'NetShift': a methodology for understanding 'driver microbes' from healthy and disease microbiome datasets

Bhusan K. Kuntal^{1,2} · Pranjal Chandrakar^{1,3} · Sudipta Sadhu¹ · Sharmila S. Mande¹

Received: 28 March 2018 / Revised: 5 September 2018 / Accepted: 14 September 2018 / Published online: 4 October 2018
© International Society for Microbial Ecology 2018

Abstract

The combined effect of mutual association within the co-inhabiting microbes in human body is known to play a major role in determining health status of individuals. The differential taxonomic abundance between healthy and disease are often used to identify microbial markers. However, in order to make a microbial community based inference, it is important not only to consider microbial abundances, but also to quantify the changes observed among inter microbial associations. In the present study, we introduce a method called 'NetShift' to quantify rewiring and community changes in microbial association networks between healthy and disease. Additionally, we devise a score to identify important microbial taxa which serve as 'drivers' from the healthy to disease. We demonstrate the validity of our score on a number of scenarios and apply our methodology on two real world metagenomic datasets. The 'NetShift' methodology is also implemented as a web-based application available at <https://web.rniapps.net/netshift>

Introduction

The human body houses an ensemble of varied groups of microbes which are estimated to be several folds higher compared to the number of our somatic cells. These microbial groups (called 'microbiome') remain in a symbiotic relationship with the human body [1]. They derive nutrition for themselves as well as for the host by harvesting the otherwise inaccessible energy from the ingested food [2]. They also help the host in transforming a variety of

products like xenobiotics [3]. The resident microbial consortia has been indicated to play significant roles in defining our physiology, development, nutritional status, health, and immunity [4–6]. Although the microbiome remains in a stable co-existence, certain aberrations in it have been reported to have potential harmful effects [7–13]. The human microbiome project [14], one of the biggest efforts to investigate the diversity of the microbiome inhabiting the different body sites, has generated significant scientific attention. Subsequently, several concurrent studies have reported the link between the human microbiome and health status as well as disease progression [15–19]. Although it is still not clear whether such microbial associations can be deemed to be a cause or an effect, these findings help us to better understand the pathogenesis of various diseases. Studies of various microbiomes therefore show promise to complement our existing knowledge of human physiology and are likely to pave way for development of new therapeutic approaches.

Although abundances of certain microbes may contribute towards conferring a specific trait in a given environment, the effect of the mutual association within the residing microbial communities plays an important role in determining several characteristics [20, 21]. Microbe(s) in a community may interact with others either by generating favorable conditions for them via mutualism and commensalism, or by helping them to evade the host immune

Enrolled as an industry sponsored PhD candidate: Bhusan K. Kuntal

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41396-018-0291-x>) contains supplementary material, which is available to authorized users.

✉ Sharmila S. Mande
sharmila.mande@tcs.com

¹ Bio-Sciences R&D Division, TCS Research, Tata Consultancy Services Ltd., 54-B Hadapsar Industrial Estate, Pune 411 013, India

² Academy of Scientific and Innovative Research (AcSIR), CSIR-National Chemical Laboratory Campus, Pune 411 008, India

³ Present address: Decision Sciences, Indian Institute of Management Bangalore, Bannerghatta Road, Bengaluru, Karnataka 560076, India

system [22]. Mutualism (also referred to as cooperation or symbiosis) corresponds to interactions between species where each member derives benefit from others, while commensalism refers to situations where some constituent species of an ecosystem derive benefit from other members without any help or harm to them [23]. On the other hand, a microbe might directly compete with another for the same nutrition source, thereby creating a competition [24, 25]. Moreover, in certain disease conditions, the host immune responses may disproportionately affect the mortality of a subset of microbial groups owing to a selective advantage, thus creating a change in their community behavior [26, 27]. In many processes, a set of key microbial groups are likely to act as 'drivers' [28, 29] for facilitating several changes and hence becomes an important factor for understanding the microbial basis of the disease.

