

# Trevor C. Allen

Trevor.Allen@Valiance.io • (614) 653-7407

Personal Website: [www.valiance.io](http://www.valiance.io) • LinkedIn Profile: [LinkedIn.com/in/tcallen07](https://www.linkedin.com/in/tcallen07)

## Summation

With multiple years' experience in data analysis, research, and programming as a part of agile cross-functional teams of Data Scientists, Data Engineers, Developers, and Research Assistants, I offer a broad range of technical skills combined with comfort and experience interacting with clients and partner organizations. Having experience as both a team leader and member, I've coordinated and participated in data integration efforts large and small, analysis and validation of new and existing predictive & prescriptive models, establishment of MVP requirements and subsequent user acceptance testing, and more. I'm actively seeking new opportunities and clients, both short- and long-term projects, in an energetic, innovative community with which to engage in meaningful, purpose-driven work.

## Work Experience

### **MatchTx**, Columbus, Ohio

*Data Scientist & Data Engineer*, May 2016 – present

MatchTx provides a cloud-based research and analytics platform for research & clinical settings to explore, identify, and capitalize on biomarkers & pattern-driven traits that link cancer patients both to personalized treatments and to "uniquely similar" patients by matching patients on genomic profile & statistically validated phenotypic traits, including clinical outcome and cancer-specific relevant metrics. Maintains client-accessible KPI dashboard.

- Solely responsible for acquisition & duration of a dozen public and private datasets of clinical & genomic data for over 12,000 patients and several thousand coding genes encompassing over 30 types of cancer.
- Developed analytic framework to implement patent-pending algorithm concepts by chief medical researcher.
- Maintain the ever-evolving codebase to generate, validate, and persist the proprietary rules-based matching system.
- Create/optimize automated & ad hoc analyses, with reporting, for massively-parallel cloud-based performance.

### **Martini Media Group**, Columbus, Ohio

*Project Manager & Business Analyst*, March 2016 – March 2019

As a Project Manager, I served as a liaison between clientele and internal Design & Dev teams, coordinating prioritization, reporting, and all else needed to ensure delivery. Additional duties included maintenance on multiple standard websites (HTML/CSS/JS) and debugging Ruby on Rails mobile app. Most projects were oriented toward new or recently-established start-up companies in the Columbus area.

### **Zenith Fitness**, Columbus, Ohio

*Founder & Developer*, January 2015 – May 2016

Revolutionary fitness technology startup applying a network of IoT sensors to capture all exercise motion using standard gym equipment, streaming real-time to integration-enabled web- and mobile-accessible platform.

- Custom-built hardware: network of multi-sensor boards with Raspberry Pi-based event-stream processing.
- Performed extensive market analysis: thousands of survey responses, dozen business manager & owner interviews.
- Selected for participation Rev1 Ventures intensive training.

### **Explorys** (now part of IBM Watson Health), Cleveland, Ohio

*Data Scientist – Data Curation Team*, February 2013 – September 2014

- Developed and delivered reports on new and existing data integration projects covering dozens of clinical data systems containing billions of data elements, including both ad hoc analysis and automated regular reports.
- Frequently led communication efforts with both technical and non-technical clients regarding progress of pilot integration efforts, preliminary user acceptance testing, and on-going validation efforts during product rollout.
- Oversaw expansion of Data Curation and Data Integration teams from a few individuals to two dozen members, including training of new-hires, interns, external contractors, and cross-team collaborative training.
- Tapped for multiple “Tiger Teams” for key clients and high-profile projects, requiring daily meetings with team leads and senior leadership with weekly company-wide updates.
- Oversaw multiple projects simultaneously and coordinated efforts across teams; largely responsible for developing and maintaining teams' use of Atlassian JIRA to track efforts, requirements, and timelines.
- Individually and jointly mapped many new and previously-encountered EMR/EHR back-ends to our internal data model, with many hours of reviewing source-to-target mapping (STM) with technical and clinical clientele, spanning overarching “health of the data” discussions to field-by-field STM reviews.

## **Kleinhenz & Associates, Cleveland, Ohio**

*Research Associate, December 2012 – September 2014*

- Regularly reported forecasts in business economics, with heavy focus on time-series estimation of quarterly and monthly national retail sales as well as modeling regional and statewide job growth in Ohio.
- Performed ad hoc research into improved modeling using basic and intermediate econometric methods such as splines, ARIMA, and comparing information criteria in STATA, EViews, and Python.

## **U.S. Peace Corps, Yeghegnadzor, Vayots Dzor, Armenia**

*English & Economics Instructor, May 2012 – November 2012 (unpaid)*

- Led classroom instruction in English, Economics, and Finance at Secondary and Post-Secondary levels.
- Individually and team-taught 13 classes of 10-30 students each from 9<sup>th</sup> grade to graduate students.

## **Department of Economics, Case Western Reserve University, Cleveland, Ohio**

*Research Assistant, July 2011 – May 2012*

- Collaborated with multiple PhD faculty on numerous small- and large-scale research projects simultaneously.
- Led data management for terabytes of data in various formats spanning years of gathering & manipulating.
- Directed a small team of research assistants – coordinated efforts by outlining workflow and delegation of projects, developed timelines, and regularly communicated progress and status updates to supervisors.
- Assisted in experimental design and gleaning new insight from existing experimental data.
- Re-coded and organized tens of thousands of lines of code representing thousands of man-hours of data analysis, automating processes and reducing run-time by an order of magnitude.
- Developed a platform for Internet-based participation in economics experiments including automated payments based on real-time results and incorporating database back-end.

## **Education**

### **Case Western Reserve University (CWRU), Cleveland, Ohio**

*Bachelor of Arts: Physics, Economics – May 2011, cum laude*

*Senior Thesis: Macroeconomic Impact of Mobile Phones in the Developing World*

## **Computer & Programming Skills**

*Languages: Python, R, git, SQL & NoSQL (Mongo, Cassandra), Ruby, bash, JavaScript/HTML/CSS, Java*  
*Platforms & Technologies: AWS, Docker, Kubernetes; Apache Hadoop, HBase, Spark, Airflow, Kafka*

## **Data Science & Statistical Analysis**

*Modeling: regression (logistic, linear, mixed-effects), time-series/forecasting, clustering, graph, AR/MA, GLM, etc.*

*Libraries: Pandas, SciPy; matplotlib, ggplot2, plotly; Django, Rails, Flask, Celery, Twisted, scikit-learn, pyspark*

*Misc.: Regular expressions, STM/Mapping, standardization, data mining & integration, ETL/ELT pipeline automation*

## **Certifications & Licenses**

- IBM Data Engineering professional certificate (Credential: [DU6VYSJK6R3A](#))
- IBM Data Science professional certificate: (Credential: [CAQDUR7NWSAH](#))

## **Clinical & Healthcare Experience**

- Extensive data mining of EHRs, including: Epic/Clarity, NextGen, Cerner, Allscripts (Enterprise, Touchworks, & Professional) and custom data warehouses, covering both ambulatory and inpatient settings.
- Familiarity with many medical standards and ontologies such as SNOMED-CT, LOINC, RxNorm, ICD, CPT.
- Ensured fidelity of clinical data acquired and standardized from dozens of systems mapped to a single unified model built atop a distributed-computing framework of Apache Hadoop + HBase + Impala.
- Frequently investigated complex data-flow and processing methods to identify and resolve data loss, corruption, or misrepresentation within terabytes of data covering tens of millions of unique patient-lives.
- Participated in expansion of clinical data model to incorporate financial and transactional (OLTP) data elements.