AAIM Perspectives

AAIM is the largest academically focused specialty organization representing departments of internal medicine at medical schools and teaching hospitals in the United States and Canada. As a consortium of five organizations, AAIM represents department chairs and chiefs; clerkship, residency, and fellowship program directors; division chiefs; and academic and business administrators as well as other faculty and staff in departments of internal medicine and their divisions.

Impact of an Interdisciplinary Computational Research Section in a Department of Medicine: An 8-Year Perspective

CrossMark

Avrum Spira, MD,[a,b] **David Coleman, MD**[c]

[a]*Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Mass;* [b]*Boston University-Boston Medical Center Cancer Center;* [c]*Department of Medicine, Boston University School of Medicine, Mass.*

INTRODUCTION

The evolving opportunities in biomedical research require that medical schools and departments continually review their respective organizational structures to optimize discovery and training. The traditional discipline-based organizations are not sufficient to stimulate the full range of new and interdisciplinary approaches to successfully address medically important questions.[1] Consequently, institutions have developed centers, institutes, cross-disciplinary departments, and programs aimed at promoting convergence in biomedical science through partnerships within the academic community and with industry.[2,3] The International Campaign to Revitalize Academic Medicine called for reinvention of academic medicine, in part, by developing emerging technology across disciplines.[4]

Departments of internal medicine have a major stake in the success of interdisciplinary research programs and are uniquely positioned to accelerate the application of new research technology targeted at human biology and disease. Importantly, the application of "personalized medicine" to diagnosis and treatment of a broad range of human diseases increasingly requires that academic clinical departments develop expertise in the application of computational and analytical methods to large "omic" datasets. The acceleration of new methods to generate and analyze genomic datasets created an important opportunity to assemble a critical mass of investigators with computational, analytical, and clinical expertise. Moreover, existing organizational structures within and outside the department did not sufficiently integrate the scientific and training opportunities with clinical medicine. Therefore, in 2009, the department of medicine at Boston University School of Medicine created the Section of Computational Biomedicine.[5]

We report herein on the 8-year follow-up of the impact of the computational biomedicine section. We will describe the scientific and translational deliverables, the challenges faced in sustaining this initiative, the factors that contributed to its success, and the impact on the broader university. On the basis of our experience, we believe that clinical departments, especially departments of internal medicine, have a vital role in catalyzing application of computational methods to medical problems and in addressing new opportunities for inter-

disciplinary biomedical research in the modern research university.

FOUNDING AND EVOLUTION OF THE SECTION OF COMPUTATIONAL BIOMEDICINE

The Department of Medicine at Boston University and Boston Medical Center has a long-standing tradition of creating research sections to address emerging research opportunities, including clinical epidemiology, preventive medicine, biomedical genetics, and vascular biology. In 2009 the accelerating growth of high-throughput molecular profiling technology, the opportunity to better understand diagnostic and disease pathways through the delineation of new biomarkers, and the emergence of an extraordinary interdisciplinary research team with computational analytic expertise converged to create a unique opportunity for a new interdisciplinary section in the department of medicine—the section of computational biomedicine.

> **PERSPECTIVES VIEWPOINTS**
>
> - Given the recent emergence of high-throughput molecular technologies and their application to clinical samples, there is an urgent need for clinical departments to develop computational infrastructure and educational programs to accelerate the translation of precision genomic medicine approaches into the clinic.
>
> - Our experience developing a section of Computational Biomedicine can serve as a roadmap for other academic medical centers that wish to accelerate development of personalized approaches to disease detection and treatment.

