Keck Foundation Biotechnology Resource Laboratory at Yale University

The Keck Laboratory brings a wide range of state-of-the-art genomic and proteomic biotechnologies within reach of >1,000 Yale and non-Yale investigators at >300 institutions annually who otherwise would not benefit from the highly sophisticated and expensive instrumentation upon which biomedical research is increasingly dependent. The eight resources that are the Keck Lab and the major technologies offered by each are summarized in Table 1. To maximize its positive impact, the Keck Lab strives to provide as many high quality services to as many investigators as possible. Although priority is given to requests from Yale investigators and to non-Yale investigators who support Center (e.g., Yale/NIDA Neuroproteomics Center), Shared Instrumentation Grant (SIG), and other grant applications submitted by the Keck Lab; the founding decision to also accept requests from other non-Yale scientists ensures that when demand by Yale investigators for a particular biotechnology is less than is needed to maintain maximal throughput, a backlog of requests will be available to maintain high productivity. This policy, which benefits *all* users of this Laboratory, minimizes operating costs and maximizes the contribution of the Keck Lab and of Yale University to biomedical research. Throughout its 33 year history since 1980, the Keck Lab has been close to "self-supporting" in that it has recovered >95% of its salary, supply, and instrument maintenance and >88% of the cost of its instrumentation from service charges, grants, contracts, and gifts. Included among its grants are 25 NIH/NSF Shared Instrumentation and five NIH center grants awarded since 1981. With 41 full time staff, 22,343 ft^2 space, 88 major instrument systems (Table 2) purchased at a cost of >$17 million dollars, and offering 113 proteomics, genomics, biophysics, biostatistics, bioinformatics, and high performance computing technologies; we believe the Keck Lab is among the largest academic biotechnology resource laboratories of its kind.

Table 1: Overview of Resources within the Keck Laboratory

Resource	Major Technologies	Notes
Bioinformatics (2008)	Free or low cost 24/7 access to software; bioinformatics/pathway analysis of MS/proteomics, next generation sequencing, microarray, genotyping data; pipelining data processing; and molecular modeling using Genespring, Lasergene, Partek Genomics Suite, Ingenuity Pathway, MetaCore, Sybyl, TRIPOS, Mol-CAD, Pipeline Pilot, VIBE, GPMAW, and DeCyder.	Ph.D. staff provide training in software use and collaborate on projects requiring longer-term commitment of time and effort, the latter is often funded by an appropriate %effort charged to a grant.
Biostatistics (2002)	Statistical analysis of proteome profiling (e.g., iTRAQ, LFQ, SILAC, SILAM, MRM) and Affymetrix, Illumina, NimblGEN, and Sequenom mRNA expression, next generation sequencing, genotyping, and DNA methylation analyses using open source and commercial software and in-house developed programs.	Biostatistics staff are playing a key role in the development and continued optimization of the workflows, visualization, and other tools that are being used to develop Targeted Proteome Assays (TPAs) as described below.
Biophysics (1999)	Microcalorimetry, HPLC SEC/laser light scattering (LS), dynamic LS, stopped flow absorption/fluorescence kinetic analysis, surface plasmon resonance, & spectrofluorometry	Quantitative analysis (e.g., kinetics, stoichiometry, thermodynamics & binding affinities) of interactions between biomolecules; many instruments available for direct use by investigators with training and support for projects' design, implementation, and data interpretation.
DNA Sequencing (1989)	Single tube & 96 well Sanger sequencing; primer walking, sequence assembly & editing, fragment analysis.	Avg read length is 650 bp, but can be as high as 700-750 bp. Sequence 6-10 x 96-well plates/day with 24 hour turn-around and some same day sequencing. Overall throughput is >300,000 templates sequenced annually

High Performance Computing (HPC) (2005)	Two Ph.D. computer scientists support users of the Yale Biomedical HPC Center by optimizing, parallelizing, and *de novo* writing of codes and by use of: message passing (MPI), Linda-like parallel languages, batch queuing (PBS) & other methodologies.	The Yale clusters & other HPC instrumentation are used to carry out many analyses: large scale BLAST comparisons, Pseudogene searches, Electron microscopy image processing, Gaussian evaluation, molecular dynamics simulations of proteins, & use of Matlab, Perl, Python and R to support distributed computing.
Mass Spectrometry (MS)/ Proteomics (1993/1980)	Peptide, protein, oligo, small molecule MS; MS/MS protein identification; *de novo* peptide sequencing; FT-ICR MS exact mass; ID of protein post-translational modifications; LC/MS/MS; proteome profiling using 4 & 8-plex Isobaric Tags for Relative and Absolute Quantitation (iTRAQ), Label Free Quantitation (LFQ), Multidimensional Protein Identification Technology (MudPIT), and Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC); LC-MRM based small molecules quantitation and MRM & SWATH based Targeted Proteomics.	Resource is involved in the development and continued optimization of the workflows, visualization, and other tools that are being used to develop Targeted Proteome Assays (TPAs). Each 90 min TPA uses scheduled LC-MRM to determine the relative concentrations of 50-200 pre-selected protein biomarkers in tissue extracts or other biological samples (e.g., human urine) of interest. The first scheduled MRM assay that has been released is the Rat/Mouse Brain Assay that interrogates the relative level of expression of 112 proteins from 24 different protein classes.
Microarray (1999)	Full service core lab that offers Affymetrix, Illumina, NimblGen, and Sequenom platforms that support gene and MicroRNA expression profiling and genotyping analysis. This Resource also offers Quantitative PCR and Sequenom's MassARRAY EpiTYPER assay for the quantitative analysis of DNA methylation and identification of differentially methylated CpG sites in any genomic region or candidate gene of interest. This technology combines base-specific enzymatic cleavage with MALDI-TOF mass spectrometry.	The Microarray Resource also offers Illumina's revolutionary Infinium Assay, the HumanMethylation27 DNA Analysis BeadChip that allows researchers to interrogate 27,578 highly informative CpG sites per sample at single-nucleotide resolution
Oligo Synthesis (1988)	DNA oligo synthesis (40nmol, 200 nmol & 1.0 micromole scales), a wide range of modified DNA oligos can be synthesized – will incorporate most commercially available and some custom synthesized modified amidites; RNA oligo synthesis (1.0 micromole); gel & RP-HPLC purification.	CY 2012 throughput was 40,039 chargeable oligos. Current charge for a 40 nmol 50-mer is $12.50 as compared to $48 in 1998. Turn-around for normal, unpurified DNA oligos is typically <24 hours for <60-mers.

