Prasad Patil

3 Blackfan Circle, CLSB 11078
Boston, MA 02115
prpatil42@gmail.com
prasadpatil.net

EDUCATION

Ph.D. in Biostatistics
Johns Hopkins University, Baltimore, MD, 2011-2016
Advisor: Jeffrey Leek
Dissertation Title: "Assessing Reproducibility and Value in Genomic Signatures"

B.A. in Mathematics
New York University, New York, NY, 2005-2008
Concentration: Applied Mathematics
Minor: Computer Science

RESEARCH EXPERIENCE

Postdoctoral Fellow 2016-present
Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

With Giovanni Parmigiani,

- Explored data and prediction combination strategies for multi-study/multi-syte machine learning
- Advised graduate and undergraduate students on projects related to multi-study learning: theoretical guidelines for when to combine studies, how to reconcile differing sets of features, transfering gene signatures from microarray to RNA-seq data.

Graduate Research 2012-2016
Department of Biostatistics, Johns Hopkins University, Baltimore, MD

With Jeff Leek,

- Identified and described bias due to data normalization in the prediction of breast cancer subtypes using gene expression information.
- Formalized feature selection and modeling with Top-Scoring Pairs for simple, decision-tree-based classifiers.
- Developed R packages for standardized analysis templating (`tdsm`) and interactive health visualizations (`healthvis`).
- Suggested 95% prediction intervals as a means of assessing whether a study result has been replicated.

With Jeff Leek and Michael Rosenblum,

- Used RCT baseline covariate adjustment methods to assess the additional value a genomic prediction can provide beyond standard clinical measurements in improving the precision of a treatment effect estimator.
- Examined the benefit of using machine learning methods to summarize the predictive value of large numbers of baseline covariates in an RCT setting.

Scientific Programmer 2009-2011
Center for Biomedical Informatics, Harvard Medical School, Boston, MA

- Developed synthetic patient simulation platform predicated on Bayesian networks, and accompanying web service.
- Built clinical trial simulation framework using synthetic patients, stochastic PK-PD models for drug clearence, and coded clinical trial protocols.
- Fosterd collaboration with Beth Israel Deaconess Medical Center and GenomeQuest, Inc. to create a commercial, clinical-grade omics analysis pipeline intended for hospital use.
- Administered a cloud computing environment via Amazon Web Services for day-to-day lab operations and omics pipeline deployement.

Bioinformatics Fellow 2008-2009
Partners Center for Personalized Genetic Medicine, Boston, MA

- Collaborated with the Center for Biomedical Informatics at Harvard Medical School to develop an open-source pipeline using next-generation sequencing (NGS) technology to detect and clincally annotate all variants in an individual human genome.
- Evaluated and tested NGS tools and restructured code for parallelization via Hadoop/MapReduce.
- Developed a comprehensive data model and modular scripted pipeline for variant annotation and reporting.

PUBLICATIONS

- **Patil P**, Parmigiani G (2018). Training replicable predictors in multiple studies. *Proceedings of the National Academy of Sciences*, 115(11), 2578-2583.
- **Patil P**, Peng RD, Leek, JT (2016). What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. *Perspectives on Psychological Science*, 11(4), 539-544.
- **Patil P**, Colantuoni E, Rosenblum MA, Leek JT (2016). Genomic and clinical predictors for improving estimator precision in randomized trials of breast cancer treatments. *Contemporary Clinical Trials Communications* 3: 48-54.
- **Patil P**, Leek JT (2015). Discussion of "Visualizing statistical models: Removing the blindfold". *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 8(4), 240-241.
- **Patil P**, Bachant-Winner PO, Haibe-Kains B, Leek JT (2015). Test set bias affects reproducibility of gene signatures. *Bioinformatics*, btv157.
- Hyland PL, Burke LS, Pfeiffer RM, Rotunno M, Sun D, **Patil P**, Wu X, Tucker MA, Goldstein AM, Yang XR (2014). Constitutional promoter methylation and risk of familial melanoma. *Epigenetics*, 9(5), 685-692.
- Fusaro, VA, **Patil P**, Chi CL, Contant CF, Tonellato PJ (2013). A Systems Approach to Designing Effective Clinical Trials Using Simulations. *Circulation*, 127(4), 517-526.
- Fusaro VA, **Patil P**, Gafni E, Wall DP, Tonellato PJ (2011). Biomedical Cloud Computing Using Amazon Web Services. *PLoS Computational Biology.* 7(8):e1002147.
- Wall DP, Kudtarkar P, Fusaro VA, Pivovarov R, **Patil P**, Tonellato PJ (2010). Cloud computing for comparative genomics. *BMC Bioinformatics* 11:259.

PRESENTATIONS AND POSTERS

Invited

- **Patil P***. Defining and evaluating reproducibility and replicability. WNAR 2017.
- **Patil P***. A statistical framework for discussing reproducibility and replicability. ENAR 2017.
- **Patil P***. Setting expectations for replication in science. Harvard Catalyst Biostatistics Symposium 2017.
- **Patil P***. Setting expectations for replication in science. AAAS 2017.