The goals of the computational biomedicine section were to provide a rich training environment and academic home for an interdisciplinary research group of MD and PhD faculty with expertise in bioinformatics, biostatistics, computational biology, and clinical medicine, with special emphasis on the analysis of whole-genome gene expression data. The concentration of expertise and equipment necessary to undertake these studies was intended to catalyze approaches to the identification of targets for prevention, diagnosis, and therapy. The initial focus of the group was on the characterization of gene expression signatures in cancer and precancerous conditions. The department invested $4.1 million to recruit faculty and learners, as well as support critical capital investments in sequencing and computational facilities. Boston University also awarded an endowed professorship to the section chief. In 2014 the space for the section was renovated by the school of medicine to facilitate interactions between computational and biologic investigators.

The interdisciplinary group of computational biomedicine faculty provides a remarkable training and collaborative environment. The section includes 12 faculty and 20 pre- and postdoctoral fellows. The faculty include 2 MDs and 10 PhDs who come from diverse academic backgrounds: bioinformatics, artificial intelligence, pulmonary medicine, biostatistics, genetic epidemiology, biology, and computer science. Five faculty were recruited from outside the institution and 7 from within the university. The fellows come from sections in the department of medicine as well as the bioinformatics, pathology, and biostatistics departments.

The computational biomedicine section faculty have rapidly responded to important opportunities for discovery and extramural funding. For example, the section has secured $26.2 million in federal grant support since 2009. Importantly, the section has developed productive ties to industry that have led to $16.2 million in sponsored research agreements. The section also has strong collaborative interactions, including joint grants, with 25 faculty from sections within and outside the department of medicine. The section's research findings have led to 11 provisional patent filings and issuing of 4 patents. To translate the discoveries made by the section into the clinic, investigators within the section have founded 4 companies that have collectively raised more than $20 million in financing.

The training mission of the computational biomedicine section has been a particularly important goal. A total of 65 pre- and postdoctoral fellows have received their primary research training in the section since 2009. Because learners must be prepared for interdisciplinary research teams in the future, the section encouraged the involvement of learners in all projects. In addition to the influence on pre- and postdoctoral students, medical residents and fellows participated in the training opportunities offered by the computational biomedicine section. The enthusiasm of the residents and fellows for computational training led to the formation of a genomic medicine research pathway in the department's internal medicine residency training program.

As Boston University and Boston Medical Center evaluated the direction and leadership of their joint cancer center, the emerging contributions of the computational biomedicine section were believed to be remarkably well aligned with the university's strategic goal to use bioinformatics to map gene expression and other molecular signatures to develop precision medicine approaches to prevent, intercept, and treat cancer. Therefore, the computational biomedicine section chief was appointed director of the Boston University-Boston Medical Center Cancer Center—a cross-campus collaboration to lead integration of the interdisciplinary expertise developed within the section with the cancer center's

programs and cores. The section faculty are also playing a leadership role in a newly formed university-wide microbiome initiative, as well as developing a new research collaborative effort in environmental carcinogenicity.

TRANSLATIONAL RESEARCH DELIVERABLES OF THE SECTION

Over its first 8 years the computational biomedicine section has developed a number of biomarkers and therapeutics (**Table 1**) as well as computational tools (**Table 2**) that are being used by the academic community or being translated into health care products by industry. A signature program developed by faculty in the section has been the creation of the OpenSESAME software.[20] Developed together with the BU Clinical and Translational Science Institute, this "Google for gene expression" database and software allows users to match a list of genes from their gene expression study with all other publically available whole-genome gene expression datasets in which those same genes are co-expressed. This opportunity can provide novel insights into the biological relevance for a given gene set, connect seemingly unrelated disease entities at the molecular level, and importantly, enable drug repurposing by matching gene expression signatures of drug treatments and disease states.[11]

   Pathoscope is a second example of a computational biomedicine software package that has had a significant impact on the biomedical genomics community. This software allows for rapid alignment of sequencing reads to microbial genomes, enabling the identification of pathogens that may be present in a variety of biospecimens.[24] The section has also developed GeneHive, a user-friendly research data warehouse that allows investigators to easily store, organize, retrieve, and share genome-scale data in a secure manner. Although this tool is currently used to manage genomic expression profiling data generated by all investigators within the section, it is available as an open-source software package that is easy to deploy to a local server or virtual machine at any hub within medical schools. A related software package, *rabbit*, has recently been developed as a modular tool for the development of robust biomarkers from genome-scale data stored within GeneHive.

