We performed an experiment where either blank (no tissue) paraffin blocks or extraction controls (no paraffin) were sequenced, and we readily detected microbial DNA reads originating from common gut, skin and oral microbes (**Figure 1**). In contrast, shavings from formalin fixed, paraffin embedded (FFPE) normal colon tissues had microbial reads originating from various gut and oral microbes, with minor contributions from skin microbes such as *Cutibacterium acnes* and *Staphylococcus epidermidis* which may be contaminants introduced during sample processing (**Figure 1**). Importantly, we still detected a substantial number of human reads even after removing reads mapping to the human genome (hg38) in a preprocessing step, underscoring the importance of having host representation in databases during taxonomic classification to reduce mis-classification artefacts (**Figure 1**). Overall, this highlights that bioinformatics filters may not be as perfect as one expects, and thus conservative approaches are needed.

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 1**. Relative abundances (y-axis) for various microbial taxa and human reads (x-axis) based on metagenomic sequencing of shavings of formalin fixed and paraffin embedded (FFPE) normal colon tissue and two types of negative controls: paraffin shavings from blank blocks and extraction controls without paraffin. DNA was extracted and Illumina metagenomic libraries were prepared using commonly used kits (PowerSoil Pro and NEBNext) and subjected to paired-end sequencing. Reads which failed to map against the human genome were classified taxonomically using Kraken2 with default parameters. Abundances were computed relative to the total number of reads classified at species level.