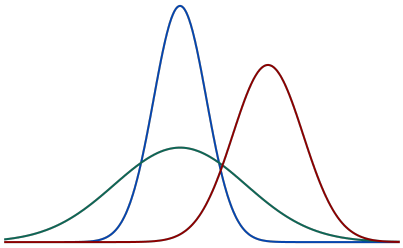
JONATAN PALLESEN

I am a data scientist with a PhD in genomics. I'm curious by nature, and enjoy the challenge of separating signal from noise to gain real insights.

Throughout my education and work I have acquired extensive theoretical and practical experience with data handling, statistics, machine learning, visualization and algorithms. I am a very skilled programmer in Python, R and Julia, with more than ten years of experience in the former two.

During my work as statistical analyst and as data scientist I have worked on a variety of different projects, for which I have been in charge of all stages of the analysis, from problem definition, data cleaning and quality control, to statistical analysis, model building and presentation.



jonatan.smp@gmail.com
jonatanpallesen
github.com/ymer
jsmp.dk

View this CV as HTML at *jsmp.dk/cv*

WORK EXPERIENCE

2019

Data scientist

Raven biosciences

- Lead data scientist
- Working with a variety of projects in education, fintech and automated machine learning.

2018 | 2015

Statistical analyst

Aarhus university

- Programming pipelines and tools, working with very large data sets, statistical analysis and machine learning.

EDUCATION

2015 | 2011

PhD, human genetics

Aarhus university

- Thesis: Association studies of psychiatric disorders: On association of genes, gene sets and runs of homozygosity.

2012

Visiting researcher

University of California, Berkeley

2011 | 2004

MSc, molecular biology and computer science

Aarhus university

SKILLS

Programming languages

R
Python
Julia

Packages

Visualization

(ggplot, shiny, matplotlib)

Data science, R

(tidyverse, tidymodels, caret, rmarkdown)

Data science, Python

(pandas, scikit-learn, tensorflow, statsmodels)

Other

Github
SQL
Unix / bash

SELECTED PUBLIC DATA SCIENCE PROJECTS

2019 **How to transform your data**

• Using simulations to determine the optimal transformation for skewed variables

2019 **Blind auditions and gender discrimination**

• Re-analysis of a seminal study. (Used for an article in the Wall Street Journal)

2016 **Project Euler**

• Computational problems solved in Python and Julia

I regularly make new analyses and visualizations on my blog

SELECTED PUBLICATIONS

2019 **Discovery of the first genome-wide significant risk loci for attention deficit hyperactivity disorder**

Nature Genetics. (link)

• Demontis et al.

2019 **Identification of common genetic risk variants for autism spectrum disorder**

Nature Genetics. (link)

• Grove et al.

2016 **LandScape: a simple method to aggregate p-values and other stochastic variables without a priori grouping**

Statistical Applications in Genetics and Molecular Biology. (link)

• Joint first author with Carsten Wiuf.