   Although development of new computational genomic tools is an important component of the section, the primary mission of the section is to translate genomic tool into the clinic to impact disease diagnosis and treatment (**Table 1**). As such, investigators within the section have established and leveraged a number of large clinical cohorts in which genomic biomarkers are discovered and validated within their intended use populations. The most striking example is the development and validation of an airway gene expression biomarker for the early detection of lung cancer, a diagnostic tool that has successfully translated from bench to bedside (Percepta).[15] This bronchial genomic classifier recently received Medicare approval and is being used in more than 50 hospitals within the United States to reduce unnecessary invasive procedures in smokers with pulmonary lesions on chest imaging that are suspicious for lung cancer. Investigators within the section recently extended this biomarker from the bronchial epithelium to cells that line the nose, enabling less invasive sampling of the airway for early lung cancer detection.[14]

   There are a number of other deliverables that are moving toward translation into the clinic. Investigators in the computational biomedicine section have developed airway gene expression biomarkers for characterizing

Table 1   Biomarkers and Therapeutic products of the CB section

Product	Description	Reference(s)	Patent(s)
Airway Biomarker for Premalignant Lesions	An airway gene expression biomarker for detecting and predicting progression of premalignant lung cancer lesions	6	Patent pending 15/ 644,721
Breast Cancer Susceptibility	A blood-based biomarker for predicting breast cancer susceptibility in high-risk populations	7,8	
COPD Airway Gene Expression Biomarkers	A bronchial airway gene expression biomarker of COPD to guide disease management	9,10	9,677,138
GHK Tripeptide	Discovery of a novel therapeutic for COPD based on lung tissue gene expression	11	9,585,930
Growth Signaling Networks Pathways	Gene expression pathway signatures for oncogenic growth, survival, and apoptotic pathways	12	
Molecular Subclass of Lymphoma	A transcriptional biomarker for the stratification of therapeutically relevant diffuse large B-cell lymphomas subtypes	13	
Nasal Genomic Classifier	A nasal gene expression biomarker for early lung cancer detection among ever-smokers with suspect lung cancer	14	Provisional patent 62/ 335,391
Percepta	Bronchial genomic classifier for lung cancer detection that reduces unnecessary invasive procedures in ever-smokers with pulmonary lesions on chest imaging	15,16	9,920,374

COPD = chronic obstructive pulmonary disease.

Table 2 Computational Tools and Software Developed by the Computational Biomedicine Section

