A. Personal Statement

I am a statistical geneticist with expertise in analysis of high dimension data for complex diseases and traits, including metabolomics (#1-#4), gene expression, methylation and DNA sequencing data from genome-wide association studies. This personal statement will focus on my research experience in metabolomics data analysis while my contributions to other areas could also be found in the section of Other Contributions to Science below.

I am the primary statistician in two international large scale metabolomics studies using samples from the landmark PREDIMED PREvención con Díeta MEDiterránea (PREDIMED) Trial, which demonstrated Mediterranean dietary pattern supplemented with either extra virgin olive oil or tree nuts significantly reduced the risk of major clinical CVD events using a prospective cohort of 7000 subjects. The two NIH funded metabolomics studies (R01 HL118264 for CVD endpoint and R01 DK 102896 for Type 2 diabetes endpoint) each selected ~1000 subjects from the PREDIMED trial using a case-cohort design and measured their metabolomic profile at baseline and later time points using the state-of-art liquid chromatography tandem mass spectrometry (LC-MS) platform at the Clary Clish lab at the Broad Institute, resulting in more than 2000 samples from each study (#1, #3, #4). In addition to advising on statistical methods to detect associations among disease outcome, dietary interventions and individual metabolites. My research group also developed methods to detect metabolomics networks and its association with disease outcomes, methods to identify metabolite artifacts from thousands of untargeted metabolites and method to normalize and combine metabolomics data generated from multiple studies. These applied and methodological papers are currently under review or being submitted.

I am also the statistician and co-first author on another large study to use metabolic profile with detailed characterization of plasma acylcarnitines to predict incident Type 2 diabetes in 6 years using two prospective cohorts in China with a total sample size of 2,130 from Beijing and Shanghai cities in China (#2). We developed a robust prediction model utilizing all metabolites from the experiment and applicable to cross-region disease prediction. We are currently extending our method to other clinical phenotypes available from these cohorts.


B. Positions and Honors

Positions and Employment

2002-2004 Teaching Assistant, Department of Statistics and Probability, Michigan State University
2004-2009 Research Assistant, Department Biostatistics, University of Michigan, Ann Arbor
2009-2015 Assistant Professor, Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health
2015-present Associate Professor, Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health

Other Experience

2008-present Member of analysis group: The 1000 Genomes Project

Honors and Awards

2002 Dr. Koo Fellowship, Michigan State University
2005-2008 Rackham Graduate School Travel Grant, University of Michigan, Ann Arbor, Michigan
2007,2008 American Society of Human Genetics, Pre-doctoral Presentation Award Finalist, 57th and 58th Annual Meeting
2008 Rackham Pre-doctoral Fellowship, University of Michigan, Ann Arbor, Michigan

C. Other Contributions to Science

(1) Epigenome-wide association study

I have been leading the data-generation, preprocessing, and statistical analyses for several large cohorts including Framingham Heart Study (FHS), MRC Asthma family panel, Saguenay–Lac-Saint-Jean asthma family panel (SLSJ) of Quebec, PAPA study (Poplogaeth Asthma Pryísgyol Abertawe: students and staff from Swansea University, UK) and the Genetics of Lipid-lowering Drugs and Diet Network (GOLDN) study (for methylation QTL mapping). In late 2013, I published a paper in *Nucleic Acids Research* (as last author) to first study cross-tissue prediction and recalibration of DNA methylation (#1). In August 2014, I published an epigenome-wide association study for fasting blood lipids in *Circulation* through my contribution in the FHS study (#2). In September 2014, I published a paper in *Human Molecular Genetics* (as first author) to discuss problems, existing statistical solutions and directions for cell heterogeneity in epigenome-wide association studies (#3). In February 2015, I published a paper in *Nature* (as first author). This study identifies novel therapeutic targets and epigenetic biomarkers for patient stratification in allergic diseases (#4). This paper also illustrates an approach to identify cell specific effect from mixture of cell types while avoid confounding due to cell heterogeneity, one of the most challenging problems in genome-wide DNA methylation studies.


