PGA Highlights

A Newsletter for the NHLBI's Programs for Genomics Applications Volume 1, Number 1

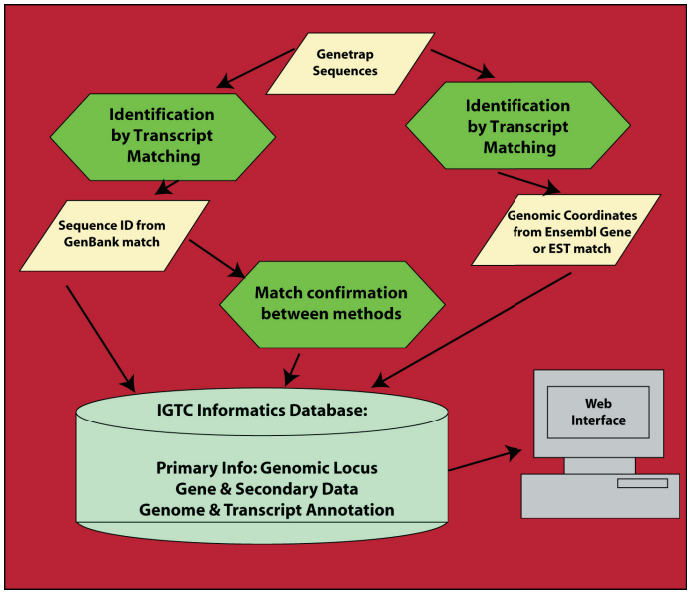
The PGA is a major NHLBI initiative to advance functional genomic research related to heart, lung, blood, and sleep health and disorders.

The goals of the PGAs include developing information, tools, and resources to link genes to biological function on a genomic scale. All the information, reagents, and tools developed in the PGAs are made freely available in a timely manner to the research community.

In addition, the PGAs provide workshops, courses, and visiting scientist programs to facilitate the training of researchers in the use of the data and related technologies developed by the PGAs.

BayGenomics: leadership in the international gene trapping community

BayGenomics, an NHLBI-funded Program for Genomic Applications, has generated more than 10,000 mutant embryonic stem (ES) cell lines, for the purpose of generating knockout mice. More than 2,500 of the ES cell lines have been distributed to investigators throughout the world. Other innovative and successful gene-trapping programs exist in England, Canada, and Germany, but BayGenomics has led the world in the numbers of mouse genes that have been inactivated. Also, by a large margin, BayGenomics has led the world in terms of the numbers of ES cell lines that have been distributed to the scientific community.



Within the past year, BayGenomics has taken a leadership role in organizing the international gene-trapping efforts, for the purpose of improving user access to the existing gene-trapping programs, and improving the annotation of the trapped genes. BayGenomics hosted an Informatics Workshop for the International Gene Trap Consortium (IGTC) in San Francisco April 14–15, 2005. Held on UCSF's new Mission Bay Campus, the workshop attracted representatives from the major national and international gene-trapping programs as well as bioinformatics experts from Ensembl, the UCSC browser group, NCBI, and the Mouse Genome Informatics (MGI) group at the Jackson Laboratory. The Workshop was initiated to spur communication and collaboration within the "informatics side" of gene trapping and to generate a cohesive plan for future IGTC informatics efforts. The specific goals of the meeting were: 1) to develop a plan for a common IGTC cell line data annotation pipeline; 2) to approve a new IGTC public resource to enable researchers to identify and obtain cell lines of interest; and 3) to initiate a mechanism for carrying this work forward. The first day was focused on the presentation and discussion of research and projects of interest to the community, including talks from each gene-trapping program. The second day was devoted to generating a plan for a new IGTC web site and informatics presence.

A prototype web site, developed in collaboration between investigators at the Sanger Institute and BayGenomics, was presented. Very importantly, the web site included a new annotation pipeline that uses both genome localization and transcript identification algorithms to provide the highest possible quality identification of trapped genes (see figure, which illustrates the dual pathways for the annotation strategy).

