Supplementary material

Supplementary material S1. Microbiome DNA Sampling and Sequencing. The data presented in this manuscript was internally generated by Neoprospecta Microbiome Technologies. Samples were anonymously collected from a hospital environment and healthcare professionals using a sterile dry cotton swab, without transport media. The swab was moistened with a sterile saline solution (0.9% NaCl) prior to sample collection. The swabs were transported to Neoprospecta laboratory facilities to DNA extraction and sequencing. Microbial DNA from the samples was obtained using AMPureXP beads (Beckman Coulter, Brea, CA) after a thermal lysis process of 96ºC for 10 min. A negative control swab sample was included since the first processing step. Amplicon sequencing library preparation was performed for bacteria using the V3-V4 16S rRNA gene primers 341F (CCTACGGGRSGCAGCAG, doi: 10.1371/journal.pone.0007401) and 806R (GGACTACHVGGGTWTCTAAT, doi: 10.1038/ismej.2012.8), with the following conditions: the first PCR primers contain the Illumina sequences based on TruSeq structure adapter (Illumina, San Diego, CA), allowing the second PCR with indexing sequences. The PCR reactions were always carried out in triplicates using Platinum Taq (Invitrogen, USA) with the conditions: 95°C for 5 min, 25 cycles of 95°C for 45s, 55°C for 30s and 72°C for 45s and a final extension of 72°C for 2 min for PCR 1. In PCR 2 the conditions were 95°C for 5 min, 10 cycles of 95°C for 45s, 66°C for 30s and 72°C for 45s and a final extension of 72°C for 2 min. The final PCR reaction was cleaned up using AMPureXP beads (Beckman Coulter, Brea, CA) and samples were pooled in the sequencing libraries for quantification. The DNA concentration of the pool amplicon was estimated with Picogreen dsDNA assays (Invitrogen, USA), and then the pooled libraries were diluted for accurate qPCR quantification using KAPA Library Quantification Kit for Illumina platforms (KAPA Biosystems, Woburn, MA). The library pool was adjusted to a final concentration of 11.5 pM and sequenced in a MiSeq system, using the standard Illumina primers provided in the kit. A single-end 300nt run was performed using a V2x300 sequencing kit.

After sequencing, the bioinformatics pipeline performs sequence demultiplexing, adaptor and primer trimming. The reads were size normalized to 283pb. Read quality filter (E) was performed

converting each nucleotide Q score in error probability (ei), that was summed and divided by read length (L):

$$e_i = 10^{\frac{-Q_i}{10}} \qquad E = \frac{\sum_{i=1}^{n} e_i}{L}$$

If E was minor or equal to 0,01 (1%) the read was considered in downstream analysis.

To increase the reliability of the read, excluding possible diversity generated by chimeric amplicons or erroneous nucleotide incorporated in PCR, 100% identical reads were clustered. If any cluster is represented by fewer than 5 reads, it is not considered in further analysis. In the pipeline, each cluster gets a unique identifier and using this approach of read clustering we stablished that our cluster is the OTU (Operational Taxonomic Unit). Clustered sequences (OTUs) were then subjected to taxonomic classification comparing them with a 16S rRNA database (NeoRefdb, Neoprospecta Microbiome Technologies, Brazil). Sequences with at least 99% of identity in the reference database were taxonomically assigned and the samples were then evaluated based on their microbiome composition, focusing on bacterial profiles and their quantity amounts. Risk map considering bacterial distribution in the hospital blue print was performed using the NeoMap tool (Neoprospeca Microbiome Techonologies).