Building Bridges to Discovery: Bioinformatics and Computational Biology at NIH

Karin A. Remington, Ph.D.

NCIBI 4th Annual Research Conference 2009

Ann Arbor, Michigan

NIGMS Center for Bioinformatics and Computational Biology (CBCB)

- NIGMS:
  - R01 portfolio in bioinformatics (interpreted very broadly), modeling, software development and maintenance…
  - Systems Biology Centers
  - MIDAS (Models of Infectious Disease Agent Study)
  - Training grants in Bioinformatics, and Biostatistics
- Trans-NIH:
  - Roadmap: National Centers for Biomedical Computing
  - BISTI: Biomedical Informatics Science and Technology Initiative
- Inter-agency Coordination

Building Bridges to discovery….

- Develop new tools and computational approaches
- Enable new science
- Promote collaboration
- Encourage Reduce/Reuse/Recycle

Building Bridges to Discovery:

The CHALLENGES of interdisciplinary funding at the NIH

…WITHIN NIGMS

CBCB sponsored Research Program Grants (R01s):

Some current and emerging interests:

- Structural and functional prediction of proteins
- Biostatistics – modeling, study design, data analysis
- Numerical linear algebra and optimization algorithms
- Network analysis/graph theory
- Systems biology approaches
- Innovative genomic analysis techniques
- Interdisciplinary training emphasizing collaborative science between Math/CS and Biology
- Biomedical data infrastructures supporting aggregation, interoperability and preservation

Fully integrating mathematics and computer science in the context of biological studies

Models of Infectious Disease Agent Study

Computational and mathematical investigations of:

- Dynamics of emergence and spread of pathogens and their products
- Identification and surveillance of infectious diseases
- Effectiveness and consequences of intervention strategies
- Host/pathogen interactions
- Ecological, climatic, and evolutionary dimensions of infectious diseases outbreaks.

NIGMS Predoctoral Research Training Program

The Vaccination Theory of education(?)

English is not history and history is not science and
science is not art and art is not music,
and art and music are minor subjects and
English, history and science major subjects, and
a subject is something you 'take' and
when you have taken it, you have 'had' it, and
if you have 'had' it, you are immune
and need not take it again.

From TEACHING AS A SUBVERSIVE ACTIVITY
Neil Postman & Charles Weingartner

NIGMS Predoctoral Research Training Program

- Major mission of NIGMS
- Predoctoral-only training grants
- 11 pre- Ph.D. areas
- 1 pre - M.D.- Ph.D. area
- Support for 3,200 trainees annually

Training Goals

- Multidisciplinary and multi-departmental training
- Faculty from different academic units provide breadth of research opportunities
- Training to master a core scientific area in depth
- Acquire skills and knowledge of related fields

TRANS-NIH

The prevailing view of NIH…

Promoting the nation's health through research.

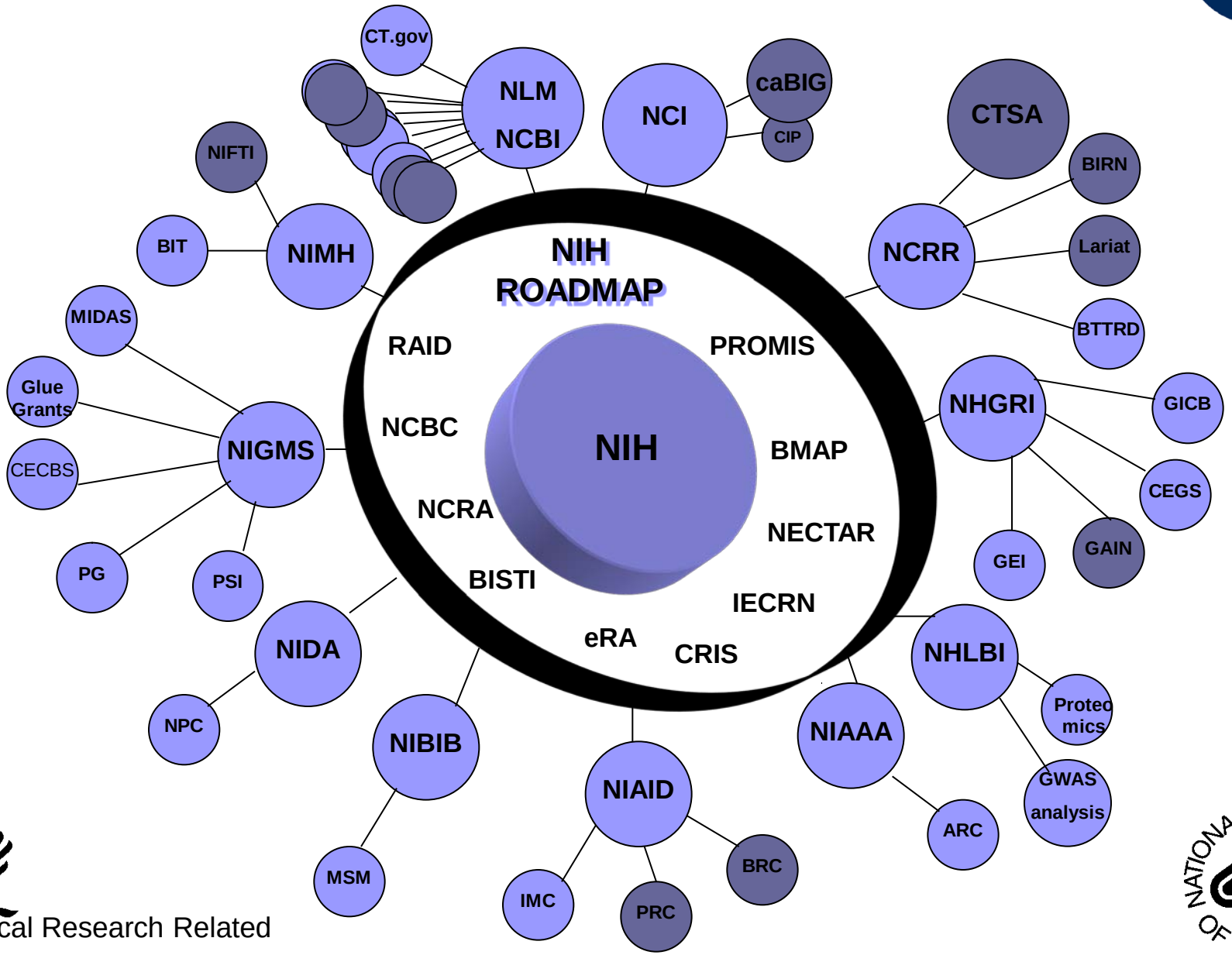
- Not a monolithic Agency - 27 Institutes and Centers
- Intramural research (NIH labs) - 10% of the budget, 6,000 scientists
- Extramural research (grants) - 80-90% of the budget