National Heart, Lung, and Blood Institute
PROGRAMS FOR GENOMIC APPLICATIONS

Summer 2005

PGA Symposium: From Genome to Disease II NIH Natcher Conference Ctr. Bethesda, MD

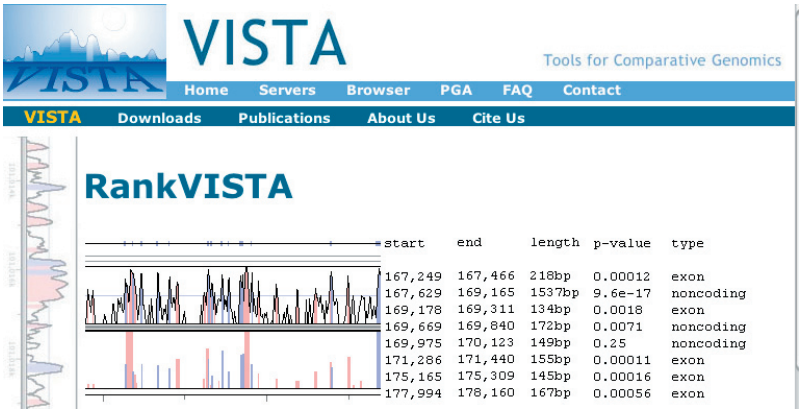
July 19-20 2005

This state-of-the-art symposium will provide an opportunity to explore the application of genomic technologies to the study of human disease. To maximize the usefulness of the symposium for both the novice and the experienced investigator, the program will couple plenary lectures by leaders in the area of genomics and in-depth, small-group tutorials on the new technologies and methodologies discussed in this symposium.

http://www.nhlbi.nih.gov/meetings/pga/

Berkeley PGA's new sequence comparison tool allows highly sensitive detection of evolutionarily conserved elements in closely related species

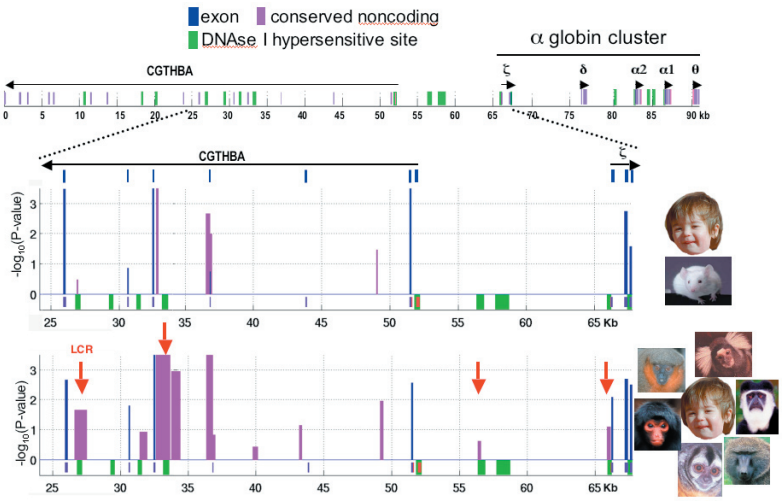
The Berkeley PGA has recently announced the release of a new comparative genomics tool – RankVISTA. This tool differs from all previously developed comparative genomics tools in that it works effectively in detecting conserved sequences among organisms that are closely related as well as those that are distantly related. Another unique feature is the display of a statistical measure for each conserved sequence to allow users to prioritize elements for functional study based on the degree of conservation, hence its name RankVISTA. The pre-computed human and mouse whole genome RankVISTA plot is now available (http://pipeline.lbl.gov). Users can also submit sequences to the mVISTA server (http://genome.lbl.gov/vista/servers.shtml) for a custom RankVISTA query.



RankVISTA homepage (http:// genome.lbl.gov/vista/rankVISTA.shtml)

RankVISTA graphs (see examples in the accompanying figure) are based on the Gumby algorithm (Prakhakar et al., in preparation), which estimates neutral evolutionary rates from non-exonic regions in the multiple sequence alignment, and then identifies local segments in the alignment that evolve more slowly than the background. The phylogenetically weighted log-odds conservation scores of conserved segments are translated into P-values using Karlin-Altschul statistics. A scale of 4 in the RankVISTA plot indicates that the probability of seeing that level of conservation by chance in a neutrally-evolving 10-kb segment of the base sequence is less than 10-4. Gumby has no window-size parameter, and no fixed percent-identity threshold either. Since the algorithm uses a more-conserved-than-background paradigm, it is particularly sensitive in performing phylogenetic shadowing analysis for closely related species such as the primate comparison [1].

Although the manuscript describing RankVISTA is still in preparation, the tool has been used in several high profile publications [2-4]. To learn more about the utilities of RankVISTA and the related bioinformatics tools, register for the hands-on workshops offered by the Berkeley PGA (http://pga.lbl.gov/Workshop/index.shtml).



