Andrey O. Kislyuk, Ph.D.

Contact Information

kislyuk@gmail.com
Tel: +1 (925) 683-0172
San Francisco, CA U.S.

kislyuk.com
github.com/kislyuk
linkedin.com/in/kislyuk

Overview

Technical leader in software engineering for scalable, data-heavy applications. Expert in design and implementation of cloud platforms for consumer applications. Open source developer ranked among top 40 in the Python ecosystem.

Specialties and areas of expertise

- Technical leadership: Drive success through effective cross-functional communication, customer focus, frequent incremental changes, testability, observability, safety, and user-focused design.
- Software engineering: Python, Typescript/Javascript, Rust, Go, Java; Linux; Terraform/HCL, SQL, Bash/Zsh.
- Testing and validation: Testability engineering, CI/CD systems, release engineering, configuration management.
- Observability: Distributed tracing (Honeycomb), alerting, post-mortem analysis.
- Distributed systems: RDBMS (PostgreSQL/MySQL)/NoSQL/document DBs, distributed queues, autoscaling, load balancing, SOA.
- Cloud computing: AWS (EC2, S3, IAM, Lambda, RDS, ECS, EKS, Batch), GCP.
- Machine learning: Application integration, deep learning fundamentals, Tensorflow, NLP, GPU optimization/MLops orchestration, probabilistic graphical models, numerical computing (NumPy, SciPy, Pandas, scikit-learn).
- Containerization and orchestration: Docker, Kubernetes, low level Linux containers (cgroups, namespaces, LXC).
- High performance computing: Cloud HPC cluster orchestration, batch processing reliability engineering.
- API design: Protobuf/gRPC/Thrift/OpenAPI, autodocs, codegen tools, SDK automation.
- Information security: AuthN/AuthZ (OIDC/OAuth2, SAML2, IAM), application security, data governance.

Genomics and Bioinformatics

- Genome sequence analysis: mapping, assembly, annotation, base calling, genome finishing, repeat masking, phylogenomics, metagenomics, consensus genome building.
- Clinical variant analysis: variant calling, phasing, haplotype resolution, variant effect prediction, variant information management, normalization and nomenclature management.
- Population genomics and epidemiology: joint variant calling, statistical analysis and development of disease risk scores.
- Genomics visualizaiton and data analysis technologies: genome browsers, report building systems, visualization techniques and libraries.
- Sequencing technology R&D: assay development and validation, application of DNA sample preparation technologies.

Employment

Color Health, May 2021 –
Principal Software Engineer: Technical lead responsible for all software infrastructure for genetic testing for hereditary disease risk (cancer, cardiovascular disease, and pharmacogenomic testing). Built high performance distributed cloud compute platforms for data processing. Led team to modernize the entire patient experience for Color genetic diagnostic products, allowing Color's product offering to lead the market in usability and accessibility. Introduced a new suite of telemetry and monitoring systems to improve product performance assessment. Rebuilt reporting system to align Color clinical reports with the latest medical guidelines. Served as Color's tech lead for contract with National Institutes of Health's All of Us program, supporting precision medicine research and the presentation of research and clinical data to participants. Managed team of 5 engineers.

Chan Zuckerberg Initiative, January 2017 – May 2021
Principal Software Engineer: Software R&D for the Human Cell Atlas and IDseq/CZ ID projects. Technical lead responsible for the development of cloud computing platforms for Human Cell Atlas, a project to map the expression profiles of all human cells; and CZ ID, an online platform for analyzing pathogen DNA in environmental and human samples. Developed novel cloud HPC orchestration systems for analysis of SARS-CoV-2 and other pathogen genomes, including phylogenetic classification of disease outbreaks and AMR analysis.

DNAnexus, Inc., October 2011 – February 2016
Director, Software Architecture: Technical lead on the full stack of the DNAnexus Platform, a global genomics service. Developed multi-tenant cloud computing architectures, new API paradigms, SDKs and services to enable at-scale processing of production genomics datasets for companies such as Natera (prenatal and cancer genetic diagnostics) and Regeneron (exome population sequencing for phenotype-genotype correlation and drug discovery). Built pioneering cloud HPC architectures utilizing EC2 and S3 services. Led developer experience team to develop SDK automation; built authentication, authorization, and cloud containerization orchestration systems; led core performance engineering efforts to scale platform services. Managed team of 3 engineers.

Pacific Biosciences of California, January 2010 – October 2011
Scientist, bioinformatics: Research, development, and support of various bioinformatics projects. Built a suite of data products and R&D software to detect DNA methylation in PacBio's single-molecule, real-time long-read DNA sequencing product. Published multiple patents and collaborations with Nobel prize-winning researchers on discovery of novel DNA methylation mechanisms. Supported genomic analysis of pathogens in epidemic outbreaks.

