MICROBES, SEWAGE, HEALTH AND DISEASE: MAPPING THE NEW YORK CITY METAGENOME

A Proposal for the NYU Grand Challenges Program

Lead PIs

NYU Center for Genomics & Systems Biology: Dr. Richard Bonneau, Co-Director of Bioinformatics
Dr. Jane Carlton, Director of Sequencing
Dr. Patrick Eichenberger, Associate Professor

NYU Center for Urban Science and Progress: Dr. Steven Koonin, Director
Dr. Ari Patrinos, Deputy Director for Research
Dr. Cláudio Silva, Head of Disciplines

NYU School of Medicine: Dr. Martin Blaser, Director Human Microbiome Program

NYU Global Institute of Public Health: Dr. Cheryl Healton, Director of GIPH and Dean of Global Public Health

Other Key Team Members

NYU Center for Genomics & Systems Biology: Mr. Paul Scheid, Manager, Genomics Core Facility

Univ. of Pittsburgh School of Medicine: Dr. Elodie Ghedin, Associate Professor, Dept. of Computational and Systems Biology

Contact information

NYU CGSB
Carrie Nygard
Center for Genomics & Systems Biology
Department of Biology
12 Waverly Place
New York, NY 10003
p: (212) 998-8341
e: cen249@nyu.edu

NYU SoM
Joyce Ying
Division of Translational Medicine
NYU Langone Medical Center
550 First Avenue BCD 689-690
New York, NY 10016
p: (646) 501-2321
e: joyce.ying@nyumc.org

NYU GIPH
Amanda Garofalo
Global Institute for Public Health
Office of the Executive Vice President for Health
240 Greene Street, Floor 2
New York, NY US 10003
p: (212) 992-2051
e: amanda.garofalo@nyu.edu

NYU CUSP
Arya Tafvizi
Center for Urban Science and Progress
1 Metro Tech Center 19FL
Brooklyn, NY 11201
p: (646) 997-0515
e: arya.tafvizi@nyu.edu

I. The Concept, Scientific Basis and Grand Challenge Nature of Mapping the New York City MetaGenome

The Concept

New York City is the most populous and cosmopolitan city in the United States, and a leading world metropolis. What happens in New York has global impact on finance, media, art, science, education, technology and entertainment. The city is home to over 8 million people on a land mass of ~300 sq miles (**Figure 1**), is the major international air gateway to the U.S, operates one of the nation's most extensive mass transit systems, and pumps 1.5 billion gallons of wastewater per day through its 7,400 miles of sewer conduits. This vast urban ecosystem is a precious resource that requires monitoring to sustain and secure it against, for example, acts of bioterrorism and environmental or epidemic threats.

We propose to use the new science of **metagenomics** (the application of genomics to the study of microbes in their natural environments) to describe, characterize and ultimately track **beneficial** and **infectious** microbial communities in NYC. Microbial populations are thought to comprise up to a billion species world-wide. The human gut alone can contain up to ~1,000 different bacterial species whose composition varies enormously according to diet and health status. Metagenomic science has exploded in recent years thanks to rapid advances in DNA sequencing technology ("**Next Generation Sequencing**", **NGS**) along with the development of sophisticated bioinformatics tools and so-called **Big Data** storage and manipulation. Metagenomics is providing insights into the life-cycles and functions of individual microbes and microbial communities that address issues in **healthcare**, **energy**, and the **environment**. By characterizing the **New York City MetaGenome** – literally **surveying the genetic material of microbes present in NYC** -- we aim to identify potential bio-threats, protect the health of New Yorkers, and provide an additional level of data that can be analyzed by the city to optimize its workings, making it a "**smart city**" [1].

Figure 1. The five boroughs of New York City. 1. Manhattan, 2. Brooklyn, 3. Queens, 4. The Bronx, 5. Staten Island. The three international airports serving New York City are indicated: North East: Laguardia Airport (LGA); South East: John F. Kennedy Airport (JFK); and West: Newark Airport (EWR).

The Scientific Basis

There have been few urban metagenomics studies. Most have been focused indoors, although at least one was an "**air genome**" project that collected and analyzed airborne DNA above New York City (J. Craig Venter Institute, *unpublished*). Another recent study sampled the microbial population in the **air of the New York City subway system** [2]. Bioaerosols on several New York City subway platforms were collected in three sampling sessions over an 18 month period, and the types and quantities of aerosolized microorganisms were determined by culture-independent analysis of 16S rRNA gene sequences. The subway bacterial composition was found to be simple, with only 26 taxonomic families contributing to ~75% of the sequences determined, a composition that remained stable over time and similar to outdoor air, as predicted given the highly efficient subway air circulation system. A mix of soil, environmental water, and human skin bacteria (*Staphylococcus* spp., especially *Staphylococcus epidermidis,* the most common commensal of the human integument) were identified, along with mainly fungal eukaryotes, but no organisms of public health concern.

These exploratory studies affirmed the feasibility of a more extensive and comprehensive project using NGS that could yield **deep insights and practical societal implications for NYC.** For example, using a combination of metagenomics, Internet search engines, and global positioning/global information systems, it should be possible to map, track and most importantly **predict flu epidemics** like the one that gripped NYC

during the Winter of 2012/2013. We will also be able to determine the immediate or long-term effects on the environment and microbial footprint of dramatic events such as **9/11 and Superstorm Sandy**.

Metagenomics has created an entirely new way of characterizing and studying microbial ecosystems, whether in **farms**, **acid mines**, **oceans**, **hospitals**, or **human bodies**. Along the way, millions of new microbial genes have been discovered and added to public databases. There are significant challenges to a successful metagenomics project compared with genome-based analysis of a single microbe. Much greater attention needs to be paid to **sampling techniques** in order to assess the true microbial diversity of the ecosystem. Other challenges will be the sheer volume of data, and discerning signal from noise. But there will be multiple payoffs to the success of our proposed study. From a basic science standpoint, establishing a **metagenomics baseline for New York City** and quantifying variability across time and space will place past and future findings in an evolutionary, biochemical, and physiological context. From a **public health and environmental protection** standpoint, we will have new ways to spot trends and shifts that can serve as early warning signals. Some of the specific questions that we hope to be able to answer are:

1. **Can we monitor the spread of community-acquired antibiotic resistance?**
2. **Can we detect agents of bioterrorism such as anthrax spores?**
3. **Can we determine the effects of storms such as Superstorm Sandy on sewage ecosystems?**
4. **Can we define the rat, cockroach and bed-bug population and distribution through sewage, and correlate with human disease?**
5. **Can we measure the NYC 'virome', and what will this teach us about circulating strains of influenza, norovirus, and bacteriophage?**
6. **Can we detect molds?**
7. **Can we take correlate the burden of microbes to pet animals (dogs, cats) in a catchment area?**

Mapping the New York City MetaGenome: A Grand Challenge

Mapping the New York City MetaGenome is an ambitious undertaking that qualifies as a "**Grand Challenge**" that will capture the public's imagination for several reasons. **First**, people are intensely curious about the silent hitchhikers – "bugs" or microbes – that live in, on and around us, and influence our daily lives. **Second**, surveying the microbial genetic material of New York City is by itself a mammoth undertaking, requiring cutting-edge approaches to sampling, data management and analysis. **Lastly**, there is the challenge of integrating this Big Data with other datasets to generate a spatial and temporal map of the NYC metagenome and use it to plan a smarter NYC.

If we are awarded seed funds, we will invest the seed funds in a pilot study that will sample the microbial diversity at four NYC Department of Environmental Protection sewer sites in each of the five NYC boroughs, in order to determine (1) if signatures of microbial diversity can be identified at each site and a metagenomics baseline determined; (2) if we can determine the immediate or long-term effects and microbial footprint of dramatic events on the environment; and (3) if we can spot trends, shifts and perturbations in the microbial patterns that can serve as early warning signals.

Ultimately our goal will be to: (1) Identifying and quantifying common microbes found in different ethnic populations and in different neighborhoods throughout NYC; (2) Identify and quantify the most common sets of microbes found in key migration choke points in the city (subway, train stations, airports); and (3) Identify specific patterns of microbial communities across space and time and their relationship to disease.

This is an exciting time for urban science and smart city planning. Our grand challenge is how to use key scientific developments in genomics and Big Data analysis to map the New York City Metagenome and make NYC -- a central gateway to the United States -- a more intelligent city.

II: Our Experimental Design, Technical Approach and Implementation Plan

Our Grand Challenge -- **to map the metagenome of New York City** -- is divided into distinct phases over a 10 year period. Since much of the work in subsequent years after the seed funding is depleted will depend upon the findings of the pilot project, and also upon future sources of funding, we outline these later phases in broad strokes below, while focusing attention on the experimental design and implementation plan for the two-year pilot project.

Phase 1: Microbes, Sewage, Health and Disease

Pilot Project, September 2013 – August 2015

Summary

This pilot project will set the stage for our future endeavours to map the NYC metagenome. Briefly, we plan to sample the sewage at four NYC Department of Environmental Protection (DEP) sewer sites in each of the five NYC boroughs, to determine (1) if signatures of microbial diversity can be identified at each site and a metagenomics baseline for New York City; (2) if we can determine the immediate or long-term effects and microbial footprint of dramatic events on the environment; and (3) if we can spot trends, shifts and perturbations in the microbial patterns that can serve as early warning signals. NYC DEP is one of the partners of NYU CUSP, allowing us unprecedented access to NYC's extensive system of sewers.

Specific Aims

Aim 1. To sample the sewage at four NYC DEP sewer sites in each of the five NYC boroughs
Aim 2. To develop standard protocols for sample extraction, genomic library preparation, and sequencing
Aim 3. To analyze the sequence data using Big Data pipelines and bioinformatics methods
Aim 4. To integrate and visualize the data through mapping against geographic and demographic information determined by the drainage area

Background

NYC sewage system

New York City is impressive in most respects, and its **sewage system** is no exception. The five boroughs are served by 7,400 miles of sewer pipes, 135,000 sewer catch basins (and their iconic curbside grates), 95 wastewater pumping stations, and 14 treatment plants. This system yearly handles over 1.3 billion gallons of wastewater from domestic, institutional, and commercial toilets and drains – so-called '**sanitary waste**' -- combined with untold billions of gallons of rainstorm runoff, which very occasionally overwhelms its capacity and leads to dumping of raw sewage into New York waterways. NYC sewage typically contains various proportions of wastewater from household and institutional toilets, bathing, laundry, food preparation, and dishwashing; dirt, trash, leaves, and animal wastes from street runoff after a storm; groundwater that has infiltrated the system through leaks in sewer system joints; and industrial waste liquids. Sewage contains large numbers of microbial organisms that feed on its organic and inorganic nutrients; bacterial action typically turns fresh domestic wastewater from gray and mildly odorous to black and foul-smelling. However, since the composition of sewage is constantly changing, the active microbial components also vary.

Components of NYC sewage: what should we be looking for?

Kingdom	Organism	In sewage?	In gut?	Kingdom	Organism	In sewage?	In gut?
bacteria	Bacillus cereus	NR	cockroach, human	archaea	Methanobrevibacter spp.	NR	human
bacteria	Bacteroides spp.	NR	human , equine	archaea	Methanosarcina spp.	Yes	NR
bacteria	Bdellovibrio bacteriovorus	Yes	mammal, avian	protist	Cryptosporidium spp.	Yes	canine
bacteria	Campylobacter spp.	NR	canine, human	protist	Entamoeba spp.	Yes	human
bacteria	Clostridium spp.	Yes	human, animal	protist	Giardia spp.	Yes	canine, human
bacteria	Enterobacter spp.	Yes	human, animal	virus	Coronavirus	NR	canine
bacteria	Enterococcus spp.	Yes	human, animal	virus	Enterovirus	Yes	human
bacteria	Escherichia spp.	Yes	human	virus	Parvovirus	NR	canine
bacteria	Klebsiella pneumoniae	Yes	human	virus	Rotavirus	Yes	human
bacteria	Lactobacillus spp.	Yes	human, mammals				
bacteria	Pseudomonas aeruginosa	Yes	human skin				
bacteria	Ruminococcus spp.	NR	equine				
bacteria	Salmonella spp.	Yes	canine, human				
bacteria	Shigella boydii	Yes	human				

Table 1. List of organisms reported in sewage and in the gut of various hosts. Taxa in red include pathogenic or parasitic species. NR: not reported.

Table 1 indicates a non-exhaustive list of organisms known to be found in sewage or the guts of humans or other hosts. Many of the organisms are bacteria, but archae, protists and viruses are also found. Our sequencing strategy will need to include methods to detect organisms from all four kingdoms as well as molds (fungi).

Preliminary data

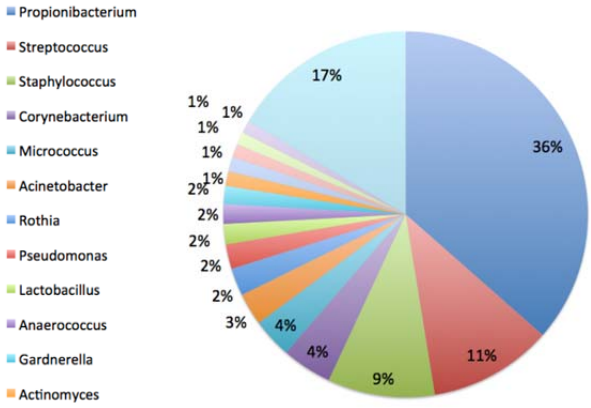
We provide below snapshots of some of the projects that several of us are involved in that are providing preliminary data, new protocols, and novel computational and analytical methods that will be used to jump-start the pilot project and foster its success.

The Monetary Microbiome (P. Scheid, J. Carlton NYU CGSB)

We carried out a proof-of-principle experiment to characterize the microbial diversity present on the surface of paper currency circulating in NYC in March 2013. Our goal was to determine if microbial material was detectable against a background of human DNA, and what bacteria or viruses might be transferred through the hand-to-hand exchange of paper money. Twenty $1 bills were obtained from a Chase Bank in Greenwich Village, Manhattan, swabbed using standard nasal swabs, and four DNA extraction methods tested for their DNA yield. Using the yield from the Mo-Bio PowerSoil Kit, a metagenomic library was constructed from trace amounts of DNA from two bills. Sequencing the library on one lane of an Illumina HiSeq2000, we generated ~50 Gb of 2x100 bp sequence data. After trimming the adapter sequences and removing human sequences (~52% of the total data), we measured non-human species abundance by aligning the sequences to MetaPhlAn and NCBI GenBank databases.