Comparison of two RankVISTA plots generated from the 5' flanking region of the alpha globin gene cluster using the human and mouse sequences (upper panel) and seven primate sequences (lower panel). Some noncoding elements conserved in primates (marked as purple vertical bars) are not visible in the human and mouse sequence alignment. Vertical arrows mark the previously confirmed cis-regulatory elements

References

1. Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM. (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science 299:1391-4.

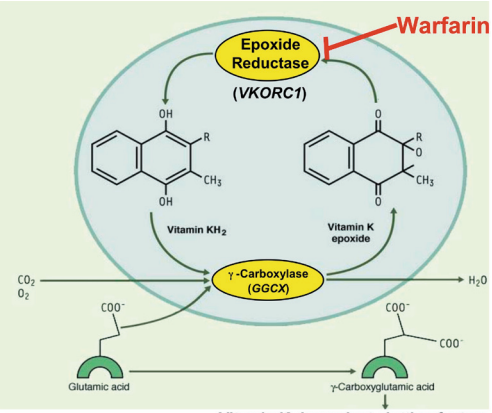
2. Schmutz J, et al., (2004) The DNA sequence and comparative analysis of human chromosome 5. Nature 431:268-74.

3. Martin J, et al. (2004) The sequence and analysis of duplication-rich human chromosome 16. Nature. 432:988-94.

4. Hughes JR, Cheng J-F, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin EM, Higgs DR. (2005) Annotation of cis-regulatory elements by identification, sub-classification and functional assessment of multi-species conserved sequences. PNAS in press.

SeattleSNPs has launched translational studies into medical resequencing

One primary goal of SeattleSNPs is to work with NHLBI investigators in translating polymorphism data into association studies with clinical applications. As reported recently in the New England Journal of Medicine (June 2nd, 2005), work carried out by the SeattleSNPs PGA led to findings that show a major association between single nucleotide polymorphisms (SNPs) in the VKORC1 (vitamin K epoxide reductase complex 1) gene, and dosing with coumarin-based anticoagulant drugs. Warfarin (a derivative of coumarin) is the most commonly employed oral anticoagulant used to treat thromboembolic disease associated with stroke and cardiovascular disease, and acts on VKORC1 to inhibit vitamin K activation of key clotting factors. Each year more than 20 million prescriptions are written for warfarin in the US alone. However, management of warfarin therapy is challenging because of the wide variation among patients in drug response which must be carefully monitored. Using the SeattleSNPs discovery samples of European and African-descent, VKORC1 was comprehensively resequenced to identify common polymorphisms in the general population. Even though the resequencing process did not identify any SNPs that alter the protein coding of the gene, it was possible to characterize the correlation between SNP sites and haplotype structure in the VKORC1 gene. These studies proved that selection of highly informative tagSNPs from this candidate gene were key to finding associations in a larger, warfarin treated clinical population. In these patients, and in a larger replication population at Washington University, we discovered a significant warfarin dose association with haplotypes in VKORC1 which classified patients into groups that would require low, intermediate or high doses of warfarin treatment. Genetic polymorphisms were found to explain approximately 25% of the variance in warfarin dose among all the patients. Futhermore, because the common SNPs were located in non-coding regions, we explored biological mechanisms related to gene expression. Our analysis revealed association between mRNA level for VKORC1, and the underlying SNPs and haplotype structure. Lastly, this study demonstrated that differences in the frequency of haplotypes in both Asian and African populations may account for inter-ethnic variability in warfarin dosing.



Schematic diagram of target pathway for warfarin adapted from Reider et al., Effect of VKORC1 Haplotypes on Transcriptional Regulation and Warfarin Dose. NEJM, 355:22, June 2005.

Seattle SNPs is translating resources into human genetic association studies

Elevated plasma levels of C-reactive protein (CRP) have emerged as an important predictor of future cardiovascular disease and metabolic abnormalities. We explored whether common genetic variants at the human CRP gene locus influence plasma CRP levels. We started with genomic resequencing to identify all common patterns of nucleotide diversity and to define common haplotypes in CRP. We then genotyped a set of polymorphisms representative of the common patterns of these SNPs and associated haplotypes in a much larger population-based study of young adults, i.e., CARDIA (Carlson et al 2005, American Journal of Human Genetics 77: 64-77—an open access article). This was accomplished through a new, SeattleSNPs-stimulated RO1 spearheaded by Drs. Alex Reiner and David Siscovick at the University of Washington. Our study revealed genetic associations between specific CRP haplotypes and plasma levels of this protein. Through a collaboration with the Stanford Genome Center, we experimentally determined that polymorphisms in the promoter region of CRP are responsible for genetic differences in plasma CRP levels in human populations.