[a](Founding year) for the individual Keck Resource.

Table 2: Major Instruments and Instrument Systems In the Keck Laboratory

Resource	Description	Manufacturer/Model	Qty
Biophysics	HPLC SEC/Laser Light Scattering System (SEC/LS#1)	Waters' HPLC & PDA UV/Vis/Wyatt DAWN MALLS & QELS LS & Optilab rEX RI Detectors	1
	HPLC SEC/Laser Light Scattering System (SEC/LS#2)	Agilent 1100' HPLC & PDA UV/Vis/Wyatt DAWN EOS & QELS LS & Optilab rEX RI Detectors	1
	Asymmetric Flow, Field-Flow Fractionation & Light Scattering System (AFFF/LS)	Agilent HPLC & PDA & FL, Wyatt Eclipse 3 AFFF, HELEOS II MALLS,QELS & Optilab rEX RI Detectors	1
	Circular Dichroism Spectrometer	Photophysics Chirascan	1
	Fluorescence Plate Reader	Molecular Devices; SpectraMax Plus	1
	Spectrofluorometer	PTI-Quantmaster	1
	Stopped Flow Absorption/Fluorescence Kinetics	Applied Photophysics Limited SX.18MV	1
	Dynamic Light Scattering Detector (DLS)	DynaPro 800 Proterion Corp.	1
	Isothermal Microcalorimeter (ITC)	MicroCal VP-ITC	1
	Surface Plasmon Resonance Sensor (SPR)	GE HealthCare/BiaCore T100	1

Table 2: Major Instruments and Instrument Systems In the Keck Laboratory

Resource	Description	Manufacturer/Model	Qty
DNA Sequencing	DNA Sequencer: 96 Capillary	Applied Biosystems 3730XL	3
	DNA Sequencer	Ion Torrent PGM Genome Machine	1
	Biorobot	Tecan Freedom Evo 150 Robotic Workstation	1
	Biorobot	Beckman BioMek NX MC Robotic Workstation	2
	Spectrophotometer	Nanodrop	1
	Real Time PCR	AB 7900HT Fast RealTime PCR	1
	Bioanalyzer	Agilent Technologies 2200 TapeStation	1
High Performance Computing	Louise Compute Cluster (3680 cores, 321 nodes)	(Dell Blades) 68 PE1955s, 48 M600s, 128 M610s, 61 M620s, 16 M915s	1
	620 TB Parallel Storage	Blue Arc Mercury	1
	Network Switch	Force 10	1
Keck ITS	Administration File Server	Dell PowerEdge 2950	2
	Lab File Server	Dell PowerEdge R710	2
	Database Server	Dell PowerEdge R410	1
	Compute Cluster	Dell PowerEdge R720 (multicore)	1
Mass Spectrometry & Proteomics	MALDI TOF/TOF MS	ABI 4800	1
	LC-Electrospray/LTQ-Orbitrap Mass Spectrometer (MS)	Waters nano-Acquity UPLC/Thermo Fisher LTQ-Orbitrap	1
	LC-Electrospray/LTQ-Orbitrap Mass Spectrometer (MS)	Waters nano-Acquity UPLC/Thermo Fisher LTQ-Orbitrap XL	1
	LC-Electrospray/LTQ-Orbitrap Mass Spectrometer (MS)	Waters nano-Acquity UPLC/Thermo Fisher LTQ-Orbitrap Elite	1
	LC-Electrospray/APCI QTrap	Waters nano-Acquity UPLC/ABI 4000 QTrap	1
	LC-Electrospray/APCI QTrap	Waters nano-Acquity UPLC/ABI 5500 QTrap	1
	LC-Electrospray/Triple Quad Tof	Waters nano-Acquity UPLC/ABSCIEX 5600 Triple Tof	1
	Electrospray FT-ICR MS	Bruker Daltonics 9.4T Apex Qe FT-ICR	1
	HPLC Systems: Analytical	Hewlett Packard 1090 & 1100	2
	Amino Acid Analyzer: Ion Exchange	Hitachi L8900A-PH	1
Microarray	Hybridization Oven	Affymetrix 640	3
	Fluidics Station	Affymetrix 400	3
	Scanner	Affymetrix 7G	1
	Micro-electrophoresis Bioanalyzer	Agilent 2100	2
	Spectrophotometer	Nanodrop	2
	HiScan Scanner with Autoloader	Illumina	1
	Tecan Robot	Illumina	2
	Sequenom MassArray	Sequenom	1
	Sequenom Liquid Handler	Sequenom	1
	Maui Hyb ovens	Maui	3
	Glass Slide Microarray Scanner	Axon GenePix 4000A, 2 fluor Capability	1
	Resonance Light Scattering Microarray Reader	Qiagen HiLight Reader	1
	Gel Imaging/Documentation System	BioRad Gel Doc	1
	Spectrophotometer	Hitachi GeneSpec II	1

