

Ding CHENG

+86 17521619529 | mkbo@me.com | github.com/dipcarbon

EDUCATION

Fudan University

Bachelor degree of Chemistry, Minor in Data Science

Shanghai, Full-time

2014 – 2018

Fudan University

Master Candidate of Software Engineering, Data Science and Application

Shanghai, Part-time

2021 – Present

EXPERIENCE

Data Engineer

Jun 2019 – Present

PD Data Sciences (under transformation)

Roche

- Use data surveillance tools and strategies to provide aggregate level reviews designed to identify patterns in data
- Work with stakeholders to assist with data-related technical issues and explore ways of automation and insights generation

Clinical Programmer

Feb 2018 – Jun 2019

PD Biometrics

Roche

- Responsible for clinical data delivery of RG6042 (Huntington's Disease), Crovalimab (Hematology), and support Tecentriq HCC, OC Filing
- Develop SAS programs for extracting and processing data with SDTM implementation and CDISC standards and lead the data collection of clinical database development with clinical data EDC programming
- Perform Spotfire programming and documentation for Risk-Based Quality Management, Data Monitoring, Medical Data Review

Research Assistant

Jan 2021 – Present

Biomedical Informatics Laboratory (Part-time)

Fudan University & Zhongshan Hospital

- Implement OHDSI OMOP common data model (SNOMED CT, MedDRA) and build up a federated analysis platform using R packages
- Improve EHR Data usability and quality based on natural language processing and machine learning (Bi-LSTM)

RELATED PROJECTS

Data Curation and Integrated Analysis | R Shiny, SQL, Python, REST API

- Implemented data curation and integration from Clinical Trial data of EDC, Omics Data and RWD within harmonized standards and control terminology for secondary use
- Supported continuous delivery using Jenkin and Docker and developed R packages (*CurateR*, *Scrubj*) and a full-stack R Shiny app (*Curation Workshop*)
- Built exploratory and analytic tools that utilize the data pipeline to provide key metrics and clinical outputs (TLGs) with R frameworks

Crovalimab Phase I/II Trials Study | SAS, R, Pinnacle 21, Spotfire

- Ensure timely and quality development and validation of datasets and reports for study-documents
- Designed and reviewed the eCRF and data structures and comply with company and industry standards (CDISC SDTM) and processes, created the review of programming specifications

PD Study Dashboard | R, Linux Server, Docker, Google Workspace, Google Data Studio, JavaScript (D3)

- Developed an E2E dataflow solution, collaborated with Clinical Scientist, Study Data Manager, CRA, and Global Study Managers to suggest features and get feedback

PUBLICATION

First Author: Ding Cheng, Liang Qiao*, Peter Horvatovich*, *Towards Spectral Library-free MALDI-TOF MS Bacterial Identification*, J. Proteome Res., 17 (2018), 2124–2130

TECHNICAL SKILLS

Languages: English (TOFEL 108), Mandarin

Programming Languages: Python, R, SQL, C#, SAS, JavaScript

Tools: Tableau (Certificated), TIBCO Spotfire, Microsoft Office, Final Cut Pro

Libraries and Frameworks: pandas, Flask, NumPy, Matplotlib, Shiny, Tidyverse, Roxygen