SeattleSNPs goes on the road, offers its first traveling tutorial

On April 19 and 20, SeattleSNPs offered a two-day tutorial on Variation Discovery and Analysis at Washington University in St. Louis. The workshop attracted 50 scientists, the majority of whom were faculty and senior scientists with clinical interests in heart, lung and blood diseases. The course highlighted use of SeattleSNPs data and its application in association studies and integration with the emerging NHGRI HapMap project. The workshop provided overviews and many hands-on tutorials on the use of human genome databases, and emphasized the latest approaches for identifying and geno-

SeattleSNPs: On the Road

As part of the PGA's commitment to delivering educational resources to NHLBI investigators, SeattleSNPs is soliciting applications from institutions willing to host a traveling Variation Workshop.

The workshops will be modeled after the one recently held in St. Louis (see story, page 3), and are offered in both one- and two-day formats.

The application information is available online at:

http://pga.gs.washington.edu/tutorial_appliation.html

New Educational Opportunities in the PGA

Courses

Workshops

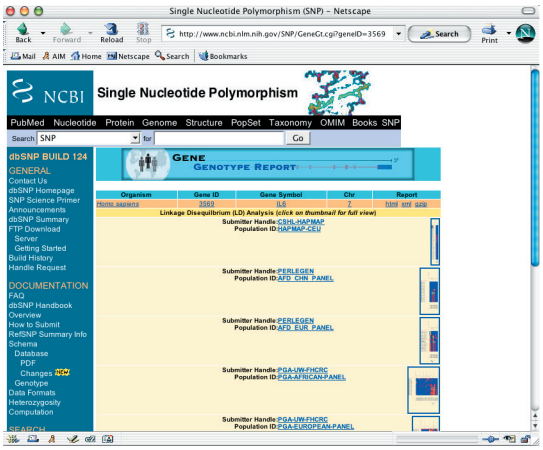
Visting Scientist Programs

http://www.nhlbi.nih.gov/resources/pga/workshop.htm

typing SNPs and evaluating associations and linkage disequilibrium in the human genome. In addition, the course featured two interactive detailed case studies, related to genetic associations with warfarin dosing, and genetic associations with CRP (see features above).

SeattleSNPs is contributing to new genomic resources

Using resources from the SeattleSNPs PGA, the first whole genome linkage disequilibrium map of the human genome has been validated and benchmarked (Hinds et al. Science 307: 1072-1079, 2005). In addition, new tools from SeattleSNPs (e.g., LD-select) played a major role in validating this new map at a population level. And the National Center for Biotechnology Information (NCBI) is applying the SeattleSNPs visual genotype display program VG2 to visually display all of the available genotypes in its public dbSNPs database. An example of the visual genotype data from interleukin 6 (http://www.ncbi.nlm.nih.gov/SNP/GeneGt.cgi?geneID=3569) is shown in the figure below.



PhysGen generates first rat ENU Knockouts for heart, lung, blood and sleep

Physiological genomics studies for heart, lung, blood, and sleep

PhysGen, the Program for Genomic Applications at the Medical College of Wisconsin (http://pga.mcw.edu), has focused on understanding the genetic basis of fundamental mechanistic pathways of the heart, lung, kidney, blood and vasculature through development and characterization of consomic rat panels, using environmental stressors. Comprehensive characterization of these strains allows for immediate mapping of traits to a particular chromosome without the need for genetic crosses. To date we have characterized blood chemistries as well as traits involved in renal, cardiac, vascular, lung, and respiratory function in 47 parental and consomic rat strains, resulting in the availability of over 400,000 mean trait values.

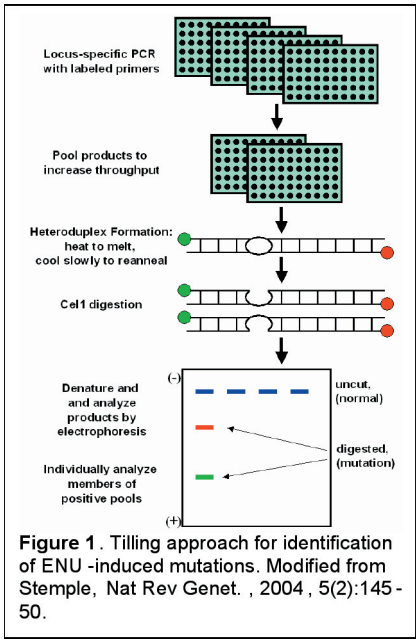


Figure 1. Tilling approach for identification of ENU -induced mutations. Modified from Stemple, Nat Rev Genet. , 2004 , 5(2):145 - 50.