Table 2: Major Instruments and Instrument Systems In the Keck Laboratory

Resource	Description	Manufacturer/Model	Qty
	Spectrofluorometer	SpectraMax Plus	1
	Pacific Biosystems Single Molecule Real-Time Sequencer	Pacific Biosystems PacBio RS	1
	Epenndorf Liquid Handling System	Eppendorf EpiMotion 5075	1
	Glass Slide Scanner	Roche NimbelGen MS 200	1
	Micro-electrophoresis Analyzer	Caliper LabChip GX	1
	Promega DNA extraction system	Promega ReliaPrep LV 32 HSM	1
	Quantitative PCR Machine	Roche LightCycler 480 II	1
Oligo Synthesis	Oligo Synthesizer: 4 Column/Parallel	Applied Biosystems 394	10
	Oligo Synthesizer: 8 Column Parallel	Azco Oligo 800 Oligo Synthesizer	1
	Oligo Synthesizer: 192 Well Based/Parallel	BioLytic Dr. Oligo-192	2
	Micro-electrophoresis Analyzer	Caliper LabChip GX	1
	Capillary Electrophoresis System	Beckman P/ACE MDQ Capillary Electrophoresis	3
	Microplate Absorbance Reader	BioTek Epoch Microplate Absorbance Reader	1
	Microplate Absorbance Reader	SpectraMax Plus Microplate Absorbance Reader	1
	Preparative HPLC	Agilent 1100 Preparative HPLC	1
Total Instrument Systems			**88**

We believe the Keck Laboratory is making an important contribution to biomedical research that extends far beyond Yale University. In CY 2012 the Keck Laboratory carried out 386,615 services for 461 Yale and 574 non-Yale (1,035 total) principal investigators at 305 institutions in 29 countries. In addition, the HPC Resource provided more than 11 million CPU-hrs of high performance computing to support the research of Yale investigators. To help users take maximum advantage of its Resources, the Keck web pages (http://keck.med.yale.edu/index.aspx) provide extensive background information on the biotechnologies it offers and on interpreting the resulting data. As judged by extensive use of the Keck Lab by non-Yale investigators, who requested 24,996 analyses and syntheses in 2012, and by comments like the following by the NIH Study Panel that reviewed its 2011 Shared Instrumentation Grant (SIG, RR0317957, which earned a perfect priority score of “10”) and which are typical of similar comments by the Study Panels that have reviewed its other 24 NIH/NSF Shared Instrumentation Grants since 1981, the Keck Laboratory provides a high level of service:

> *“This is a highly productive facility which is used over its capacity.” “This is an outstanding proposal that will meet the research needs of multiple, highly productive investigators. The Keck Facility at Yale has a long (30 year) history of providing high quality, cutting-edge mass spectrometry resources to Yale researchers and to many other researchers around the world.*

The following sections provide additional information on individual Keck Resources.

Mass Spectrometry/Proteomics Resource (http://keck.med.yale.edu/proteomics/)

The Mass Spectrometry and Proteomics Resource was founded in 1993 by Kathryn Stone who is now the Associate Director of Proteomics in the Keck Laboratory. Ms. Stone has been recognized for her high level of MS expertise by the Association of Biomolecular Resource Facilities (ABRF), who chose her to form and to become the first Chair of its Proteomics Research Group. The Keck MS/Proteomics Resource has 9 staff including 3 Ph.D. and 3 M.S. level appointments with >100 years of MS experience. The MS/Proteomic Resource has 8 tandem mass spectrometry systems (Table 3) that provide a comprehensive range of biotechnologies to support proteomics research. These instruments include a Bruker 9.4T Apex-Qe FT-ICR MS [purchased with a $1.5 million high end instrumentation grant (NIH RR17266, PI; K. Williams)]; three Thermo LTQ Orbitraps including XL and Elite models, an Applied Biosystems (AB) Model 4800 MALDI-Tof/Tof system, AB 4000 and 5500 Q-Trap MS instrument platforms, and an AB Model 5600 Triple TOF. As indicated in Table 3, four of these mass spectrometers have been acquired at a cost of more than $2.5 million since

2009. Using these mass spectrometers and other instrumentation, which includes a Hitachi L8900A-PH Amino Acid Analyzer that is used to quantify and match the amounts of protein in control versus experimental samples, the MS/Proteomics Resource provides investigators with access to several quantitative proteome analysis technologies. These technologies include LC-MS Label Free Quantitation (LFQ), iTRAQ, SILAC, MudPIT, phosphoproteome profiling, as well as two targeted approaches: LC-MRM and SWATH. In addition to carrying out many customized projects to quantify selected proteins and their post-translational modifications of interest, the MS/Proteomics Resource has developed and continues to optimize the workflows, visualization, and other tools that are being used to design and implement Targeted Proteome Assays (TPAs). Each 90 min TPA uses scheduled LC-MRM to determine the relative concentrations of 50-200 pre-selected protein biomarkers in tissue extracts or other biological samples (e.g., human urine) of interest. The first scheduled MRM assay that has been released is the Rat/Mouse Brain Assay that interrogates the relative level of expression of 112 proteins from 24 different protein classes. A unique advantage of the complementary SWATH technology is that it allows retrospective interrogations of stored MS/MS datasets.

