



**Supplementary Figure 1. Overall bacterial composition in NAT and OSCC patient samples:** Sunburst plots showing inferred bacterial composition in NAT (left) and OSCC (right) patient samples. Bacteria were inferred by at least 2 proteins, with those showing  $\log_2$  fold-changes of  $\leq -1$  and  $\geq 1$  being allocated to NAT and OSCC conditions, respectively. Species inferred by proteins within this range were considered present in both conditions. Classification level from inner to outermost radials: Phylum, Class, Order. Colours represent species' phylum while percentage values represent proportion within a given classification level. Venn diagram (middle) indicates number of bacterial species per condition.