The complex nature of inter microbial interactions in an ecological community can be best characterized by networks using graphs. Various components of such networks constitute nodes representing the resident microbial members of the community and a connection or edge quantifying a known or predicted microbial association. In order to understand the changes in microbial community structure pertaining to a disease, a set of affected individuals constituting the disease 'case' is used to compare the same with a set of healthy 'controls'. A comparison of the microbial community structure in the disease with the healthy can hence be achieved by construction of microbial association networks corresponding to each state, known as the 'case' and 'control' network respectively. This 'case' and 'control' networks can be further evaluated and compared based on their constituent network properties [21, 30–34]. However, meaningful insights can only be obtained when such pair of networks is comparable. A comparable network pair can be constructed when the participating samples (or set of individuals) belonging to the healthy ('control') set share matched. Microbial association networks (namely the 'case' network and 'control' network) can be constructed using the microbial abundance profile utilizing the mutual co-occurrence information of the constituent microbes in each set. In order to decipher the microbial community level changes, it becomes important to analyze the properties of these association networks as well as quantify the changes in various types of associations between a healthy ('control') and diseased ('case') [35–37]. In this communication, we propose a new strategy called network shift (or 'NetShift') to characterize important community level changes between two association networks ('control' and 'case') that are obtained for a particular disease or condition and subsequently identify key microbial taxa. NetShift addresses three key questions: (1) is there an overall significant change in the interaction or community pattern between the healthy ('control') and diseased ('case')? (2) Are there any

major changes in associations of each constituent node (taxon) between the two states? (3) If yes, has the node (taxon) been an important member of the community and is there an increase in its importance in the disease ('case')? We demonstrate the utility of the methodology using two publicly available metagenomic datasets consisting of healthy ('control') and disease ('case') samples.

Methods

Microbial association networks provide a means to quantify interactions based on individual microbial abundance counts coupled with several statistical measures. In a typical workflow, the microbial abundance of the samples at a particular phylogenetic level is obtained and an optional step to filter out rare taxa is performed. Following this, statistically significant correlations (co-occurrence) are calculated between the constituent taxa using various techniques utilizing their abundance information to determine qualifying associations. Several methods have been developed in the past to create such networks from the microbial abundance data [38–41]. The CCREPE tool [38], for example, calculates a statistically significant correlation between a given pair of taxa by repeated sampling of a subset of their abundance values using a technique known as bootstrapping. The process is repeated for all possible taxa pairs in a given dataset and the resultant set of edges constitutes the microbial association network. The various components of such association networks (represented as a graph) constitute nodes representing the resident microbes and edges representing their direct associations. The set of edges connecting two nodes which are not connected directly represents a path. Such a graph has a set of global and local properties having specific utilities for deciphering the microbial community structure. Global graph properties like density, cluster coefficient and average path length provide insights on the overall organization of the network and enables assessment of its modularity. In a microbial community, density corresponds to the proportion of observed microbial associations (edges) out of all theoretically possible associations (considering all the nodes in the network). Therefore, a greater density value indicates a higher cross talk between the resident microbes which are represented in the network. Cluster coefficient quantifies the tendency of the graph to be divided into sub-units. In other words, a microbial network with a higher number of independent units of associated microbes is expected to have a higher clustering coefficient value. Average path length indicates the average number of steps which would be required to reach from one node to another in the network. This value in the present situation represents how compact is the microbial community structure. In short, the global