Our reliminary results demonstrate that the "**Monetary Microbiome**" contains genetic material from prokaryotic and eukaryotic organisms, many of which are relevant to human health (**Figure 2**). We found >15 non-human species each at an abundance of >1%. We identified two bacterial species found in the human vagina, *Gardnerella vaginalis*, a pathogenic bacterium that causes vaginosis, and *Lactobacillus iners*. Other high abundant bacterial species included *Propionibacterium acnes*, linked to skin acne. These data indicate the potential of environmental sequencing, even on such "inert" material as paper money.

Figure 2. Preliminary results from metagenomic sequencing of the DNA found on two $1 bills. The pie chart indicates the MetaPhlAn genus level results, and the table indicates the retative abundance and association of particular organisms with human flora.



Organism	Relative Abundance	Associated with Human Flora
Propionibacterium acnes	36.4	Gram+, anaerobic, linked to skin acne
Streptococcus mitis	3.9	Gram+, inhabits the mouth
Staphylococcus epidermidis	3.8	Gram-, associated with human skin
Staphylococcus hominis	3.7	Gram-, associated with human skin
Micrococcus luteus	3.5	Gram+, aerobe, found in soil and skin
Gardnerella vaginalis	1.5	Assoicated with disruption of normal vagina flora
Streptococcus parasanguinis	1.4	Gram+, inhabits the mouth
Rothia mucilaginosa	1.4	Gram+, associated with mouth and respiratory tract
Lactobacillus iners	1.0	Gram-,Associated with a healthy vagina flora
Rothia dentocariosa	1.0	Gram+, resides in healthy mouth and respiratory tract
Staphylococcus capitis	1.0	Gram-, associated with human skin
Other Bacteria (586 total)	37.20	

Antibiotics (M. Blaser, NYU SoM)

We have considerable experience generating and analyzing data from 16S rRNA studies of samples from humans and mice. Studies in humans have involved specimens from skin, esophagus, stomach, and colon, and we have extensive experience with mouse fecal samples [3]. One of our studies explored the phenomenon of widespread sub-therapeutic antibiotic use in livestock. We examined the effects of antibiotic exposures in mice to understand how physiology was being perturbed. We showed that each of several subtherapeutic regimens of antibiotics (termed STAT) changed body composition, with increased adiposity, increased production of short chain fatty acids by the intestinal microbiota, and induction of hepatic genes responsible for

lipogenesis [3]. In subsequent experiments, we showed that even short-term use of antibiotics could have long-term consequences, when we compared mice exposed to penicillin for 4, 8, or 28 weeks to untreated controls, since all had increased total, lean, and fat mass in relation to the controls. Over the 4 weeks of treatment, the composition of the microbiota changed substantially, but after stopping it, the composition normalized, indicating that the transient microbial perturbation was accompanied by permanent metabolic effects. Other experiments included transfer of cecal contents to germ-free mice. Mice that received STAT microbiota had more total and fat mass than mice that received control microbiota, showing that the altered microbiota, not the antibiotics, increased adiposity. In addition, the transferred microbial communities remained distinct between control and STAT recipients over time, despite removal of the antibiotic selective pressure, and that there was a time-dependent reassembly of microbial communities (data unpublished).

Our studies on the human microbiome and antibiotics will be of direct relevance to this project, in particular when dealing with the question of whether we can **monitor the spread of community-acquired antibiotic resistance**. Antibiotic resistance, such as MRSA (methicillin resistant *S. aureus*) has moved from being a hospital-acquired infection to community-acquired, through the spread of ESBL (extended-spectrum beta lactamase) producing bacteria that provide resistance to beta-lactam antibiotics like penicillins.

Computational methods to explore and learn from species abundance data (R. Bonneau, NYU CGSB)
The abundance of species in a body or city site is measured by counting sequence reads. However, the complexity of the environment and the microbes living on, in and around us, overwhelms the current methods for counting and organizing these data. One way we deal with that is to express species abundance as the relative abundance (% of total measurements for a site at a given time) or clusters of species (**operational taxonomic units**, OTUs). The use of relative abundance (or compositional) data presents several challenges that has resulted in the formation of the CoDa sub-field of modern statistics (**co**mpositional **da**ta analysis). We are adapting our methods for dynamic network inference [4] to decipher dynamic OTU-OTU and OTU-environment interactions. We are also developing multiple CoDA based distance metrics for clustering microbiome data and detecting relationships between OTUs and between OTUs and clinical meta-data. This will result in a new set of tools for microbiome-environment predictive network inference.

Technical Approach
Aim 1. To sample the sewage at four NYC DEP sewer sites in each of the five NYC boroughs
In most areas of NYC, sanitary and industrial wastewater, rainwater and street runoff are collected in the same sewers and then conveyed together to the City's treatment plants. This is known as a **combined sewer system**, which describes approximately 70% of the City sewers. In some New York City neighborhoods, sanitary waste and stormwater runoff are channeled in **separate sewer systems**: sanitary waste is carried to wastewater treatment plants while stormwater is channeled directly to local streams, rivers, and bays. In unsewered areas, such as parks and wetlands, this water is absorbed into the ground or channeled into waterways. A map of NYC sewer systems is shown in **Figure 3**.

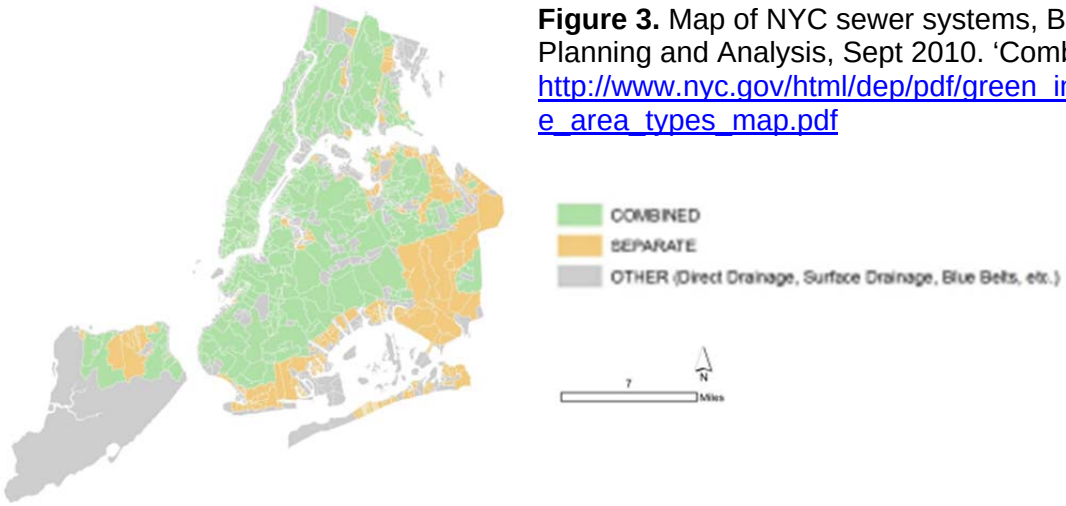


Figure 3. Map of NYC sewer systems, Bureau of Environmental Planning and Analysis, Sept 2010. 'Combined' sewers Adapted from http://www.nyc.gov/html/dep/pdf/green_infrastructure/sewer_drainage_area_types_map.pdf

We plan to sample four sewer sites, one commercial, two residential and one industrial in each of the five boroughs. The exact sites will be chosen in consultation with the NYC DEP; NYU CUSP are already in discussions with Christopher M. Hawkins, DEP Chief of Staff, and a Memorandum of Understanding between the two organizations is being generated. The two residential sites will be chosen from different residences as outlined in **Table 2**.

Table 2. Suggested residential sites from the five boroughs. All boroughs contain low-rise (defined as 3-6 story high) buildings, whereas high-rise buildings (defined as >10 stories) are predominant in Manhattan. Attached houses are defined as homes containing 2-4 families, whereas private homes contain a single family.

Manhattan	Brooklyn	Queens	Bronx	Staten Island
High-rise	Attached	Private home	Attached	Private home
Low-rise	Low-rise	Low-rise	Low-rise	Low-rise

Once the representative sites have been chosen, a sampling schedule will be developed. Our sampling plan may change once preliminary datasets have been analyzed. Initially, since we can multiplex up to 96 samples on a single lane of the Illumina HiSeq2500, giving ~780,000 reads per sample, and we have sufficient funds for 22 lanes, we can plan to collect as many as 2,000 samples over two years: approximately 40 samples per week, or 8 samples from each of the 5 boroughs. Our initial sampling will be to determine the type of effluent to collect, *i.e.,* consistency and color. Subsequent collections will be from before a storm and just after the same storm. We will also develop a time series of samples collected over a 24-hr period to assess fluctuations that may occur during the daily work week and at the weekend.

Although we will employ a **Postdoctoral Fellow** and **PhD student** for the project, the collection of sewage samples is likely to be done by NYC DEP staff, at their request. We will provide **NYU MetaGenome Kits** to the staff, containing a standardized set of bar-coded storage tubes, reagents and collection devices, as follows:
(1) 50 ml conical tube for collection of 25 ml sample for DNA extraction
(2) 50 ml conical tube containing RNA Later for collection of 25 ml sample for RNA extraction
(3) 50 ml conical tube for collection of 25 ml sample for archiving
We will also train the staff in the sampling protocol and the method of storing samples until the MetaGenome Kits can be brought back to the lab. All samples will be entered into a FreezerWorks database for sample tracking and storage. [N.B. The total volume of sewage to be collected may alter once preliminary sequence analysis is performed.]

Aim 2. To develop standard protocols for sample extraction, genomic library preparation, and sequencing
We are fortunate that the **Human Microbiome Project** (**HMP**) has an extensive selection of DNA extraction and sample processing SOPs (http://www.hmpdacc.org/tools_protocols/tools_protocols.php); we will adapt some of these in addition to attempting other environmental sample protocols (see for example the recent publication by [5]). we will adapt some of these in addition to trying other environmental sample protocols (see for example the recent publication by [5]). Since our samples may contain bacterial, viral, protist, archaeal and fungal material, we will have to modify certain nucleic acid extraction procedures. For example, one issue with spore-forming organisms (such as *Bacillus anthracis*) is that DNA cannot be easily extracted from spores. We will enrich for such microbes by adding a heat-resistance step in the protocol, as only spores will be resistant to boiling. Spores can then be germinated and DNA can be easily recovered from vegetative cells. For viral genomes, collaborator Dr. Elodie Ghedin (see **Appendix 1**) will provide us with her extensive protocol list for DNA and RNA viral isolation and nucleic acid extraction.

We will initially employ high-throughput amplicon sequencing, *i.e*., 16S rRNA for bacteria and archaea and 18S rRNA for protists and fungi, to determine baseline data. However, whole genome shotgun metagenomic analysis will determine gene content and is the ultimate goal. One confounding factor will be the amount of human or multi-cellular organism DNA since this will swamp the signal from smaller microbial genomes. Our preliminary data collected as part of the Monetary Microbiome project identified half of the DNA as human, a surprisingly small amount. However, we don't know if this will hold true for sewage samples.

We will multiplex samples using Illumina's dual indexing system. During sample preparation, up to 12 unique 'Index 1' sequences are arrayed across the columns of a 96-well plate and up to eight unique 'Index 2' sequences are arrayed down the rows, which creates up to 96 uniquely dual-indexed adapters. A total of ~780,000 reads per library will be generated if each HiSeq 2500 lane is 96-plexed, accounting for the PhiX spike-in of close to 50% that will be necessary to correct for lack of base diversity. Metagenomic sequencing will also be performed using the HiSeq 2x150 paired-end protocol, allowing for multiplexing of 1-4 samples depending on the degree of human DNA contamination. The exact multiplexing strategy will be determined empirically to optimize quality and sequence yield for the specimens.

Aim 3. To analyze and interpret the genomic sequence data using Big Data pipelines and bioinformatics methods.

Once the data have come off the HiSeq2500, sequence processing on the NYU CGSB Bowery cluster will depend on sequence type. For 16S rRNA sequences, the raw paired sequence reads will be de-noised, stitched together and segregated by barcode, using our in-house Big Data pipeline. De-multiplexed sequences will be subjected to standard processing consisting of (1) filtering potential chimeric artifacts, short sequences, low quality sequences, followed by (2) clustering into operational taxonomical units (OTUs) and assigning taxonomy using the *QIIME* pipeline [6]. In addition, QIIME will be used to calculate alpha- and beta-diversity using UniFrac statistics for descriptive characterization of microbial types. For shotgun metagenomics sequences, the NIH Roadmap HMP project has established *HUMAnN* [7] as the protocol of choice for upstream processing of shotgun sequencing reads for bacterial metagenomics. *HUMAnN* was developed to address the challenges in determining biological function within microbial communities using short sequence reads from metagenomic high-throughput sequencing. Statistical Analyses: statistical analysis of microbiome and metagenome composition will be performed in the R statistical programming environment [8] using packages such as ade4, vegan, and *phyloseq* [9]. Descriptive analysis. For 16S data, we will evaluate the adequacy of sequencing efforts using rarefaction plots. Alpha-diversity of each sample will be characterized through dominance, equitability, richness, evenness, Shannon, and Simpson diversity indices. The diversity metrics will be calculated at OTU and higher taxonomical levels to best characterize the community structure. We will utilize skyline plots to visualize the patterns of community structure in terms of relative abundances in the collected samples by clinical and phenotypic groups. Similarly, for metagenomic data, skyline plots will be used to reveal functional compositions of the samples. Heat-maps will be plotted to visualize clustering patterns in the data.

All sequence data will be deposited in public sequence databases such as those supported by GenBank.

Aim 4. To integrate and visualize the data through mapping against geographic and demographic information determined by the drainage area.

We will plan to integrate the microbial sequence data at each of the sites in the five boroughs with geographic and demographic information determined by the drainage area, in order to provide empirical-based models of the change in microbial diversity over time and space. However, multi-level datasets are difficult to analyze for meaning, due to their complexity and the non-linear nature of the relationships between levels. Visualization tools that group such datasets around relevant pathways are critical to the process of extracting relevant results. Although tools exist to generate data visualizations (e.g. *Excel* and *Tableau* for statistical plotting; *Cytoscape*, *GeneMAPP*, and *Pathway Studio* for network visualization), these tools are ineffective when applied to complex datasets that span experimental data types and public databases.

We will use a new tool called **defog** (*Data Exploration with Faces, Objects, and Groups*), developed by one of us (Dr. Claudio Silva, see **Appendix 1**). The main characteristic of defog is that it takes an integrated approach to data visualization, so visualization (e.g. network drawing, statistical plot, text annotation, image, free hand drawing) may represent any data in any level of abstraction (e.g. a pathway, a reaction, a metabolite). A consequence of defog's design is that it is easy to customize its functionalities to handle complex specific datasets and obtain as output a powerful visual exploratory tool for such datasets.