Contributed

- **Patil P***, Parmigiani G. Training replicable predictors in multiple studies. ENAR 2018.
- **Patil P***. Replicating scientific studies. Harvard Biostatistics Cancer Training Grant Working Group 2017.
- **Patil P***, Alquicira J, Leek JT. Measuring the value of GWAS results in a clinical trial setting. ASHG 2015 [poster].
- **Patil P***. What to expect when you're replicating. Hopkins Biostatistics Journal Club 2015.
- **Patil P***, Leek JT. Reproducibility and value of genomic signatures. JSM 2015.
- **Patil P***. On organization. Hopkins Biostatistics Computing Club 2015].
- **Patil P***, Leek JT. Assessing the reproducibility and value of genomic signatures. ENAR 2015.
- **Patil P***, Haibe-Kains B, Leek JT. Cross-platform gene signature development using Top-Scoring Pairs. JSM 2014.
- **Patil P***. Cross-validation in the presence of many features. Hopkins Biostatistics Journal Club 2014.
- **Patil P***, Haibe-Kains B, Leek JT. Cross-platform gene signature development using Top-Scoring Pairs. ENAR 2014.
- Chi CL, Fusaro VA, **Patil P**, Crawford MA, Contant CF, Tonellato PJ. An approach to optimal individualized warfarin treatment through clinical trial simulations. Proceedings of IEEE Cairo International Biomedical Engineering Conference (CIBEC) 2010, Cairo, Egypt.
- Chi, CL, **Patil P**, Fusaro VA, Kos PJ, Pivovarov R, Contant CF, Tonellato PJ. A Simulation Platform to Examine Heterogeneity Influence on Treatment. Proceedings of the 2010 American Medical Informatics Association Annual Symposium, Washington, DC.
- Chi CL, Kos PJ, Fusaro VA, Pivovarov R, **Patil P**, Tonellato PJ. Mining personalized medicine algorithms with surrogate algorithm tags. Proceedings of the First ACM International Health Informatics Symposium 2010 [poster].
- **Patil P***, Heus H, Arnaout R, Tonellato PJ. Refining a method for processing an individuals whole genome to clinical utility. CSHL Personal Genomes Meeting 2010. Cold Spring Harbor, N.Y.

- **Patil P***, Tonellato PJ. Individual Whole Genome Mapping: from NGS reads to clinical variants, American Medical Informatics Association Clinical Research Informatics Summit 2010, San Francisco, CA [poster].
- Fusaro VA, Kos PJ, Tector M, Tector A, **Patil P**, Tonellato PJ. Electronic Medical Record Analysis Using Cloud Computing, American Medical Informatics Association Clinical Research Informatics Summit 2010, San Francisco, CA [poster].
- **Patil P*.**Clinical algorithms for whole genome data Partners Health Care Information Systems Research Council Symposium 2009, Harvard Medical School, Boston, MA.
- **Patil P*.** Clinical annotation of an individual whole genome assembly, Center for Biomedical Informatics Research Day 2009, Harvard Medical School, Boston, MA.

* - Presenter

TEACHING

Guest lecturer, BIO622 2015
Conducted daily lecture for 500+ student course

Lead TA, BIO621-623 2014-2015
Prepared and held 2-3 sections per week, 40-60 students each
Beta-tested, proctored, and graded exams

TA, BIO611-612, 615, 621-624, AS.280.35 2012-2014
Graded homework and exams, held small sections and office hours

SERVICE

Refereeing: Biometrics, Biostatistics, Computational Statistics and Data Analysis, Genome Biology, Nature Neuroscience, PLOS ONE, IEEE Sensors, Canadian Journal of Statistics, The American Statistician

Organization
- Organizer (2017), Harvard Biostatistics Cancer Training Grant Working Group.
- Session chair, Next Generation Sequencing. ENAR 2014.
- Co-Organizer (2012), Hopkins Biostatistics Computing Club.

AWARDS

Jane and Steve Dykacz Award 2016
Departmental award for best student paper in medcial statistics
Awarded for "Genomic and clinical predictors for improving estimator precision in randomized trials of breast cancer treatments"

Helen Abbey Award 2016
Departmental award for teaching

JHSPH Student Assembly Teaching Assistant Recognition Award 2015
One of two voted on by students across all courses in JHSPH

NYU Honors Scholar 2008

NYU Deans List 2006-2007

National Merit Scholarship 2005-2008

SOFTWARE

R Packages:
`scifigure` (https://cran.r-project.org/package=scifigure)
Visualize Reproducibility and Replicability in a Comparison of Scientific Studies
`healthvis` (https://github.com/prpatil/healthvis)

Interactive health visualizations. Built using `d3`, `shiny`, `htmlWidgets`.

`tdsm` (https://github.com/prpatil/tdsm)

Templated Deterministic Statistical Machines. Automated analysis templates and standardized reports that can be edited and compared.

Languages:
R, Javascript, C/C++, Java, Perl, MATLAB, Stata, Hadoop, Shell scripting

REFERENCES Available upon request.