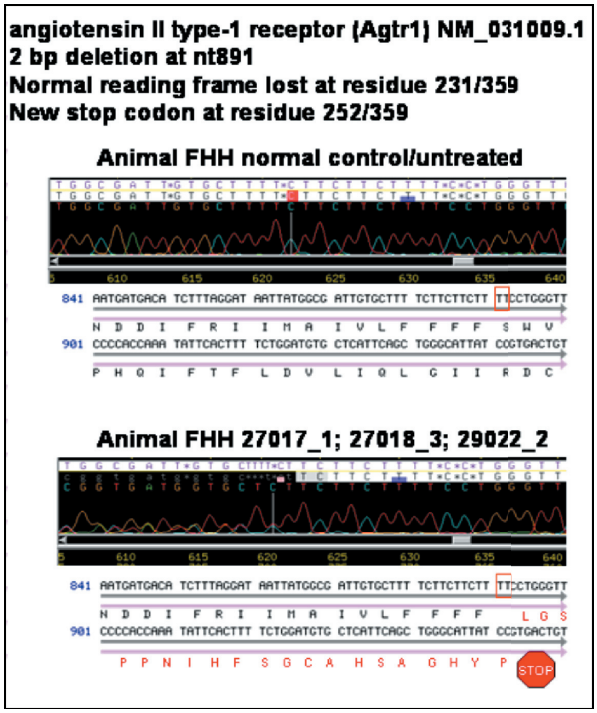
Identifying causal genes for complex traits

Although mapping of quantitative trait loci (QTL) involved in NHLBI disorders has become easier with the availability of the PhysGen resources, identifying the genes that underlie the QTL that are associated with common phenotypes has proven to be a challenge. Too many genetic variants in humans, too few alleles in inbred animal models and the potential involvement of many genes in a single QTL have hindered our ability to unequivocally prove the causal role of a gene in a QTL. One way that is traditionally used to show gene causality of a phenotype is to alter the gene's function, e.g. by gene knockout, knock-in, transgenesis, etc. With the complete draft genomic sequence, the only remaining major weakness of the rat is inability to generate gene knockouts. Recently, methods have been developed to knock out genes by linking random chemical mutagenesis (ENU) with high-throughput screening for functional mutations in targeted genes, and subsequent breeding of selected mutants to homozygosity.

Generating rat knockouts by TILLING

We have adopted one of these methods (TILLING) to generate rat knockouts in PhysGen. By the TILLING approach, male rats are subject to ENU (N-ethyl-N-nitrosourea), a chemical that causes DNA mutations in developing sperm. The males are subsequently bred to healthy females to generate F1 offspring heterozygous for an average

of 36 random functional mutations per pup. Because large numbers of offspring are required to find mutation in the genes of interest, the F1 offspring are screened prior to phenotypic screening by a high-throughput method called TILLING (see figure). Fluorescently labeled gene-specific PCR products, from genomic DNA of the F1 animals (heterozygous mutants), are generated and subject to heteroduplex formation, creating a mismatch at the site of mutation. An enzyme derived from celery (Cel1) recognizes and cleaves the heteroduplex generated at the mutation site, resulting in a cleaved product detected by simple electrophoresis (a sequencing gel). TILLING has already been proven successful in Arabidopsis, C. elegans, and zebrafish, in addition to the rat. The advantage of the TILLING approach over traditional ENU mutagenesis screens is that a targeted gene mutation can be identified in offspring before weaning, eliminating the need to grow the animals and evaluate with a phenotypic screen.



PhysGen rat knockouts

The genes selected for KO have been derived from those that have been linked to heart, lung and blood disorders using molecular genetic, genomic and expression technologies during the first four years of all PGA programs. As genomic background is shown to play a major role in the expression of phenotypes in KO mice, this PGA will attempt to make the KOs in 3 genetic backgrounds in the rat, the SS, BN, and FHH strains. The selection of these 3 strains will also enable PhysGen to leverage the enormous amount of phenotypic data generated on these strains during the first four years, notably 8,610 physiological data points per strain, including baseline and stressors, providing an unprecedented level of baseline data for a mutagenesis screen.