Other Big Data integrations and manipulations will be indertaken to see if signatures of microbial diversity can be identified at each site and a metagenomics baseline determined. We will also see if we can determine the immediate or long-term effects and microbial footprint of dramatic events on the environment, such a pre- and post-storm; and we will see if we can spot trends, shifts and perturbations in the microbial patterns that can serve as early warning signals for the health of New Yorkers.

Phase 2: Biosensing the New York City Environment
Timeline: Summer/Fall 2013

We have made plans to convene a **workshop** with various teams to solicit input and hear presentations from colleagues who are also interested in biosensing the NYC environment. Approximately 25-30 attendees are expected for the 2-3 day workshop. Leaders at the newly formed **NYU CUSP** will provide expertise concerning New York's transport, sewage, and housing systems; a team from NYU's **Center for Genomics and Systems Biology** will provide NGS, systems biology and informatics knowledge; colleagues from **NYU School of Medicine** will provide expertise in the human microbiome. Collaborators such as **Dr. Christopher Mason**, Assistant Professor in the Department of Physiology and Biophysics and the Institute for Computational Biomedicine at Weill Cornell Medical College and **Dr. Eric Schadt**, Director of the Institute for Genomics and Multiscale Biology at Mt Sinai Hospital, will also be in attendance. A **research plan** will be developed starting with a series of pilot projects, and break-out groups will provide discussion concerning the best technical approaches, sources of funding and data analysis and visualization techniques.

Phase 3: Mapping the NYC MetaGenome
Timeline: 2015 and beyond

Due to the public interest in this Grand Challenge, we are hopeful of continued funding after depletion of an NYU seed grant. In these subsequent years, we will use the protocols, approaches and data analyses generated during this pilot study to take a focused approach at different areas of the NYC metagenome, in collaboration with other members of the wider scientific community. For example, we may consider the use of **high volume air collectors** used mostly for measuring atmospheric concentrations of particulates, which can be applied to the collection of airborne DNA on properly sized filters. Others have suggested collecting DNA from swabs of **subway poles, traffic change or elevator buttons**.

References

1 Siegele, L. (2012) Mining the urban data. In *The Economist*

2 Robertson, C.E.*, et al.* (2013) Culture-Independent Analysis of Aerosol Microbiology in a Metropolitan Subway System. *Applied and environmental microbiology*

3 Cho, I.*, et al.* (2012) Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* 488, 621-626

4 Bonneau, R.*, et al.* (2006) The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol* 7, R36

5 Cai, L. and Zhang, T. (2013) Detecting Human Bacterial Pathogens in Wastewater Treatment Plants by a High-Throughput Shotgun Sequencing Technique. *Environmental science & technology*

6 Caporaso, J.G.*, et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7, 335-336

7 Abubucker, S.*, et al.* (2012) Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. *PLoS Comput Biol* 8, e1002358

8 R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing

9 McMurdie PJ and Holmes S (2012) Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing. *http://bioconductor.org/packages/devel/bioc/vignettes/phyloseq/inst/doc/phyloseq_analysis.pdf*

III: The Milestones and Quantifiable Metrics of Success Used to Assess Progress

Metrics of success

Our broad metrics of success can be divided into five categories (**Figure 4**). Briefly, by sampling sewage at four different sites in five boroughs over time and before and after traumatic events, we plan to uncover microbial signatures that correlate with each of the five categories. Identifying any microbial signature with one of these categories will be considered a success.

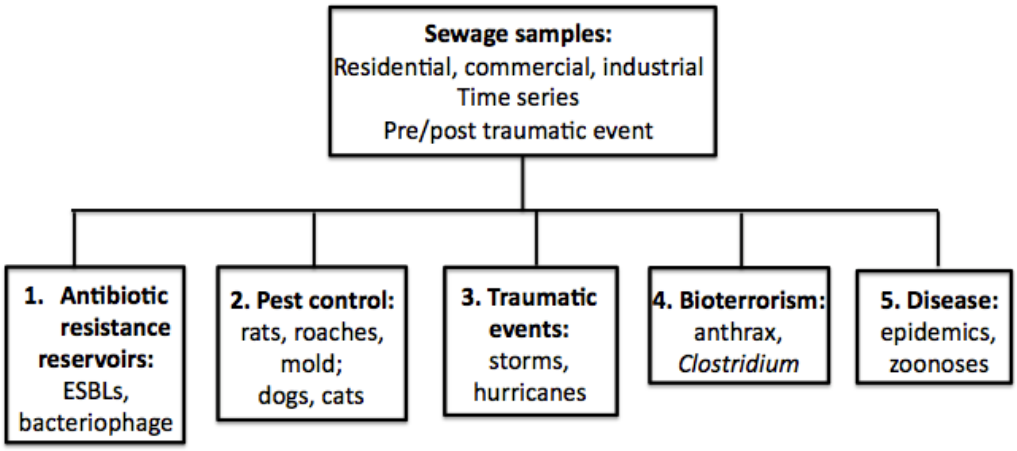


Figure 4. Five categories that we plan to associate with microbial signatures.

Milestones and timeline

Milestone 1: developing a sewage sampling strategy (first quarter Year 1)

Milestone 2: developig a standard set of lab protocols for sample processing, nucleic acid extraction, and library preparation (first quarter Year 1)

Milestone 3: preliminary sample sequencing & analysis, and refinement of Milestones 1 & 2 (third & fourth quarters Year 1)

Milestone 4: development of statistical analyses for species abundance (Year 1)

Milestone 5: development of visualization methods for integrated data (Year 1)

Milestone 6: initial research publication and press release (first quarter Year 2)

Milestone 7: submission of grant proposals to foundations and federal sources (first/second quarter Year 2)

Milestone 8: meta-analysis of all data, final high profile publication (third/fourth quarters Year 2)

Examples of quantifiable success criteria understandable by the general public include (1) Identifying and quantifying common microbes found in different ethnic populations and in different neighborhoods throughout NYC; and (2) Identifying specific patterns of microbial communities across space and time and their relationship to diseases, such as asthma.

IV: Our External Partners, Potential Funding Sources and Plan to Mobilize the External Community

External partners
Members of the NYC MetaGenome Project have elicited considerable interest in the project through discussion with several of their close collaborators and partner institutions. For example, NYU CGSB has strong connections to the **New York Genome Center**, a non-profit organization founded by several NYC academic medical centers including NYU to support large-scale whole genome sequencing projects, while NYU CUSP has ties to the **DOE Joint Genome Institute (JGI)** in Walnut Creek California, a leader in the sequencing of microbial genomes and metagenomic studies; there is a potential to engage both these institutions in the study of the NYC samples at a later stage. NYU CUSP has an extensive network of partners and associates, not limited to the NYC Dept of Environmental Protection mentioned in this proposal, but also to the **Dept. of City Planning**, the **Dept. of Parks & Recreation**, and the **Major's Office for Operations**. NYU GIPH has an extensive network with NYC health agencies in particular the NYC Department of Health and Mental Hygiene, should our pilot research findings be implementable.

Potential funding sources
Several NYU CGSB faculty have been in discussions with the **Simons Foundation** (https://simonsfoundation.org), one of the country's leading private funders of basic scientific research, because of their interest in Big Data. The mission of the Simons Foundation is "to advance the frontiers of research in mathematics and the basic sciences". The Foundation sponsors a range of programs that aim to promote a deeper understanding of the world, and supported grants to the value of $120 million in 2011. Our proposal to map the NYC metagenome will be part of several projects being discussed with Simons personnel at a second meeting in mid-May.

In addition, we anticipate the interest of several funding agencies and other foundations. For example, the possibility that the results of our study would provide bioremediation solutions to NYC Department of Environmental Protection problems such as the clogging of channels from the illegal dumping of grease may interest the environmental remediation programs of the **U.S. Environmental Protection Agency, U.S. Dept. of Energy**, and the **NIH National Institute of Environmental Health Sciences** (NIEHS). Urban metagenomics should interest the **National Science Foundation** ecological programs particularly those involving urban settings such the National Ecological Observatory Network (NEON).

Mobilizing the external community
We have started discussions with several **crowd-funding organizations** that mobilize public support of innovative and eye-catching science projects. Crowd-funding utilizes existing social networks to build a user base with a main goal of funding research projects and facilitating better collaboration, communication and participation in the research process. This approach has been used to raise funds and public awareness by several human microbiome projects, for example the American Gut Project (http://www.indiegogo.com/projects/american-gut-what-s-in-your-gut--7) and uBiome (http://www.indiegogo.com/projects/ubiome-sequencing-your-microbiome?website_name=ubiome).

Finally, our Grand Challenge has the potential to excite the public at large because its very nature – identifying and tracking our "bugs" and how they may influence our health -- is fascinating. We have already started to engage the public through various outreach initiatives. For example, with NYU Tisch School's Interactive Telecommunications Program (ITP), where "citizen cyber science" initiatives are actively encouraged, we have joined a project called **Invisible Exchange** where citizens upload pictures of money-derived microbes they have cultured at home to an interactive website in an attempt to identify and share microbes present on their currency. We are investigating using their platform to engage the public in the Monetary Microbiome project. We have given several presentations concerning next generation sequencing and Big Data to recruit computer scientists to the effort, including one at *Science and the City Hackfest*, and one at ITP's weekly seminar series. We are also planning to give a presentation at the *World Science Festival* "hack day" in June 2013 (http://www.worldsciencefestival.com/events/science_hack_day).

APPENDIX 1: MANAGEMENT AND STAFFING PLAN

The NYC MetaGenome Team
We are a diverse set of NYU clinicians, genomicists, Big Data analysts, science policy experts and public health scholars. Our collective expertise will ensure the success of this pilot project, while providing a strong base for strategic pitches to federal agencies and foundations for funds to support a full Grand Challenge.

NYU Center for Genomics and Systems Biology (NYU CGSB)
Dr Jane Carlton is the lead PI at NYU Biology, and her area of expertise is genomics of eukaryotic microbes. She is leading a project called the *Monetary Microbiome* that aims to uncover the microbial diversity being transmitted on paper currency in New York City. She is the program director of a seven-year, NIH-funded $11 million *International Center of Excellence for Malaria Research*, and leads a team of more than 60 investigators based in the US and India. She will be directing all aspects of the NYC MetaGenome Project, including facilitating interactions between team members, driving the research agenda, developing and writing research papers from the project, and soliciting financial support from federal and private sources.

Dr Richard Bonneau is Co-Director of Informatics at CGSB, and a computational biologist used to grappling with Big Data. His recent paper in Cell on a regulatory network for immune cell specification received wide acclaim for its use of network inference and modeling of the human immune system. Dr. Bonneau and his group are developing computational methods to explore and learn from species abundance data, and they will build models to describe and predict changes in microbial composition in response to city-wide metadata.

Dr Patrick Eichenberger is an Associate Professor with expertise in microbial genomics, and uses genomics and systems biology approaches to investigate bacterial sporulation. Spore-forming pathogens such as *Bacillus anthracis*, the causative agent of anthrax, are among the most dangerous infection agents. Dr. Eichenberger will provide expertise on the topic of bioterrorism agents and how spore-forming pathogens can be detected through metagenomic methods.

NYU School of Medicine
Dr. Martin Blaser is director of the *Human Microbiome Project* at NYU School of Medicine. His lab has been studying the dynamics of microbial populations in humans for more than a decade, and more recently been exploring microbiome changes associated with the obesity and diabetes epidemics. He is an internationally reknowned member of the microbiome research community, and leads several multimillion dollar NIH-funded projects studying the human microbiome in health and disease. Dr. Blaser will lead the NYC MetaGenome Project at NYU School of Medicine and lend his expertise in the study of human microbial communities to the project. The postdoc fellow for our pilot project will be integrated in his group, where he will learn the various sampling and analysis protocols that Dr Blaser's group have developed.

NYU Center for Urban Science and Planning (NYU CUSP)
Dr. Steven Koonin is the Director of NYU CUSP and oversees a consortium of partners including six international universities, multiple Fortune 500 corporate research offices, four national labs, and 15 New York City and State agencies. Dr. Koonin will facilitate access to the NYU CUSP consortium of partners.

Dr Ari Patrinos is the Deputy Director for Research at NYU CUSP and a Professor at NYU Poly. Dr. Patrinos served as Director of the Biological and Environmental Research Program in the Department of Energy, and helped create the DOE Joint Genome Institute. He also served as President of Synthetic Genomics Inc., a company employing synthetic biology methods to produce sustainable fuels, foods, vaccines and pharmaceuticals. Dr. Patrinos will lead the NYU CUSP involvement in the project.

Dr. Cláudio T. Silva is the Head of Disciplines at NYU CUSP, Professor of Computer Science and Engineering at NYU Poly, and affiliate faculty at NYU Courant. He is an active member of the visualization, graphics, and geometric computing research communities, and his research interests include Big Data and

Urban Systems. Dr. Silva will provide expertise in the areas of scientific data management and data visualization for the pilot project data.

NYU Global Institute for Public Health

Dr Cheryl Healton is the Director of NYU's Global Institute of Public Health and Dean of Global Public Health. She is an active member of the public health community, serving on several boards including the Betty Ford Institute, and Lung Cancer Alliance. She has extensive working knowledge of the NYC Department of Health and Mental Hygiene. Dr Healton and her team at GIPH will facilitate interactions with NYC health agencies and provide a conduit to other health departments or government authorities, for example to implement our pilot research findings.

Other key personnel

NYU CGSB Genomic Sequencing Core Team

The Genomic Sequencing Core (GenCore) in CGSB provides access to a range of next generation sequencing technology to assist ~40 NYU Biology faculty labs and collaborators with their genomics research. GenCore is housed in a custom-built, state-of-the-art facility on the 8th floor of CGSB, and currently houses two NGS platforms, an Illumina HiSeq 2000 (to be upgraded to a HiSeq 2500 in May 2013), and an Ion Torrent Personal Genome Machine. The wet-lab staff (**Manager Paul Scheid**, see below, and a Laboratory Technician) have expertise in experimental design, nucleic acid extraction and purification, NGS library construction, and NGS applications. The bioinformatics staff (a Computational Biologist and a Programmer) provide raw data processing, pipeline analysis and storage of the terabytes of data that are generated during each NGS machine run. A designated High Performance Computing (HPC) Specialist helps maintain a 64-node Intel Xeon compute cluster "Bowery" with 50TB storage, for use by faculty for data storage and analysis. All the sewage samples for our pilot project will be processed and sequenced using the GenCore facility, and the raw data will be manipulated, stored and backed-up using the Bowery HPC.

