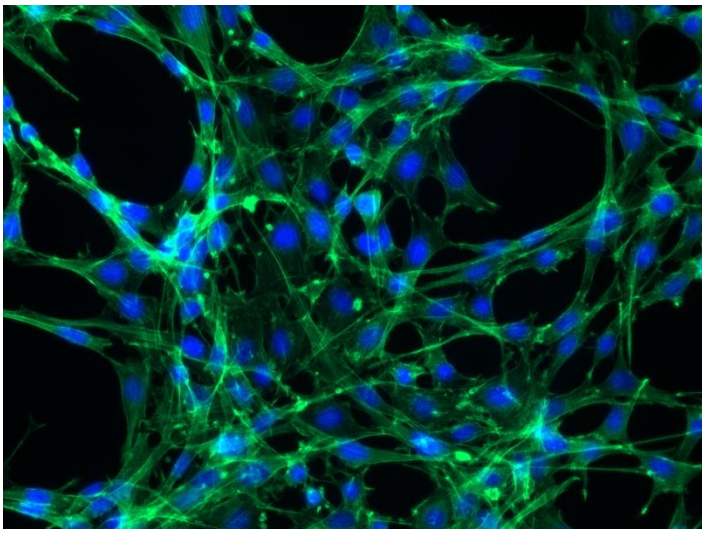
Study Session 6: Inferential Statistics with Pandas (extra practice)

Lecturers: Dr. Rylie Green, Dr. Christopher Chapman
GTAs: Martina Genta, Hendrick Beck, Mikolaj Kegler, Sergio Mena, Alexey Novikov, Frederik Puffel



Problem:
You have been hard at work new materials for bioelectronics interfaces with the body. You have fabricated five (5) different formulations of the most promising material and have tested their cytocompatibility *in vitro* through a cell viability assay.

The resulting data set is in the "extra_data_example.xlsx" spreadsheet. Using Python you must analyze this data for any significant differences between the *in vitro* cytocompatibility of the materials. In order to fully analyze this data set you will need to provide plots of the data and conduct appropriate inferential statistics to determine if any significant differences were seen. Specifically, the following aspects should be included:

- A bar plot of the *in vitro* cell viability data means with appropriate standard deviations
- Tests of normality for each material
- ANOVA and (correct) post-hoc tests to demonstrate potential significance

Please provide justifications for the statistical tests that you choose based off of the properties of the data (i.e. normality, sample size, etc.). These justifications should be made in your Jupyter Notebook. At the end of this study session please comment on aspects of this study design that you think could be improved. Also, what are some potential challenges that could make this data very hard to interpret and how could you better incorporate those variables into this analysis?