properties of a microbial association network can provide insights on the overall differences in the community structure between a ‘case’ and a ‘control’ network. In order to understand the individual node level changes in a network, the local graph properties like degree, betweenness centrality and coreness centrality can serve as important quantification metrics [42]. While the degree of a node corresponds to the number of its direct connections with other nodes, betweenness centrality of a node indicates its importance measured by its frequency of occurrence in the paths connecting other nodes. For example, in a microbial association network, the degree measures the number of direct associations of microbes with other microbes. On the other hand, a higher betweenness centrality value of the same microbe, measured by its involvement in connecting other microbes, highlights its importance as a preferred member of the microbial community. The coreness centrality value of a node corresponds to the number of maximally connected neighboring nodes which has at least that many number of connections. In other words, nodes with high coreness centrality values are indicators of microbes showing higher “colony forming capability” with other microbes in the community. Colonization refers to establishment of a mutualistic or commensal relationship of the microbes inside the host and thereby resulting in formation of tightly linked groups [43]. Hence, a set of nodes (microbes) in a microbial community (or ‘hub’) having high coreness centrality values represents ‘core areas of the network’ and the community can be referred to as a ‘core hub community’. Local property measures like degree, betweenness centrality and coreness centrality of a node (microbe) can thus efficiently identify the key nodes in a microbial association network. Further, two networks having a comparable set of nodes, are said to have undergone rewiring, if a large number of new edges appear in one network which are completely absent in the other. A rewiring event between a pair of association network essentially indicates a remodeling the entire microbial community structure. The NetShift methodology applies a set of steps to quantify such rewiring and then utilizes some pre-existing local graph properties coupled with a new score to identify ‘driver’ taxa between a ‘case’ and a ‘control’ microbial association network. An overview of the method is presented in Fig. 1 and the details of the steps are discussed below.

Quantifying overall changes in interaction pattern and community structure

A comparison of graph properties between the association networks corresponding to the ‘control’ and ‘case’ for a particular physiological condition can be done efficiently if the two networks are comparable (as described earlier in the

introduction). Additionally, in order to make the comparison on a comparable set of nodes, a preprocessing step is employed to extract the most common sub-network (Fig. 1b). A sub-network is a smaller network which is a part of the original network derived from the subset of nodes common to both the ‘case’ and ‘control’ networks. This sub-network extraction step is implemented in the NetShift tool and is automatically executed after a user uploads the ‘case’ and ‘control’ networks. A step by step description of the methodology is provided in Supplementary material S1 Fig. 1. It should be noted that NetShift compares only the most common sub-networks between the two input networks and not the full input networks. All references hereafter to ‘control’ and ‘case’ networks essentially imply the most common sub-networks.

In order to understand the salient features of any biological interaction network, it is important to first evaluate the inherent graph properties like clustering co-efficient, density, average path length, etc [30]. Although a comparison of these graph properties between two association networks can yield valuable insights, it cannot capture the associations of individual nodes as well as the extent to which they change upon transition from one state to another. Changes in taxa associations between healthy and disease are important for assessing their role as markers for the disease. Thus, in order to quantify changes in the microbial association networks between two states (‘case’ and ‘control’), it is necessary to consider the participating nodes as well as the changes in their connection in the two states. Such changes can be partly quantified using global properties like clustering co-efficient, density, average path length, total nodes, total edges, and exclusive edges. Although the exclusive edge count between two networks is an indicator of rewiring, it is also valuable to consider the Jaccard Edge Index (JEI) of the compared networks.

$$\text{Jaccard edge index} = \frac{A_E \cap B_E}{A_E \cup B_E} \quad (1)$$

Where, A_E and B_E represent the edge set in network A and B respectively. As the compared networks in this case are the common sub-networks, biases of JEI arising due to unequal node content are minimized.

Community detection algorithms have been extensively used in biological network analysis to detect closely linked hubs of interacting nodes [44–46]. Each community (hub) may represent a group of nodes in a network performing a function with some degree of independence with respect to other communities (hubs). In a microbial association network, these modular hubs may constitute microbial communities that facilitate stable occurrence of a group of taxa processing attributes desirable for inter dependent growth or to meet other ecological requirements [47]. Any changes in such communities would hence be of great interest for