NCI $4.7B	NIAID $4.3B	NHLBI $2.9B	NIDDK $1.8B	NINDS $1.5B	NIMH $1.3
NICHD $1.2B	NIA $1.0B	NIDA $0.9B	NEI $0.6B	NIEHS $0.6B	NIAMS $0.5B
NIAAA $0.4B	NIDCD $0.3B	NIDCR $0.3B	NCMHD $0.1B	NCCAM $0.1B	NINR $0.1B
NIGMS $1.9B	NCRR $1.0B	NHGRI $0.4B	NLM $0.3B	NIBIB $0.2B	
CC	CIT	CSR	FIC	OD	

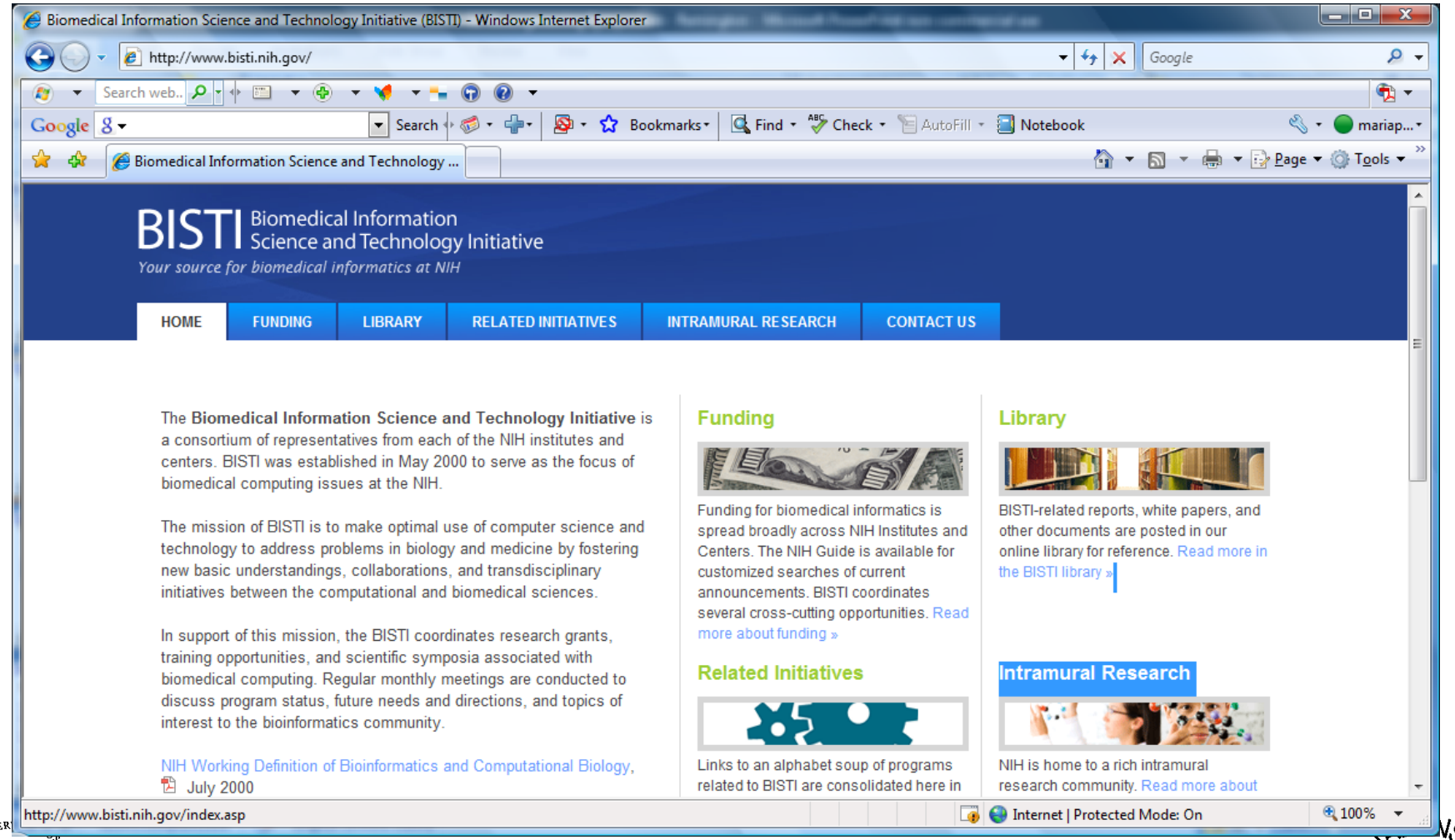
“BISTI”

