



# ***Covariate Regression Illustrated***



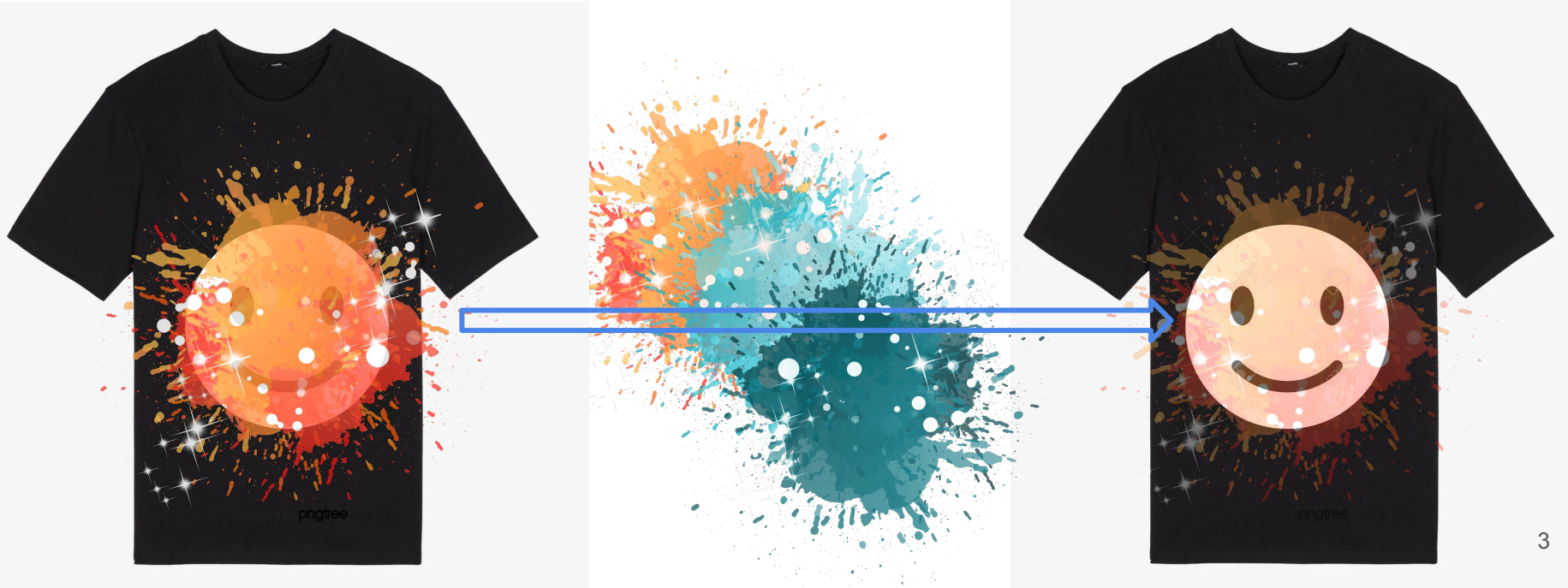
# HO NO !

- Some close relative offered you your favorite gear : a nice T-shirt with a cool design (*that's your SC data*)
- Unfortunately your 5-years old inadvertently tainted it with its paint flask... (*that's the biases in your measures*)
- Hopefully, you know a skilled team of cleaners right around the corner (*who actually are bioinformaticians in disguise ! \*wink wink\**)
- They explain you they learned how to clean stains by **accumulating knowledge**, having cleaned tons of them in several contexts, thanks to friends that screened the best detergents ever (*psst, they're biologists who made discoveries in the field*).



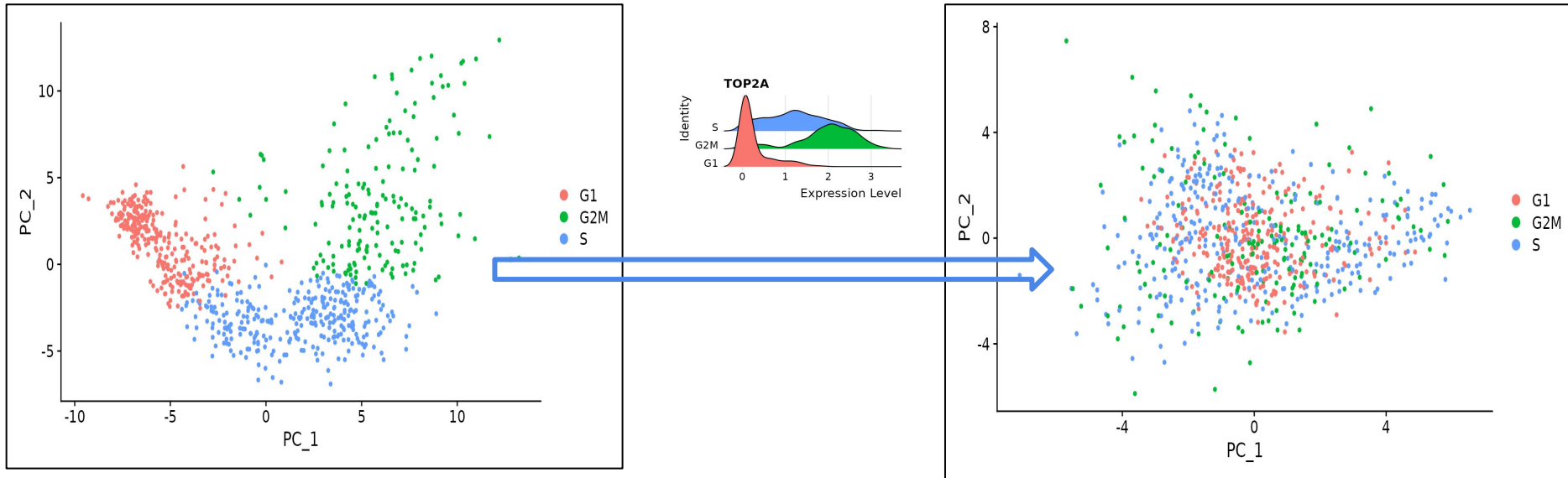
# Cleaners to the rescue

- By compiling this amount of **knowledge** (*eg : gene signatures*), they could produce a specific cleaning protocol (*regression*) to the kind of staining (*biases*) affecting your Tee (*data*), subtracted the former to the latter, revealing the object of desire (*the biological signal*)



# Let's get back to reality

- This is what is performed in the cell cycle example, thanks to acquired data on the cell-cycle -specific gene expression, and regression
- When data is **continuous**, a linear **regression** is used
- When data is **discrete**, a **block-based** regression is used



# The devil may be in the bottle

- But what would happen if by bad luck, the cleaner team trained themselves on stains that unfortunately **look a bit to much like** your pattern ?
- The same way you **increased** (biological) **signal by removing some noise** on the former example, one may perform the **complete opposite** if one doesn't pay attention
- *Probably because in this specific context, what you thought was noise, wasn't*