Fig. 1 This diagram shows an overview of how NetShift workflow is applied, **a** beginning with microbial abundance matrix consisting of common taxa from a set of healthy ('control') and disease ('case') samples which is used to create the corresponding microbial association networks, **b** extraction of the most common sub-network between

the two association networks. **c** Quantification of changes in associations of a node between the two networks based on degree cannot distinguish the exact changes. **d** Need to quantify each node for its change from 'healthy' to 'disease' in terms of association and importance relative to the network

understanding the alterations in interaction patterns of the resident microbes (community structure). For example, since the human microbiome has been indicated to be affected in the disease, one would expect changes in the community structure of the resident microbes in the two states (healthy and disease). Such changes in the community structure could be due to microbial rewiring in response to the perturbations [48]. In order to understand such changes, we present a simple yet effective plot of 'community shuffling' to highlight the changes in the communities between the two compared networks. To generate the plot, a community detection algorithm [49] is first used to separately assign a community membership to each node of the 'case' and 'control' networks. Subsequently, each community in the 'control' network is compared with every other community in the 'case' network. The similarities between the communities identified in the 'case' and 'control' networks is calculated in terms of shared nodes and represented as a matrix as well as a network diagram called 'community shuffling plots' (Fig. 2). The coreness

centrality values of the nodes in the 'case' and 'control' network are then used to identify the 'core communities' having the highest distribution 'core' nodes and referred to as 'core hub communities' [50].

Quantifying changes in associations of a single taxon

During the process of rewiring, several nodes (taxa in this case) in the network undergo changes in their individual associations in the two networks. Although centrality measure like node degree can be used to quantify these changes, it fails to score the exact associations. For example, a node (microbial taxon) may have a similar degree (total connections with other resident taxa) in both 'case' and 'control', but may have an entirely different set of associations (the members which are involved in the connections) as demonstrated in Fig. 1b, c. The individual associations (quantified as first neighbor) of a node is an important graph property for analyzing microbial

Fig. 2 The changes in community structure between the two association networks corresponding to the healthy ('control') and disease ('case') is captured using 'community shuffling plots' in form of a network and heatmap generated using the NetShift web tool. **a** The network view constitutes the nodes belonging to the 'control' network plotted along the left half of a circle and the ones belonging to the 'case' plotted along the right half. An edge is assigned between the similar nodes pertaining to the two networks to enable an easy view of community shuffling. Node labels in the network plot are assigned a random color based on the identified community affiliation for assisting easy visual identification. Grayed out node labels corresponds to the ones which do not interact directly with the common sub-

network. **b** The heatmap plot represents the 'control' network communities along the vertical axis and the 'case' network communities along the horizontal axis with a count of the total nodes in the community placed inside a bracket. The identified communities (represented by a community id: C1, C2, etc.) in the 'case' and 'control' networks are prefixed with a 'case_' and 'control_' identifier. The numerical values inside the heatmap show the split in a community from the 'control' to a 'case' network. **c** The labels convention followed in the network view of the community shuffle plot is similar to the one used for the heatmap with an additional suffix containing the labels corresponding to the taxon belonging to a community in a given state ('control' or 'case')

associations. A Jaccard index can be used to quantify such changes in the interacting partners for each node between two graphs. However, Jaccard index being non-directional, may not be effective in this case (see Results) since our primary objective is to quantify exclusively enriched interacting partners in the disease ('case') over the healthy ('control'). A new scoring scheme called 'Neighbor Shift (NESH) Score' has been developed to quantify this directional change in the individual node associations.

Calculate the Neighbor shift (NESH) score

The NESH score for a node common to the 'control' and 'case' networks is calculated using the formula given below:

$$\text{NESH}_{A \rightarrow B} = 1 - \left\{ \left(\frac{[\text{Neighbors}]^A \cap [\text{Neighbors}]^B}{[\text{Neighbors}]^A \cup [\text{Neighbors}]^B} \right) - \left(\frac{[\text{Neighbors}]^B - [\text{Neighbors}]^A}{\text{Max degree in B}} + \left(\frac{[\text{Neighbors}]^B - [\text{Neighbors}]^A}{[\text{Neighbors}]^B \cup [\text{Neighbors}]^A} \right) \right) \right\} \quad (2)$$