(2) Gene expression quantitative trait loci (eQTL) mapping

In 2007, I published a widely cited eQTL paper in *Nature Genetics* (#1). In 2013, I published a paper in *Genome Research* (as first author) that used a new statistical method to improve power for eQTL mapping and developed an updated eQTL category for more than 14,000 genes (#2). In November 2014, I had a paper accepted by *Nature Communications* (as last author) (#4). This is the first study using genotype imputation to examine the effect of short insertion and deletion (indel), the second most abundant genetic variant, genome-wide on gene expression across tissues. In 2014, I published a paper in *Database* (as last author) to develop a new online tool for real-time two-SNP eQTL analysis (#3). This is the first on-demand tool to analyze arbitrary selection of SNPs
for eQTL analysis in real-time; the same analysis has been a standard analysis protocol for many other GWAS utilizing eQTL information, including a series of GIANT consortium papers.


(3) **Combining eQTLs with GWAS for complex diseases and traits**

I also successfully combined eQTL data with other GWAS, resulting in publications in very high impact journals (3 in Nature, 8 in Nature Genetics, 1 in Nature Review Genetics, 2 in American Journal of Human Genetics, 3 in PLoS Genetics, 1 in Human Molecular Genetics, 1 in Human Genetics). These are very large scale studies, where I led the eQTL analysis part combining our eQTL results with the GWAS findings, including length in early life, human adult height, body mass index, waist-hip ratio, osteoporosis-related traits, the 1000 Genomes Project, Grave’s disease, pancreatic cancer, diabetes, esophageal squamous cell carcinoma, pathway analysis in basal cell carcinoma, follicular lymphoma, B cell lymphoma, leukemia, and eQTL comparision across tissues (4 typical papers are listed below. For a complete list of these papers please see http://www.hsph.harvard.edu/liming-liang/publications/)


#2. Li, L. (other authors) and Liang, L. Using eQTL weights to improve power for genome-wide association studies: a genetic study of childhood asthma. *Front Genet* **4**, 103 (2013). PMCID3668139.


(4) **Genome-wide association study and sequencing data analyses**

I have developed efficient statistical methods and open source software packages and eQTL online databases to tackle challenges in genome-wide association studies including population stratification (#1), whole genome simulation and efficient association analysis of large datasets (family and unrelated sample). I am also actively participating in large scale genome-wide association studies and sequencing data analyses. In 2013, I published a paper in *Human Molecular Genetics* (as lead statistician) to report five novel Graves’ disease risk loci in Asian populations (#2). In 2013, I published a paper in *Bioinformatics* (as last author) introducing a tool my postdocs and I developed to identify sample swapping and contamination in RNA sequencing data (#3). In 2014, I published a paper in *Human Molecular Genetics* (as co-first author) to report 17 novel genetic loci for adult height in Asian populations (#4). This study is the largest genome-wide association study for human height in Asian populations involving more than 93,000 samples and explored population-specific genetic architecture of adult height in Asian vs. European populations. For a complete list of my GWAS papers, please see http://www.hsph.harvard.edu/liming-liang/publications/


A more complete list of my peer-reviewed publications can be found at:

D. Research Support

**ACTIVE**

<table>
<thead>
<tr>
<th>Grant ID</th>
<th>Title</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>R01 HL118264 (Hu)</td>
<td>Mediterranean diet, Metabolites, and cardiovascular Disease</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>NIH/NHLBI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R01 DK102896 (Hu)</td>
<td>Dietary Interventions, Metabolites, and Risk of Type 2 Diabetes</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>NIH/NIDDK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R01 ES022981 (Qi)</td>
<td>Environmental Obesogens and Weight Change in the POUNDS LOST Trial</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>NIH/NIEHS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8UL1TR000170 (Nadler)</td>
<td>Harvard Clinical and Translational Science Center (UL1)</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>NIH/NCRR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R01 GM105857-01A1 (Sunyaev)</td>
<td>Improving polygenic prediction using large next-generation data sets</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>BWH (NIH/NIGMS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U01 AI110397 (Phipatanakul/Gold)</td>
<td>School Inner-City Asthma Intervention Study</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>CHB (CBD/NIH/NIAID)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPA (Albert)</td>
<td>Biostatistics and Bioinformatics Branch</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>NIH/NICHD/DIPHR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**In-Progress**