Education

Georgia Institute of Technology, Atlanta, Georgia — Advisor: Joshua S. Weitz
Ph.D. 2010, Bioinformatics

University of California, Berkeley — Advisor: Inna Dubchak
B.A. 2004, Computer Science, Mathematics, and Statistics

Projects

argcomplete (github.com/kislyuk/argcomplete) - a Python command-line autocompletion framework (1.4K GitHub stars)

yq (github.com/kislyuk/yq) - a command-line YAML/XML/TOML processor (2.7K stars)

pyotp (github.com/pyauth/pyotp) - the Python One-Time Password Library (3K stars)

watchtower (github.com/kislyuk/watchtower) - a Python logging observability framework for AWS CloudWatch (700 stars)

SignXML (github.com/XML-Security/signxml) - the Python XML Signature and XAdES library (140 stars)

Domovoi (github.com/cloud-utils/domovoi) - AWS Lambda event handler manager (170 stars)

Taxoniq (github.com/taxoniq/taxoniq) - Taxon Information Query - query engine for NCBI Taxonomy and related data (50 stars)

Over 20 of these and other highly successful open source software packages are used widely in the Python and other software ecosystems. In a recent algorithmic ranking of Python software developers, I was ranked #40.

Publications

See Google Scholar Profile for details.

K.L. Kalantar, T. Carvalho, C.F.A. de Bourcy, B. Dimitrov, G. Dingle, R. Egger, J. Han, O.B. Holmes, Y.F. Juan, R. King, **A.O. Kislyuk**, M.F. Lin, M. Mariano, T. Morse, L.V. Reynoso, D.R. Cruz, J. Sheu, J. Tang, J. Wang, M.A. Zhang, E. Zhong, V. Ahyong, S. Lay, S. Chea, J.A. Bohl, J.E. Manning, C.M. Tato, and J.L. DeRisi, "IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring", GigaScience, doi:10.1093/gigascience/giaa111, 2020.

D. A. Rasko, D. R. Webster, J. W. Sahl, A. Bashir, N. Boisen, F. Scheutz, E. E. Paxinos, R. Sebra, C.-S. Chin, D. Iliopoulos, A. Klammer, P. Peluso, L. Lee, **A.O. Kislyuk**, J. Bullard, A. Kasarskis, S. Wang, J. Eid, D. Rank, J. C. Redman, S. R. Steyert, J. Frimodt-Møller, C. Struve, A. M. Petersen, K. A. Krogfelt, J. P. Nataro, E. E. Schadt, and M. K. Waldor, "Origins of the E. coli Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany." NEJM, doi:10.1056/NEJMoa1106920, 2011.

L. S. Katz, J. C. Humphrey, A. B. Conley, V. Nelakuditi, **A.O. Kislyuk**, S. Agrawal, P. Jayaraman, B. H. Harcourt, M. A. Olsen-Rasmussen, M. Frace, N. V. Sharma, L. W. Mayer, and I. K. Jordan, "Neisseria Base: a comparative genomics database for Neisseria meningitidis." Database, doi:10.1093/database/bar035, 2011.

A.O. Kislyuk, B. Haegeman, N. H. Bergman, and J. S. Weitz, "Genomic fluidity: an integrative view of gene diversity within microbial populations." BMC Genomics, doi:10.1186/1471-2164-12-32, 2011.

A.O. Kislyuk, L. S. Katz, S. Agrawal, M. S. Hagen, A. B. Conley, P. Jayaraman, V. Nelakuditi, J. C. Humphrey, S. A. Sammons, D. Govil, R. D. Mair, K. M. Tatti, M. L. Tondella, B. H. Harcourt, L. W. Mayer, and I. K. Jordan, "A computational genomics pipeline for prokaryotic sequencing projects," Bioinformatics, doi:10.1093/bioinformatics/btq284, 2010.

A.O. Kislyuk, A. Lomsadze, A.L. Lapidus, and M. Borodovsky. "Frameshift detection in prokaryotic genomic sequences." International Journal of Bioinformatics Research and Applications, vol. 5, no. 4, pp. 458-477, 2009.

A.O. Kislyuk, S. Bhatnagar, J. Dushoff, and J.S. Weitz. "Unsupervised Statistical Clustering of Environmental Shotgun Sequences." BMC Bioinformatics, doi:10.1186/1471-2105-10-316, 2009.

L.S. Katz, C. Bolen, B. Harcourt, S. Schmink, X. Wang, **A.O. Kislyuk**, R. Taylor, L. Mayer, and I.K. Jordan. "Meningococcus Genome Informatics Platform: a System for Analyzing Multilocus Sequence Typing Data." Nucleic Acids Research, vol. 37, pp. W606-W611, 2009.

I. Dubchak, A. Poliakov, **A.O. Kislyuk**, and M. Brudno. "Multiple whole genome alignments without a reference organism." Genome Research, vol. 19, pp. 682-689, 2009.

D. Papatsenko, **A.O. Kislyuk**, M. Levine, and I. Dubchak. "Conservation patterns in different functional sequence categories of divergent drosophila species." Genomics, vol. 88, no. 4, pp. 431-442, 2006.