Mr. Paul Scheid has been the Manager of GenCore since 2009 and has extensive experience in DNA sequencing methods and lab management. He manages all aspects of running the HiSeq 2000 and PGM, including standardizing experimental protocols and core workflows, conducting trial runs, and interacting with the computational members of GenCore for data processing. Mr. Scheid will be responsible for sequencing all sewage samples in GenCore and providing the raw data to the team for further analyses.

Dr. Elodie Ghedin is an Associate Professor and MacArthur Fellow at the University of Pittsbugh, and was previously on the faculty at the J. Craig Venter Institute. She developed a high-throughput pipeline to perform rapid sequencing and analysis of viral genomes from various host species specimens, and she is the PI of a lung microbiome study to analyze infections present in the lungs of people with HIV. Dr. Ghedin will be an external collaborator on this project and provide her expertise in viral metagenomics and the evolution and transmission of viruses. (She is currently a top candidate for an open faculty position at NYU Biology.)

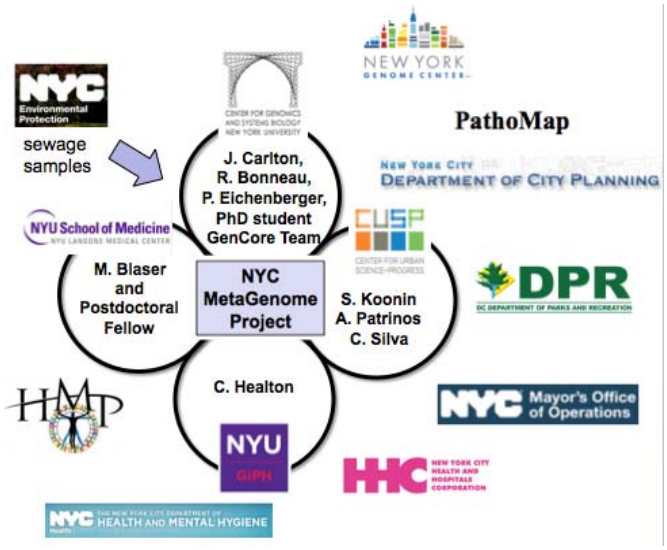


Figure 5. Schematic of linkages between the four NYU institutions involved in the NYC MetaGenome Project (open circles), flanked by several potential external organizations.

Abbreviations: NYU GIPH: NYU Global Institute of Public Health; HMP: Human Microbiome Project; PathoMap: joint U. Columbia/Mt. Sinai initiative to develop an interactive database of 15 humans and their microbes over 1 year.

APPENDIX 2: BIOGRAPHICAL INFORMATION

Two-page biosketches are provided for:

A. Lead PIs:

NYU Biology:	Dr. Jane Carlton Dr. Richard Bonneau Dr. Patrick Eichenberger
NYU SoM:	Dr. Martin Blaser
NYU CUSP:	Dr. Steven Koonin Dr. Ari Patrinos Dr. Claudio Silva
NYU GIPH:	Dr. Cheryl Healton

B. Two key personnel:

GenCore:	Mr. Paul Scheid
Collaborator:	Dr. Elodie Ghedin, U. Pittsburgh

BIOGRAPHICAL SKETCH	
NAME **Carlton, Jane**	POSITION TITLE **Faculty Director of Genomic Sequencing, NYU Center for Genomics & Systems Biology, and Professor, NYU Biology**
eRA COMMONS USER NAME Carlton	

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEAR	FIELD OF STUDY
U Edinburgh, Edinburgh, Scotland, UK	B.Sc.	1990	Genetics
U Edinburgh, Edinburgh, Scotland, UK	Ph.D.	1995	Parasite genetics
U Edinburgh, Edinburgh, Scotland, UK	Postdoc	1995-1997	Molecular parasitology
U Florida, Gainesville, FL, USA	Postdoc	1997-1999	Genomics/bioinformatics

A. Personal Statement

My area of expertise is whole genome sequencing, bioinformatics and comparative evolutionary genomics of **eukaryotic microbes**. My lab members are involved in the sequencing and analysis of several important parasite genomes, and we also scan the genomes for genetic variation to understand their evolution and impact on human health and disease. More recently our project to characterize the female microbiome associated with an STD parasite was published, and we have ongoing studies to characterize the microbial diversity present on the surface of paper currency circulating in New York City.

B. Positions and Honors

Professional Positions

1999-2000 Assistant Scientist (non-tenured track faculty), University of Florida, Gainesville, FL
2000-2001 Research Fellow (VP), National Center for Biotechnology Information, NIH, Bethesda, MD
2001-2006 Associate Investigator (faculty), The Institute for Genomic Research, Rockville, MD
2006-2011 Associate Professor, Dept Medical Parasitology, NYU School of Medicine, New York, NY
2009-2011 Director of Genomics, NYU School of Medicine, New York, NY
2006-present Adjunct Research Associate, American Museum of Natural History, New York, NY
2011-present Affiliate, Division of Medical Parasitology/Dept of Microbiology, NYU School of Medicine
2011-present Professor and Faculty Director of Genomic Sequencing, Center for Genomics and Systems Biology, Dept of Biology, New York University, New York, NY

Other Experience and Professional Memberships

1995-present Member, American Society of Tropical Medicine and Hygiene
2006-present Member American Society Microbiology
2001-present *Ad hoc* member various study sections including: NIAID Genomics, Computational Biology and Technology 2008; NHGRI Special Emphasis Panel ZRG1 GGG-J (52) Human Microbiome Project 2008; NIH Genomics, Computational Biology and Technology (GCAT) study section 2008 & 2009; NIH Pathogenic Eukaryotes (PTHE) study section 2009; Development of New tools for Computational Analysis of Human Microbiome Project Data (ZRG1 GGG-N) 2010.
2007-2012 Co-chair NHGRI/NIAID Eukaryotic Pathogens and Disease Vectors Sequencing Target Selection working group
1996-present Reviewer various journals including: Nature, Nature Genetics, Science, Gene, Genomics, BMC Genomics, PLoS Pathogens, PloS Biology, PloS Computational Biology, Heredity, PNAS
2003-present Editorial Board member of: BMC Genomics, Database, Trends in Parasitology, PLoS NTD
2010-present Scientific Advisor, BBC/Public Radio International radio program 'The World'

Awards and Nominations

1993 Young Investigator Award, Scottish Universities Molecular Parasitology Group
2004 Martin J. Rodbell Award, The Institute for Genomic Research, Rockville, MD
2006/10 NYU School of Medicine Dean's Honors
2009 Nominated for 2009 Blavatnik Awards for Young Scientists
2010 Stoll-Stunkard Award, American Society of Parasitologists
2011 Nominated for 2011 Dan David Prize in Genome Research
2012 Elected Fellow of the American Association for the Advancement of Science

Scientific Meetings Organized and Coordinated
Co-organizer with George Weinstock, Julian Parkhill, Matt Berriman, joint bi-annual Cold Spring Harbor/Wellcome Trust Sanger Institute conference 'Infectious Disease Genomics and Global Health', Hinxton Campus, Cambridge, UK: September 2008, September 2010, October 2012, October 2013.

C. Selected peer-reviewed publications (selected from >90 publications)

1. Melissa D. Conrad, Patricia Kissinger, Norine Schmidt, David H Martin, and **Jane M. Carlton**. Genetic diversity of *Trichomonas vaginalis* reinfection in HIV$^+$ women. Sex Transm Infect 2013, in press.

2. Brotman RM, Bradford LL, Conrad M, Gajer P, Ault K, Peralta L, Forney LJ, **Carlton JM**, Abdo Z, Ravel J. Association Between Trichomonas vaginalis and Vaginal Bacterial Community Composition Among Reproductive-Age Women. Sex Transm Dis. 2012 Oct;39(10):807-812.

3. Neafsey DE, Galinsky K, Jiang RH, Young L, Sykes SM, Saif S, Gujja S, Goldberg JM, Young S, Zeng Q, Chapman SB, Dash AP, Anvikar AR, Sutton PL, Birren BW, Escalante AA, Barnwell JW, **Carlton JM**. The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. Nat Genet. 2012 Sep;44(9):1046-50. **Front cover article.**

4. Conrad MD, Gorman AW, Schillinger JA, Fiori PL, Arroyo R, Malla N, Dubey ML, Gonzalez J, Blank S, Secor WE, **Carlton JM**. Extensive Genetic Diversity, Unique Population Structure and Evidence of Genetic Exchange in the Sexually Transmitted Parasite Trichomonas vaginalis. PLoS Negl Trop Dis. 2012 Mar;6(3).

5. Reed DL, Currier RW, Walton SF, Conrad M, Sullivan SA, **Carlton JM**, et al. The evolution of infectious agents in relation to sex in animals and humans: brief discussions of some individual organisms. Ann N Y Acad Sci. 2011 Aug;1230(1):74-107.

6. Malik SB, Brochu CD, Bilic I, Yuan J, Hess M, Logsdon JM Jr, **Carlton JM**. Phylogeny of parasitic parabasalia and free-living relatives inferred from conventional markers vs. Rpb1, a single-copy gene. PLoS One. 2011;6(6):e20774

7. **Carlton JM**, Adams JH, Silva JC, Bidwell SL, Lorenzi H, et al. Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature*. 2008 Oct 9;455(7214):757-63. **Front cover article.**

8. **Carlton JM**, Hirt RP, Silva JC, Delcher AL, Schatz M, et al. Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. *Science*. 2007 Jan 12;315(5809):207-12. **Front cover article.**

9. **Carlton JM.** Toward a malaria haplotype map. *Nat Genet.* 2007 Jan;39(1):5-6.

10. 'The nuts and bolts of sequencing protist genomes.' Daniella Bartholomeu, Neil Hall, **Jane Carlton**. In *Genomics and Evolution of Microbial Eukaryotes*, Oxford University Press, September 2006, edited by Laura A. Katz and Debashish Bhattacharya.

D. Other Support

ACTIVE

1 R01 AI097080-01 P. Kissinger (PI) 07/01/12-06/30/17
NIH/NIAID "*Trichomonas vaginalis* repeat infections among HIV negative women"

1U19AI089676-01 J. Carlton (PI) 07/01/10-06/30/17
NIH/NIAID "Center for the Study of Complex Malaria in India"

COMPLETED (relevant awards completed during last 3 years only)

S10 RR026950-01 J. Carlton (PI) 04/01/10-03/31/11
NIH/NCRR "Roche 454 Next Generation Sequencer for Human Microbiome and Infectious Disease Research"

BIOGRAPHICAL SKETCH	
NAME **Bonneau, Richard**	POSITION TITLE **Associate Professor and Co-Director of Informatics, Center for Genomics and Systems Biology, New York University**
eRA COMMONS USER NAME RB133.NYU	

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEAR	FIELD OF STUDY
Florida State University, Florida	B.A.	1997	Biochemistry
University of Washington, Washington	Ph.D.	2001	Ab Initio Structure Prediction

A. Personal Statement

I focus on two main categories of computational biology: learning networks from functional genomics data and predicting and designing protein and peptoid structure. In both areas I have played key roles in achieving critical field-wide milestones. In the area of structure prediction I was one of the early authors on the Rosetta code, which was one of the first codes to demonstrate accurate and comprehensive ability to predict protein structure in the absence of sequence homology. My lab has also made key contributions to the areas of genomics data analysis. We have also started a new project with political scientists and experimental psychologists to apply methods for learning network structure from time series to social media time series data using Twitter, online blogs about politics, and Facebook as our initial data sources (recently funded by NSF INSPIRE).

B. Positions and Honors

Professional Positions

1994-1996	Board Member, FSU Center for Participant Education.
1994-1996	Undergraduate Research. Florida State University, w/ Tim Logan.
1997	Research Assistant. National High Magnetic Field Laboratory, Tallahasse.
1997-1998	High School Teacher. NOVA High, Seattle.
1997-2001	University of Washington. Graduate Student David Baker lab.
2001	Structural GenomiX, Inc. San Diego. Sr. Scientist.
2001-2006	Institute for Systems Biology, Seattle, WA. Sr. Research Scientist.
2007-2010	Senior Advisor. Tacitus, LLC, Philadelphia, PA, USA.
2005-Present	Associate Professor, New York University Center for Comparative Functional Genomics and New York University, Courant Institute, Dept. of Computer Science

Awards and Fellowships

1993	Florida Academic Scholars Award & International Baccalaureate Diploma
1996	American Cancer Society – James Jay Fisher Fellowship
1998	Howard Hughes Medical Institute pre-doctoral Fellowship in the Biological Sciences
2008	One of top 20 scientists under 40, Discover magazine

C. Selected Peer-reviewed publications

For a current list see: http://scholar.google.com/citations?user=Wq8XTykAAAAJ&hl

1. Youngs N, Penfold-Brown D, Drew K, Shasha D, **Bonneau R**. Parametric Bayesian priors and better choice of negative examples improve protein function prediction. Bioinformatics. 2013 May 1;29(9):1190-8. PubMed PMID: 23511543.

2. Greenfield A, Hafemeister C, **Bonneau R**. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. Bioinformatics. 2013 Apr 15;29(8):1060-7.PubMed PMID: 23525069.

3. Ciofani M, Madar A, Galan C, Sellars M, Mace K, Kirigin FK, Birchmeier C, Wagner EF, Murphy KM, Myers RM, **Bonneau R***, Littman DR* (2012) . A validated regulatory network for Th17 cell specification Cell. 2012 Oct 12;151(2):289-303. PMID: 23021777; PubMed Central PMCID: PMC3503487. * denotes co-corresponding authors.

4. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, **Bonneau R**, Selbach M, Dieterich C, Landthaler M. The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. (2012) Mol Cell. 46(5):674-90. PMID: 22681889.

5. Renfrew PD, Choi EJ, **Bonneau R**, Kuhlman B (2012) Incorporation of Noncanonical Amino Acids into Rosetta and Use in Computational Protein-Peptide Interface Design. PLoS ONE 7(3): e32637. doi:10.1371/journal.pone.0032637.

6. Drew K, Winters P, Butterfoss GL, Berstis V, Uplinger K, Armstrong J, Riffle M, Schweighofer E, Bovermann B, Goodlett DR, Davis TN, Shasha D, Malmström L, **Bonneau R**. (2011) The proteome folding project: proteome-scale prediction of structure and function. Genome Res. 21(11):1981-94.

7. Waltman P, Kacmarczyk T, Bate AR, Kearns DB, Reiss DJ, Eichenberger P, **Bonneau R**. Multi-species integrative biclustering. Genome Biol. 2010;11(9):R96. Epub 2010, Sep 29. PubMed PMID: 20920250.

