

# Dilsher Singh Dhillon

## Contact Information

dhillon.dilsher@gmail.com  
979-676-2387

3001 Murworth drive  
Houston, TX 77025

## Education

### Master's in Statistics

Texas A&M University, College Station, TX

3.91/4.00

December 2017

### Master's in Biotechnology

Texas A&M University, College Station, TX

3.42/4.00

May 2012

### Bachelor of Engineering in Biotechnology

Panjab University, Chandigarh, India

3.50/4.00

April 2010

## Applied Skills

- Multivariate Regression, penalized regression methods, tree-based regression techniques, time series regression, mixed effects modeling, resampling techniques, Unsupervised learning methods
- Design of experiments, survival analysis of censored data, analysis of longitudinal data, imputation techniques for missing data
- Develop fixed and mixed effects models to analyze experimental data
- Analysis of biological data using Bio-conductor, DNA sequence alignment, pre-processing and analyzing NGS data
- Extensive experience with programming in R, Python and SAS
  - Data 'tidying' and management using *tidyverse*, *data.table*, *pandas* and *numpy* libraries in R and Python
  - Visualization techniques using *ggplot* and *seaborn*
  - Parallel computing in R to execute programs for implementing computationally intensive procedures
  - Enabling reproducible analysis and documentation using R markdown and jupyter notebooks
  - Version control using github

## Experience

### Shell International Exploration and Production (Houston, TX)

#### Statistical Consultant

March 2019 - Present

- Provide Statistical support to researchers and engineers in Shell's fuels and lubricants business
  - Implement designs accounting for physical and financial constraints of the experiment
  - Automated a processing pipeline that takes in gigabytes of data generated by engine tests, merges files for relevant parameters and encodes meta-data information from filenames into the dataset.
  - Effectively communicate results to stakeholders using dynamic Rmarkdown reports
  - Developed a R package for use internally by the team that implements an opinionated project workflow and pre-designed Rmarkdown templates

### Baylor College of Medicine (Houston, TX)

#### Statistician

January 2018-March 2019

##### *Identifying risk factors for poor outcomes in co-infected TB-HIV patients*

- Data extraction and processing
  - Worked with monitoring and evaluation officers in Africa to develop MySQL script identify key variables to extract
  - Extracted longitudinal data for ~25000 subjects from 7 clinics in Sub-Saharan Africa in the time period 2013-2018
- Statistical Analysis
  - Developed a regression model to identify risk factors associated with poor TB outcomes
  - Presented findings of the project in the network wide meeting held in Johannesburg, South Africa (November 2018) – manuscript in preparation

##### *Studying underlying epigenetic mechanism of TB infection*

- Processed methylation data from EPIC arrays consisting of ~850,000 individual probes using the minfi package from Bioconductor
- Analyze flow cytometry data to understand mechanistic interactions

##### *Managing Clinical Research Database*

- Automated a pipeline in Python for merging of datasets from clinical research forms(CRFs) entered in eSwatini
- Worked with database management team to remove redundancies across datasets reducing the size and number of variables to be extracted

**Texas A&M University** (College Station, TX)

**Summer Project**

**June 2017 – November 2017**

*Develop statistical model to predict Red Light Violations in the City of Chicago*

- Extracted data available publicly on red light violations for 354 intersections in the city of Chicago for the time period 2014-2017
- Downloaded daily weather data available using the NOAA API for both airports in Chicago which includes potential predictors such as snowfall, precipitation, wind speed, visibility in addition to temperature conditions on a daily basis from the start of 2014
- Developed an ARIMA model to predict daily traffic violations

**Statistical Consulting**

**Jan 2017- May 2017**

*Evaluating efficacy of pesticides in reducing horn fly counts in cattle*

- Cleaned and pre-processed dataset obtained from client for horn fly counts over 26 weeks in cattle across 15 counties in Texas
- Developed a repeated measures mixed effects model to analyze outcomes of pesticides treatment under the assumption that the counts are correlated with each other
- Prepared a client report summarizing research goals, exploratory data analysis, statistical methods and key findings of the analysis. In addition, provided suggestions on how to improve experimental design in order to have greater power to detect differences

