Jasvinder Ahuja, Ph.D.

Portland, OR · (734) 233-8896 · jasvinderahuja@gmail.com · https://www.linkedin.com/in/jasvinderahuja/ · GitHub · Citizenship: USA

Data analysis expert with 7+ years of experience using Linux based high performance clusters (HPCs; similar to AWS or Azure) for distributed computing using task managers (Snakemake), R, Python, and SQL for applying analysis techniques and machine learning to uncover patterns and identify trends.
Effective communicator experience in presenting progress to scientific advisory boards, company stakeholders (100+), at international conferences (10+) in addition to years of classroom teaching experience.
Lead author of publications in top scientific journals including *Science* (2017).
**Note– My biology research accomplishments are listed to highlight my data gathering, analysis, collaboration, and presentation skills.*

TECHNICAL PROFICIENCY

Data Exploration– Univariate and bivariate statistics, **Modeling**– Regression, classification & segmentation.
R– Rstudio, R-Markdown, tidyverse, ggplot2, dbplyr, R-Shiny, genomic-ranges, etc.
Python (Jupyter, NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, etc)
OS and resource management – Linux, Mac, PC, bash, conda, snakemake, and cluster computing.
Bioinformatics Suite– Sequence-QC, alignment, variant-calling, Heatmaps, Samtools, Seurat, Qiime2, etc.
MS Office Suite– PowerPoint, Word & Excel (VLOOKUP, HLOOKUP, PIVOT_TABLES, charts, etc.)

EDUCATION AND PROFESSIONAL DEVELOPMENT

Business analysis, Business analytics in Excel, SQL, Tableau and Power-BI (Purdue Univ.)	Expected 2024
Bioinformatic Analysis of Next Generation Sequencing Data (NIH)	2016
Practical R (NIH)	2016
Ph.D. Regulatory Biology – Cleveland State University	2014

PROFESSIONAL MEMBERSHIP

International Institute of Business Analysis (Portland OR, Chapter)

PROFESSIONAL EXPERIENCE

Refreshing and Developing skills to transition to data analysis role in industry ***March 2024-present***
Purdue University - Post-Graduate Program in Business Analysis (IIBA accredited; expected August 2024)
AnalytixLabs - Machine Learning in Python with real world case studies (expected June 2024)

Research Assistant Professor - Oregon Health & Science University ***May 2022 – Feb 2024***

- Co-managed two postdoctoral fellows to conduct research projects ($1M/year NIH funding).
- Co-managed sequencing of DNA from patients with infertility (n>1500) from 20 global centers and its analysis.
- Deployed Population Sampling Probabilities (PSAP) modeling to identify causal mutations in infertility patients –
  - Managed DNA preparation, library preparation and sequencing
  - Only 2% (60M base-pairs, bp) of our genome codes for proteins (exome), ~2-5GB of sequence data per person.
  - This data is in fragments of strings of length = 150 bp (bp='A', 'T', 'G' or 'C'). These strings are then aligned to the reference genome.
  - Human reference genome is 3 billion (B) bp long organized in 26 fragments of string (chromosomes).
  - Per individual ~1-3B, 150 bp strings are matched to the reference genome, for variant identification.
  - Each person has about 40,000 distinct variations from the published reference genome!
  - We used PSAP modeling developed in-house to identify infertility causing variations (mutations).
  - The premise is that disease causing variations absent in genomes of healthy controls (ref. GenomAD with >190,000 healthy exome sequences). Furthermore, the causal variation (mutation) needs to make sense based on function of the gene that carries it.
- to identify systematic, or batch effects samples were segmented using PCA, and K-means clustering etc.
- Defects in genome segregation into sperm causes non-obstructive azoospermia (NOA; untreatable by IVF). I secured NIH funding for functional analysis of causal mutations – as a step to development of therapy.
- Presented results to national scientific advisory boards, global collaborators, and at international conferences.
- IMPACT: Individuals with mutations, who could not be helped by in *vitro* fertilization (IVF) methods were identified, sparing them the agony of going through multiple painful IVF cycles.

