Sanjar Hudaiberdiev, PhD

Staff Scientist, NCBI/NIH

URL: https://okurman.github.io
Email: kyrgyzbala@gmail.com, hudaiber@nih.gov
Phone: +1 443 758 1688

Scientific research areas

Genomics, Gene Regulation, Disease Genetics (Type 2 Diabetes)
Evolution, Origin of Life
Machine/Deep Learning, Statistics

Technical Skills & Domain Knowledges

Biology	Gene regulation (enhancers, silencers), 3D chromatin organization, condensates
	Type 2 Diabetes: pancreatic islet (β) cells, islet-specific TFs, risk loci (GWAS)
	PopGen: common/rare variant analysis, LD blocks, GWAS, PRS
	Evolution: non-coding regions, ultra-conserved regions, human-specific evolution (HARs, hCONDELs), protein(domain) evolution, phylogenetics/phylogenomics genomic synteny, host-parasite arms race
ML/stats	Statistical Inference, Biostatistics
	Bayesian modeling, Probabilistic Reasoning, Causality
	ConvNets, LSTMs, Transformers
ML tools	Tensorflow, PyTorch, Keras
	Scikit-learn, NumPy/SciPy, R, MATLAB/Octave
Bioinformatics	GWAS, fine-mapping, e/s/raQTLs, MPRA
	ChIP-seq, RNA-seq, ATAC-seq, Hi-C
	ENCODE, Roadmap, HCA, 4D Nucleome, UK Biobank
	Containers (Docker, Singularity), Workflows (Nextflow, Snakemake)
	Massive data processing, parallel computing, HPC (SLURM, SGE)
Soft. Dev.	OOP, functional programming, Unit tests, Git, CI/CD
Languages	**Actively using**: Python, R, Rust (learning), Bash
	Have used: MATLAB, C#, C/C++, Perl
Databases	MySQL, PostgreSQL, SQLite, SQL Server 2008/2010, Oracle 10g.
OS	Linux, MacOS

Work Experience

2023 –	**Staff Scientist. NCBI, NIH, Bethesda, MD**
	- Integrating multi-omics (TF/Histone ChIP-seq, DHS, ATAC-seq, HiC, RNA-binding proteins' binding sites, methylome) data to model disease transcriptional codes using multi-modal fusion deep neural networks to detect disease-causal genetic variation.
	- Using Deep Learning to assess the phenotypic impacts of human-specific (vs. Macaques and Chimpanzees) INDELs in enhancers.

2020 – 2023 | Research Fellow. NCBI, NIH, Bethesda, MD

- Developed TREDNet: A novel Deep Learning method for enhancer detection.
- Lead a large multidisciplinary collaboration project to apply TREDNet to detect GWAS candidate causal mutations linked to Type 2 Diabetes and experimentally verified.
- Developed a model for explaining the high-occupancy target (HOT) regions in humans using transcriptional condensates.

2018 – 2020 | Postdoctoral Fellow. NCBI, NIH, Bethesda, MD

Mentor: Ivan Ovcharenko, PhD

- Spearheaded the transition to Deep Learning methods to study the enhancers.
- Automated the distributed training of large models (Keras+Tensorflow) on Google Cloud and Biowulf (NIH's HPC cluster).
- Applied ConvNets and LSTMs to model gene regulation and decompose LD blocks.
- Managed and implemented the deployment of decades-worth codebase developed in-house to containers (Docker, singularity) and workflows management systems (Nextflow, Snakemake).

2015 – 2018 | Postdoctoral Fellow. NCBI, NIH, Bethesda, MD

Mentor: Eugene Koonin, PhD

- Studied the general evolution of CRISPR/Cas systems, as well as deep focus on protein CRISPR-associated 4 (Cas4). Paper is cited >55 times, including by Prof. Jennifer Doudna, and has given me an Erdös number of 3.
- Analyzed evolutionary links between microbial adaptive immune systems and other selfish elements.
- Gained an expert understanding of microbial genetics, comparative genomics, molecular evolutionary analysis.