Where A and B corresponds to the healthy ('control') and disease ('case') networks respectively. $[Neighbors]^A$: A set of first neighbors of a node A; $[Neighbors]^A \cap [Neighbors]^B$: Intersection of set A and set B, i.e., objects belonging to both set A and set B; $[Neighbors]^A \cup [Neighbors]^B$: Union of set A and set B, i.e., objects belonging to set A or set B; $[Neighbors]^B - [Neighbors]^A$: Relative complement of set B and set A, i.e., objects belonging to set B and not to set A

The score can be broken down into three components, namely, X, Y and Z.

$$\text{Where, } X = \frac{[Neighbors]^A \cap [Neighbors]^B}{[Neighbors]^A \cup [Neighbors]^B};$$

$$Y = \frac{[Neighbors]^B - [Neighbors]^A}{\text{Max degree in B}}$$

$$\text{and, } Z = \frac{[Neighbors]^B - [Neighbors]^A}{[Neighbors]^B \cup [Neighbors]^A}$$

Such that,

$$\text{NESH} = 1 - (X - (Y + Z))$$

The component X provides a measure of the extent of neighborhood similarity irrespective of the direction of change, while component Y and Z penalizes X over exclusive enrichment in the set of first neighbors corresponding to the disease ('case') set over the healthy ('control'). The component Y quantifies the exclusive enrichments over the maximum interacting partners a node can have in the disease (given by the maximum degree in the network). On the other hand, the component Z quantifies the exclusive set enrichments over the union of the interacting partners of the node in the two compared networks. For ease of analysis the score can be easily translated to a different scale by subtracting the value of $(X - (Y + Z))$ from 1 such that a higher NESH score would account for a higher neighborhood shift for a compared node.

Computation of relative increase in importance of a node and prediction of 'drivers'

The scaled betweenness ' B_{scaled} ' for each of the common set of nodes in both of the networks is first calculated using the formula:

$$B_{\text{scaled}} = \frac{B_{\text{calculated}} - B_{\text{min}}}{B_{\text{max}} - B_{\text{min}}} \quad (3)$$

Where, $B_{\text{calculated}}$, B_{min} , and B_{max} correspond to the calculated, minimum, and maximum betweenness values, respectively. The delta betweenness (ΔB) for each of the common nodes from healthy ('control') network (A) to the disease ('case') network (B) is computed as:

$$\Delta B = B_{\text{scaled}(B)} - B_{\text{scaled}(A)} \quad (4)$$

An increase in betweenness of a node in the disease ('case') is further used as a measure to quantify the increase

in importance of the considered node. A taxon in the disease ('case') with an altered set of associations (identified by a high NESH value), while still being increasingly important (identified with a positive ΔB) for the whole network, necessarily holds a key significance in microbial interplay and is predicted as a 'driver'.

Implementation of the NetShift as a web-application

The NetShift methodology has been implemented as a web-application for assisting easy analysis of microbial association networks obtained from 'case-control' studies available at <https://web.rniapps.net/netshift>. The tool requires two microbial association networks ('case' and 'control') as inputs in a simple 'edge-list' format. 'Edge-list' refers to a text file containing a list of connected node pairs (separated by a 'tab' delimiter) in each line. On successful execution with the input files, the output can be analyzed interactively to visualize changes in community structure and to identify the 'driver' taxa. The NetShift output contains four components:

Changes in global graph properties

The first component of the output is designed to display the global graph properties, namely density, clustering coefficient, average path length, total nodes, total edges, and exclusive edges along with reporting the Jaccard edge index (JEI) between the two common sub-networks. These network properties can be used to get an overall idea of the relative change in network topology.