The Biomedical Information Science and Technology Initiative

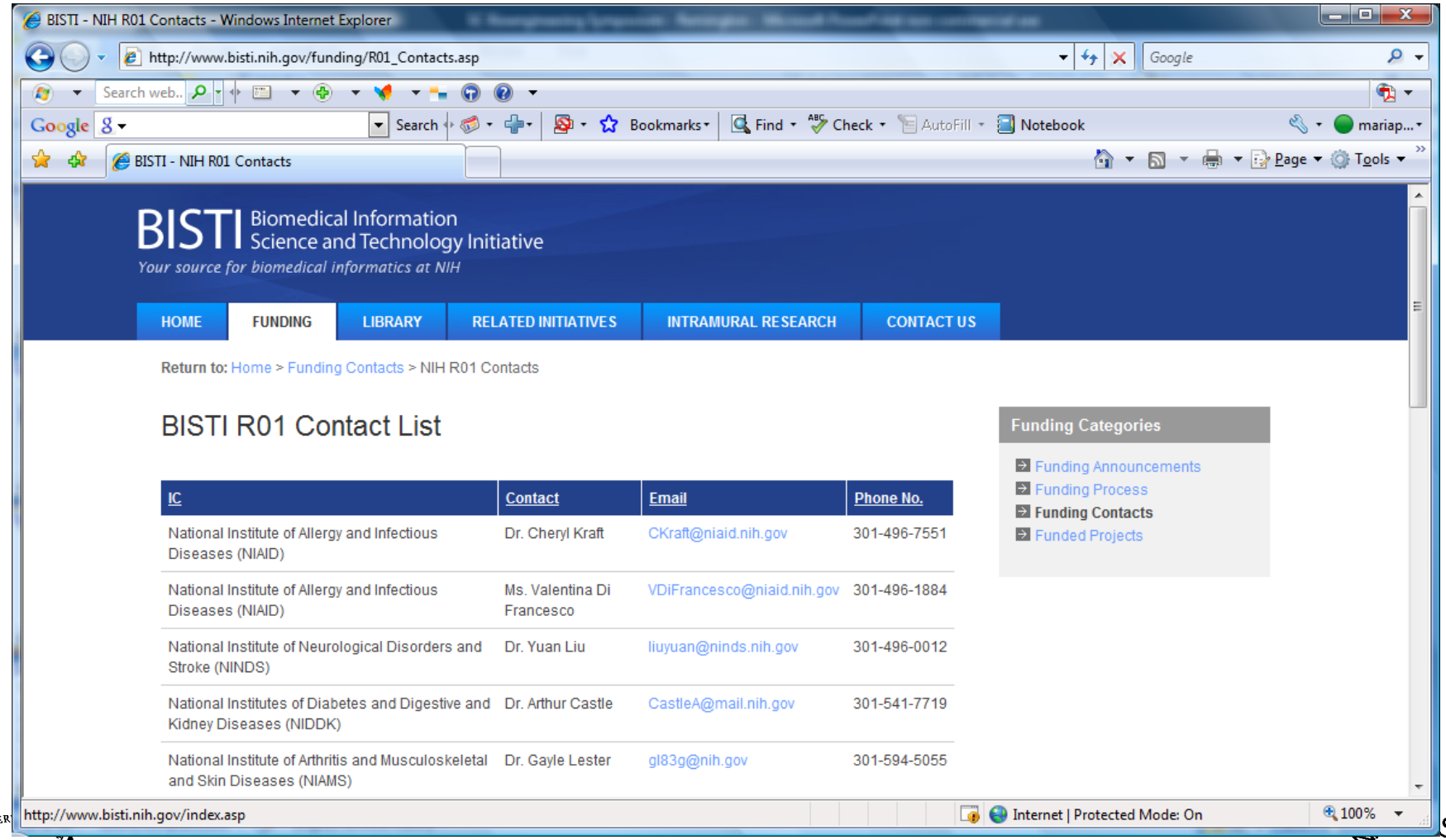
BISTI related NIH activities…



BISTI Website: http://www.bisti.nih.gov

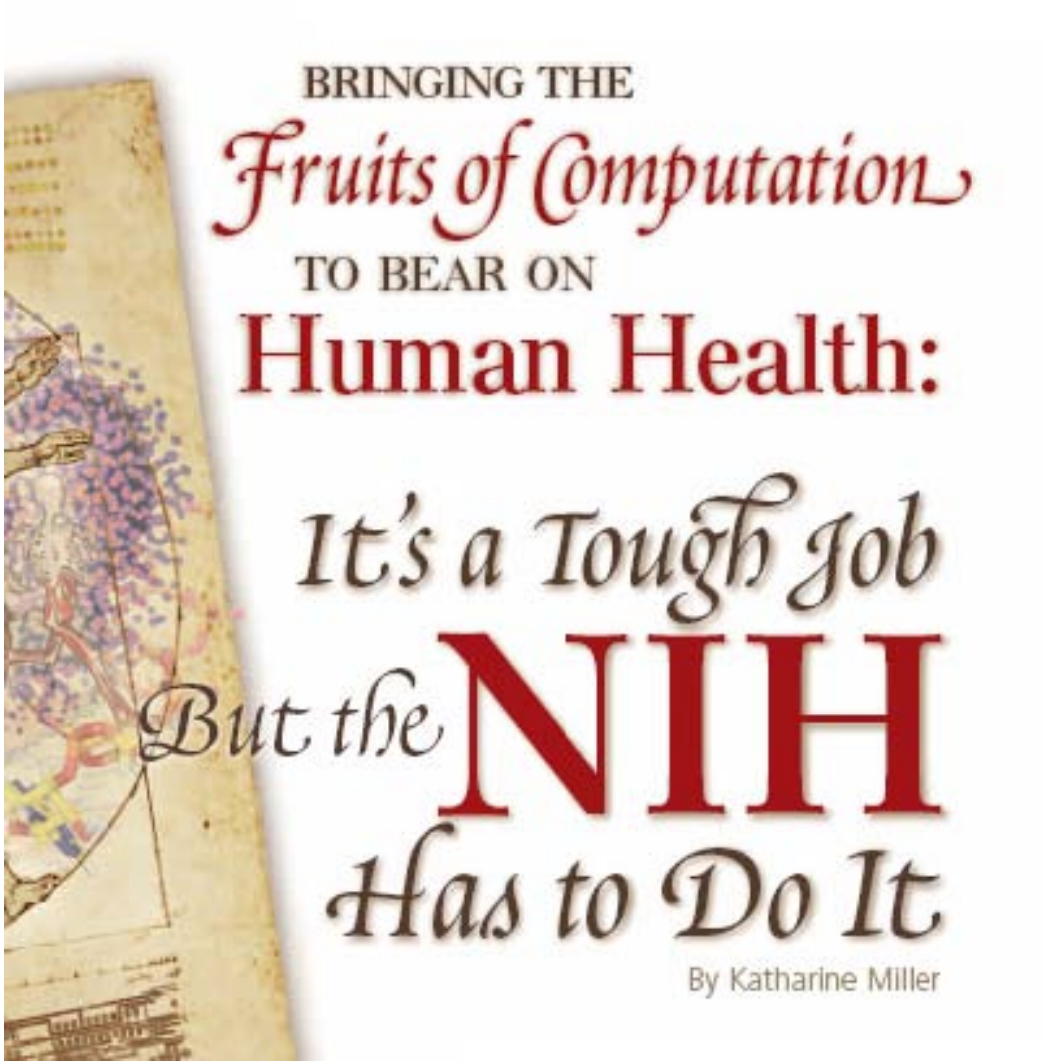


BISTI Website: http://www.bisti.nih.gov



DIVERSE DISCIPLINES, ONE COMMUNITY

BiomedicalComputation REVIEW



BRINGING THE

Fruits of Computation

TO BEAR ON

Human Health:

It's a Tough Job But the NIH *Has to Do It*

By Katharine Miller

DIVERSE DISCIPLINES, ONE COMMUNITY

BiomedicalComputation REVIEW

"As functional as [BISTI] has been over the years, it has really been unable to look across institutes in a real data-driven way, to analyze across NIH where our investments are going," says Remington.

