Sean, Soonweng, Cho, PhD

Fremont CA 94536 • soonwengcho@gmail.com • +1(909)-618-7354
github.com/sean-cho • linkedin.com/in/seanswcho • sean-cho.github.io

Professional Summary

- Effective bioinformatician with 9 years computational biology and 8 years experimental biology experience in the oncology, immunology, and molecular biology space.
- Excels in a dynamic, interdisciplinary setting with clinicians, biostatisticians, and experimental scientists.
- Developed numerous R packages, Shiny apps, and Python modules for genomics analysis and visualization.
- Led NGS pipeline development for local HPC and cloud computing with version control through Github.
- Integrated multi-omics findings with biological models for target discovery, resulting in numerous publications, research projects, and grants.
- Spearheaded development of a harmonized multi-omics database to enable efficient bioinformatics discovery.

Relevant skills

Research Topics:	Cancer genomics, biomarkers, immuno-oncology, mosaic and somatic variant calling
Programming:	R, Bioconductor & Shiny, Python & conda, Bash, SQL, Git, Github
Data Analysis:	Biostatistics, survival analysis, machine learning, NGS pipeline development (Snakemake, Nextflow, CWL), multi-omics (genomics, epigenomics, transcriptomics, 10x Genomics)
Data mining:	TCGA/ICGC, 1000 Genomes, GTEx, GEO
Platforms:	SGE, Slurm, Amazon Web Services, Seven Bridges Genomics, Docker

Education

Ph.D.	2016	Cellular and Molecular Medicine	Johns Hopkins School of Medicine
B.Sc.	2008	Biotechnology	California State Polytechnic University Pomona

Professional Experience

Arcus Biosciences — Hayward, CA

Scientist, Bioinformatics — *June 2019 – Current*

Broad bioinformatics functionalities to support data-driven drug discovery.

- Lead bioinformatics functions of early stage, preclinical immuno-oncology projects in multi-functional teams.
- Analyze Phase 1/2 clinical trial NGS data for biomarker discovery.
- Create data harmonization framework and internal database of over 8TB of -omics data to support drug discovery and translational decisions, such as hypothesis testing, model selection, and indication assessment.
- Develop analysis frameworks, computational methods, and software for -omics data analysis.
- Spearhead efforts for data-driven discovery of new drug targets using bioinformatics approaches.
- Design and implement infrastructure, NGS pipelines, and Shiny applications.
- Manage and mentor direct reports and interns to successful completion of several projects.
- Establish relationships with vendors for bioinformatics needs and NGS applications.

Kennedy Krieger Institute & Johns Hopkins University — Baltimore, MD

Postdoctoral Fellow, Lab of Dr. Jonathan Pevsner — *Dec 2016 – May 2019*

Characterization of somatic mosaicism in neurobehavioral disorders

- Analyzed NGS (genomic, 10x Genomics, and transcriptomic) data to identify variants associated with disease.
- Designed phase-aware computational methods for somatic variant discovery and prioritization that led to increased specificity in variant calling.
- Developed and maintained Dockerfiles and bioinformatics pipelines for high performance cluster or cloud computing to enable efficient genomic analysis and promote reproducibility.
- Generated a 14TB public resource of 2,500 genomic variant call format (GVCF) files from 1000 Genomes to increase variant calling sensitivity using the Seven Bridges Genomics cloud computing platform.

Johns Hopkins University School of Medicine — Baltimore, MD

PhD Candidate, Labs of Drs. Christopher Umbricht and Saraswati Sukumar — *Sept 2010 – Nov 2016*

Thesis: Scarcity and sparsity – Genomics studies in a low resource setting

Multi-omics analysis of archival tissue for biomarker discovery in breast and thyroid cancer

- Designed statistical framework & computational tools for integrative genomic & epigenomic analysis.

- Developed the Epicopy R package for estimating copy number variation from methylation microarrays, effectively increasing the amount and type of data obtained.
- Optimized protocols for DNA/RNA extraction from archival tissue and microarray analysis leading to greater nucleic acid yield and microarray quality.
- Actively collaborated on interdisciplinary teams to make decisions and discoveries for grants and research projects, resulting in thirteen publications.
- Led the development of bioinformatics research strategy, generated pilot data, and successfully co-wrote two competitively funded grants totaling $300,000.

Pfizer La Jolla, CA

Graduate Intern, Oncology: Computational Biology Group *Aug 2015 – Oct 2015*

Multi-omics analysis of breast cancer

- Led the analysis of genomic and transcriptomic data of a unique breast cancer cohort, resulting in a peer-reviewed publication in Nature Communications.

City of Hope National Cancer Center Duarte, CA

Research Associate, Lab of Dr. Michael Jensen *June 2008 – May 2010*

Immuno-oncology research in chimeric antigen receptor (CAR) T-cells

- Designed and constructed lentiviral vectors of CARs for T-cell therapy as a translational research team member.

California State Polytechnic University, Pomona Pomona, CA

Undergraduate Research, Lab of Dr. Wei-Jen Lin *Oct 2007 – June 2008*

Development of a transposon system for gene delivery in Gram-positive bacteria

- Modified and optimized the use of transposon Tn5 for large gene delivery through electroporation.

Selected Publications (5 out of 19; complete list at tinyurl.com/seanswcho)

Cho S., Kim H., Cope L., Umbricht C. (2019) Measuring DNA copy number variation using high-density methylation microarrays. Journal of Computational Biology

Merino V.F.*, Cho S.* (equal contribution), Nguyen N., et al. (2018) Induction of cell cycle arrest and inflammatory genes by combined treatment with epigenetic, differentiating, and chemotherapeutic agents in triple-negative breast cancer. Breast Cancer Research

Kan Z., Ding Y., Kim J., Jung H.H., Chung W., Lal S., Cho S., et al. (2018) Multi-omics profiling of younger Asian breast cancers reveal distinctive molecular signatures. Nature Communications

Kim HS, Umbricht CB, Illei PB, Cimino-Matthews A, Cho S., Chowdhury N., et al. (2016) Optimizing the use of gene expression profiling in early stage breast cancer. JCO

Fackler M.J., Bujanda Z.L., Umbricht C., Teo W.W., Cho S., Zhang Z., …, Sukumar S. (2014) Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. Cancer Research

Invited Talks

Cho, S. (2014) "Epicopy: Measuring DNA copy number using Illumina 450K methylation microarrays". Computational Genomics Symposium, Johns Hopkins Hospital

Selected Posters

Cho, S., et al. (2016) "Characterization of metastatic follicular thyroid cancer by RNA-seq." 2016 Annual Meeting, American Thyroid Association, Denver CO

Cho, S., et al. (2015) "Epicopy: Measuring DNA copy number variation using Illumina high density methylation microarrays." AACR Annual Meeting 2015, Philadelphia PA

Cho, S., et al. (2015) "Epicopy: Measuring DNA copy number variation using Illumina high density methylation arrays." Safeway Breast Cancer Retreat 2015, Baltimore MD

Teaching Experience

Bio-Trac, Bioscience Education Center Germantown, MD

Guest Lecturer, Advances in Epigenetics: Epigenomics Lecture and Lab *2018*

Johns Hopkins University Baltimore, MD

Teaching Assistant, Practical Genomics Workshop *2018*

Tutor (Cellular & Molecular Medicine), Molecular Biology and Genomics *2012*