8. Greenfield A, Madar A, **Bonneau R**. DREAM4: Combining Genetic and Dynamic Information to Identify Biological Networks and Dynamical Models. DREAM4 top performers special collection. PLoS ONE 2010, 5(10): e13397. doi:10.1371/journal.pone.0013397.

9. **Bonneau R***, Facciotti MT, Reiss DJ, Madar A, Baliga NS*, et al. A predictive model for transcriptional control of physiology in a free living cell. (2007) Cell. Dec 131:1354-1365.

10. **Bonneau R**, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CEM, Baker D. (2001) Rosetta in CASP4: Progress in ab initio protein structure prediction. Proteins. 45(S5)119-126.

D. Other Support

ACTIVE

SES-1248077 R. Bonneau, J. Tucker, J. Jost and J. Nagler (co-PIs) 09/15/2012-08/31/2015
NSF “INSPIRE: Computer Learning of Dynamic Systems to Investigate Cognitive and Motivational Effects of Social Media Use on Political Participation”

CHE-1151554 R. Bonneau and P. Arora (co-PIs) 04/01/2012-03/31/2015
NSF “A Systematic Approach to Targeting Protein Interfaces with Nonpeptidic Helix Mimetics”

IOS-1126971 R. Bonneau and M. Purugganan (co-PIs) 09/01/2011-08/31/2015
NSF “Environmental Gene Regulatory Interaction Networks in Rice” (Co-PI, with Michael Puruggannan)

1 RC4 AI092765-01 R. Bonneau and D. Littman (co-PIs) 09/30/2010-09/29/2013 29/09
NIH “Elucidation of the transcriptional network underlying the Th17 lineage program”

I U54CA143907-01 09/28/2009-07/31/2014
NIH “Physical Sciences Oncology Center”

7 PN2 EY016586-06 M. Dustin (PI) 09/30/2009-09/29/2014
NIH “Nano Medicine Center for Mechanical Biology”

BIOGRAPHICAL SKETCH

NAME	POSITION TITLE
Eichenberger, Patrick Yves	**Associate Professor, Center for Genomics and Systems Biology, Department of Biology, New York University**
eRA COMMONS USER NAME EICHENBERGER	

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEAR	FIELD OF STUDY
University of Geneva, Switzerland	B.Sc.	1991	Biochemistry
University of Geneva, Switzerland	Diploma	09/1992	Biochemistry
University of Geneva, Switzerland	M.Sc.	1996	Molecular Biology
University of Geneva, Switzerland	Ph.D.	1992-1997	Biology
Harvard University, Cambridge, MA, USA	Postdoc	1998-2004	Molecular and Cellular Biology

A. Personal Statement

I have been a Faculty member of the Department of Biology at New York University since 2004. I use genomics and systems biology-influenced approaches to investigate the process of bacterial sporulation, particularly in the model organism *Bacillus subtilis*. Spore-forming pathogens, such as *Bacillus anthracis*, the causative agent of anthrax, and various *Clostridium* spp., including *Clostridium botulinum*, *Clostridium tetani* and *Clostridium difficile*, are among the most dangerous infectious agents, in part because of their extraordinary resistance properties. **An additional concern is that *B. anthracis* spores have been used as a bioterrorism weapon.** At NYU Center for Genomics and Systems Biology, I collaborate with Prof. Richard Bonneau, a computational biologist, on the characterization of the *B. subtilis* and *B. anthracis* global transcriptional regulatory networks. In addition to my expertise in **systems biology**, I have extensive knowledge of **molecular and cellular microbiology**, in particular the use of GFP fusions to investigate sub-cellular localization of sporulation proteins. In the past few years, my lab has published several papers investigating mechanisms controlling the assembly of the *B. subtilis* spore envelope, in particular the outermost layer of the spore known as the crust.

B. Positions and Honors

Professional Positions

1991-1992 Diploma Student, University of Geneva, Switzerland. Advisor: Prof. Jean-Paul Giacobino.
1992-1998 Graduate Student (M.S. and Ph.D), University of Geneva. Advisor: Prof. Johannes Geiselmann.
1998-2004 Post-doctoral Fellow, Harvard University, USA. Advisor: Prof. Richard Losick
2004-2010 Assistant Professor. Center for Genomics and Systems Biology, New York University, USA.
2010-Present Associate Professor. Center for Genomics and Systems Biology, New York University, USA.

Other Experience and Professional Memberships

2010-2015 Editorial Board, Journal of Bacteriology
2013-2015 Editorial Board, Applied and Environmental Microbiology
2004-Present Member, American Society for Microbiology (ASM)
Present Book editor, The Bacterial Spore: From Molecules to Systems (ASM press). To be published in 2014
Present Meeting organizer: Annual NYBIG meeting (New York Bacillus Interest Group)

Honors

1998 Post-Doctoral Fellowship (Human Frontier Science Program).
2000 Fellowship for Advanced Researchers (Swiss National Science Foundation).
2003 Merck Award for Genome-Related Research.
2005 Whitehead Fellowship for Junior Faculty in Biomedical and Biological Sciences.

C. Selected Peer-reviewed Publications (selected from 30 publications)

1. McKenney PT, Driks A and **Eichenberger P** (2013). The *Bacillus subtilis* endospore: assembly and functions of the multi-layered coat. *Nature Reviews Microbiology* **11**: 33-44.
2. McKenney PT and **Eichenberger P** (2012). Dynamics of spore coat morphogenesis in *Bacillus subtilis*. *Molecular Microbiology* 83(2):245-60.
3. de Francesco M, Jacobs JZ, Nunes F, Serrano M, McKenney PT, Chua MH, Adriano O. Henriques AO and **Eichenberger P** (2012). Physical interaction between coat morphogenetic proteins SpoVID and CotE is necessary for spore encasement in *Bacillus subtilis*. *Journal of Bacteriology* 194(18):4941-50.
4. Kacmarczyk T, Waltman P, Bate AR, **Eichenberger P** and Bonneau R (2011). Comparative microbial modules resource: generation and visualization of multi-species biclusters. *PLoS Comp Bio* 7(12): e1002228.
5. McKenney PT, Driks A, Eskandarian HA, Grabowski P, Guberman J, Wang KH, Gitai Z and **Eichenberger P** (2010). A distance-weighted interaction map reveals a previously uncharacterized layer of the *B. subtilis* spore coat. *Current Biology* 20(10):934-8.
6. Waltman P, Kacmarczyk T, Bate AR, Kearns DB, Reiss DJ, **Eichenberger P**, Bonneau R (2010). Multi-species integrative biclustering. *Genome Biology* 11(9):R96.
7. de Hoon MJ, **Eichenberger P**, Vitkup D (2010). Hierarchical evolution of the bacterial sporulation network. *Current Biology* 20(17):R735-45.
8. Wang KH, Isidro AL, Domingues L, Eskandarian HA, McKenney PT, Drew K, Grabowski P, Chua MH, Barry SN, Guan M, Bonneau R, Henriques AO and **Eichenberger P** (2009). The coat morphogenetic protein SpoVID is necessary for spore encasement in *Bacillus subtilis*. *Molecular Microbiology* 74(3): 634-649.
9. Mallozzi M, Bozue J, Giorno R, Moody KS, Slack A, Cote C, Qiu D, Wang R, McKenney P, Lai EM, Maddock JR, Friedlander A, Welkos S**, Eichenberger P** and Driks A (2008). Characterization of a *Bacillus anthracis* spore coat-surface protein that influences coat surface morphology. *FEMS Microbiol Lett* 289:110-7.
10. Wang ST, Setlow B, Conlon EM, Lyon JL, Imamura D, Sato T, Setlow P, Losick R and **Eichenberger P** (2006). The forespore line of gene expression in *Bacillus subtilis*. *Journal of Molecular Biology* 358(1):16-37.
11. **Eichenberger P**, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS and Losick R (2004). The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *Public Library of Science Biology* 2(10):e328.

D. Other Support

ACTIVE

R01GM081571-05 Eichenberger (PI) 09/01/08-07/31/2013
A systems level analysis of spore coat assembly in *Bacillus subtilis*
The goal of this study is to integrate the genetic, biochemical and genomic information concerning all genes involved in spore coat formation in *B. subtilis* into a spatiotemporal interaction network. By identifying the principal hubs in this network, we will uncover new targets for control of spore formation/viability.
Role: Principal Investigator

COMPLETED

RC2GM092616 Rudner (PI) 09/30/09-08/31/2011
A multidisciplinary approach to elucidating gene function in *B. subtilis*, a model Gram-positive bacterium.
The goal of this study is to create and make available genome-scale tools for the analysis of *B. subtilis*, including mutant libraries. The Eichenberger lab has carried out transcriptional profiling analyses in time series over various genetic and environmental perturbations (> 200 microarray experiments) in order to infer the global transcriptional regulatory network of *B. subtilis* (in collaboration with Richard Bonneau's group).
Role: co-PI

W81XWH-04-01-0307 Stein (PI) 02/23/04-07/31/09
Establishing a Fundamental Research Program in Prokaryotic Genomics at New York University
Department of Defense, United States Army Medical Research
Start-up funds
Role: Investigator

BIOGRAPHICAL SKETCH

NAME **Blaser, Martin J.**	POSITION TITLE **Professor, Medicine and Microbiology, NYU School of Medicine**
eRA COMMONS USER NAME blasem01	

INSTITUTION AND LOCATION	DEGREE	YEAR(s)	FIELD OF STUDY
Univ. Pennsylvania, Philadelphia, PA	BA	1969	Economics
New York University, New York, NY	MD	1973	Medicine

A. Personal Statement. My lab has been studying the human microbiome since 2002. We conducted early 16S rRNA surveys of the esophagus, stomach, and forearm skin. Such studies establish the baseline present in health that then can be used to assess pathologic relationships. We have been studying psoriasis for the past five years to address whether there has been disruption of the cutaneous microbiome and whether that could have a causative role in the disease. More recently, we have been exploring microbiome changes that could be fueling the obesity and diabetes epidemics, as well as childhood-onset asthma and allergic disorders.

B. Positions and Honors

1972-1973 New York University Student Travel Fellowship (Ethiopia).
1973-1979 Intern, Resident, Fellow in Infectious Diseases, Univ. Colorado Health Sci. Ctr., Denver, CO
1979-1981 Epidemic Intelligence Service Officer, Enteric Diseases Branch, Bacterial Diseases Division, Center for Infectious Diseases, Centers for Disease Control, Atlanta, GA
1980 Consultant to International Centre for Diarrheal Disease Research, Dacca, Bangladesh
1981-1989 Assistant and Associate Professor of Medicine, Div of Infect Diseases, Univ. Colorado School of Medicine; Chief, Infect Disease Section, VA Medical Center, Denver, CO
1987-1988 Guest Investigator, Laboratory for Bacteriology & Immunology, Rockefeller Univ., NY, NY
1989-2000 Addison B. Scoville Professor of Medicine, Director, Division of Infectious Diseases, and Professor of Microbiology and Immunology, Vanderbilt Univ. School of Medicine,TN
1991,92,94,96 Professeur Invité, Institut Pasteur, Paris, France
2000-2012 Frederick H. King Professor of Internal Medicine, Chair, Department of Medicine, Professor of Microbiology, New York University School of Medicine (NYUSOM), NY, NY
2008-present Affiliated member, Dept. of Biology, Faculty of Arts and Sci., New York University, NY, NY
2012-present George and Muriel Singer Professor of Medicine; Professor of Microbiology; Director, Human Microbiome Program, NYUSOM, New York, NY

Other Experience and Professional Memberships (since 2005)

2005-2010 Board of Scientific Counselors, National Cancer Institute; Chair 2009-2010
2005-present Editorial Board, *FASEBJ*
2006-2012 Associate Editor, *Microbes and Infection*
2006-present Doris Duke Medical Foundation Scientific Advisory Committee
2006-present Editorial Board, *Cell Host and Microbe*
2008-present Senior Editor, *Cancer Prevention Research*
2012-present Editorial Board, *Microbiome*
2013-present Editorial Board, *mBio*

Honors

1988 Western Society for Clinical Investigation Young Investigator Award
1992 Squibb (Oswald Avery) Award, Infectious Disease Society of America
1995 Association of American Physicians
1996 Alpha Omega Alpha, New York University School of Medicine
2001 Wade Hampton Frost Award, American Public Health Association
2003 AACR-American Cancer Society Award for Research Excellence in Cancer Epidemiology
2004 Master, American College of Physicians
2005-2006 President, Infectious Diseases Society of America
2009-2013 Advisory Board for Clinical Research, National Institutes of Health; Chair 2012-present
2011 Institute of Medicine

C. Selected peer-reviewed publications (in chronological order since 2010, of 508 reviewed articles)

1. **Blaser MJ**. Harnessing the power of the human microbiome. PNAS 2010;107:6125-6126.
   Dominguez-Bello MG, **Blaser MJ**, Ley RE, Knight R. Development of the human gut microbiota: insights from high-throughput sequencing. Gastroenterology 2011; 140: 1713-1719.
2. Plottel CP, **Blaser MJ**. Microbiome and malignancy. Cell Host Microbe 2011; 10:324-335.
   Cho I, **Blaser MJ**. The human microbiome: at the interface of health and disease. Nature Rev Genetics 2012; 13; 260-270.
3. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012; 486; 207-214.
4. The Human Microbiome Project Consortium. A framework for human microbiome research. Nature 2012; 486; 215-221.
5. **Blaser MJ**, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Estrada I, Gao Z, Clemente JC, Costello EK, Knight R. Distinct cutaneous bacterial assemblages in a sampling of South American Amerindians and United States Residents. ISME Journal 2012; doi: 10.1038/ismej.2012.81.
6. Cho L, Yamanishi S, Cox L, Methe BA, Zavadil J, Li K, Gao Z, Mahana D, Raju K, Teitler I, Li H, Alekseyenko AV, **Blaser, MJ**. Early-life antibiotics alter the murine colonic microbiome and adiposity. Nature 2012; 488:621-6.
7. Trasande L, Blustein J, Liu M, Corwin E, Cox LM, **Blaser MJ**. Infant antibiotic exposures and early life body mass. Int J Obesity 2012; doi: 10.1038/ijo.2012.132.
8. Cox LM, Cho I, Young SA, Anderson WHK, Waters BJ, Hung SC, Gao Z, Mahana D, Bihan M, Alekseyenko AV, Methe B, **Blaser MJ**. The non-fermentable dietary fiber hydroxpropyl methylcellulose modulates intestinal microbiota. FASEB J 2013; 27:692-702.
9. Redel H, Gao Z, Li H, Alekseyenko AV, Zhou Y, Perez-Perez GI, Weinstock G, Sodergren E, **Blaser MJ**. Quantitation and composition of cutaneous microbiota in diabetic and non-diabetic men. J Infect Dis 2013; 207:1105-1114.