NIH Roadmap Strategy

Building Blocks Pathways
Molecular Libraries
Structural Biology
Nanomedicine
Bioinformatics and Computational Biology

Translational Research Initiatives

Clinical Research Informatics

Clinical and Translational Science Awards

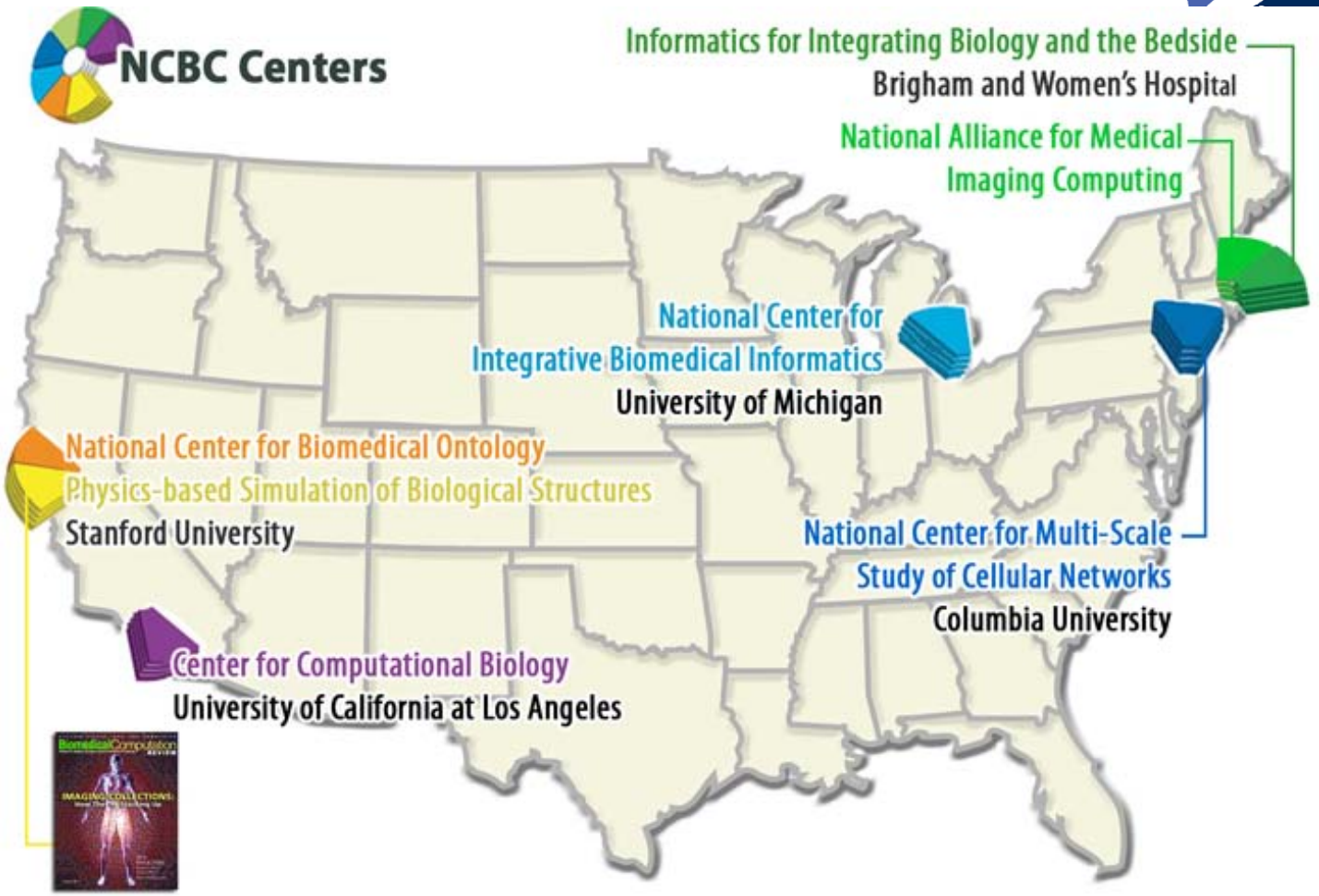
Clinical Research Training
National Clinical Research Associates

Bench ⟷ **Bedside** ⟷ **Practice**

Interdisciplinary Research
Pioneer Award
Nanomedicine

Public Private Partnerships

Integrated Research Networks
Clinical Outcomes



www.ncbcs.org

National Centers for Biomedical Computing

NCBC Goals

- ***Development*** of cutting edge computer science
- ***Translation*** of this computer science into biomedical computation, i.e., effective algorithms and environments for solving real biological problems
  - Enable the analysis, modeling, understanding, and prediction of dynamic and complex biomedical systems across time and distance scales
  - Allow the integration of biomedical and behavioral data and knowledge at all levels of organization

NCBC Organization by Core

- **Core 1:** Algorithm Development
  - (computer science)
- **Core 2:** Software Engineering
  - (biomedical computation)
- **Core 3:** Driving Biological Problems

Core 3: *Driving Biological Problems*

- Collaborative science
- Investigator(s) with challenging biomedical problems which focus the NCBC's computational research.
- Problems selected for their broad biomedical significance and compatibility with the core computational expertise of the specific NCBC.