Table 3: Keck Mass Spectrometry Instrumentation

#	Date[1]	Model	Cost	Funding/Notes
1	Sep-2004	FT-ICR 9.4T Qe	$1,536,500	NIH High End SIG, RR017266, PI: K. Williams
2	Dec-2006	4800 Tof/Tof	$390,000	Applied Biosystems, PI: K. Williams
3	Oct-2007	LTQ-Orbitrap XL	$694,295	NIH SIG, RR024617, PI: E. Gulcicek
4	Jan-2008	4000 Q-Trap,	$563,375	NIH CTSA,UL1 RR024139, PI: R. Sherwin
5	Apr-2010	5500 Q-Trap Pro	$491,747	NIH SIG, RR026707, PI: C. Colangelo
6	May-2010	LTQ-Orbitrap	$665,432	System, which included a CapLC, was purchased by Yale in 2007 and donated to Keck Lab in 2010.
7	Jan-2012	5600 Triple Tof	$625,600	Instrument loaned from AB SCIEX from 1/11-5/11 then leased until purchased with equal funding from YCCI/CTSA grant (UL1 RR024139, PI: R. Sherwin) and YSM.
8	Jan-2012	LTQ-Orbitrap Elite	$777,269	NIH SIG, RR031795, PI: E. Gulcicek (Priority Score: "10"), instrument brought fully "on-line" during January, 2012.
Total Purchase Cost			**$5,744,218**	

The Keck MS/Proteomics Resource collaborates closely with the Biostatistics, Bioinformatics, and High Performance Computing Resources for analysis and computational needs and holds bi-monthly meetings which include extensive discussion of the best approaches for meeting the changing research needs of its users. The web-based Yale Protein Expression Database (YPED, see below for additional information about this resource) transmits protein profiling data to users while at the same time archiving, and providing an associated repository to facilitate public dissemination of published data. YPED is under constant refinement and development, and has the capability to integrate, analyze, visualize, and archive data and results from each of the proteomic analysis platforms in the Center.

Yale Protein Expression Database (YPED, http://medicine.yale.edu/keck/proteomics/yped/)

All of the proteomics data from analyses carried out in the MS/Proteomics Resource is archived in the Yale Protein Expression Database (YPED) that has been designed and is continually being improved to meet the need for the storage, retrieval, visualization, integrated analysis, and dissemination of high throughput proteomic analyses (Shifman et al, 2007). YPED was initiated as a collaborative effort between the previously funded NHLBI Proteomics Center, the currently funded Yale/NIDA Neuroproteomics Center, the Keck Laboratory and the Center for Medical Informatics. YPED contains >15,000 datasets from >1,300 users and has >3 million unique peptides identified (with a Mascot score greater than or equal to the homology score) from >650,000 unique proteins, including 20,843 human and 20,059 mouse. YPED provides a powerful resource for supporting MRM and SWATH proteome technologies and MS/MS based protein identifications.

YPED handles all types of proteomic and small molecule MS analyses including LC-MS/MS protein identifications, and identification/quantitation results from proteome profiling experiments such as DIGE, iTRAQ, ICAT, and SILAC. YPED also contains a repository for public access to proteomics data that supports publications. Sample descriptions are compatible with the evolving MIAPE standards. Tools are available to facilitate sample comparison, viewing of phosphoproteins, calculation of the mass defect of phosphopeptides (and therefore the probability that the peptide is a phosphopeptide) and links to the PANTHER Classification System. For Targeted Proteomics, reports of peptide concentrations are listed and a synthetic peptide database is maintained of the “AQUA” peptides that were utilized in the MRM analysis. Users Log into YPED with a unique login and password that allows them to easily view their proteomics and small molecule data.

Biophysical Resource (http://keck.med.yale.edu/biophysics/)

The Biophysical Resource, directed by Dr. Ewa Folta-Stogniew, provides technologies for characterizing interactions between biomolecules including the oligomeric state of the interacting species; thermodynamic parameters that govern the interactions including binding constants, the enthalpic and entropic contributions to complex formation; and kinetic information: k_a and k_d rates.