The Network-shift plot

The most common sub-network corresponding to the 'case' and 'control' is displayed using a circular graph. The plot consists of nodes arranged along the periphery of a circle and ordered by their community affiliations in the disease ('case') [49]. The associations between the nodes are shown as connecting edges within the circle which can be selectively highlighted upon clicking on a node. Edges in the network are colored as per their affiliation to the networks (green for edges exclusive to the 'control', red for ones exclusive to the 'case' and blue for edges common to both). The diameter of the nodes are mapped to the NESH score while a red color is assigned to nodes having a positive increase in betweenness from 'control' to 'case' (ΔB). Hence, nodes colored red and bigger in size are the predicted 'drivers' (Supplementary material S1 Figure 4–6).

Tabulated summary of network properties

Local network properties like node degree (in 'case' and 'control'), union, intersection and exclusive partner count,

coreness, delta (or increase in) betweenness, NESH score, and community affiliations are displayed in an interactive table. Each column can be sorted by their respective values on clicking over the column name. An interactive search box is also provided to assist users to look for a specific node (or taxon) name.

Community shuffle plots

The all versus all similarities between the ‘control’ and ‘case’ networks are represented as ‘community shuffling plots’ in form of heatmap as well as a network view. The heatmap displays the identified communities in the ‘control’ and ‘case’ networks as rows (i) and columns (j) of a matrix with each cell representing the similarity (S_{ij}) between the *i*th ‘control’ and *j*th ‘case’ community. The value of S_{ij} is obtained as the intersecting nodes between the node contents of the two communities and assigned a color based on the value as shown in Fig. 2. More community splits (from ‘control’ to ‘case’) represent increased changes in microbial assembly rules. Hence, plots having less shuffling will show less horizontal splits (in the blocks of the matrix) and individual blocks will have a higher cell value (S_{ij}). The network view of community shuffling on the other hand allows visualizing the exact changes in the community members between the two networks. In the network view, the nodes belonging to the ‘control’ network are plotted along the left half of a circle and the ones belonging to the ‘case’ are plotted along the right half. Each community in the two networks is assigned a distinct color and an edge is assigned between the similar nodes pertaining to the two networks, thereby enabling an easy visualization of the community shuffling. Mouse over on a community label highlights its edge connection. Nodes in the network are mapped to their ‘coreness’ values by default allowing visualization of ‘core communities’ which can be changed to degree or betweenness using the available option. The currently available web server implementation of NetShift is expected to undergo continuous development; future plans include development of a Cytoscape [51] plugin for the application to grant its usage and diffusion to the larger community of computational biologists.

Interpretation of NetShift outputs

After loading a ‘case’ and a ‘control’ network, the first and foremost step pertains to obtaining an overall idea of the differences in the association networks. While the bar plots of global graph properties provide information regarding the differences in the overall connectivity of the networks (density), modular differences (cluster coefficient) and the compactness of the networks (average path length), the plots for total node, total edge and exclusive edge help to

decipher the network similarity and extent of rewiring. The exclusive edge count and the Jaccard edge index can be further used to assess rewiring between the two networks. Although higher exclusive edge count indicates higher rewiring, smaller values of Jaccard edge index provide a more concrete evidence for the same. Once a rewiring event is confirmed, as expected, there will be a change in both the local and global community structure of the two networks owing to the changes in microbial associations. The local or individual node level changes can be deciphered using the Network-shift plot and the tabulated summary as described in the earlier section. Nodes highlighted in red and big in size are particularly the ones which are identified to be the ‘driver’ nodes by the NetShift program. A user may list the identified driver names for further validation; interactively query their associative differences in the two states using the network and probe deeper into other local graph properties like degree, betweenness and NESH score using the tabulated summary. Following this, the global changes in the community structure can be deciphered using the community shuffle plots. The identified communities (represented by a community id: C1, C2, etc.) in the ‘case’ and ‘control’ networks are prefixed with a ‘case_’ and ‘control_’ identifier [49]. The similar convention is also followed in the network view of the community shuffle plot, where in addition to prefixing the network and community identifier to each taxon name, the labels corresponding to each taxon belonging to a same community in a state are assigned a unique color. For example a node labeled ‘case_C1_Akkermansia’ represents the microbial genera *Akkermansia* in the ‘case’ network belonging to the community assigned an id ‘C1’ (Fig. 2c). The extent to which a community is conserved can be easily identified using the community shuffle plots. For example, if a ‘control’ specific community shows a high change (little conservation) in the ‘case’ network, the corresponding row for that ‘control’ community in the heatmap will have high number of shaded cells with low values, indicating a split of the community members in the ‘case’ network. In order to enquire the exact changes, the network view of community shuffling plot can be interactively queried to visualize the changes using a mouse over operation on any member of the query ‘control’ community. Further, using the ‘highlight nodes’ feature available in NetShift, users can easily map the coreness centrality value of each node to the node sizes in the community shuffling plot. A community having greater number of nodes with high coreness centrality values is called a ‘core hub community’. Such communities serve as units of increased relevance. Thus, changes in community structure of ‘core hub communities’ between healthy (‘control’) and disease (‘case’) can help identification of events like pathogen colonization as demonstrated in one of the case studies later in the manuscript.

