintire , Jennifer Jones	jcjones@leland.stanford.edu	Umetsu and DeKruyff Labs Stanford University			v							
Mehrabian , Margarete	mmehrabi@ucla.edu		v				v					
Meisler ,	meislerm@umi	Department of Human	v									

Miriam	ch.edu	Genetics University of Michigan Ann Arbor, MI										
Mock, Beverly	bev@helix.nih.gov	Investigator, LG, NCI, NIH Associate Director, Scientific Policy CCR, NCI			v		v					
Moen, CJA	cmoen@lumc.nl	Dept. of Human Genetics, Leiden University Medical Center Leiden The Netherlands										v
Nadeau , Joseph	jhn4@po.cwru.ed											
Noben-Trauth , Konrad		Section on Neurogenetics NIDCD, National Institutes of Health Rockville, MD	v									
Noyes, Harry	harry@liv.ac.uk				v	v						
Riblet , Roy	rriblet@tpims.org	Torrey Pines Institute for Molecular Studies San Diego, CA			v	v	v					
Schalkwyk , Leo	L.Schalkwyk@iop.kcl.ac.uk						v					
Schwartzberg , Pam	pams@nhgri.nih.gov	National Human Genome Research Institute National Institutes of Health Bethesda, MD			v							
Southard-Smith , Michelle	michelle.southard-smith@mcmail.vanderbilt.edu	Vanderbilt University Medical Center Division of Genetic Medicine Nashville, TN						v				
Spearow , Jimmy	jlspearow@ucdavis.edu	Section of Neurobiology, Physiology and Behavior RoomUniversity of California at Davis Davis, CA				v						
Staelens , Jan	jan@dmb.rug.ac.be	Mouse Molecular Genetics Unit Gent Belgium	v									
Teuscher , Cory	cteusche@uiuc.edu	University of Vermont School of Medicine				v			v			

Van Molle , Wim	wimvm@dmb.rug.ac.be	Department of Molecular Biology Ghent University-VIB K.L. Gent Belgium		v								
Vidal, Silvia	svidal@uottawa.ca		v	v		v	v					
Wielockx, Ben	Ben.Wielockx@lmb001.rug.ac.be	Mouse Molecular Genetics Unit Gent Belgium		v								
			13	6	10	8	7	4	2	1	1	1

Table 2: Investigators interested in BAC libraries: research projects.

Avner, Philip	We are really interested in C3H/He for transgenesis work on our diabetes congenics.
Beier, David	Analysis of modifying loci affecting progression of PKD and affecting airway hypperesponsiveness
Bilger, Andrea	We are interested in a C3H BAC library. It would allow us to manipulate (transgenics, k.o.'s) a C3H liver cancer susceptibility gene.
Bix , Mark	We are currently trying to fine map an interleukin-4 regulatory QTL on mouse chromosome 16 (Bix et al. Journal of Experimental Medicine (1998) 188:1-11) that segregated in a (BALB/c x C57BL/6)F1 x B6 backcross. We are considering an in vivo functional complementation approach using BAC transgenic mice to identify the gene once we have narrowed the QTL interval to below a megabase. Toward this end a BALB/c BAC library would be of immense utility. A CAST BAC library would also be of utility to us for the same reason.
Carlson, George	Our most urgent need is a BAC library from CAST/Ei for use in identifying the genes underlying prion incubation time QTLs (Stephenson et al., Genomics 2000, 69:47-53.
Denny , Paul	BAC transgenic complementation of ENU-induced mutations
Disteche, C.	My laboratory is currently funded by the NIH to pursue the study of *Clc4* in inbred strains of mice and in *Mus spretus.* Large-scale genomic sequencing of the *Clc4* region in both species will be done to compare sequences around *Clc4* and to define the breakpoints of the evolutionary rearrangement. *In vitro* activity of the *Clc4* promoters will then be compared between species to identify elements, which may explain differences in gene expression. In addition, our plans are to insert *Clc4* into an autosome and into the X chromosome in transgenic mice to recapitulate the evolutionary rearrangement and to determine whether active X upregulation or position effects may explain the changes in *Clc4* expression. Thus, this study has the potential to reveal specific elements that modify gene expression within the defined environment of the X chromosome or the autosome.
Everett, Eric T.	The A/J strain has great potential in helping us to understand oral clefts
Fisher, Elizabeth M C	CAST/Ei has been used in a number of QTL studies, including ours (Lloyd et al, PNAS 98: 6279 (2001)), and a BAC library would potentially aid our QTL finding and validation.
Foley, Kevin	BALB/cJ BAC library for isogenic KOs in BALB/c ES cells for inflammation models.
Forejt , Jiri	PWD. The reason is that we are at Bc4 / Bc7 (N5 to N8) of preparing all 21 B6-PWD consomic strains. We are trying also to prepare ES cell lines, but this is obviously a high risk project. PWD is pure, highly inbred (F64-F68) Mus m. musculus strain with high frequencies of DNA polymorphisms compared to laboratory strains, and lot of phenotypic differences, including behaviour. Still, it is easy to breed these mice inter se and with laboratory strains. The consomics will become available to mouse community as soon as they will be ready.
Iraqi, Fuad	A/J and BALB/c mouse strains. These mouse strains are being used in positional cloning of QTLs resistance to infectious disease, such as sleeping sickness (trypanosomosis), Malaria.
Libert, Claude	I am extremely interested in a SPRET/Ei BAC library, more than a SEG library. We need such a library for the positional cloning of TNF protection genen from SPRET. We also want to use BAC clones to generate transgenic mice.
Mao, Jian-Hua	the SPRET/EI strain, because it would be helpful to map a modifier gene of cancer.
Mcintire , Jennifer Jones	We are interested primarily in a BALB/c BAC library, as the BALB/c mouse is the primary strain used in our studies of helper T cell differentiation, allergen-induced airway reactivity (asthma), and other immune responses
Mehrabian , Margarete	CAST is very diverse and phenotypically quite different from the other strains in a number of complex metabolic disorders. We have generated several congenic lines related to atherosclerosis and now the BACs would be extremely helpful.
Meisler , Miriam	Our lab is mapping a modifier-of-myo5A locus using a sensitive allele in CAST/Ei that causes lethality of an otherwise viable neurological mutant, flailer (Jones et al (2000) Human Molecular Genetics 9: 821-828.) Genetic mapping is in progress through CIDR, and we would use a CAST library of BAC clones to identify the molecular lesion in the CAST allele.
Mock, Beverly	I am in favor of DBA/2 and BALB/cByJ BAC libraries. I would isolate clones from these libraries for the Pctr1,2,3,4 loci to use in making BAC transgenics and as the starting material for modifying the sequences for knock-in constructs. These loci control the susceptibility/resistance of mice to the development of B cell tumors, specifically plasmacytomas which share several pathogenetic features in common with human