The instrumentation in the Biophysics Resource (Tables 2 and 4) allows determination of the oligomeric state of the interacting species and of the resulting complex utilizing size exclusion chromatography/multiangle laser light scattering (SEC/MALLS) or dynamic laser light scattering (DLS); binding constants are determined using isothermal calorimetry (ITC), surface plasmon resonance (SPR) or a PTI-QuantaMaster spectrofluorometer; kinetics using stopped-flow and SPR; and the enthalpy and entropy of binding reactions is determined using ITC – which is an almost universal technology for studying macromolecular interactions that is based on the heat that interactions give off or take up upon complex formation, hence, this approach does not require labeling of the interacting species. Similarly, SPR also is used to study label-free macromolecular interactions. SPR detects binding in real time by monitoring changes in mass concentration at the chip surface with the association and dissociation rate constants then being determined from the reaction traces. Samples ranging from small molecules to crude extracts, lipid vesicles, viruses, bacteria and eukaryotic cells can be studied in real-time with little or no sample preparation. The Biophysical Resource is equipped with a state-of-the-art BiaCore T100 instrument that was funded with an NIH SIG (Table 4) in 2009. The T100 can run analytes serially so the same aliquot of analyte can be applied to both the reference and the "active" cell, which decreases the amount of analyte needed and the experimental time by about half. The T100 also has a 10-fold lower detection threshold (i.e., 100 rather than 1,000 Da on the older BiaCore Model 1000) and can run 384 well plates – making it more suitable for high throughput studies. Lastly, the Chirascan spectrometer provides a sensitive circular dichroism (CD) spectrometer for directly probing the secondary structure, or conformation, of biomolecules. The Chirascan can also be used to investigate protein tertiary structure by measuring the CD induced in the aromatic side-chains and disulfide bonds of the peptide backbone. Protein secondary and tertiary structures often change as a function of temperature or pH, and circular dichroism may be used to assess conformational stability under varying conditions.

Table 4: Keck Biophysics Instrumentation

#	Date	Model	Cost	Funding/Notes
1	9/07	Asymmetric flow FFF and Composition Gradient/Light Scattering System	$271.000	SIG (RR023748) PI: E. Folta-Stogniew),
2	9/09	BiaCore T100	$363,515	SIG (RR026992, received perfect priority score of "10", PI: E. Folta-Stogniew), instrument had been leased prior to purchase

An NIH SIG (RR023748, PI: E. Folta-Stogniew) also funded the purchase of an Asymmetric Flow Field-Flow Fractionation (AFFF) and Automated Composition Gradient syringe delivery systems that share Light Scattering (LS) and other detectors. AFFF uses single phase chromatography to separate samples from 1nm to >20 microns. High-resolution separation by size occurs within a thin channel against a perpendicular flow force. The separation is gentle, rapid, and non-disruptive - without a stationary phase that may degrade, bind, or otherwise alter the sample. AFFF fractionation and light scattering (LS) detection allows sample fractionation and determination of size and molar mass in a single experiment. Coupling LS measurement to a fractionation step enables determination of the molar masses and oligomeric states of diverse macromolecules (e.g., proteins and their complexes, nucleic acids, liposomes, and polysaccharides). In contrast to ultracentrifugation, AFFF/LS fractionation and sizing allows the facile collection of the resulting fractions for further analyses.

The best methods for each project are recommended based on the questions being addressed, the amount of available sample, and the spectroscopic properties of the interacting molecules. Users can either request that their samples be analyzed by Biophysics staff or through “open access” whereby center investigators have *direct* use of the needed instrumentation. The Biophysics Resource provides training for interested investigators and support for each projects’ design, execution, and interpretation of the resulting data. This approach allows investigators with little knowledge of biophysics to successfully complete advanced biophysical analyses that have been designed to best address their research challenges.

Biostatistics Resource (http://keck.med.yale.edu/biostatistics/)

Biostatistics Resource was founded by Dr. Hongyu Zhao in 2002 and has three Ph.D. level biostatisticians on its staff. This resource provides statistical analysis of quantitative proteomics (e.g., LFQ, iTRAQ, SILAC, MRM, SWATH) and genomics (e.g. gene and MicroRNA expression profiling and genotyping analysis on Affymetrix, Illumina, NimblGen, and Sequenom platforms) research data that often is directed at identifying biomarkers for disease diagnosis and prognosis. In addition, as described above, Biostatistics staff are playing a key role in the development and continued optimization of the workflows, visualization, and other tools that are being used to develop Targeted Proteome Assays (TPAs) that are based on scheduled LC-MRM. To accomplish these objectives the Resource uses and integrates open source, in-house developed, as well as commercial software. The Biostatistics Resource collaborated with the Keck MS/Proteomics and Microarray Resources and the Center for Medical Informatics to build the institutional, web-based Yale Microarray Database (YMD) and Yale Protein Expression Database (YPED). The Biostatistics Resource also collaborates with the Yale Center for Statistical Genomics and Proteomics to develop pathway and protein interaction database and visualization tools, and collaborates with and complements the Keck Bioinformatics Resource. Also, the Biostatistics Resource works closely with the Keck High Performance Computing (HPC) Resource on those challenges amenable to an HPC solution.

Bioinformatics Resource (http://keck.med.yale.edu/bioinformatics/)

The Bioinformatics Resource is Co-directed by Hongyu Zhao [Professor, Public Health (Biostatistics)] and Mark Gerstein (Professor, Molecular Biophysics and Biochemistry) and has Ph.D. level staff who provide support at three levels. First, free or subscription-based access is provided to widely used commercial and open source bioinformatics programs on computers located in the Resource. Some software is loaded on PCs in the Bioinformatics Resource while other software can be used remotely, either through client programs or with a Web browser. The Bioinformatics staff are available to recommend the most appropriate software to meet a given research objective and to provide training on the use of these resources. Secondly, the staff provide service charge-based consultation services for well–defined bioinformatics analyses. Thirdly, the staff are available for collaborative projects requiring long-term commitment of time and effort. Often, the best mechanism for funding these collaborations is via an appropriate %effort charged to a grant.