D. Other Support

ACTIVE

3UH2AR057506-01S1 (NIH/NIAMS) Blaser (PI) 09/15/2011 – 08/31/2013
Evaluation of the cutaneous microbiome in psoriasis
The major goal of this work is to examine the relationship of the cutaneous microbiome to the inflammatory disorder psoriasis. High throughput sequencing technologies will be used to understand the diversity and abundance of microbiome constituents.

1R01 GM63270-11 (NIH/NIGMS) Blaser (PI) 09/8/2000 – 08/31/2013
Mathematical models of *H. pylori* gastric colonization
The goal of this project is to understand the biological basis of *H. pylori* persistence in the human stomach.

1R01DK090989-02 (NIH/NIDDK) Blaser (PI) 9/25/2010 – 8/31/2015
Disappearing gastrointestinal microbiota in epidemic obesity
The major goal of this Transformative R0-1 (T-R01) project is to test the hypothesis that changes in the gastrointestinal tract microbiota, particularly disappearances (extinctions), is fueling epidemic human obesity.

W81XWH-11-1-0739 Blaser (PI) 09/26/2011-09/25/2014
US Army Medical Research and Materials Command
The initiative in the human microbiome and infectious diseases
This project is designed to study the microbiome in human and animal model skin infections to develop new approaches to treatment

JDRF Award 17-2012-360 Blaser (PI) 02/01/2012 – 01/31/2014
Juvenile Diabetes Research Foundation
Effect of early life antibiotic exposure on type 1 diabetes in NOD mice.
This Strategic Research Agreement is aimed to determine whether early life antibiotic exposure (STAT) accelerates the pathophysiology and onset of type 1 DM in NOD mice.

BIOGRAPHICAL SKETCH			
NAME **Koonin, Steven E.**		POSITION TITLE **Director, NYU CUSP; Professor of Information, Operations & Management Sciences, Stern School of Business; Professor of Civil and Urban Engineering, Polytechnic Institute of NYU**	
eRA COMMONS USER NAME			
EDUCATION/TRAINING			
INSTITUTION AND LOCATION	DEGREE	YEAR	FIELD OF STUDY
California Institute of Technology, Pasadena Massachusetts Institute of Technology , Boston	B.S. Ph.D.	1972 1975	Physics Theoretical Physics

A. Personal Statement

My research interests have included nuclear astrophysics; theoretical nuclear, computational, and many-body physics; and global environmental science. I have been involved in scientific computing throughout my career and am a strong advocate for research into renewable energies and alternate fuel sources. My academic research in computational and nuclear physics has impacted the direction of science both nationally and internationally.

B. Positions and Honors

Professional Positions

1975 – 1981	Assistant Professor, Caltech
1976 – 1977	Research Fellow, Neils Bohr Institute
1981 – 2004	Professor, Theoretical Physics, Caltech
1989 – 1991	Chair of Faculty, Caltech
1995 – 2004	Provost, Caltech
2004 – 2009	Chief Scientist, BP p.l.c.
2009 – 2011	Under Secretary for Science, U.S. Department of Energy
2011 – 2012	Adjunct Staff Member, Institute for Defense Analyses
2012 – Present	Founding Director, Center for Urban Science and Progress, New York University
2012 – Present	Professor, Civil and Urban Engineering, Polytechnic Institute of New York University
2012 – Present	Professor, Information, Operations & Management Sciences, Stern School of Business, New York University

Other Experience and Professional Memberships

Former Advisory Committee Member, Department of Energy, National Science Foundation, and the Department of Defense
Former Member, Trilateral Commission
Member, U.S. National Academy of Sciences
Member and Past Chair, JASON Study Group
Member, Council on Foreign Relations
Fellow, American Physical Society
Fellow, American Association for the Advancement of Science
Fellow, American Academy of Arts and Sciences
Chief Scientist, BP: I developed the long-range technology strategy for alternative and renewable energy sources. I managed the firm's university-based research programs and played a central role in establishing the Energy Biosciences Institute at the University of California Berkeley, the Lawrence Berkeley National Laboratory, and the University of Illinois at Urbana-Champaign.
Under Secretary for Science, U.S. Department of Energy: I functioned as the Department's Chief Scientific Officer, coordinating and overseeing research across the DOE. I led the preparation of the Department's

2011 Strategic Plan and was the principal author of its Quadrennial Technology Review. In particular I championed research programs in High Performance Simulation, Exascale Computing, Inertial Fusion Energy, and Greenhouse Gas Monitoring, Reporting, and Verification. I also provided technical counsel on diverse nuclear security matters.

Center for Urban Science and Progress, NYU: As Director of CUSP, I oversees an innovative consortium of partners that currently includes six international universities, multiple Fortune 500 corporate research offices, four national labs, and 15 New York City and State agencies.

Awards and Nominations

1972	George Green Prize for Creative Scholarship, Caltech
1972	National Science Foundation Graduate Fellowship
1977 – 1979	Alfred P. Sloan Foundation Fellow
1985	Alexander von Humboldt Foundation Senior U.S. Scientist Award
1988	Department of Energy's E. O. Lawrence Award
1994	Fusion Power Associates Leadership Award
1999	Ernest Orland Lawrence Award in Physics

C. Selected peer-reviewed publications

I recently returned to academia after eight years in the private and government sectors. During my career at Caltech, I produced more than 200 peer-reviewed research publications, and authored or edited three books, including a pioneering textbook on Computational Physics in 1985. As Under Secretary of Science at the U.S. Department of Energy, I was the lead author on several major agency strategy documents (listed below) and co-authored an article for the National Academies of Science.

1. U.S. Department of Energy, Report on the First Quadrennial Technology Review: Technology Assessments (DOE/S-0002), Washington, DC.: 2012

2. U.S. Department of Energy, QTR: Report on the First Quadrennial Technology Review (DOE/S-0001), Washington, DC: 2011.

3. U.S. Department of Energy, *Strategic Plan*, Washington, DC: 2011.

4. **Koonin SE**, and Gopstein AM. Accelerating the Pace of Energy Change*,* Issues in Science & Technology, Washington, DC: National Academies of Science (Winter 2011).

BIOGRAPHICAL SKETCH	
NAME **Patrinos, Aristides A.N.**	POSITION TITLE **Professor, NYU-Poly and Deputy Director for Research, NYU Center for Urban Science and Progress**
eRA COMMONS USER NAME	

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR	FIELD OF STUDY
National Technical University of Athens, Athens, Greece	Diploma	1970	Mechanical & Electrical Engineering
Northwestern University, Evanston, Illinois	Ph.D.	1975	Mechanical Engineering and Astronautical Science

A. Personal Statement

I am a Professor of Mechanical and Biomolecular Engineering at NYU-Poly and the Deputy Director for Research at NYU Center for Urban Science and Progress.

My career path has spanned several scientific disciplines such as aerodynamics, atmospheric flows, environmental chemistry, climate change, radiation biology, and genomics. My involvement in the Human Genome Project led me to biological research, including genomics and synthetic biology. At Synthetic Genomics, Inc., among other R&D activities I led the metagenomics efforts for enhancing methane production in coal beds.

B. Positions and Honors

Professional Positions

1975 - 1976 Assistant Professor, Department of Mechanical and Aerospace Sciences, University of Rochester, Rochester, New York
1976 - 1980 Research Scientist, Engineering Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee
1980 - 1988 Research Scientist, Dept. of Applied Sciences, Brookhaven National Laboratory, Upton, NY
1988 - 1993 Research Scientist, U.S. Department of Energy, Washington D.C.
1993 - 2006 Director for Biological and Environmental Research, Office of Science, U.S. Department of Energy, Germantown, Maryland
2006-2012 President, Synthetic Genomics, Inc., La Jolla, CA
2006-Present Distinguished Investigator, J. Craig Venter Institute, Rockville, Maryland
2012-Present Deputy Director for Research, NYU Center for Urban Science and Progress and Distinguished Industry Professor of Mechanical Engineering and Biomolecular Engineering, NYU-Poly, Brooklyn, NY

Other Experience and Professional Memberships

2006-Present Member, Board of Directors, Tsakos Energy Navigation, TNP listed NYSE, Athens, Greece
2007-2009 Member, Committee on “America’s Energy Future,” National Academy of Sciences, Washington, DC
2007 - 2009 Member, Committee Providing Strategic Advice to the Climate Change Science Program, National Academy of Sciences, Washington, DC
2007-Present Member, External Advisory Board, University of Pittsburgh Clinical and Translational Science Institute, Pittsburgh, Pennsylvania
2007-Present Member, External Advisory Board, University of Pittsburgh Clinical and Translational Science Institute, Pittsburgh, Pennsylvania
2010-2011 Member, National Academy of Sciences Committee on the Economic and Environmental Impacts of Increasing Biofuels Production

Awards and Nominations

1999 ComputerWorld Smithsonian Platinum Technology Award
1999 Presidential Rank Award at Meritorious Level
2001 U.S. Secretary of Energy's Gold Medal
2001 Presidential Rank Award at Distinguished Level
2003 U.S. Secretary of Energy's Gold Medal
2004 Presidential Rank Award at Meritorious Level

Societies

Fellow of the American Association for the Advancement of Science
American Geophysical Union
American Society of Mechanical Engineers
National Technical Society of Greece
Fellow of the American Meteorological Society

C. Selected publications

1. **Patrinos AAN** and Bradley RA. Energy and Technology Policies for Managing Carbon Risk. Science, Aug 2009; 325: 949 - 950.

2. Viola MV and **Patrinos AA**. A neuroprosthesis for restoring sight. Acta Neurochir Suppl, Jan 2007; 97(Pt 2): 481-6.

3. **Patrinos AAN**. Biotechnology reenergized. The Scientist. Vol. 19; pp. 20-21. March 14, 2005.

4. Frazier M, Johnson G, Thomassen D, Oliver C and **Patrinos AAN**. Realizing the potential of the genome revolution: the genomes to life program. Science. Vol. 300; pp. 290-293. April 11, 2003.

5. Collins F, Morgan M and **Patrinos AAN**. The human genome project: lessons from large-scale biology. Science Vol. 300; pp. 286-290. April 11, 2003.

6. The International Genome Consortium, Initial Sequencing and Analysis of the Human Genome. Nature. Vol. 409; pp. 860-921. February 15, 2001.

7. **Patrinos AA**. The Human Genome Project: interaction of the physical sciences with biology. J Law Med Ethics. 2000 Winter; 28(4 Suppl):54-64.

8. Collins F, **Patrinos AAN**, Jordan E, Chakravarti A, Gesteland R, Walters L and DOE and NIH Planning Group Members. New goals for the U.S. human genome project: 1998-2003. Science. Vol. 282; pp. 682-689. October 23, 1998.

9. **Patrinos AAN** and Drell DW. Introducing the human genome project: its relevance, triumphs, and challenges. Judges' Journal Vol. 36; No. 3. Summer 1997.

10. **Patrinos AAN** and Drell DW. The human genome project: view from the Department of Energy. Journal of the American Medical Women's Association. Vol. 52; No. 1. Winter 1997.

11. **Patrinos AAN**, Leach MJ, Brown RM, Tanner RL and Binkowski FS. An acid rain study in the Washington, DC area. Journal of Applied Meteorology. 28 948-968, 1989.

12. **Patrinos AAN**. The impact of urban and industrial emissions on mesoscale precipitation quality. Journal of the Air Pollution Control Association. 35 No. 7; 719-727, July 1985.

BIOGRAPHICAL SKETCH			
NAME **Silva, Cláudio T.**		POSITION TITLE **Professor, Polytechnic Institute of NYU and Head of Disciplines, CUSP**	
eRA COMMONS USER NAME			
EDUCATION/TRAINING			
INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR	FIELD OF STUDY
Federal University of Ceará , Brazil	B.S.	1990	Mathematics
State University of New York at Stony Brook	M.S.	1993	Applied Mathematics
State University of New York at Stony Brook	Ph.D.	1996	Computer Science
State University of New York at Stony Brook	Postdoc	1997	Applied Mathematics

A. Personal Statement

I am an expert on the data analysis and visualization of Big Data. Over the last several years, I have developed new tools to enable the study of complex diverse datasets of the type that will be generated by this project. These include VisTrails, a workflow and provenance management system, and DEFOG, a new data analysis framework, both of which have been used to enable genomics studies. I will work closely with the rest of the team to develop tools and techniques for data management, analysis, and visualization for the project.

B. Positions and Honors

Professional Positions

1997-1999 Research Staff Member, IBM T. J. Watson Research Center
1998-2000 Adjunct Assistant Professor, State University of New York at Stony Brook
1999-2002 Senior MTS, AT&T Labs-Research
2002 Principal MTS, AT&T Labs-Research
2003 Faculty Scholar, Lawrence Livermore National Laboratory
2002-2006 Associate Professor, OGI School of Science & Engineering at OHSU
2003-2010 Associate Professor, University of Utah
2010-2011 Professor, University of Utah
2011-present Professor, Polytechnic Institute of New York University
2012-present Head of Disciplines, NYU Center for Urban Science and Progress

Other Experience and Professional Memberships

2002 DIMACS Workshop on Visualization and Data Mining
2002-2006 IEEE Transactions on Visualization and Computer Graphics Editorial Board
2003 DIMACS Implementation Challenge on Surface Reconstruction
2004 Papers co-chair, IEEE/SIGGRAPH Symposium on Volume Visualization and Graphics
2005-2006 Papers co-chair, IEEE Visualization
2006-present IEEE Computing in Science & Engineering Editorial Board
2008-present Computer & Graphics Editorial Board
2010 General co-chair, IEEE Visualization
2010-present Graphical Models Editorial Board
2011-present The Visual Computer Editorial Board
2013 Computer Graphics Forum Editorial Board
2013 Co-chair, IEEE Symposium on Large-Scale Data Analysis and Visualization

Program Committee Member (over 100 conferences, including)

IEEE Visualization; SIGGRAPH; Eurographics; Pacific Graphics; Symposium on Point-Based Graphics; Shape Modelling International; Symposium on Geometry Processing.