	Burkitt's lymphoma, multiple myeloma and certain Non-Hodgkins' lymphomas.
Moen, CJA	We use FVB (in combination with C57BL/6) in identifying modifiers involved in cardiovascular diseases. Regarding our research a FVB BAC library would be very helpfull.
Nadeau , Joseph	The B6.A-Chr consomic panel is now complete, (N10+++F++), the strains are now being validated, and they are eing prepared for shipment to Jackson Laboratory for preservation and ditribution. One of the progenitors is A/J and many uses of an A/J BAC library could be imagined.
Noben-Trauth , Konrad	For a number of diseases the CAST/Ei strain is more resistant than the B6 strains, i.e. hearing loss. For rescue-experiments, it would be a fine resource to have a CAST-derived BAC library. thank you
Riblet , Roy	DBA, BALB, and A/J for determination of Igh structure (with B6, extensive captured polymorphism and decades of antibody response data) and analysis of the quantitative genetics of hematopoietic stem cell frequencies and behavior.
Southard-Smith , Michelle	We are characterizing modifiers between C57BL/6 and C3H so having a BAC library available would help us out tremendously and allow us to perform some transgenic experiments that currently are limited by the lack of this reagent.
Spearow , Jimmy	We are examining strain differences in ovarian response to gonadotropins between A/J and B6. We have mapped and confirmed several Reproductive QTL controlling Ovulation Rate hormone Induced (ORI) and Aromatase Activity Induced (AAI) and are working toward positonally cloning these reproductive QTL.
Staelens , Jan	I would like to show my interest in a BAC library derived from SPRET/Ei. Being a strain derived from Mus spretus, it has been used a lot for mapping of genes and interesting phenotypes. In our lab we have identified ourselves a lot of interesting phenotypes in SPRET/Ei, some of which we do not find back in other Mus spretus-derived inbred lines.
Teuscher , Cory	We have constructed A/J and SJL/J BAC libraries. They are currently only 2X and I have contacted Dr. Peterson at NIH about providing us with additional money to expand them or have them taken over by one of the centers for this purpose. We have done pooling at this point for quality control and they seem to be fine. We can readily isolate clones by PCR for either expressed genes (Hrh1) or SSLPs as long as they are mid to large in fragment size.
Vidal, Silvia	A/J > DBA/2 > CAST/Ei > SPRET/Ei. We are interested in the genetic basis of susceptibility to infection. We have experimental models involving cytomegalovirus, coxsackievirus (these two pathogens are ubiquitous in humans, severe to fatal effects in immunocompromised patients, potential autoimmune responses in normal people) and cutaneous leishmaniasis, involving the above mentioned strains. In addition, Emil Skamene at the McGill Center for Host Resistance has recently produced a novel AcB/BcA panel that is being extensively phenotyped for susceptibility to infection with many micro-organisms as well as many other immunological, behavioral, etc. traits.