Bioinformatics staff will assist in all phases of biomedical research. They will help obtain grant support by writing letters of support and by working with the biostatistics staff and users to optimally design experiments, edit the bioinformatics sections of grant applications, and by suggesting the most appropriate bioinformatics approaches for helping to interpret and derive the most benefit from the high throughput technologies available through the Keck Lab. The Bioinformatics Resource makes available a wide range of software to help users optimally analyze and fully interpret their data. Some of the software includes (but is not limited to) Lasergene and Gene Construction Kit for a number of protein/DNA sequence analysis tasks; Genespring GX and Partek Genomics Suite for the analysis of microarray expression data and other types of microarray data such as that produced from the use of protein/peptide substrates; Genespring GT and HelixTree for genotyping analyses; Ingenuity Pathway Analysis and MetaCore for pathway and network analysis; Sybyl for protein structure-modeling and visualization, GPMAW for interpreting mass spectral data from peptide/protein analysis (particularly for manual interpretation of protein PTM analyses); Pipeline Pilot and VIBE (with user-friendly graphical interfaces and specialized scripting languages) for creating data-processing “pipelines” consisting of modules that perform separate tasks either on local computers or at remote Web sites.

High Performance Computing (HPC) Resource and the Yale Biomedical Center for HPC (http://medicine.yale.edu/keck/hpc/)

The High Performance Computing (HPC) Resource plays a critically important role in the Yale Biomedical Center for High Performance Computing. The HPC Resource is Co-Directed by two Ph.D. level computer scientists (Nick Carriero and Robert Bjornson) who provide the parallel programming knowledge for bringing HPC to bear on research. Interactions between users and the Resource varies widely and depends upon the ability of each investigator. Experienced investigators are given accounts and they use the computer clusters and other instrumentation on their own. Usually, however, the HPC Center's staff consults with users to some degree. Typically, one of the Center's computer scientists examines and benchmarks user's codes, often finding ways to improve the serial performance or to parallelize the code or both. From 2010-2012 the HPC Resource, provided high performance computing to >400 researchers from 107 research groups and 33 Yale departments. In 2012 this Resource provided more than 11 million CPU-hrs of high performance computing. The staff in the HPC Resource oversee the use of ~3,600 HPC cores and 600 TB of high performance parallel storage with most of the cost of the previous and current equipment being covered by an NIH High End SIG awarded in 2004 (RR19895, PI: K. Williams, $1.6 million) and an NIH SIG awarded in 2011 (Table 5).

Table 5: Keck HPC Instrumentation

#	Date	Model	Cost	Funding/Notes
1	6/10	BlueArc Storage (300 TB) and Dell Compute Nodes (112 M610 nodes, totaling 896 cores)	$533,107	Yale School of Medicine and individual researchers
2	9/11	BlueArc Storage (300 TB) and Dell compute nodes (45 M620 and 16 M915 nodes, totaling 1744 cores) and associated hardware	$590,099	SIG (RR029676, PI: C. Colangelo and then R. Bjornson)

The Biomedical Cluster is housed in a custom renovated Data Center that opened in 2006 in the 300 George St. Technology Building. The Data Center has 4,215 ft^2 of raised floor space for housing computing equipment including the 15 racks @1,600 pounds/rack needed for the Biomedical Cluster. The raised floor with vented tiles was the first generation cooling method used in the Data Center. When “In-Row” cooling technology became available, the Data Center also installed Liebert In-Row XDC/XDH HVAC equipment to further enhance the rack CFM air flow to the cluster equipment. There is 210 Tons (739 kW) of cooling capacity in the Data Center that provides N + 1 redundancy. Cluster Storage and Master Servers are powered from a Mitsubishi UPS that is further backed up by gas-powered generators. The Data Center has a Vesda “very early warning smoke detection” system that is monitored through a Siemens notification application.

In addition to assisting individual investigators, the staff in the HPC Resource continue to develop and support systems now in production that were reported in previous renewals, in particular: 1) A system that converts MS data from a vendor specific output format to a more generic format, 2) A processing pipeline to periodically submit all of the distinct peptides in YPED to BLAST for comparison against current standard references such as UniSwiss and to transfer the results to YPED. These data can then be used to select peptides that are specific to a particular protein target, 3) X!!Tandem (Bjornson et al, 2008), which is a parallelized version of the popular, open source X!Tandem search algorithm that matches tandem mass spectra to peptide sequences (http://www.thegpm.org/tandem/). The success of the HPC Resource at parallelizing X!Tandem and the clusters in the Biomedical HPC Center provide the opportunity of searching all peptide MS/MS spectra for the presence of all common PTMs which has the potential to further increase the fraction of proteins identified. 4) A tool is being designed to process records describing peptides and associated candidate phosphorylation sites into expanded amino acid sequence data in a format suitable for motif detection, and 5) Exploratory work is being carried out on deploying OpenMS/OpenSWATH (http://open-ms.sourceforge.net/) on our HPC cluster.

The HPC Center serves as a focal point for the staff and physical infrastructure needed to bring computer science to bear on challenging, yet amenable problems that slow biomedical research.