Awards

2005-2007 IBM Faculty Award (awarded three times)
2007 Dean's teaching commendation
2007 Best paper award at IEEE Visualization
2008 Best paper award at IEEE Shape Modeling International
2010 Best paper award at Eurographics EDUCATOR program
2011 Best paper award at ACM Eurographics Symposium on Parallel Graphics and Visualization
2013 IEEE Fellow

C. Selected peer-reviewed publications (selected from over 200 publications and 11 U.S. patents)

1. Morisette J, Jarnevich C, Holcombe T, Talbert C, Ignizio D, Talbert M, **Silva C**, Koop D, Swanson A, and Young N. VisTrails SAHM: visualization and workflow management for species habitat modeling. Ecography, to appear.

2. Santos E, Poco J, Wei Y, Liu S, Cook R, Williams D, and **Silva C**. UV-CDAT: Analyzing Climate Data sets from a User's Perspective. Computing in Science and Engineering, 2013.

3. Fekete J and **Silva C**. Managing Data for Visual Analytics: Opportunities and Challenges. IEEE Data Eng. Bull, 35(3):27–36, 2012.

4. Freire J and **Silva C**. Making Computations and Publications Reproducible with VisTrails. Computing in Science and Engineering, 14(4):18–25, 2012.

5. Ha L, Prastawa M, Gerig G, Gilmore J, **Silva C** and Joshi S. Efficient Probabilistic and Geometric Anatomical Mapping using Particle Mesh Approximation on GPUs. International Journal of Biomedical Imaging, 2011.

6. Ferreira N, Lins L, Fink D, Kelling S, Wood C, Freire J, and **Silva C**. BirdVis: Visualizing and Understanding Bird Populations. IEEE Transactions on Visualization and Computer Graphics (Proceedings of InfoVIS 2011), 17(12):2374–2383, 2011.

7. Bauer B, Carr LD, Evertz HG, Feiguin A, Freire J, Fuchs S, Gamper L, Gukelberger J, Gull E, Guertler S, Hehn A, Igarashi R, Isakov SV, Koop D, Ma PN, Mates P, Matsuo H, Parcollet O, Pawlowski G, Picon JD, Pollet L, Santos E, Scarola VW, Schollwck U, **Silva C**, Surer B, Todo S, Trebst S, Troyer M, Wall ML, Werner P, and Wessel S. The ALPS project release 2.0: open source software for strongly correlated systems. Journal of Statistical Mechanics: Theory and Experiment (JSTAT), 2011.

8. Ha L, **Silva C**, Krueger J, Comba J, and Joshi S. Optimal Multi-Image Processing Streaming Framework on Parallel Heterogeneous Systems. 11th Eurographics Workshop on Parallel Graphics and Visualization (EGPGV 2011), 2011. **Best paper award**.

9. Koop D, Santos E, Mates P, Vo HT, Bonnet P, Bauer B, Surer B, Troyer M, Williams DN, Tohline JE, Freire J, and **Silva C**. A Provenance-Based Infrastructure for Creating Executable Papers. Procedia Computer Science, 2011. ICCS 2011. **Grand Challenge Finalist.**

10. Anderson E, Potter K, Matzen L, Shepherd J, Preston G, and **Silva C**. A User Study of Visualization Effectiveness Using EEG and Cognitive Load. Computer Graphics Forum (Proceedings of EuroVis 2011). **Best paper award – 2nd prize.**

11. Vo H, Summa B, Osmari D, Comba J, Pascucci V, and **Silva C**. Streaming-Enabled Parallel Dataflow Architecture for Multicore Systems. Computer Graphics Forum (Proceedings of EuroVis 2010).

12. **Silva C**, Anderson E, Santos E, and Freire J. Using VisTrails and Provenance for Teaching Scientific Visualization. Computer Graphics Forum, 30(1):75–84, 2011.(Presented at EUROGRAPHICS 2010 Educator Program, 2010). **Best paper award.**

13. Scheidegger CE, Vo HT, Koop D, Freire J, and **Silva C**. Querying and Creating Visualizations by Analogy. IEEE Transactions on Visualization and Computer Graphics, 13(6):1560-1567, 2007. **Best paper award at IEEE Visualization 2007.**

BIOGRAPHICAL SKETCH	
NAME **Healton, Cheryl**	POSITION TITLE **Dean, Global Public Health and Professor Wagner School of Public Service, New York University**
eRA COMMONS USER NAME	

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEAR	FIELD OF STUDY
New England College, Henniker, NH	BA	1975	Sociology/Psychology
New York University, Robert Wagner School of Public Administration, New York, NY	MPA	1978	Health Policy and Management
Columbia University, Mailman School of Public Health, New York, NY (with distinction)	Dr.P.H.	1991	Sociomedical Sciences

A. Personal Statement

I am Dean of Global Public Health and Professor at Wagner School of Public Service, New York University. I am an active member of the public health community, serving on several boards including the Betty Ford Institute, the Lung Cancer Alliance, and the National Board of Public Health Examiners. As founding president and CEO of Legacy, a public charity created out of the tobacco Master Settlement Agreement, I have extensive experience in tobacco control issues. Prior to coming to Legacy, I served as PI and co-PI on a number of tobacco related grants at Columbia University where I was Professor of Clinical Public Health. I have authored numerous tobacco related peer reviewed publications and oversee a broad program of public education, research and policy related activities. I have extensive working knowledge of the NYC Department of Health and Mental Hygiene. As part of this proposal to map the NYC metagenone, I will facilitate interactions with NYC health agencies and provide a conduit to other health departments or government authorities, for example to implement pilot research findings.

B. Positions and Honors

Professional Positions

1982-1987 Associate Dean and Assistant Vice President, Columbia College of Physicians and Surgeons, Health Sciences Div., New York, NY.
1987-1995 Associate Dean, Mailman School of Public Health at Columbia University, New York, NY.
1995-2000 Associate Dean and Chair of Department of Sociomedical Sciences, Mailman School of Public Health, Columbia University, New York, NY.
1999-present Professor of Clinical Public Health, Columbia University, Mailman School of Public Health, New York, NY.
2003-present Adjunct Professor in the School of Nursing and Health Studies, Georgetown University, Washington, DC.
2000-present President & CEO, American Legacy Foundation, Washington, DC
2013-present Dean, Global Public Health and Professor Wagner School of Public Service, New York University, New York, NY.

Other Experience and Professional Memberships

2001-present Advisory Committee- University of California, San Francisco (UCSF) Center for Tobacco Control Research & Education,
2002-present Commission on Substance Abuse at Colleges and Universities, The National Center on Addiction and Substance Abuse at Columbia University (CASA)
2002-2003 Member, Secretary of Health and Human Services Interagency Task Force on Tobacco, Subcommittee on Cessation.
2004-present Treasurer, C-Change (formerly National Dialogue on Cancer), Board of Directors.
2004-present Member, Lung Cancer Alliance, Board of Directors.
2005-present Secretary/Treasurer, National Board of Public Health Examiners.
2008-present Board of Directors, Betty Ford Institute

2008-present Board of Directors, National Coalition on Health Care
2008-present Board Ex-Officio, American Legacy Foundation

Honors and Awards

1991 Marisa de Castro Benton Prize for Dissertation Excellence, Columbia University
1992 APHA NYC Public Health Association Media Award – Citywide health education campaign
1997 HHS Secretary Public Health Award Perinatal HIV Transmission Prevention Campaign, HHS
2000 Public Health Achievement Award, NYC DOH
2003 American Lung Association Honoree, American Lung Association
2003 Social Justice Award, State of Hawaii
2004 New York Women's Agenda Star Honoree, New York Women's Agenda
2008 Recipient of the Troy R. Westmeyer Distinguished Alumnus/Alumna Award, New York University
2008 Donald A. Berreth Lecturer, Presented by the National Public Health Information Coalition
2009 LGBTQ Tobacco Control Leadership Award, Presented by The National LGBT Tobacco Control Network

C. Selected Peer-reviewed Publications (selected from 55 peer-reviewed publications)

1. Pierce, J, Messer, K, James, L, White, M, Kealey, S, Vallone, D, **Healton, C.** Camel No. 9 Cigarette Marketing Campaign Targeted Young Teenage Girls. Pediatrics. December 2009. (IN PRESS)
2. Holtgrave DR, Wunderink KA, Vallone DM, **Healton CG**. Cost-utility analysis of the National truth campaign to prevent youth smoking. American Journal of Preventive Medicine. 2009 May; 36(5):385-8. Epub 2009 Feb 11.
3. Gritz ER, Sarna L, Dresler C, **Healton CG**. Reducing carcinogen levels in cigarette smoke. Cancer Epidemiology, Biomarkers & Prevention. 2007 Oct; 16(10):2171.
4. Messeri PA, Allen JA, Mowery PD, **Healton CG**, Haviland ML, Gable JM, Pedrazzani SD. Do Tobacco Countermarketing Campaigns Increase Adolescent Under-reporting of Smoking? Addictive Behavior. 2007 July; 32(7):1532-6.
5. **Healton C**, Gritz ER, Davis KC, Homsi G, McCausland KL, Haviland ML, Vallone D. Women's Knowledge of the Leading Causes of Cancer Death. Nicotine & Tobacco Research, July 2007; 9(7): 761–768.
6. Gritz ER, Sarna L, Dresler C, **Healton CG**. Building a United Front: Aligning the Agendas for Tobacco Control, Lung Cancer Research, and Policy. Cancer Epidemiology, Biomarkers & Prevention, May 2007; 16(5).
7. Hund LM, Farrelly MC, Allen JA, Chou RH, St Claire AW, Vallone DM, **Healton C**. Findings and Implications from a National Study on Potential Reduced Exposure Products (PREPs). Nicotine and Tobacco Research, 2006; 8(6): 791-797.
8. **Healton C**, Vallone D, McCausland K, Xiao J, Green M. Smoking, Obesity, and their Co-occurrence in the United States: Cross Sectional Analysis. British Medical Journal, 2006; 1(333): 25-26.
9. **Healton C**, Watson-Stryker E, Allen J, Messeri P, Graham P, Stewart A, Dobbins D, Glantz S. Televised Movie Trailers: Undermining Restrictions on Advertising Tobacco to Youth, Archives of Pediatrics & Adolescent Medicine, 2006; 160(9): 885-858.
10. **Healton C**, Farrelly M, Weitzenkamp D, Lindsey D, Haviland L. Youth Smoking and Tobacco Industry Revenue. Tobacco Control,(15(2): 103-106, April 2006).

D. Other Support

COMPLETED

No ongoing or completed research support is available for the past three years. All previous research/grant support where I served as Principal Investigator (PI)/Program Director (37 total) were completed as of 2000.

BIOGRAPHICAL SKETCH

NAME	POSITION TITLE
Scheid, Paul E.	**Manager, Genomics Core Facility, Center for Genomics & Systems Biology, Dept. of Biology, New York University**
eRA COMMONS USER NAME	

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEAR	FIELD OF STUDY
New York University, New York, NY	BA	2010	Media Studies

A. Personal Statement

As the Genomics Core (GenCore) Facility Manager, I (1) consult with PIs and their lab members in conjunction with GenCore Facility bioinformatics staff to design, execute and analyze next-generation sequencing projects; (2) provide advice to labs concerning library construction; and (3) oversee next-generation sequencing operation and maintenance, including run monitoring, data-quality assessment, troubleshooting failed runs, and supervision of core technical staff. The GenCore Facility has successfully sequenced over 1600 samples with our HiSeq 2000 since implementing the system in October 2011. Frequent sequencing applications include: Directional RNA-seq, Single-Cell RNA-seq, Re-seq, ChIP-seq, *de novo*-seq, and more recently amplicon-seq.

For the past two years, I have been involved in all aspects of sequencing sample preparation, library construction and machine operation for three next-generation sequencing technologies at NYU: **LifeTech SOLiD** (instrument retired), **Illumina HiSeq 2000** (HiSeq 2500 upgrade due May 2013) and **Ion Torrent PGM**. I have also implemented multiple next-generation sequencing methods including low-input DNA-sequencing, PCR-free DNA-sequencing, single-cell RNA sequencing and whole transcript directional RNA-sequencing.

In addition to my GenCore management role, I have overseen several experimental design efforts in large-scale genomics consortia including modENCODE's *C. elegans* transcriptome and 3' UTR projects and the Center for the Study of Complex Malaria in India (CSCMi) amplicon sequencing project. **Most recently, I have spearheaded a project to characterize the microbial diversity present on the surface of paper currency. We have already optimized a method to extract, sequence, and analyze trace amounts of DNA from paper money. Preliminary results demonstrate that the "Monetary Microbiome" contains genetic material from an array of prokaryotic and eukaryotic organisms, many of which are relevant to human health. I am eager to apply the methods from the Monetary Microbiome project to sequence samples collected as a part of the Grand Challenge Project.**

B. Positions and Honors

Positions and Employment

2003 – 2004	Student Research Assistant, Department of Neuroscience, Cleveland Clinic Foundation, Cleveland, OH
2004 – 2005	Student Research Assistant, Department of Hematology/Oncology, Case Western Reserve University Medical School, Cleveland, OH
2006 – 2007	Research Assistant, Department of Hematology/Oncology, Case Western Reserve University Medical School, Cleveland, OH
2007 – 2009	Lab Manager and Assistant Research Scientist, Department of Biology, New York University, New York, NY
2007 – Present	Instructor and Tutor, Bespoke Education, New York, NY
2009 – Present	Manager, Genomics Core Facility, Center for Genomics and Systems Biology, New York University, New York, NY

Meetings and conferences attended and presented (* 1st author on abstract, ** co-author on abstract)

2005	Society for Neuroscience Meeting, Washington, DC, USA*
2006	International Society of Stem Cell Research (ISSR), Toronto, ON, CAN**
2006	American Society of Hematology (ASH) Meeting, Orlando, FL, USA**

2009 Life Technologies SOLiD User Meeting, Redwood City, CA, USA
2010 Life Technologies 'Sequencing at the Tipping Point' Meeting, San Diego, CA, USA
2012 Advances in Genome Biology and Technology (AGBT), Marco Island, FL*
2013 Advances in Genome Biology and Technology (AGBT), Marco Island, FL

C. Peer-reviewed Publications

1. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K, Agarwal A, Alexander RP, Barber G, Brdlik CM, Brennan J, Brouillet JJ, Carr A, Cheung MS, Clawson H, Contrino S, Dannenberg LO, Dernburg AF, Desai A, Dick L, Dosé AC, Du J, Egelhofer T, Ercan S, Euskirchen G, Ewing B, Feingold EA, Gassmann R, Good PJ, Green P, Gullier F, Gutwein M, Guyer MS, Habegger L, Han T, Henikoff JG, Henz SR, Hinrichs A, Holster H, Hyman T, Iniguez AL, Janette J, Jensen M, Kato M, Kent WJ, Kephart E, Khivansara V, Khurana E, Kim JK, Kolasinska-Zwierz P, Lai EC, Latorre I, Leahey A, Lewis S, Lloyd P, Lochovsky L, Lowdon RF, Lubling Y, Lyne R, MacCoss M, Mackowiak SD, Mangone M, McKay S, Mecenas D, Merrihew G, Miller DM 3rd, Muroyama A, Murray JI, Ooi SL, Pham H, Phippen T, Preston EA, Rajewsky N, Rätsch G, Rosenbaum H, Rozowsky J, Rutherford K, Ruzanov P, Sarov M, Sasidharan R, Sboner A, **Scheid P**, Segal E, Shin H, Shou C, Slack FJ, Slightam C, Smith R, Spencer WC, Stinson EO, Taing S, Takasaki T, Vafeados D, Voronina K, Wang G, Washington NL, Whittle CM, Wu B, Yan KK, Zeller G, Zha Z, Zhong M, Zhou X; modENCODE Consortium, Ahringer J, Strome S, Gunsalus KC, Micklem G, Liu XS, Reinke V, Kim SK, Hillier LW, Henikoff S, Piano F, Snyder M, Stein L, Lieb JD, Waterston RH. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science, 330(6012), 1775-87.