Cores 4 - 6

- (4) providing ***infrastructure*** to serve the needs of the broad community of biomedical and behavioral researchers;
- (5) enhancing the ***training*** for a new generation of biomedical researchers in appropriate computational tools and techniques;
- (6) ***disseminating*** newly developed tools and techniques to the broader biomedical research community;

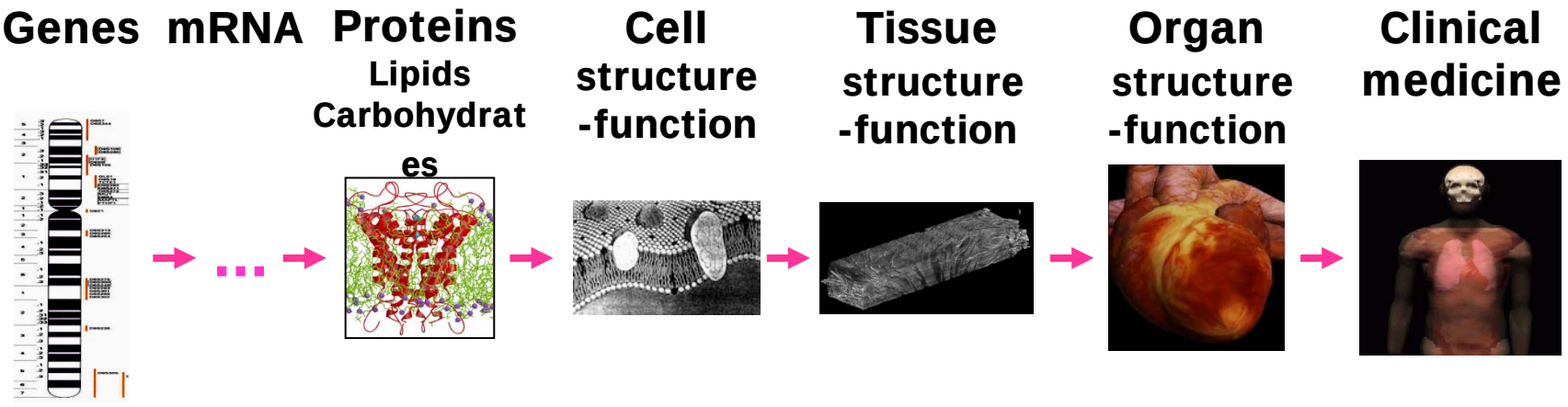
…ACROSS AGENCIES

also….Many Related Inter-Agency Activities:

- Networking and Information Technology Research and Development (NITRD)
- Interagency Modeling and Analysis Group (IMAG) NSF 04 607
- Collaborative Research in Computational Neurosciences (CRCNS) NSF 04 514
- Dynamic Data Driven Applications Systems NSF 05 570
- Multi-Agency Tissue Engineering Science (MATES) Working Group
- The Biomaterials and Medical Implant Science (BMIS) Coordinating Committee
- Roundtable on Biomedical Engineering Materials and Applications (BEMA)
- Interagency Image Guided Interventions (IGI)
- Joint NSF-NIH Initiative to Support Research in the Areas of Mathematical Biology NSF 04 572, NSF 06 607

Multiscale Modeling

Spanning scales from molecular to population, designed to illuminate biological processes and further the predictive capability in biological, biomedical and environmental systems. Multi-scale modeling encompasses concepts of space, time and state space.



Courtesy of Peter Hunter

Biological Scales

Interagency Modeling and Analysis Group

IMAG Multiscale Modeling (MSM) Consortium Working Groups

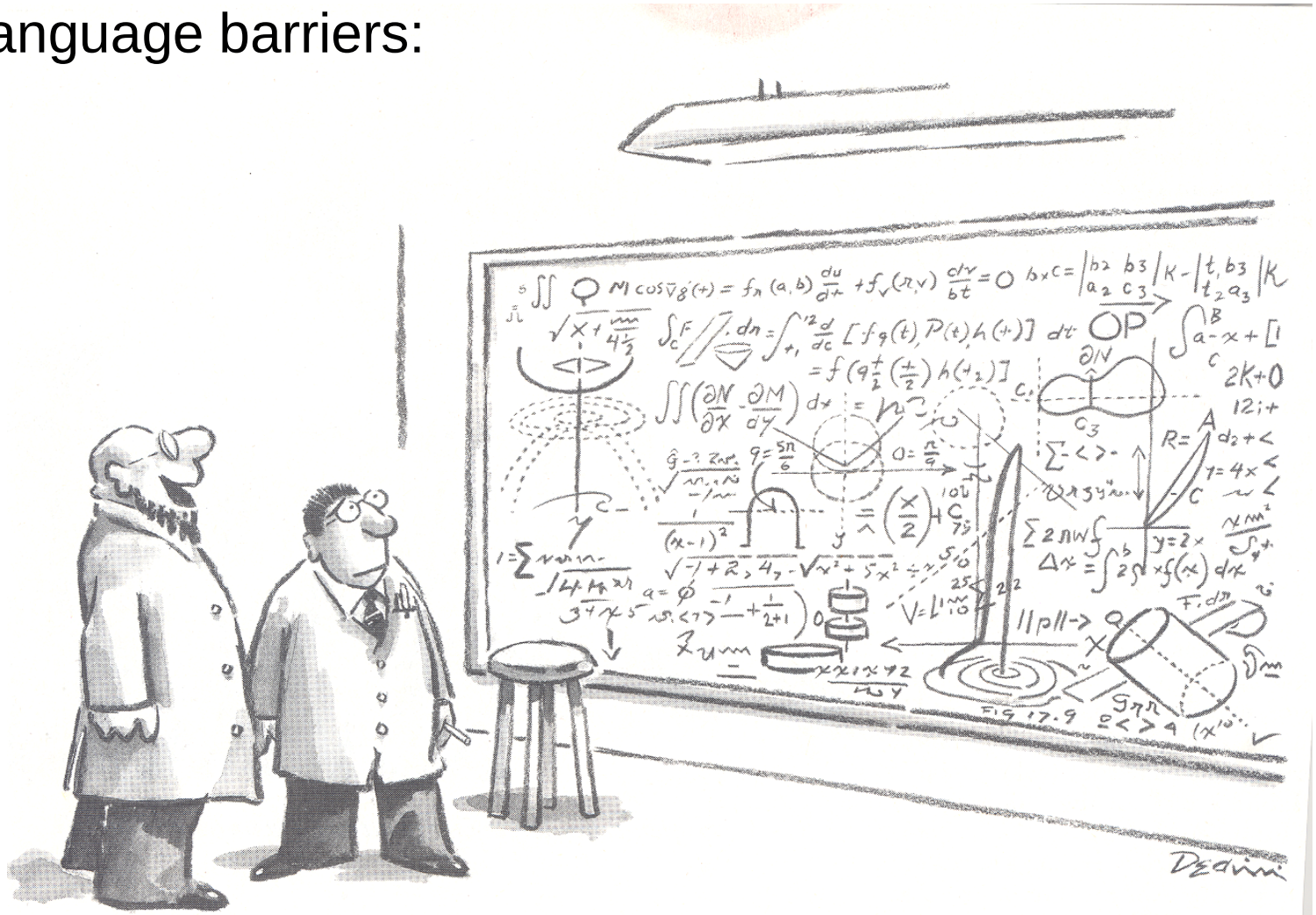
- Filament Dynamics and Simulation (FDS)
- Cardiac and Skeletal Muscle Physiology
- Macro-To-Micro Scale Transport in Human Systems
- Cell Level Modeling
- High Performance Computing, Computational Issues and Algorithms
- Tissue Mechanics
- Multiscale Imaging
- Theoretical Methods
- Nano-modeling
- Model Sharing

