Todd H. Creasy

Bioinformatics Pipelines | Software Engineering | Applied AI
Gaithersburg, MD | 301-802-5509 | todd.creasy@pilotbio.io | linkedin.com/in/todd-creasy-66a7043 | github.com/toddcreasy

SUMMARY

Hands-on bioinformatics and AI engineer. I build NGS and multi-omics pipelines, and I design, deploy, and monitor Python services and LLM agents on AWS. I run PilotBio, an independent consultancy that takes biotech and pharma teams from the first question to documented, working software. Before that I shipped agentic AI systems for clinical trial document analysis that cut manual review from months to weeks. Available for contract, consulting, or full-time work.

SKILLS

Bioinformatics and Pipelines: Nextflow, nf-core, Docker, NGS analysis, bulk RNA-seq, scRNA-seq, proteomics, multi-omics integration, oncology biomarker analysis, survival analysis, cBioPortal, real-world evidence, RWE, digital pathology, FAIR data principles, DNA methylation and bisulfite sequencing, nf-core/methylseq, Bismark, DSS, differential methylation, cfDNA.

Cloud and AWS: AWS CDK, Python, Bedrock, Bedrock AgentCore (Runtime, Gateway, Evaluations), Bedrock Guardrails, Titan embeddings, Aurora Serverless v2, RDS Data API, Lambda, EventBridge Scheduler, EC2, ECR, S3, App Runner, ECS, IAM, OIDC, Secrets Manager, CloudWatch, SNS, AWS Batch, Redshift.

Software Engineering: Python, FastAPI, Pydantic, SQLAlchemy, Alembic, Django, Dash, TypeScript, React, Next.js, Tailwind CSS, R, SQL, PostgreSQL, pgvector, Redshift, REST APIs, OpenAPI, pytest, ruff, pre-commit, uv, Docker, Git, GitHub Actions, CI/CD, Linux, shell scripting.

Architecture and Practices: Infrastructure as code, REST API design, schema migrations, idempotent ingestion, structured logging, alerting, keyless CI/CD (GitHub OIDC), cost controls, Agile, Scrum, requirements gathering, technical documentation.

AI and ML: Agentic workflows, Strands Agents, Pydantic-AI, custom MCP servers, Claude API, OpenAI API, RAG, hybrid search, LLM evaluation harnesses, HuggingFace Transformers, PyTorch, prompt engineering.

EXPERIENCE

Founder and Principal Consultant
PilotBio LLC, Gaithersburg, MD | pilotbio.io | October 2026 - Present

- Independent consultancy delivering reproducible Nextflow and nf-core omics pipelines (RNA-seq, DNA methylation) for AWS, local, or Slurm, plus retrieval and agent tools over client documents.
- **methyl-flow** (github.com/toddcreasy/methyl-flow): raw bisulfite reads to 10,947 differentially methylated regions in colorectal cancer (GSE95654, 18 samples) with nf-core/methylseq, Bismark, and DSS on AWS Batch Spot, defined in a Python CDK app.
- **GliaNav** (github.com/toddcreasy/glia-nav): public search and chat navigator over every glioblastoma trial and 62,000+ PubMed papers. Hybrid search on Aurora, Claude on Bedrock for eligibility criteria, an AgentCore agent that cites only retrieved IDs, six CDK stacks.
- **ehr-rwe-eln** (in development, source on request): real-world evidence platform on synthetic data in an OMOP CDM v5.4 database. Cohort pipeline, data quality checks, target trial emulation, LLM extraction of lab notebook entries into OMOP and FHIR R4.
- **histostate** (in development, source on request): colorectal H&E tissue classifier, a small CNN trained from scratch in PyTorch. Validated on a second clinical center and scored once on test (94.7%); caught a stain color shortcut.
- **methyl-floor** (in development, source on request): leakage-aware cfDNA methylation classifier and detection-limit benchmark, with a check that fails when train and test share a batch.

Associate Director of AI and Bioinformatics Platforms
AstraZeneca, Gaithersburg, MD | August 2015 - July 2026

- Designed and built Contract Accelerator, a stateful agentic AI platform (Python, Pydantic-AI, MCP, A2A, Claude/ChatGPT) that parses 300+ page clinical trial protocols, extracts requirements and statistical strategy, and checks vendor contracts against them. Went from proof of concept to production in six months; a Python Dash app lets reviewers drill into flagged inconsistencies, cutting manual review from months to weeks.
- Built custom MCP servers for parsing clinical protocols and vendor documents (PDF/Word), connected through A2A to AstraZeneca's enterprise agentic platform for reuse across workflows.
- Migrated HPC biomarker workflows from Slurm to AWS Batch using Nextflow and Singularity, so hundreds of biomarker-outcome analyses across clinical endpoints and treatment arms run in parallel without manual oversight.
- Developed Python pipelines for a multi-modal oncology biomarker platform that run survival analysis across clinical trial datasets, process genomic and proteomic biomarkers, and write results to AWS Redshift. Extended it for PyTorch and HuggingFace work, including fine-tuning and embedding generation across genomic, proteomic, pathology, and clinical data.

- Built the pipeline that ingests Tempus AI real-world evidence into cBioPortal, harmonizing it with internal clinical and genomic datasets, and managed sequencing data flows.
- Engineered the Xyall Broker FastAPI middleware connecting the Xyall Tissector instrument, CP-Portal pathology platform, and LIMS to automate high-throughput tissue sampling; 15+ Pydantic models and 100+ automated tests in GitHub Actions.
- Re-architected a digital pathology imaging pipeline to pull annotations through the CP-Portal API instead of processing large SVS files on disk. Gathered requirements from Late Oncology Clinical Operations leaders and mentored junior data scientists.

Bioinformatics Scientist and Account Lead
5AM Solutions, Inc., Rockville, MD | January 2012 - August 2015

- Led development of the first production web application that pulled multi-modal data from MedImmune (later AstraZeneca) prior clinical trials for internal analysis.
- Built web-based visualization tools (Python, Flask, Django) for 10X Genomics, Thermo Fisher, Schrödinger, MedImmune/AstraZeneca, and NIH. As ScrumMaster, launched seven bioinformatics tools for four clients in 18 months.

Manager of Bioinformatics Software Engineering
Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD | August 2007 - December 2011

- Developed automated pipelines in the Transcriptomics Analysis Group for managing large NGS and microarray datasets, and built tools for interpreting and visualizing the results.

Bioinformatics Engineer Manager
The Institute for Genomic Research / J. Craig Venter Institute, Rockville, MD | January 2001 - August 2007

- Built pipelines and visualization tools for prokaryotic and eukaryotic gene annotation; senior technical liaison to NIAID and NHGRI program managers on consortium collaborations.

AWARDS

AstraZeneca R&D Award, 2022: Surface Proteomics platform supporting patient-centric, targeted discovery.

PUBLICATIONS

25 publications. Full list on PubMed: https://www.ncbi.nlm.nih.gov/myncbi/18sNm-jpk3WQG/bibliography/public/

EDUCATION

Johns Hopkins University | M.S. Bioinformatics
University of Maryland College Park | B.S. Molecular Biology