Research and Postdoc Fellow – National Cancer Institute (NCI, NIH) – Bethesda, MD *Feb 2015 – May 2022*

- Most DNA breaks are repaired by homologous recombination (HR, which is scarless) including ones caused by modern therapeutic reagents like *CRISPR*-Cas9.
- Repair via HR can be easily studied using the physiological process of meiosis where chromosomal DNA inherited from mom and dad breaks and recombines leading to formation of gametes (egg and sperm). To achieve this, I-
  - Designed 50 variants (~one string long) in a DNA span of 6Kb flanking a frequent DNA break site in the yeast model system.
  - Then I analyzed ~600 gametes per experiment where I extracted DNA and determined sequence combination of the 50 polymorphic markers (designed variants).
  - Using sequencing >100M sequences (length = 150), >1200 molecules (length 6000) were reconstructed.
  - Developed custom scripts in R, Python, and bash to determine combination of 50 variations among millions of DNA reads.
  - Wrote scalable code to run on Biowulf cluster at NIH (with >95,000 cores and RedHat Linux Enterprise).
- In conclusion, based on our results we proposed a radically improved model for DNA repair via homologous recombination, where we found that DNA template selection is a dynamic and reversible process – published in *Molecular Cell (2021)*.
- IMPACT: This research improved our understanding of how DNA from mom and dad is recombined before being packaged into sperm and egg.

Graduate Assistant - Cleveland State University & Cleveland Clinic – Cleveland, OH Jan 2006 – Dec 2014

- For my Ph.D., I explored the mechanisms that ensure inheritance of exactly one copy of genome via gametes (egg or sperm) which is vital for maintenance of two copies of genome among most eukaryotes including human beings.
- IMPACT: We discovered that the protein degradation machinery localizes to chromosomes during meiosis and is essential for gamete formation. This discovery was published in top STEM journal *Science (2017).*
- Mentored 11+ undergrad and grad students in molecular genetics research setting.
- Classroom Teaching: undergraduate level (>20 credit hrs.) and graduate level (9 credit hrs.), avg. class ≥ 20.
- Developed plugins for ImageJ, used excel-VBA, and R – to analyze microscopy images for pattern detection.

SELECTED PODIUM PRESENTATIONS

2020 - "Mechanisms of homologous recombination during meiosis." Department of Reproductive and Development Sciences & Department of Genetics at Oregon National Primate Research Center, Portland, OR.
2018 - "Fine-structure analysis of meiotic recombinants reveals that divergent mechanisms lead to formation of noncrossovers and crossovers" at Gordon Research Seminar, Colby Sawyer College, New London NH.
2018 - "Fine-structure analysis of meiotic recombination products" Mid Atlantic Mitosis and Meiosis Meeting, John Hopkins University, Baltimore, MD.

SCIENTIFIC PUBLICATIONS (Complete list @PubMed)

Ahuja, J.S., Sandhu, R., Huang L., Klein F., Börner G.V. (2024). Temporal and Functional Relationship between Synaptonemal Complex Morphogenesis and Recombination during Meiosis. *bioRxiv*, 2024.01.11.575218.

Ahuja, J.S., Harvey C., Wheeler, D.L. and Lichten, M.J. (2021). Repeated strand invasion and extensive branch migration are hallmarks of meiotic recombination. *Mol. Cell.* 81:4258–4270.e4. PMID: 34453891.

Ahuja, J.S., Sandhu, R., Mainpal, R., Lawson, C., Henley, H., Hunt, P.A., Yanowitz, J.L., Borner, G.V. (2017). Control of meiotic pairing and recombination by chromosomally tethered 26S proteasome. *Science*. 355: 408-11. PMID: 28059715.

Ahuja, J.S., and Börner, G.V. (2011). Analysis of meiotic recombination intermediates by two-dimensional gel electrophoresis. *Methods Mol Biol*. 745:99-116. PMID: 21660691.