Contact: Grace Peng, NIBIB, email: penggr@mail.nih.gov

Building Bridges to Discovery:

The CHALLENGES of interdisciplinary funding at the NIH

Language barriers:



"Hey, no problem!"

NIGMS Predoctoral Research Training Program

The Vaccination Theory of education(?)

English is not history and history is not science and
science is not art and art is not music,
and art and music are minor subjects and
English, history and science major subjects, and
a subject is something you 'take' and
when you have taken it, you have 'had' it, and
if you have 'had' it, you are immune
and need not take it again.

From TEACHING AS A SUBVERSIVE ACTIVITY
Neil Postman & Charles Weingartner

A portrait of NIH Study Sections, when it comes to interdisciplinary teams and emerging areas:

Culture barriers!

To encourage fair review of the desired collaborations:

- Special Emphasis Panels for targeted FOAs
- Multiple PI applications
- Encouragement of “New” Investigators

  IC-level policies to boost likelihood of funding
- “Enhancing Peer-Review” effort underway

Culture barriers:

A portrait of tenure and promotion, when it comes to interdisciplinary teams and emerging areas:

The genome sequence of Drosophila melanogaster.

Science. 2000 Mar 24;287(5461):2185-95

elniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Ho e RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wort andell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod M, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, SM, WodageT, Wo KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhao Q, Zheng L, XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith bs RA, Myers EW, Rubin GM, Venter JC.

.

Remington K.

Another challenge: Appropriate funding mechanisms

- Interdisciplinary science – interdisciplinary teams
- Multiple PIs - strong leadership and shared vision
- Sustainable funding plans

Cooperative Agreements in a Competitive Science World

Models of Infectious Disease Agent Study

Computational and mathematical investigations of:

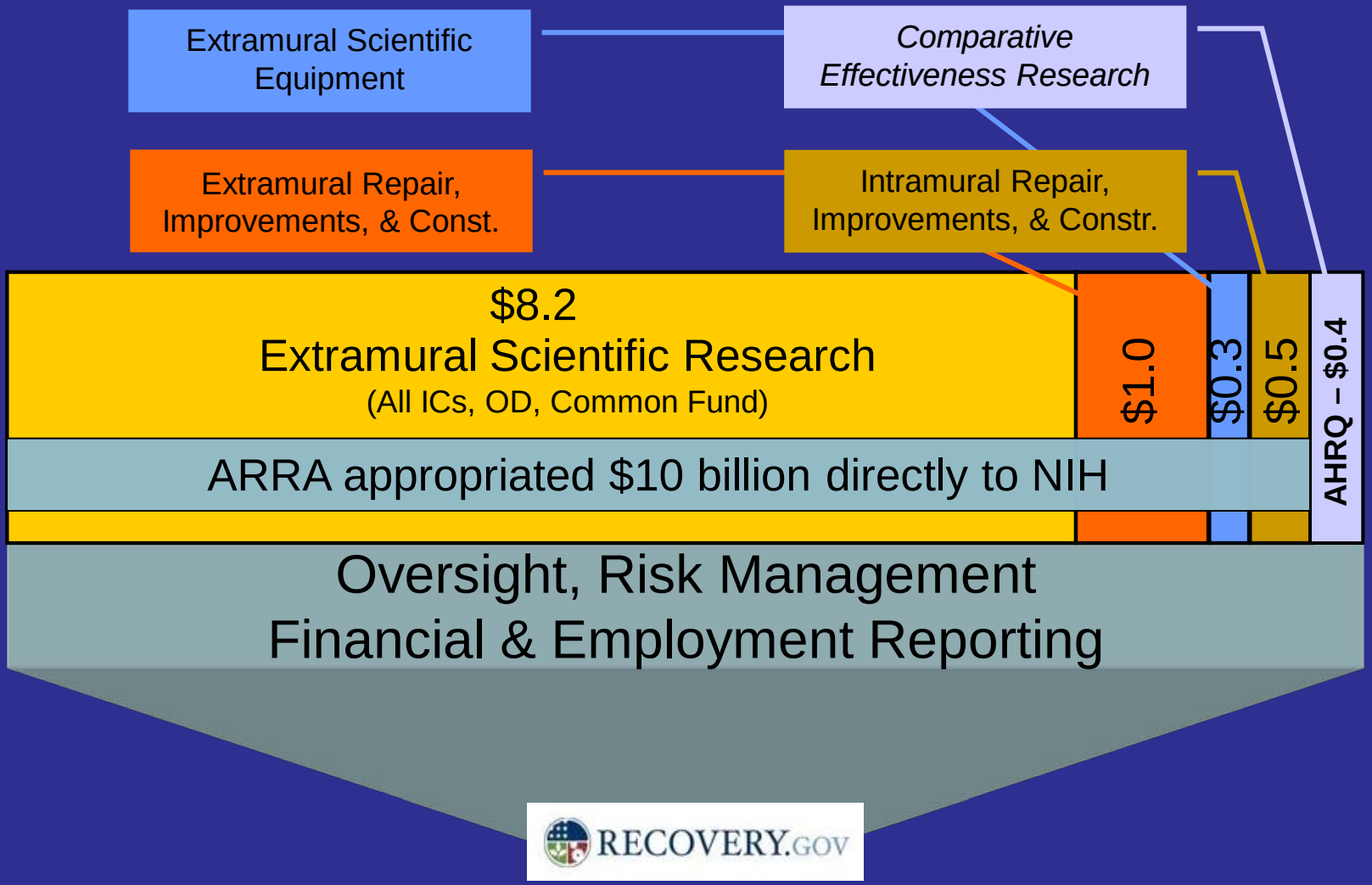
- Dynamics of emergence and spread of pathogens and their products
- Identification and surveillance of infectious diseases
- Effectiveness and consequences of intervention strategies
- Host/pathogen interactions
- Ecological, climatic, and evolutionary dimensions of infectious diseases outbreaks.