Computational Tool	Description/Application	Reference(s)	Website
ASSIGN	Adaptive Signature Selection and InteGratioN	Bayesian factor regression model to identify genomic biomarkers for applications in pathway profiling, drug responsiveness, environmental exposure, and infectious disease diagnosis	[17]	
BatchQC	BatchQC is a user interface for interactive evaluation of batch effects in -omic data	[18]	
CaDrA	*Ca*ndidate *Dr*iver *A*nalysis	An R tool for the identification of cancer drivers from the integrative analysis of multi-omics data		www.github.com/ montilab/CaDrA
GeneHive	User-friendly research data warehouse that allows investigators to easily store, organize, retrieve, and share genome-scale data in a secure manner		
GNUMAP	Genomic Next-generation Universal MAPper	Software suite for aligning next sequencing data from DNA-seq, BS-seq, and RNA-seq (including small RNAs, RNA editing) using a probabilistic alignment approach	[19]	
Hydra	A flexible pipeline for the parallel processing and QC-ing of RNA sequencing data		www.github.com/ montilab/hydra
iEDGE	*i*ntegrative (*E*pi)*D*NA and *G*ene *E*xpression	An R package and html interface for the analysis of multi-omics datasets		www.github.com/ montilab/iEDGE
ISR	Tool to Infer Splicing Regulator from transcriptome data	Unpublished data	github.com/montilab/ISR
openSESAME	A search engine for gene expression data, which allows users to identify publicly available whole-genome gene expression datasets in which a set of genes are co-regulated in the same manner as in their own gene expression study	[20]	http://opensesame .bu.edu
PathoScope/PathoStat	Rapid and Accurate Sequence-based Infectious Disease Diagnostics	PathoScope is a complete bioinformatics framework for the metagenomic analysis of data from clinical or environmental sequencing samples. PathoStat is a graphic user interface for multi-sample metagenomic data analysis	[21-24]	
rabbit	Modular tool for the development of robust biomarkers from genome-scale data		
SCAN-UPC	Single Channel Array Normalization and Universal Probability Codes	SCAN is a microarray normalization method that removes background noise using only data from within each array individually, and UPC produces barcode that estimates gene activity in data from microarray and RNA sequencing platforms	[25,26]	
The Carcinogenome Project	A Web-based portal for querying chemical carcinogens' transcriptional signatures and modes of actions.		carcinogenome.org
Cellular Latent Dirichlet Allocation (Celda)	A discrete Bayesian hierarchical model that simultaneously clusters co-expressed genes into transcriptional states and cells into subpopulations.		https://github.com/ compbiomed/celda

QC = quality control; UPC= universal probability codes.

molecular phenotypes of chronic obstructive pulmonary disease to guide disease management.[9,10] Additionally, they have leveraged gene expression signatures of emphysema progression to reposition the tripeptide GHK as a novel therapeutic for chronic obstructive pulmonary disease.[11] This therapeutic and the associated companion diagnostics are being further developed by a university start-up company before entering early-phase clinical trials. Investigators have also discovered molecular subclasses of diffuse B-cell lymphomas that

hold the potential to personalize their treatment.[13] Furthermore, section faculty have supported the development of a blood-based genomic biomarker for predicting breast cancer susceptibility in high-risk populations.[7,8]

LESSONS FROM THE COMPUTATIONAL BIOMEDICINE INITIATIVE

The section is a successful example of how interdisciplinary research initiatives based in the department of

medicine can stimulate new research collaborations within the department as well as more broadly across a medicine school and university. The success of the section led to a strategic and programmatic integration with the cancer center. The initiative benefited from an inclusive leadership approach, reaching across boundaries of traditional disciplines and departments; investment in high-risk and high-reward initial studies; creation of a robust instrumentation infrastructure; and an unwavering commitment to mentoring and advising faculty and learners.

The training component of the section was particularly important in its success. The section became a very popular training site for graduate students in bioinformatics, postdoctoral fellows, medical residents, and subspecialty fellows. Moreover, the students and residents played a vital role in connecting interdisciplinary research groups through their own social and scientific networks. The section provided important examples of the potential impact of collaborations across disciplines on clinical problems. Importantly, the opportunities for individuals with strong computational and analytical backgrounds to impact clinical medicine are increasing dramatically. The training functions of the computational biomedicine section are intended to prepare clinician and nonclinician faculty to address this opportunity.

A key metric for the success of any interdisciplinary medical research initiative is the ability to translate discovery from the bench to the bedside and ultimately impact patient care. As such, collaborations with industry are a critical step toward overcoming the traditional "valley of death" for medical sciences. Using a number of models, the section has successfully partnered with biotechnology and pharmaceutical partners to facilitate the translational goals of their respective research programs. One of the key paradigm shifts in these models has been the engagement with industry partners early in the discovery process (as opposed to licensing existing technology). The section has been particularly successful in developing substantial collaborations with partners from biomedical companies to develop diagnostic and therapeutic biomarkers for immune-related and malignant diseases. Further, the department and university have provided support to faculty to enable the startup of a number of companies, which has already led to the commercialization of the early detection biomarker for lung cancer described.[15] Importantly, the institutional Conflict of Interest Committee was very helpful in the management of the potential conflicts of interest that arose in the relationships with industry.