2011 – 2014 | Graduate research program ICGEB, Trieste, Italy

- **Thesis project**: Functional annotation of *Quorum Sensing* systems in bacteria using subsystem based approach.
- Implemented a wide range of algorithms and tools of computational genomics and sequence analysis.

2010 – 2011 | Software developer, Octopus, Bishkek, Kyrgyzstan

- Developed a concept for data analysis module for an existing commercial software for micro-financing.
- Implemented financial data analysis module in *.NET* environment using *C#* and *MS SQL Server*.
- Developed an ETL module to consolidate the data from government and third-party vendors using *Pentaho* and *Python*.

2009 – 2010 | Oracle Database Developer, Demirbank, Bishkek, Kyrgyzstan

- Designed and implemented backup strategies of Oracle database and decreased the backup time by a factor of 10.
- Lead the project of transitioning the existing infrastructure to Oracle Blade Modules.
- Conducted SQL tuning and optimization which lead to reduction in ATM transaction latencies during the peak times from 10 seconds to < 1 second.

2008 - 2009 | Research assistant, Istanbul Technical University, Turkey

Mentor: Prof. Zehra CATALTEPE.

- Designed and implemented Machine Learning experiments on protein function prediction using profile-HMMs.
- Obtained a formal graduate training on classical Machine Learning (PRML by Christopher Bishop)

Education

09/2011 - 09/2014	PhD in Bioinformatics, **ICGEB**, Trieste, Italy Thesis: "Computational analysis of quorum sensing systems in bacterial genomes: Developing automated annotation tools." Supervisor: Prof. Sándor PONGOR
09/2006 - 09/2008	MSc in Machine Learning, **Istanbul Technical University**, Istanbul, Turkey
09/2002 - 09/2006	BSc in Computer Engineering, **Erciyes University**, Kayseri, Turkey

Publications

C. Hill, **S. Hudaiberdiev**, I. Ovcharenko
ChromDL: A Next-Generation Regulatory DNA Classifier ISMB/ECCB 2023 (in submission)
bioRxiv (https://doi.org/10.1101/2023.01.27.525971)

S. Hudaiberdiev, I Ovcharenko
Sequence characteristics and an accurate model of abundant hyperactive loci in the human genome
eLife 2023 (in submission)
bioRxiv (https://doi.org/10.1101/2023.02.05.527203)

S. Hudaiberdiev, D. Taylor, W Song, N Narisu, RM. Bhuiyan, H. Taylor, T Yan, A Swift, L Bonnycastle, DIAMANTE Consortium, M Stitzel, M Erdos, I Ovcharenko, F. S. Collins
Modeling islet enhancers using deep learning identifies candidate causal variants at loci associated with T2D and glycemic traits *PNAS. 2023 (being reviewed)*
medRxiv (https://doi.org/10.1101/2022.05.13.22275035)

EZ Kvon, Y Zhu, G Kelman, CS Novak, I Plajzer-Frick, M Kato, T H Garvin, Q Pham, AN Harrington, R Hunter, J Godoy, E Meky, J Akiyama, V Afzal, S Tran, F Escande, B Gilbert-Dussardier, N Jean-Marçais, **S Hudaiberdiev**, I Ovcharenko, MB Dobbs, CA Gurnett, S Manouvrier-Hanu, F Petit, A Visel, DE Dickel, LA Pennacchio
Comprehensive in vivo interrogation reveals phenotypic impact of human enhancer variants. *Cell. 2020*

Hudaiberdiev S., Shmakov S, Wolf YI, Terns M, Makarova KS and Koonin EV **2017**.
Phylogenomics of Cas4 family nucleases. BMC Evol. Biol. 2017 PMID: 29179671

Hudaiberdiev, S., Choudhary, K.S., Vera, R., Gelencser, Z., Lamba, D. and Pongor, S. **2015**.
Census of solo luxR genes in bacteria. Frontiers in Cell. Infect. Micriobiol. 2015. doi:10.3389

Dogsa, I; Choudhary K.S; Marsetic, Z; **Hudaiberdiev, S**; Vera, R; Pongor, S, Mandic-Mule, I **2014**.
ComQXPA quorum sensing systems may not be unique to Bacillus subtilis: a census in prokaryotic genomes. PLoS One. 9(5):e96122. doi: 10.1371