2. Finney MR, Fanning LR, Joseph ME, Goldberg JL, Greco NJ, Bhakta S, Winter DG, Forster M, **Scheid PE**, Sabe M, Pompili VJ, Laughlin MJ. (2010). Umbilical Cord Blood Selected CD133+ Hematopoietic Stem Cells Exhibit Vasculogenic Functionality In Vitro and In Vivo. Cytotherapy, 12(1), 67-78.

3. Gross JB, Protas M, Conrad M, **Scheid PE**, Vidal O, Jeffery WR, Borowsky R, Tabin CJ. (2008). Synteny and candidate gene prediction using an anchored linkage map of Astyanax mexicanus. Proceedings for the National Academy of Science, 105(51), 20106-11.

4. Bhakta S, Greco NJ, Finney MR, **Scheid PE**, Hoffman RD, Joseph ME, Banks JJ, Laughlin MJ, Pompili VJ. (2006). The safety of autologous intracoronary stem cell injections in a porcine model of chronic myocardial ischemia. Journal of Invasive Cardiology, 18(5), 212-8.

D. Other Support:

ACTIVE

1U19AI089676-01 J. Carlton (PI) 07/01/10-06/30/17
NIH/NIAID

Center for the Study of Complex Malaria in India. Malaria in India remains an enormous public health problem. The Center for the Study of Complex Malaria in India (CSCMi) will bring together a group of leading US experts in the fields of malaria parasite research to work in tandem with faculty at the National Institute of Malaria Research (NIMR) in New Delhi, the only institute in India dedicated to finding solutions to the problem of malaria through basic, applied, and operational field research. Indian malaria complexity, its variation by location, and its consequences for malaria severity and transmission, are the research focus of the CSCMi, which aims to develop the knowledge, tools, and evidence-based strategies needed to support the intervention and control programs of Indian government organizations, and to build research capacity in India and help train its next generation of malaria and mosquito vector biologists. Role: Assistant Research Scientist.

BIOGRAPHICAL SKETCH

NAME **Ghedin, Elodie**	POSITION TITLE **Associate Professor of Computational & Systems Biology**
eRA COMMONS USER NAME eghedin	

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEAR	FIELD OF STUDY
McGill University, Montreal, Canada	B.S.	1989	Biology
University of Quebec in Montreal	M.S.	1993	Environmental Sciences
McGill University, Institute of Parasitology	Ph.D.	1998	Molecular Parasitology
Laboratory of Parasitic Diseases, NIAID/NIH	Postdoctoral	1998-2000	Molecular Parasitology
Eukaryotic Genomics, TIGR	Postdoctoral	2000-2001	Genomics

A. Personal Statement

My research is aimed at generating critical insights about host-pathogen interaction, microbial and viral population structures, and how these impact emerging infectious diseases. I use functional and comparative genomics, computational and evolutionary biology, and molecular virology techniques. I am PI of a lung microbiome study to analyze infections present in the lungs of people with HIV and determine if there are correlations between specific microbiota and lung diseases. During my six years at The Institute for Genomic Research (TIGR) I initiated and led the Viral Genomics Group. There, my laboratory was the first to develop a high-throughput pipeline to perform rapid sequencing and analysis of influenza virus and coronavirus genomes from large specimen collections from various host species, allowing for a comprehensive understanding of the transmission and evolution of these viruses. My group's current focus is pathogen emergence and adaptation.

B. Positions and Honors.

Professional Positions

2001-2004 Collaborative Investigator, Parasite Genomics Group, TIGR
2004-2006 Assist. Investigator, Parasite Genomics Group and Viral Genomics Group, TIGR
2006-2009 Assist. Prof., Division of Infectious Diseases, Dept. of Med., U. Pittsburgh School of Med.
2009-present Secondary appointment, Dept. of Microbiology and Infectious Diseases, Graduate School of Public Health, U. Pittsburgh
2010-2012 Asst. Prof., Dept. of Computational and Systems Biology, U. Pittsburgh School of Med.
2010-present Member, Center for Vaccine Research
2012-present Assoc.Prof. (tenure), Dept. of Computational and Systems Biology, U. Pittsburgh SoM

Other Experience and Professional Memberships

2008-present Associated Editor ('08-'11), Deputy Editor ('11-present), *PLoS Neglected Tropical Diseases*
2009-present Member, American Society for Tropical Med. & Hygiene and American Society for Microbiology
2009-present Scientific advisory board member NIAID-funded Influenza Research Database
2009-present Editorial board member, *Genes*
2010-present Member, International Society for Influenza and Other Respiratory Virus (ISIRV)
2011 Co-Editor, Viral Genomics section, *Current Opinion in Virology*
2011-present Member, International Society for Computational Biology (ISCB)
2011-present Board member, Rosalind Franklin Society
2012-present Editorial board member, *Microbiome*

Honors & Awards

2010 Chancellor's Distinguished Research Award, Junior Scholar (U. Pittsburgh)
2011 MacArthur Foundation Fellow
2012 Kavli Frontiers of Science Fellow

Scientific Meetings Organized and Coordinated

Nematode Symbiosis (NemaSym), November 2010, Tucson, Arizona and July 2011, Corvallis, Oregon
Great Lakes Bioinformatics Conference 2012, Ann Arbor, Michigan and May 14-16, 2013, Pittsburgh, PA

C. Selected peer-reviewed publications (selected from 80 publications)

1. Saira K, Lin X, DePasse JV, Halpin R, Twaddle A, Stockwell T, Angus B, Cozzi-Lepri A, Delfino M, Dugan V, Dwyer DE, Freiberg M, Horban A, Losso M, Lynfield R, Wentworth DN, Holmes EC, Davey R, Wentworth D, **Ghedin E***. (2013) Sequence analysis of *in vivo* defective-interfering (DI)-like RNA of influenza A H1N1 pandemic virus. *J Virol*., accepted.
2. Lozupone C, Cota-Gomez A, Palmer BE, Linderman DJ, Charlson ES, Sodergren E, Mitreva M, Abubucker S, Martin J, Yao G, Campbell TB, Flores SC, Ackerman G, Stombaugh J, Ursell L, Beck JM, Curtis JL, Young VB, Lynch SV, Huang L, Weinstock GM, Knox KS, Twigg H, Morris A, **Ghedin E**, Bushman FD, Collman RG, Knight R, Fontenot AP; for the Lung HIV Microbiome Project (2013) Widespread colonization of the lung by *Tropheryma whipplei* in HIV infection. *Am J Respir Crit Care Med* [Epub ahead of print].
3. **Ghedin E***, Holmes EC, DePasse JV, Pinilla LT, Fitch A, Hamelin M-E, Papenburg J, Boivin G. (2012) Presence of oseltamivir-resistant pandemic A/H1N1 minor variants before drug therapy with subsequent selection and transmission. J. Inf. Dis. 206(10):1504-1511. PMC3475640
4. Kerr PJ, **Ghedin E**, DePasse JV, Fitch A, Cattadori I, Hudson PJ, Tscharke DC, Read AF, Holmes EC. (2012) The evolutionary history and attenuation of myxoma virus on two continents. PLoS Pathogens 8(10):e1002950 PMC3464225
5. **Ghedin E***, Laplante J, DePasse J, Wentworth DE, Santos RP, Lepow ML, Porter J, Stellrecht K, Lin X, Operario D, Griesemer S, Fitch A, Halpin RA, Stockwell TB, Spiro DJ, Holmes EC, St.George K. (2011) Deep sequencing reveals mixed infection of pandemic H1N1/2009 viruses and the emergence of oseltamivir resistance. J. Inf. Dis. 203(2):168-174. PMC3071067
6. **Ghedin E***, Fitch A, Boyne A, Griesemer S, DePasse J, Bera J, Zhang X, Halpin RA, Smit M, Jennings L, St George K, Holmes EC, Spiro DJ. (2009) Mixed Infection and the Genesis of Influenza Virus Diversity. J. Virol. Sep;83(17): 8832-8841. PMC2738154
7. **Ghedin E***, Wang S, Spiro DJ, …[+63 co-authors]. (2007) Draft Genome Sequence of the Filarial Nematode Parasite Brugia malayi Science 317:1756-1760. PMC2613796 Faculty of 1000 evaluation

D. Other Support

ACTIVE

U01 HL098962 E. Ghedin/A. Morris (MPI) 09/30/2009 – 09/29/2014
NIH/NHLBI **Pathogenesis of Obstruction/Emphysema and the Microbiome (POEM) in HIV**

Bill & Melinda Gates Foundation Project Y. Kawaoka (PI) 11/01/2009 – 10/31/2014
Subcontract from Board of Regents of the U. of Wisconsin System
High Throughput Identification of Influenza Virus Amino Acids Responsible for Human to Human Transmission

R01 AI093804-01A1 E.C. Holmes (PI) 12/01/2011 – 11/30/2013
NIH/NIAID **Genomic analysis of the canonical case of virulence evolution: Myxomatosis in Australia**

U54 GM088491 D. Burke (PI) 07/01/2009 – 06/30/2014
NIH/NIAID **Computational Models of Infectious Disease Threats – Models of Infectious Disease Agent Study (MIDAS)**

N01-AI-3-0071 K. Nelson (PI) 04/01/2012 – 03/31/2014
NIH/NIAID **Genome Sequencing Center (Influenza Project)**

U01 N. Kaminski (PI) 04/01/2012 – 03/31/2015
NIH/NHLBI **Sarcoidosis and AAT genomics and informatics center (SAGIC)**

1R56AI093930-01A1 G. Clermont/T. Ross (co-PIs) 04/01/2012 – 03/31/2013 (NCE)
NIH/NIAID **Predictive biosignatures of complicated Influenza A infection**

NSF P. Stock/H.Goodrich-Blair/D. Bird/E. Ghedin (co-PIs) 03/31/2009 – 02/28/2014
Nematode-Bacteria Symbioses Research Coordination Network

APPENDIX 3: BUDGET

A two year budget for the pilot project is shown below. Fringe costs for the Postdoc Fellow only have been included, and no F&A since this is a grant application for internal NYU funds.

PROPOSED BUDGET	FY 2013 Base Salary	Effort Months			2013-2014	2014-2015	Cumulative
		C	A	S	Year 1	Year 2	
PERSONNEL							
Postdoc Fellow (to be named)	50,000	12.00			50,000	51,250	101,250
GRADUATE STUDENT							
PhD Student (Julia Maritz)			9.00	3.00	8,403	20,569	28,972
TOTAL SALARY					58,403	71,819	130,222
FRINGE					14,500	14,863	29,363
TOTAL SALARY & FRINGE					**72,903**	**86,682**	**159,584**
EQUIPMENT					4,000	0	4,000
TRAVEL (domestic)					3,000	3,000	6,000
MATERIALS & SUPPLIES					37,724	42,692	80,416
TOTAL OTHER DIRECT COSTS					**44,724**	**45,692**	**90,416**
TOTAL DIRECT COST					**117,627**	**132,374**	**250,000**
TOTAL COST					**$117,627**	**$132,374**	**$250,000**

APPENDIX 4: BUDGET JUSTIFICATION

Personnel
Funds to cover the salary/stipend of one **Postdoctoral Fellow** and one **PhD student** for 2 years are requested.

Postdoctoral Fellow (to be named), 12 months effort, Year 1 and Year 2
A to-be-named Postdoc Fellow will be based in Dr. Martin Blaser's lab at NYU School of Medicine. This lab has extensive working knowledge of high-throughput bacterial sequencing and analysis of 16S rRNA amplicons from human samples.

PhD Student (Julia Maritz), 3 Summer months Year 1 and 9 Academic months Year 2
The PhD student (Julia Maritz, NYU Biology Year 1), will be based in Dr. Jane Carlton's lab at NYU Biology. This lab has extensive experience with amplicon sequencing of eukaryotic microbes, and all sequencing will occur at the Genomic Sequencing Core at NYU CGSB, where Dr. Carlton is the Director.

The Postdoc and PhD student will be expected to coordinate sample collection, processing and sequencing with members of the NYU CUSP team and with the Genomics Core at NYU CGSB. This arrangement will foster cross-school collaborations and an interdisciplinary approach to the project. They will also work closely with the teams of Drs Bonneau and Silva to use computational methods to explore and learn from species abundance data, and to manage and visualize the data mapped against geographic and demographic information determined by the drainage area.

Travel
$3,000 per year for a total of $6,000 is requested to cover the costs of travel associated with trips to the NYC Department of Environmental Protection for sample collection.

Materials & Supplies
$37,724 is requested in Year 1 and $42,692 in Year 2 for:

(1) supplies associated with the collection, storage and processing of sewage samples, including stocking of MetaGenome Kits for distribution to NYC Department of Environmental Protection, @ ~$6,000 in both years
(2) Illumina HiSeq sequencing reagents, Illumina index sequences, library preparation/DNA quantification materials, to be used for multiplexing samples on 10 lanes in Year 1 and 12 lanes in Year 2 of the Illumina HiSeq2500 in the Genomic Sequencing Core at NYU CGSB, @ ~$2750 per lane
(3) storage of NGS data generated from the 24 lanes of sequencing, and to be stored on the NYU high performance computer "Bowery" @ $2000 for 2 TB data, for a total of $4000 in Year 1 and $4000 in Year 2

Equipment
$4,000 is requested for one compute node (12 cores, 48GB RAM) to process the ~24 sequencing libraries on the NYU high performance computer "Bowery".