NIH/NIGMS Center for Bioinformatics and Computational Biology

Karin A. Remington, Ph.D.
remingka@nigms.nih.gov

NIH Implementation of ARRA

NIH Allocation of ARRA Funds
Dollars In Billions



Funding Impact

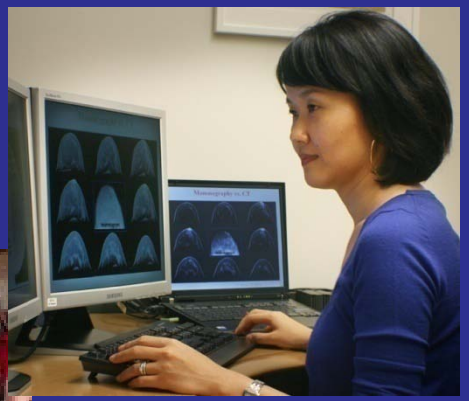
- Stimulate the economy
- Create and preserve jobs
- Advance biomedical research

http://www.nih.gov/recovery

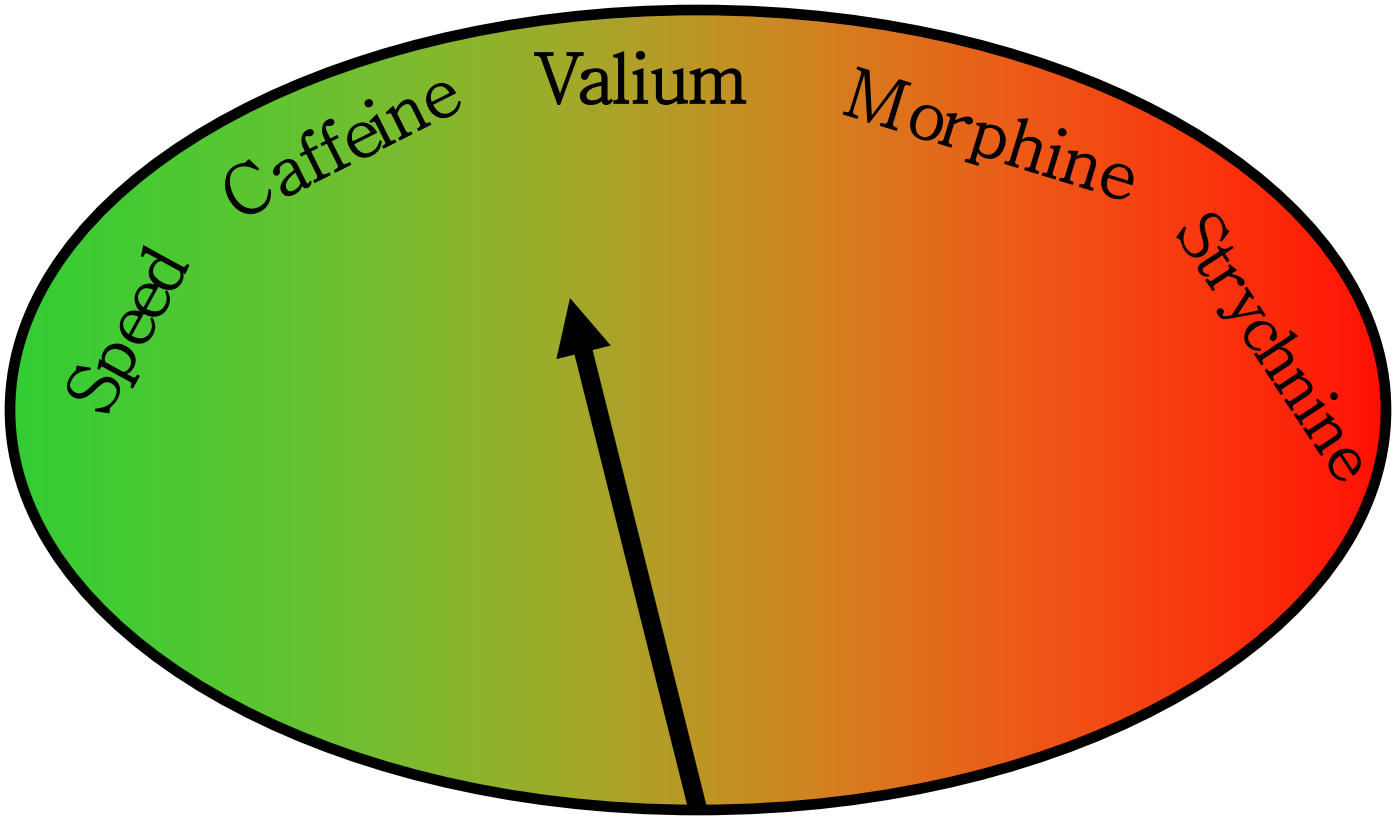
ARRA Grant Funding Opps and Info at **Grants.nih.gov/Recovery**

Scientific Research Approach

- Stimulate and accelerate biomedical research with existing mechanisms
- Expand science with new programs



NIGMS alone has $500M over two years (…really 18 months)



ARRA Meter

Culture barriers:

A portrait of tenure and promotion, when it comes to interdisciplinary teams and emerging areas:

First Genome Sequenced 1995

Haemophilus influenzae

1.8 million basepairs

The genome sequence of Drosophila melanogaster.

Science. 2000 Mar 24;287(5461):2185-95

elniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Ho e RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wort andell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod M, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, SM, WodageT, Wo KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhao Q, Zheng L, XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith bs RA, Myers EW, Rubin GM, Venter JC.

.

Remington K.