Microarray Resource (http://ycga.yale.edu/microarrays/)

The Microarray Resource is a full service core laboratory that is very closely associated with the Yale Center for Genome Analysis (YCGA, see below) and that offers Affymetrix, Illumina, NimblGen, and Sequenom platforms that support gene and MicroRNA expression profiling and genotyping analysis. The Microarray Resource also offers Illumina's Infinium Assay, the HumanMethylation27 DNA Analysis BeadChip that allows researchers to interrogate 27,578 highly informative CpG sites per sample at single-nucleotide resolution. In addition, this Resource offers Sequenom's MassARRAY EpiTYPER assay for the quantitative analysis of DNA methylation and identification of differentially methylated CpG sites in any genomic region or candidate gene of interest. This technology combines base-specific enzymatic cleavage with MALDI-TOF mass spectrometry. The Microarray Resource was strengthened substantially by the NIH Center Grant (U24 NS051869, T.C. $6,525,884, PI Shrikant Mane) that funded the Yale Microarray Center for Research on the Nervous System (2005-2010) and by an NIH/NCRR High End Shared Instrumentation Grant (2008, PI: Shrikant Mane, DC: $1,015,809) that brought the first Next Generation Genome Sequencer to Yale University. In 2010 the Keck Microarray Resource was merged into the Yale Center for Genome Analysis (YCGA) at Yale's West Campus to create a centralized facility for comprehensive, cutting-edge genomic services. The Microarray Resource offers a wide range of services that include:

1) Gene expression analysis using Affymetrix, Illumina and in house spotted arrays for both whole genome as well as custom applications.
2) SNP genotyping analysis using Affymetrix, Illumina, and Sequenom platforms.
3) Advanced biostatistical analysis using GeneSpring, dChip, RMA, GProcessor, Partek, XRAY, Bioconductor, and MetaCore pathways analysis packages.

The Microarray Resource is actively involved in the biotechnology research needed to evaluate and optimize the application of DNA microarray technology to biomedical research. The Resource also provides advice on experimental design, manuscript and grant preparation, and works in close collaboration with the Keck Biostatistics, Bioinformatics, and HPC Resources.

The Keck Laboratory Is Very Closely Associated with Several NIH and Other Centers

The value of the Keck Laboratory's centralized instrumentation and expertise has been very positively leveraged by the awarding of center and other grants that provide funding for biotechnological research, subsidized access for many investigators (e.g., in the Neuroproteomics Center) to new technologies, and state-of-the-art instrumentation. Since 1981 Keck Laboratory has been awarded 25 NIH/NSF Shared Instrumentation and five Center grants and contracts. As new and improved technologies and databases are developed in these Centers, they are rapidly published, offered as services, and made available to the hundreds of users of the Keck Lab at hundreds of institutions – thus deriving the maximum possible benefit from the funding. Examples include development and implementation of many new protein profiling technologies (e.g., the Rat/Mouse Brain LC-MRM assay for 112 proteins in Post Synaptic Density (PSD) and two major institutional databases [with one for protein expression (YPED) and the other for mRNA expression data (Yale Microarray Database, YMD)] built in collaboration with the Yale Center for Medical Informatics. Having multiple proteomics centers co-located greatly benefits the Keck Laboratory by bringing together a larger nucleus of mass spectrometrists and proteomics specialists that synergistically benefits all, as there are more staff immediately available to share and contribute ideas and solutions to help quickly and efficiently solve the often complex problems presented by the research of users of the Keck Laboratory. Centers associated with the Keck Laboratory include:

- Yale Center for Clinical Investigation (http://ycci.yale.edu/)
  - Established from an NIH clinical research initiative, the Clinical and Translational Science Awards (CTSA). Yale was among 12 other national institutions to receive the first of these awards.
  - CTSA mission is to bring new treatments more efficiently and quickly to patients by:
    - Encouraging development of new methods and approaches to clinical and translational research

    - ♦ Improving training and mentoring to ensure that investigators are able to take maximum advantage of advances in research biotechnologies
    - ♦ Designing new and improved clinical research informatics tools
    - ♦ Assembling interdisciplinary teams that cover the complete spectrum of medical research
    - ♦ Forging new partnerships with private and public health care organizations
  - – The Keck lab maintains close partnership with CTSA as evident by CTSA support for Keck instrument purchases to improve clinical research. CTSA has provided >$1.0 million dollars in equipment funding to purchase an AB SCIEX 4000 QTRAP, (2) Waters nanoACQUITY capLC systems, a Sequenom Mass Array system, and 50% of the funding needed in 2012 to purchase an AB SCIEX 5600 Triple TOF instrument. The AB SCIEX 4000 QTRAP enabled the MS/Proteomics Resource to carry out its first LC-MRM analyses and the AB SCIEX 5600 Triple TOF is an integral part of the current Targeted LC-MRM Proteome Analysis workflow that resulted in the Mouse/Rat Brain Proteome Assay that interrogates the level of expression of 112 proteins from 24 different protein classes: http://keck.med.yale.edu/proteomics/technologies/targetedproteomics/TargetedProteomeAssays.aspx

- ➢ Northeast Biodefense Center (NBC) Proteomics Core (http://www.nbc.columbia.edu/proteomics.html)
  - – One of 8 Regional Centers of Excellence established in 2003.
  - – NBC has 26 member institutions that support research in 5 major areas.
  - – NBC Proteomics Core, which is within the Keck Laboratory's MS/Proteomics Resource, supports basic science, clinical biodefense research programs, and first responders in the event of a biodefense emergency.