CHALLENGES AND BROADER APPLICATION OF THE COMPUTATIONAL BIOMEDICINE INITIATIVE

The development of the computational biomedicine section encountered important practical challenges. One critically important challenge was securing the resources necessary to fund the section. The financial investment in the section was nearly $5,000,000 over an 8-year period. The investment reflected a strategic decision by the department to invest in new high-risk, high-reward interdisciplinary research, in addition to investments in traditional organ-based sections. The successful involvement of many faculty across the department and the extraordinary productivity that arose from a relatively modest investment represents a very strong return on investment. For example, the sum of grants and contracts secured over the 8-year history of the section to date reflects an approximate 10:1 yield on the department's investment.

The Department of Medicine at Boston University School of Medicine has access to endowment funds for investment in initiatives such as the computational biomedicine section. However, other academic departments of internal medicine with less flexible resources or different revenue sources can use the organizational principles and approach described to catalyze interdisciplinary research and application of new research methods to clinical problems. Many academic departments invest in their respective research programs through startup packages, bridge funding, instrumentation, cores, and pilot grants for their faculty. These funds typically originate from endowments, royalty income, reserve funds, support from the school or heath center, and to a modest and declining degree, from clinical revenue. On the basis of our experience, the research programs of the department benefit enormously from diversion of a portion of these resources to new interdisciplinary research structures that respond to new scientific and clinical opportunities. For example, departments could create programs in computational biomedicine through a few key recruitments that enucleate collaborations among sections and departments across the university. In essence, supplementing, and in some cases partially replacing, the research program development of traditional organ-based sections with interdisciplinary approaches to clinical problems that connect many disciplines from across the university can create a "win-win" for departments of internal medicine and the university.

SUMMARY

The need to assemble rigorous and "user-friendly" tools for the analysis of large genomic datasets has emerged as a critical determinant in personalized approaches to medical care. Departments of internal medicine are central to the development and application of new insights derived from this emerging technology. In response, the computational biomedicine section was established to accelerate the development of interdisciplinary computational approaches to clinical problems in the department of medicine. The section combined faculty with expertise from a broad range of analytical, clinical, and

bioinformatics backgrounds. The section was developed with collaborative expertise of faculty across campuses, without reorganizing existing departments.[27] The section's remarkable success has led to durable increases in the research and training capacity in and outside the department of medicine. In addition, the section has been able to establish highly productive collaborations with industry that are enabling the scientific findings to impact patient care. Importantly, the section has become a highly sought-after training site for graduate students, postdoctoral fellows, and medical residents and fellows. As such, the computational biomedicine section represents an important roadmap for the interdisciplinary application of emerging technologies to clinical problems that benefits both the department of medicine as well as the broader university.