The Dog Genome: Survey Sequencing and Comparative Analysis

Ewen F. Kirkness,[1] Vineet Bafna,[2*] Aaron L. Halpern,[2*] Samuel Levy,[2*] Karin Remington,[2*] Douglas B. Rusch,[2*] Arthur L. Delcher,[1] Mihai Pop,[1] Wei Wang,[1] Claire M. Fraser,[1] J. Craig Venter[2]

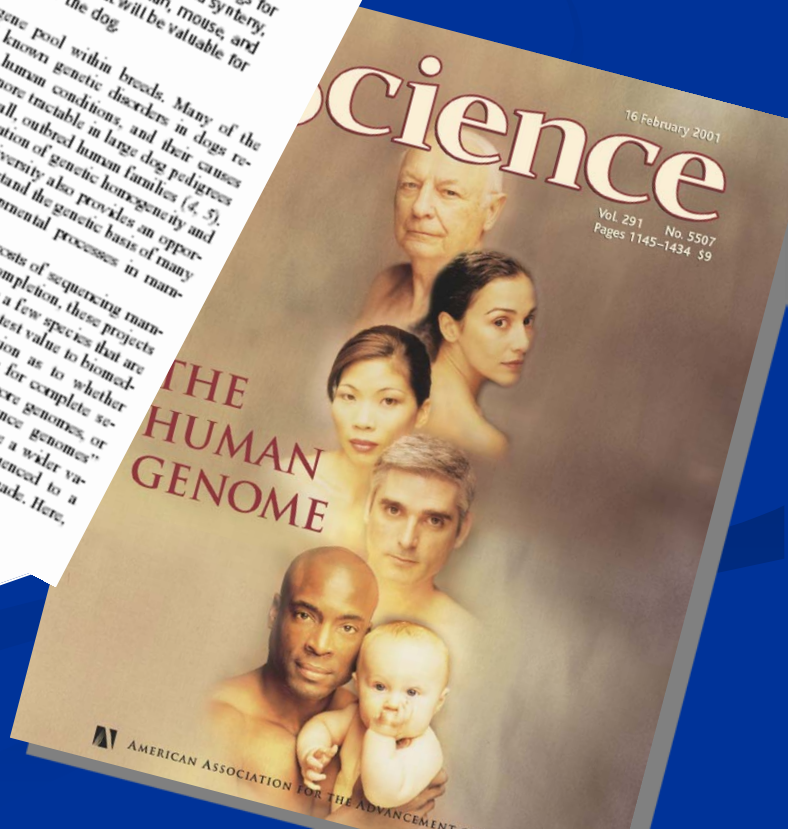
A survey of the dog genome sequence (6.22 million sequence reads; 1.5× coverage) demonstrates the power of sample sequencing for comparative analysis of mammalian genomes and the generation of species-specific resources. More than 650 million base pairs (>25%) of dog sequence align uniquely to the human genome, including fragments of putative orthologs for 18,473 of 24,567 annotated human genes. Mutation rates, conserved synteny, repeat content, and phylogeny can be compared among human, mouse, and dog. A variety of polymorphic elements are identified that will be valuable for mapping the genetic basis of diseases and traits in the dog.

Our understanding of how the human genome functions in health and disease will benefit from comparison of its structure with the genomes of other species (*1*, *2*). The domestic dog is a particularly good example, where an unusual population structure offers unique opportunities for understanding the genetic basis of morphology, behaviors, and disease susceptibility (*3*, *4*). The physical and behavioral characteristics of ~360 dog "breeds" are maintained by restricting gene flow between breeds. Many modern breeds are derived from few founders and have been inbred for desired characteristics. This has led to a species with enormous phenotypic diversity, but with significant homogenization of the gene pool within breeds. Many of the ~360 known genetic disorders in dogs resemble human conditions, and their causes may be more tractable in large dog pedigrees than in small, outbred human families (*4*, *5*). The combination of genetic homogeneity and phenotypic diversity also provides an opportunity to understand the genetic basis of many complex, developmental processes in mammals (*6*).

Because of the costs of sequencing mammalian genomes to completion, these projects have been restricted to a few species that are considered to be of greatest value to biomedical research. The decision as to whether future projects should aim for complete sequence coverage of a few more genomes, or whether the existing "reference genomes" can be exploited to characterize a wider variety of genomes that are sequenced to a lower level of coverage, must be made. Here,

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: kirkness@tigr.org
‡Present address: TIGR Center for Bioinformatics, The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA.
§Present address: Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA.
¶Present address: Center for Bioinformatics, University of Maryland, College Park, MD 20742, USA.
#Present address: Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, USA.
||Present address: Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA.
**Present address: Center for Human Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA.
††Present address: Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA.
‡‡Present address: Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, USA.

www.sciencemag.org SCIENCE VOL 296 31 MAY 2002



REVIEW

A Whole-Genome Assembly of *Drosophila*

Eugene W. Myers,[1]* Granger G. Sutton,[1] Art L. Delcher,[1] Ian M. Dew,[1] Dan P. Fasulo,[1] Michael J. Flanigan,[1] Saul A. Kravitz,[1] Clark M. Mobarry,[1] Knut H. J. Reinert,[1] Karin A. Remington,[1] Eric L. Anson,[1] Randall A. Bolanos,[1] Hui-Hsien Chou,[1] Catherine M. Jordan,[1] Aaron L. Halpern,[1] Stefano Lonardi,[1] Ellen M. Beasley,[1] Rhonda C. Brandon,[1] Lin Chen,[1] Patrick J. Dunn,[1] Zhongwu Lai,[1] Yong Liang,[1] Deborah R. Nusskern,[1] Ming Zhan,[1] Qing Zhang,[1] Xiangqun Zheng,[1] Gerald M. Rubin,[2] Mark D. Adams,[1] J. Craig Venter[1]

We report on the quality of a whole-genome assembly of *Drosophila melanogaster* and the nature of the computer algorithms that accomplished it. Three independent external data sources essentially agree with and support the assembly's sequence and ordering of contigs across the

end to sequence next in an interactive walk across the genome. Weber and Myers then proposed the whole-genome shotgun sequencing of the human genome in 1997 (*8*,


USA-80

NIH Center for Bioinformatics and Computational Biology

Karin A. Remington, Ph.D.
remingka@nigms.nih.gov