Thus far, we have successfully identified 12 mutations for targeted genes in F1 offspring of the Fawn Hooded Hypertensive (FHH) rat strain. Figure 2 shows a mutation detected in the angiotensin II type 1 receptor (Agtr1) gene in an F1 pup using the TILLING approach. In pup #27017_1, two fragments resulted from the Cel1 heteroduplex cleavage, one of approximately 400bp and the other of approximately 600bp. Sequencing of this pup identified a 2 bp deletion in the gene at nucleotide 891, creating a new translation reading frame at predicted residue 231/359, which prematurely terminates at residue 252/359 after the addition of 20 "new" amino acids. An additional knockout has been identified and confirmed by sequencing in the transcription factor nuclear receptor 4A1 (NR4A1) gene. This knockout arose through a C_A substitution at nt500, changing the encoded residue from a tyrosine (TAC) to a stop codon (TAA). Adopting the ENU knockout technology ushers in a new era of rat functional genomic investigations, allowing us to validate genes identified in the rat and determine downstream mechanisms leading to heart, lung, blood and sleep disorders.

References:
Hrabe de Angelis M, B.R. Large scale ENU screens in the mouse: genetics meets genomics. Mutation Research 400, 25-32 (1998).
Justice, M.J., Noveroske, J.K., Weber, J.S., Zheng, B. & Bradley, A. Mouse ENU mutagenesis. Hum Mol Genet 8, 1955-63 (1999).
Nolan, P.M. et al. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nat Genet 25, 440-3 (2000).

New Genomic Resources from the PGA

Animal Models & Phenotypes

Clinical & Physiologial Studies

Databases & Software Tools

Expression Profiles

Mutagenesis

Proteomics

SNPs and Genotypes

Comparative Sequence Analysis

http://pga.lbl.gov/PGA/PGA_inventory.html

JAX PGA at the cutting edge: New mouse models for heart, lung, blood and sleep disorders

With close to 12,000 mice tested and over one million data points collected to date, the JAX PGA has used high-throughput, non-invasive phenotypic screens developed specifically for this program to produce some important new models of human disease. We draw on two significant resources to develop new models using a phenotype-driven approach. First, we produce and test 4000 third-generation chemically-induced mutant mice per year to identify novel functional recessive mutations. Secondly, we have conducted comprehensive phenotyping of 42 commonly used inbred mouse strains to further develop these as viable disease models. Our efforts in mutagenesis have produced 65 heritable mutants thus far for disorders in blood, lipid, glucose, blood pressure, obesity, sleep, lung function, bone density, and cardiac function phenotypes. We are currently testing trait heritability in an additional 100 phenotypic deviants. Our web site (http://pga.jax.org) features 35 established mutant lines that are currently available for distribution as breeding pairs. Chromosomal mapping has been completed for 8 of these mutants.
Our strain characterization effort has revealed broad diversity in every phenotype measured among the large set of strains we have tested. Data from this effort are routinely deposited to the Mouse Phenome Database (http://www.jax.org/phenome) as testing is completed. Data sets may be analyzed in the context of the MPD, or downloaded for analysis by users. This important resource enables the scientific community to gain insight into disease pathways and allows users to choose in silico the most appropriate inbred strains with one or more features of human disease for further study, without having to conduct broad surveys in their own labs. We are especially excited about having developed and validated a novel method to assess sleep non-invasively in mice and have subsequently used this approach to identify both chemical mutants and inbred strains with sleep disorders.

Courses planned at the JAX PGA

Like the other PGAs, JAX also provides training in the use of its resources. In particular the workshop on Integrative Systems Approaches to the Analysis of Heart, Lung, Blood and Sleep Disorders addresses contemporary approaches to using both expression and phenotype data together to identify new genes and networks implicated in disease, and will be offered in the Spring of 2006. The annual short course,

Genetic Approaches to Complex heart lung and Blood Disease, will take place September 23-October 2. Details and updates regarding schedules, faculty lists and application procedures are available on-line at http://www.jax.org/courses/events/coursedetails.do?id=128.

PGAs hold their first traveling tutorial

The PGAs held their first joint Traveling Tutorial at the Morehouse School of Medicine in Atlanta on May 2, 2005. The day-long tutorial featured presentations by the BayGenomics, Berkeley, PhysGen and SeattleSNPs PGAs, to an audience composed mostly of academic scientists interested in cardiology and hypertension. The tutorial drew nearly

100 scientists, and was well received. The PGA intends to continue to offer Traveling Tutorials, and plans to advertise this resource at national meetings as well as at the upcoming PGA Symposium.