**University of Texas, MD Anderson Cancer Center** (Houston, TX)

**Research Investigator**

**Aug 2016-Jan 2018**

*Assay Development Platform*

- Promoted to a supervisory capacity to oversee assay development of novel biomarkers and testing of in-house and commercial assays in clinical samples for validation
- Manage the Specimen Manger database for recording subject characteristics, biospecimen banking and inventory and multi-center study inventory for the Lung Cancer Early Detection (LCED) trial

*Statistical Analysis*

- Assist scientists and postdocs with statistical analysis of proteomic and genomic data
  - Apply statistical methods such as hierarchical clustering, principal component analysis (PCA) and regression on multivariate data to identify features (genes/proteins) associated with a particular phenotype
  - Prepare publication quality figures and tables for manuscripts
- Integrated genomic data for normal and cancer tissues/cell lines from CCLE, TCGA and BioGPS with proteomic data generated for 30 breast cancer cell lines on the Mass Spectrometry

**Sr. Research Assistant**

**Jan2013-Aug2015**

*Biomarker panel development for early detection of lung cancer*

- Translated two ELISA assays for proteins differentially expressed in lung cancer into a bead-based assay using the Luminex MagPix platform and achieved higher sensitivity and better inter-assay precision through development and optimization
- Executed an expanded panel of 34 protein biomarkers for lung cancer assembled into in-house as well as MilliPlex designed panels on various pre-diagnostic and diagnostic lung cancer subjects

**Publications**

Guida F, Sun N, Bantis LE, Muller DC, Li P, Taguchi A, **Dhillon D**, Kundnani DL, Patel NJ, Yan Q, Byrnes G, Moons KGM, Tjønneland A, Panico S, Agnoli C, Vineis P, Palli D, Bueno-de-Mesquita B, Peeters PH, Agudo A, Huerta JM, Dorronsoro M, Barranco MR, Ardanaz E, Travis RC, Byrne KS, Boeing H, Steffen A, Kaaks R, Hüsing A, Trichopoulou A, Lagiou P, La Vecchia C, Severi G, Boutron-Ruault MC, Sandanger TM, Weiderpass E, Nøst TH, Tsilidis K, Riboli E, Grankvist K, Johansson M, Goodman GE, Feng Z, Brennan P, Johansson M, Hanash SM *Assessment of Lung Cancer Risk on the Basis of a Biomarker Panel of Circulating Proteins* (JAMA Oncology 2018)

Shiels MS, Kirk GD, Drummond MB, **Dhillon D**, Hanash SM, Taguchi A, Engels EA *HIV Infection and Circulating Levels of Prosurfactant Protein B and Surfactant Protein D* (Journal of Infectious Diseases 2018)

Çeliktas M, Tanaka I, Tripathi SC, Fahrman JF, Aguilar-Bonavides C, Villalobos P, Delgado O, **Dhillon D**, Dennison JB, Ostrin EJ, Wang H, Behrens C, Do KA, Gazdar AF, Hanash SM, Taguchi A *A Role of CPS1 in Cell Growth, Metabolism and Prognosis in LKB1-Inactivated Lung Adenocarcinoma* (JNCI 2017)

Capello M, Bantis LE, Scelo G, Zhao Y, Li P, **Dhillon D**, Patel NJ, Kundnani DL, Wang H, Abbruzzese JL, Maitra A, Tempero MA, Brand R, Firpo MA, Mulvihill SJ, Katz MH, Brennan P, Feng Z, Taguchi A, Hanash SM *Sequential Validation of Blood-Based Protein Biomarker Candidates for Early-Stage Pancreatic Cancer* (JNCI 2017)

**Honors**

|   |                  |
|---|------------------|
| <b>International Education Scholarship</b> , Texas A&M University | <b>2012</b>      |
| <b>Regents Fellowship</b> , Texas A&M University                  | <b>2010-2011</b> |