- ➢ Yale Cancer Center (YCC) shares and co-sponsors the Microarray Resource with the Keck Laboratory.
  - – A National Cancer Institute (NCI) designated comprehensive cancer center for over 38 years, the Yale Cancer Center is one of only 41 Centers in the nation

Yale Center for Genome Analysis (http://ycga.yale.edu/**)**

The Yale Center for Genome Analysis (YCGA) is a full service facility that was launched in 2010 on Yale's West Campus and that is dedicated to providing a broad range of microarray and high-throughput DNA sequence analysis services. Incorporation of the Keck Microarray Resource (see above) into YCGA created a centralized facility for cutting-edge genomic services. YCGA provides RNA expression profiling, DNA genotyping, and high-throughput sequencing analyses as well as the biostatistical and bioinformatics support needed to fully interpret genomics data on a fee-for-service basis. The Center serves both Yale and non-Yale investigators. YCGA has substantially expanded Yale's capacity for sequence generation and analysis, bringing further depth to Yale's expertise in the field of genomics. YCGA provides microarray and high-throughput DNA sequence analysis services using technologies that include Affymetrix, Illumina, Pacific Bioscience, Roche/454, Sequenom, Ion torrent, and Nimblegen. YCGA, which is one of the premier genome centers in the country, occupies 7,000 sq. ft. of custom renovated laboratory space and is equipped with 10 Illumina HiSeqs, and one each of MiSeq, PacBio-RS, and Ion Torrent sequencing systems. During CY 2012 YCGA carried out more than 4,700 sequence analyses for 60 Yale and 23 non-Yale investigators from 21 institutions including Harvard, Duke, and Boston Universities, Broad Institute, UCLA, and the Universities of Chicago, Wisconsin, Florida, and Connecticut. YCGA supports multiple sequencing applications including whole exome and targeted sequencing, transcriptome analysis (mRNA-Seq), DNA-protein interactions (ChIP-Seq), DNA methylation analysis (methyl-Seq), microRNA analyses and *de novo* sequencing. The genomic core staff have over five years of experience and are very well trained in sample processing as well as in the operation and maintenance of the equipment. All necessary infrastructures such as high performance computing, an adequate data storage facility, and bioinformatics support are already in place. YCGA has developed a WIKI LIM System for tracking sample submission and progress, billing, and it also enables users to view raw data as well as archive the final output files.

High Performance Computing (HPC): Genomics data generated in YCGA is transferred to a dedicated, high performance cluster for further analysis. Data is retained on the cluster for at least 18 months, after which it is transferred to a tape backup system managed by Yale's Information Technology Services (ITS). The cluster runs on a Linux operating system and it consists of 140 nodes with approximately 1200 cores/CPUs and approximately 2.5 Petabytes of high performance parallel storage (Panasas Inc.). All machines are connected via gigabit Ethernet. YCGA's HPC instrumentation is housed in a secure, climate-controlled room with hardware and software support provided by Yale ITS. Two Ph.D. computer scientists and one M.S. level staff support the IT and High Performance Computing needs of the Center.

Biostatistics and Bioinformatics: Technical assistance in the analysis of data generated at YCGA is provided by four Ph.D.-level scientists and the Keck Laboratory's Bioinformatics and Biostatistics Resources. The bioinformatics staff have extensive experience developing new programs for the analysis of data generated by microarray as well as by next gene sequencing platforms and they are actively involved not only in carrying out high level analyses but also in developing new algorithms and data analysis tools.

YCGA has emerged as one of the leaders in the field of identification of disease-associated genetic factors, as evidenced by the first successful report with genome-wide association studies, identifying genes impacting the risk of developing age related macular degeneration. This initial landmark success was followed by the identification of genetic factors associated with cardiovascular disease, inflammatory bowel disease, intracranial aneurism, SeSAME syndrome, and Tourette syndrome. The Center's direct impact on medicine is evidenced by a recent high-profile discovery that allowed the first clinical diagnosis and subsequent tailored treatment based on whole-exome sequencing. Since 2010 the Center has >85 publications with many in highly regarded journals.

The Yale Center for Mendelian Genomics (http://www.mendelian.org/) , which is one of three national centers created by the NIH to study the genetics of rare inherited diseases, is located within YCGA. Together with researchers at the University of Washington, and a center operated jointly by Baylor and Johns Hopkins University; YCGA is analyzing the genomes of thousands of patients who suffer from more than 6,000 rare Mendelian disorders affecting more than 25 million individuals in US. The Yale Center for Mendelian Genomics received $11.2 million under this four-year program. Exome sequencing and analysis on qualified phenotypes is carried out on a collaborative basis at no cost to YCGA investigators. The principal investigators for the Yale Center are Richard Lifton, M.D., Ph.D., Chair and Sterling Professor of Genetics, HHMI investigator; Murat Gunel, M.D., the Nixdorff-German Professor of Neurosurgery and professor of genetics and neurobiology; Shrikant Mane, Ph.D., Director of Yale Center for Genome Analysis and Co-Director of the Keck Foundation Biotechnology Resource Laboratory; and Mark Gerstein, Ph.D., the Albert L. Williams Professor of Molecular Biophysics and biochemistry and Co-director of the Yale Computational Biology and Informatics. The Center for Mendelian Genomics applies next-generation sequencing and computational approaches to discover the genes and variants that underlie Mendelian disorders. The discovery of new genes that cause Mendelian conditions expands our understanding about the biology of these diseases, facilitates their diagnosis, and potentially opens the way to developing new treatments.