ACKNOWLEDGMENTS

References

1. Hall P, Weaver L. Interdisciplinary education and teamwork: a long and winding road. *Med Educ*. 2001;35(9):867-875.
2. Sharp P, Jacks T, Hockfield S. Biomedical engineering. Capitalizing on convergence for health care. *Science*. 2016;352(6293):1522-1523.
3. Sharp PA, Langer R. Research agenda. Promoting convergence in biomedical science. *Science*. 2011;333(6042):527.
4. Awasthi S, Beardmore J, Clark J, et al. Five futures for academic medicine. *PLoS Med*. 2005;2(7):e207.
5. Coleman DL, Spira A, Ravid K. Promoting interdisciplinary research in departments of medicine: results from two models at Boston University School of Medicine. *Trans Am Clin Climatol Assoc*. 2013;124:275-282.
6. Beane J, Mazzilli SA, Tassinari AM, et al. Detecting the presence and progression of premalignant lung lesions via airway gene expression. *Clin Cancer Res*. 2017;23:5091-5100.
7. Piccolo SR, Andrulis IL, Cohen AL, et al. Gene-expression patterns in peripheral blood classify familial breast cancer susceptibility. *BMC Med Genomics*. 2015;8:72.
8. Piccolo SR, Hoffman LM, Conner T, et al. Integrative analyses reveal signaling pathways underlying familial breast cancer susceptibility. *Mol Syst Biol*. 2016;12(3):860.
9. Steiling K, van den Berge M, Hijazi K, et al. A dynamic bronchial airway gene Expression signature of COPD and lung function impairment. *Am J Respir Crit Care Med*. 2013;187(9):933-942.
10. van den Berge M, Steiling K, Timens W, et al. Airway gene Expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity. *Thorax*. 2014;69(1):14-23.
11. Campbell JD, McDonough JE, Zeskind JE, et al. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome Med*. 2012;4(8):67.
12. Rahman M, MacNeil SM, Jenkins DF, et al. Activity of distinct growth factor receptor network components in breast tumors uncovers two biologically relevant subtypes. *Genome Med*. 2017;9(1):40.
13. Monti S, Savage KJ, Kutok JL, et al. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood*. 2015;105(5):1851-1861.
14. Perez-Rogers JF, Gerrein J, Anderlind C, et al. for the AEGIS Study Team. Shared gene expression alterations in nasal and bronchial epithelium for lung cancer detection. *J Natl Cancer Inst*. 2017;109(7).
15. Silvestri GA, Vachani A, Whitney D, et al. A bronchial genomic classifier for the diagnostic evaluation of lung cancer. *N Engl J Med*. 2015;373(3):243-251.
16. Whitney DH, Elashoff MR, Porta K, et al. Derivation of a genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy. *BMC Med Genomics*. 2015;8:18.
17. Shen Y, Rahman M, Piccolo SR, et al. Assign: context-specific genomic profiling of multiple heterogeneous biological pathways. *Bioinformatics*. 2015;31(11):1745-1753.
18. Manimaran S, Selby HM, Okrah K, et al. BatchQC: interactive software for evaluating sample and batch effects in genomic data. *Bioinformatics*. 2016;32(24):3836-3838.
19. Hong C, Clement NL, Clement S, et al. Probabilistic alignment leads to improved accuracy and read coverage for bisulfite sequencing data. *BMC Bioinformatics*. 2013;14:337.
20. Gower AC, Spira A, Lenburg ME. Discovering biological connections between experimental conditions based on common patters of differential gene expression. *BMC Bioinformatics*. 2011;12:381.
21. Francis OE, Bendall M, Manimaran S, et al. Pathoscope: species identification and strain attribution with unassembled sequencing data. *Genome Res*. 2013;23(10):1721-1729.
22. Byrd AL, Perez-Rogers JF, Manimaran S, et al. Clinical PathoScope: rapid alignment and filtration for accurate pathogen identification in clinical samples using unassembled sequencing data. *BMC Bioinformatics*. 2014;15:262.
23. Hong C, Manimaran S, Johnson WE. PathoQC: computationally efficient read preprocessing and quality control for high-throughput sequencing data sets. *Cancer Inform*. 2015;13(suppl 1):167-176.
24. Hong C, Manimaran S, Shen Y, et al. PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome*. 2014;2:33.
25. Piccolo SR, Sun Y, Campbell JD, Lenburg ME, Bild AH, Johnson WE. A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics*. 2012;100(6):337-344.
26. Piccolo SR, Withers MR, Francis OE, Bild AH, Johnson WE. Multiplatform single-sample estimates of transcriptional activation. *Proc Natl Acad Sci U S A*. 2013;110(44):17778-17783.
27. Mallon WT, Biebuyck JF, Jones RF. The reorganization of basic science departments in U.S. Medical Schools, 1980-1999. *Acad Med*. 2003;78(3):302-306.