



## Data Science Engineer

Join us at Break *Through* Cancer for a once in a lifetime opportunity with the potential to impact millions of lives. A distinctly new type of foundation, Break *Through* Cancer empowers outstanding researchers and physicians to intercept, as well as find cures, for the deadliest cancers by stimulating *radical collaboration*. It represents a first-in-kind partnership between five of the top cancer research centers in the world: Dana-Farber Cancer Institute, Johns Hopkins Sidney Kimmel Comprehensive Cancer Center, MIT's Koch Institute for Integrative Cancer Research, The University of Texas MD Anderson Cancer Center, and Memorial Sloan Kettering Cancer Center.

Break *Through* Cancer is focused on diseases with poor prognoses for which progress has been very slow: initially, glioblastoma, AML, ovarian, and pancreas cancers. We are looking beyond conventional therapies, utilizing new strategies, structures, and thinking from across disciplines, with the goal of enabling discoveries at an unprecedented pace and supporting *radical collaboration* between academic institutions as well as industry partners.

### Position Summary

Alongside some of the finest minds in cancer research, you will tackle some of the hardest open problems in areas of greatest unmet need and apply the results directly to patient treatment across the nation. Part of what makes Break *Through* Cancer compelling is our founding principle that computational methods and data are an integral part of the entire research, development, and treatment process—not merely a quantitative afterthought. This ethos is codified into the design of our first-of-kind *Data Science Hub*, which aims to rapidly unify the most advanced patient and molecular data with analysis methods and visual tools to analyze and explore them, making all of it accessible in a timely, frictionless, and simple manner as possible—regardless of one's level of technical sophistication. As an enabler of biological discovery and facilitator of patient treatment, the *Data Science Hub* will be a vital strategic asset in our portfolio and bedrock element of all disease research teams and clinical efforts.

In this role you will provide data engineering, bioinformatics, and software development expertise to interdisciplinary projects ranging from basic research & discovery to clinical sample processing & analysis and machine learning models. Motivated by the chance to help decipher fundamental biological processes that directly impact patient outcomes, the ideal candidate will exude a deep passion for science and through this

- co-develop or maintain pipelines, software, databases and UIs
- embrace governance of & access to internal & external data
- continuously improve scalability, software & data quality, and SOPs
- harden, benchmark, and deploy novel bio-informatic methods
- iterate with stakeholders on requirements gathering
- help identify & fill gaps in data, software tools and documentation
- review and/or contribute to papers and presentations

## Key Experience and Skills

- B.S. in Bioinformatics, CS, Math, Biology (M.S. or Ph.D. a strong plus)
- 2-5 years' experience doing academic/clinical/biotech bioinformatics
- Coding skill in Python, R, SQL/NoSQL, HTML, JavaScript, UNIX shell
- Extracting insight from rapidly evolving and/or unstructured data
- APIs and tools to warehouse, process & compute upon data
- Root cause of failure analysis on complex systems
- Visual portals/dashboards with Jupyter, PANDAS, Streamlit, Shiny
- High-throughput computation on the cloud (GCP, AWS, Azure)
- Strong attention to detail, oral & written communication skills
- Working from incomplete information, without micromanagement
- Systematically prioritize deliverables across multiple projects
- Positive attitude, strong drive & resilience in face of challenges
- Integrate external software tools, data repositories or research into concrete deliverables aligned with organizational goals

## Bonus Experience and Skills

- NextFlow, CWL, WDL workflow orchestration languages
- Terra, Synapse, Code Ocean, Foundry data / analysis systems
- Storage, analysis & interpretation of: DNA+RNA sequence data (bulk, single cell), spatial profiling & pathology images, clinical data
- Multi-omic bioinformatics, data cleansing, integration & analysis
- Biomarker discovery, clinical trial sample processing and analysis
- Cloud and/or big data tools: e.g. Docker, Kubernetes, Spark, Kafka
- Oncology, immunology, immunotherapies for cancer

Break Through Cancer is committed to providing equal opportunities in employment and prohibits discrimination and harassment of any kind. We treat our fellow Break Through Cancer colleagues and Break Through Cancer applicants fairly and respectfully. We seek to employ people with skill and integrity, and provide them with the means to develop professionally. We hire without regard to race, color, religion, creed, citizenship, national origin, age, sex, gender, pregnancy, gender identity/expression, sexual orientation, marital status, disability (including neurodiversity), genetic information, veteran status, and any other legally protected group, in accordance with applicable federal, state, or local law.

To apply, please submit a cover letter and cv to: [info@breakthroughcancer.org](mailto:info@breakthroughcancer.org)