# PATRICK CHERRY, PHD

pcherry [at] pm dot me | upon request | Senior Scientist | Oncology | BillionToOne | San Francisco, California | pdcherry.github.io | github.com/pdcherry | linkedin.com/in/p-cherry | United States Citizen | Updated on 2024-12-20

PhD scientist and professional critical thinker, problem-solver, and science communicator. At the interface of data science, genomics, and statistics; skilled at data visualization, bioinformatics, next-generation sequencing (NGS), and tool-building. I've coded reproducible pipelines for high-throughput experimental designs and genomic analyses. I've launched best-in-class oncology reference standards and analyzed NGS data from Illumina, MGI, and Element. Originally trained in Molecular Biology, I am passionate about advancing data science and bioinformatics to improve human health and to understand biology.

## INDUSTRY EXPERIENCE

Current | Sep 2024

**Senior Scientist, Oncology**
BillionToOne — Menlo Park, CA

- Generated 500+ automated ddPCR primer / probe designs with pipeline in `snakemake` using `s3` database of variant data and `primer3` primer selection
- Informed key product decisions with historical analyses using PostgreSQL / ssh and s3 data sources and linear mixed-effects modeling

Apr 2024 | Jan 2021

**Senior Scientist, Genomics**
Twist Bioscience — South San Francisco, California

- Invented and introduced multiple new products to market, yielding $1M+ in new revenue as *Tech Lead*, including: *Pan-cancer cfDNA* (v1 & v2), *CNV Controls*, *RNA Fusion Controls*, *Fragmentome Controls*, RNA-seq, and primer / adapter removal methods for highly-multiplexed dsDNA pools
- Answered biological and business questions with DoE and data analysis using R and the libraries: `dplyr`, `dbplyr`, `DBI`, `purrr`, `ggplot2`, `Bioconductor`, `DEseq2`, `seurat`, `tidymodels`, `glm`, `nls`, `lme4`, `AlgDesign`, and more
- Edited and maintained production QC pipelines with automated reporting using `Python` and the libraries: `NumPy`, `pandas`, `polars`, `seaborn`, `statsmodels`, `biopython`, `pybedtools`, `pysam`, `vcfpy`, `scanpy`, `Jinja2` (with `CSS`), `WeasyPrint`,and others
- Maintained production code (`git` version control with code review) with CD (github actions); operated in high-performance compute environment `aws s3` and `Databricks` using Unix shell / `bash` tools, like `ssh`, `STAR-fusion`, `GATK`, BaseSpace CLI, `bwa`, `bedtools`, `samtools`, `vcftools`, `UMI-tools/fgbio`; Confluence documentation
- Authored internal R package "`twistcolorpal`" on Github with help files to help scientists style plots to Twist branding
- Answered production and business questions with original `SQL` queries for `Snowflake`-based database

Jan 2021 | Apr 2019

**Scientist I, NGS Core & Strain-Build Process Development**
Zymergen, Inc. — Emeryville, California

- Boosted *NGS Core* genotyping success by 45% using DoE experimentation, data-driven decision-making, and teaching; Guided demanding and diverse internal customers on complex NGS experiments
- Built and disseminated `Rmarkdown` and `JMP` statistical templates for autonomous NGS data exploration
- Coded plasmid and strain build and QC experiments using Zymergen's alembic Python API to LIMS database
- Rendered strain build and plasmid build reports from LIMS `SQL` database in `MySQL` Workbench

## EDUCATION

May 2019 | Aug 2013

**PhD**
University of Colorado School of Medicine — Aurora/Denver, Colorado

- Ph.D. in Molecular Biology. Advisor: Jay Hesselberth, PhD
- Thesis: RNA terminus chemistry impact decay events that target *HAC1* mRNA during the Unfolded Protein Response

May 2013 | Aug 2009

**BA**
Hendrix College — Conway, Arkansas

- Biochemistry and Molecular Biology, with Distinction. Advisor: Andres Caro, PhD
- Senior Capstone Project showing key stress response gene expression changes to oxidative stress in liver cells
- Minor in Mathematics; PI: Lars Seme; Project: Newton's method as a fractal chaotic dynamical system

## SELECTED INTELLECTUAL PROPERTY

Apr 2021 — **Libraries for mutational analysis**
Twist Bioscience — South San Francisco, California
- US Prov. Pat. Ref No 823.101
- Configuration and fabrication of synthetic DNA & RNA reference standards and synthetic variant sequences

Mar 2021 — **Method for counterselection in microorganisms**
Zymergen, Inc. — Emeryville, California
- US 2021_0087586 A1
- Demonstration of novel counterselection mechanism in non-model *Bacillus* microbe for genetic modification

## SELECTED PUBLICATIONS

Jan 2024 — **Twist cfDNA Pan-Cancer Reference Standard v2 Technical Guidance**
Twist Bioscience Product Sheet & FAQ — South San Francisco, California
- Patrick Cherry, Lydia Bonar, & Mike Bocek

Apr 2022 — **Characteristics and specificity of the wild-type / 0% VAF reference material**
Twist Bioscience — South San Francisco, California
- Patrick Cherry & Mike Bocek

Mar 2019 — **Multiple decay events target HAC1 mRNA during splicing to regulate the unfolded protein response**
eLife — Aurora/Denver, Colorado
- Cherry, P., Peach, S., & Hesselberth, J.

Dec 2017 — **Genetic bypass of essential RNA repair enzymes in budding yeast**
RNA — Aurora/Denver, Colorado
- Cherry, P., White, L., York, K., & Hesselberth, J.

## SELECTED PRESENTATIONS & TALKS

Feb 2024 — **Twist pan-cancer reference standard V2: Enhanced precision and reduced errors in ctDNA analysis**
Advances in Genome Biology and Technology, Twist Bioscience — Orlando, Florida

Apr 2023 — **Twist pan-cancer synthetic RNA fusion control for assay development**
American Association for Cancer Researchers, Twist Bioscience — Orlando, Florida

Feb 2023 — **Use of synthetic CNV fragments to mimic copy number alterations for ctDNA reference standards**
Advances in Genome Biology and Technology, Twist Bioscience — Hollywood, Florida

Apr 2022 — **Twist pan-cancer synthetic reference materials for cell-free DNA (cfDNA) assay development**
American Association for Cancer Researchers, Twist Bioscience — New Orleans, Louisiana

Jun 2020 — **R use at Zymergen**
Z-Tech Talk — Emeryville, CA

Apr 2020 — **Data-driven troubleshooting of NGS experiments**
Data Science Talk — Emeryville, CA

Aug 2017 — **RNA processing regulates the unfolded protein response**
CSHL: mRNA Processing Meeting Talk — Cold Spring Harbor, New York