<table>
<thead>
<tr>
<th>Grant ID</th>
<th>Title</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>R21 ES032700 (Qi)</td>
<td>Improving polygenic prediction using large next-generation data sets</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>NIH/NIEHS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Completed**

<table>
<thead>
<tr>
<th>Grant ID</th>
<th>Title</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>R25 GM078148 (Liang)</td>
<td>Determining the possible contribution by PFCs to adipogenesis</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>BWH (NIH/NIGMS)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Additional Funding**

<table>
<thead>
<tr>
<th>Grant ID</th>
<th>Title</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>R03 AI126433 (Liang)</td>
<td>Providing professional service in leading biostatistics, epidemiology, and bioinformatics societies</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>NIH/NICHD/DIPHR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Major Goals**

As a statistical/bioinformatics scientist, Professor Liang's primary duties are to collaborate with faculty members on the design and analysis of "omics" outcomes in large cohort studies, lead an independent research program developing new statistical methods for the analysis of "omics" data, and to teach and mentor graduate students in the department. Professor Liang is also expected to perform professional service in leading biostatistics, epidemiology, and bioinformatics societies.
Investigating the Link between Cancer and Neurodegenerative Disease

Major Goals: To further investigate the inverse epidemiologic association between cancer and Alzheimer’s Disease, to investigate the association between metformin use and the risk of both diseases, and to characterize the genetic overlap between them using available GWAS data.

R01 ES026317 (Mazumdar, M) 01/01/16-12/31/20 Role: Co-Investigator
NIH/NIEHS via Boston Children’s Hospital
“Does Arsenic Increase Risk of Neural Tube in a Highly-Exposed Population?”

The major goal of this R01 application is to conduct a population-based case control study in Bangladesh to investigate the associations between environmental arsenic exposure and neural tube defects.

R56 HL134356 (Christiani) 09/16/16-08/31/17 Role: Co-Investigator
NIH/NHLBI
Whole Blood MicroRNAs as Risk and Survival Biomarkers for ARDS

This project’s objective is to develop and validate miRNA biomarkers in whole blood of potential utility in early diagnosis, prognosis and treatment of ARDS.

U34 AG051418 (Murabito) 06/1/15-05/31/17 Role: Co-Investigator
Boston Univ (NIH)
Identifying epigenetic mechanisms underlying age-related disease risk in CHARGE

Liming will concentrate his work on Specific Aim 2: Perform pilot studies to identify the most useful epigenetic measurements for study. This will include pilot studies of emerging technologies to assess different epigenetic measurements, and their potential as mechanisms underlying age-dependent disease risk, and to develop new methods to predict epigenomic marks in different cell and tissue types.

Completed:
R01 GM104411 (Xiong) 04/01/13-01/31/17 Role: Consortium PI
NIH/NIGMS
Unified Methods for Sequence-based Association Studies

The goals of this study are to develop novel and powerful statistical methods for sequence-based association studies and QTL (eQTL) analysis which leverage high dimensional data reduction and functional data analysis techniques to integrate multiple variants across a given genomic region or a pathway, and unify family and population-based design to use various types of family and unrelated individual data sampled from any population structure.

R21 ES 024236 (Hauser) 07/02/14-06/30/16 Role: Co-Investigator
NIH/NIEHS
Environmental Chemicals, Exosomal miRNAs in Ovarian Follicles, and IVF Success

The objective of the proposed study is to determine if follicular fluid (FF) levels of BPA and DEHP metabolites are associated with miRNA profile in FF isolated exosomes and with oocyte maturation and day 3 embryo quality. We will also use mediation analysis to estimate the indirect effects of BPA and DEHP metabolites on oocyte maturation and embryo quality mediated through FF isolated exosomes miRNA expression.

R01 HL086601 12/1/11-11/30/15 Role: Co-Investigator
NIH (Raby)
Genetics and Gene Expression Profiling in Asthma

We propose to develop and apply novel statistical methods to expression quantitative trait locus (eQTL) mapping and subsequently combine the eQTL maps with asthma GWAS.