Choudhary, K.S.; **Hudaiberdiev, S.**; Gelencsér, Z.; Coutinho, B.G.; Venturi, V.; Pongor., S. **2013**.
The organization of the Quorum Sensing luxI/R Family Genes in Burkholderia. Int J Mol Sci. 13727

Gelencsér, Z.; Choudhary, K.S.; Coutinho, B.G.; **Hudaiberdiev, S.**; Galbáts, B.; Venturi, V.; Pongor., S. **2012** *Classifying the topology of AHL-driven quorum sensing circuits in proteobacterial genomes.* Sensors. 12(5), 5432-5444.

Gelencsér, Z.; Galbáts, B.; Gonzalez, J.F.; Choudhary, K.S.; **Hudaiberdiev, S.**; Venturi, V.; Pongor, S. **2012**. Chromosomal arrangement of AHL.driven quorum sensing circuits in Pseudomonas. ISRN Microbiology. 2012, 6.

Oral presentations

2023	*Sequence characteristics and an accurate model of abundant hyperactive loci in the human genome* NLM IRP Seminar
2022	*Transcriptional regulation and disease genetics using Deep Learning* NLM IRP Seminar
2021	*Deep causality in non-coding GWAS SNPs.* NLM Summer Lectures Series. NLM
2020	*Using Deep Learning to to infer causality of GWAS SNPs.* METU, Ankara, Turkey
2019	*Recurrent Language model for enhancers.* NLM Summer Lectures Series. NLM
2019	*Understanding the grammar of enhancers using Deep Learning.* NHGRI
2019	*Causal mutations in ZRS enhancer region using Deep Learning.* LBNL, Berkeley
2017	*Phylogenomics of Cas4 family nucleases.* AiChE. Raleigh, NC
2016	*Current topics in biological sequence analysis.* ICGEB. Trieste, Italy

Posters

2021	Hyperactive loci within non-coding regulatory regions of human genome *Mechanisms of Eukaryotic Transcription*, CSHL, NY.
2019	Two-stage Deep Learning cross-tissue predictor of enhancers shows unprecedented accuracy in detection of causal regulatory variants. *Mechanisms of Eukaryotic Transcription*, CSHL, NY.
2019	Using Deep Learning to understand the grammar of enhancers. *RECOMB*, Washington DC
2017	Phylogenomics of Cas4 family nucleases. *International conference on CRISPR Technologies*, Raleigh, NC
2013	Computational approaches to microbial communication/quorum sensing signaling. *BAGECO*, Ljubljana, Slovenia

Awards

- Fellows Award for Research Excellence 2019, NIH, Bethesda, MD
- ICGEB PhD fellowship 2011, ICGEB, Trieste, Italy
- Scholarship for research in Machine Learning/Bioinformatics. 2008, TUBITAK, Istanbul, Turkey
- Scholarship for BS final year project. 2007, TUBITAK, Kayseri, Turkey
- Scholarship of Turkish ministry of education for undergraduate studies. 2002/2007, Turkey
- Bronze medals in national physics olympiads for highschool students. 2001/2002, Kyrgyzstan

Teaching & Mentorship

- Co-mentor, Post-bac student, 2020 NIH
- Co-mentor, Summer research student, 2019 NIH
- Lead instructor, Summer intern journal club, 2018 NIH
- Co-mentor, Summer research student, 2018 NIH

Peer-reviewing for Scientific Journals

BMC Bioinformatics, NAR Database, Frontiers in Microbiology

Organization & Leadership

Commitee
2023	Fellow Award for Research Excellence. Committee member
2022	Fellow Award for Research Excellence. Committee member

Organised
2013	Bioinformatics: Computer Methods in Molecular Biology –ICGEB, Trieste, Italy
2014	Bioinformatics: Computer Methods in Molecular Biology –ICGEB, Trieste, Italy

Languages

Kyrgyz	Native
English, Turkish	Fluent
Italian, Russian	Advanced