Fig. 3 Comparison of NESH (neighbor shift) score with Jaccard index on several scenarios. The NESH score is able to assign a correct score in all the cases. Higher NESH score signifies greater difference

Representative scenario 'control' → 'case'		Biological insight	Jaccard Index	NESH score
<div>1</div> <div>2</div> <div>3</div>		Node has similar count of neighbors with varying interacting partners (which may or may not be overlapping)	1	0
			0	2.33
			0.43	1.19
<div>4</div> <div>5</div> <div>6</div>		Node has higher number of neighbors in 'case' network with varying interacting partners (which may or may not be overlapping)	0.5	1.5
			0.6	1.13
			0.14	2.10
<div>7</div> <div>8</div> <div>9</div>		Node has lower number of neighbors in 'case' network with varying interacting partners (which may or may not be overlapping)	0	1.83
			0.125	1.46
			0.5	0.5

Creation of microbial association networks for the case studies

Datasets used for the case studies were obtained from the EMBL Microbiome database [52]. The genus level taxonomic abundance files were downloaded and filtered to include genera present in at least 30% of the samples. Similar to a differential abundance analysis, genera exclusive to either the 'case' or 'control' datasets were excluded. The final abundance matrices were used to generate the microbial association networks for each datasets using a bootstrap and correlation based approach using CCREPE tool [38, 41]. This approach calculates the statistical significance of a correlation (Spearman in this case) between a given pair of genera using a re-sampling and null distributions of correlation values (obtained by 1000 iterations). An edge is assigned between a pair of

genera for every predicted significant positive correlation ($p < 0.005$).

Results

Validation of the NESH score

In order to characterize the importance of the NESH score in the context of actual networks, we created three categories (each consisting of three representative conditions). Every condition represents a biologically relevant scenario where a node has a change in interacting partners between a 'control' and 'case' network. As shown in Fig. 3, the three categories correspond to conditions where a node has similar, lower or higher number of interacting partners between the compared networks. With respect to microbial

association networks, these correspond to cases where a microbe shows an altered set of interacting partners in a diseased/perturbed ('case') state. The new partners in the disease as a result of the altered interactions hold a greater significance in quantifying the importance of a microbe. When a microbe is associated with a completely different set of partners in the perturbed state, it essentially contributes in remodeling the entire microbial community structure. These types of microbes can be assumed to be a key member or a potential 'driver' of the infection. Earlier studies have referred to such microbes as 'alpha-bugs' and implicated their role in inducing carcinogenesis [29]. Scenarios 2, 6, and 7 illustrate such cases and have been successfully captured by the NESH score (with high values). Similarly, a lower value is assigned when the altered set of associated nodes in the disease ('case') shares a higher overlap with those in the healthy ('control') (as in scenarios 1, 9, 5, and 3). Merely losing a set of associated partners is also assigned a lower score value (scenario 9) as it signifies lowering of importance. Nodes with exactly similar associations are considered neutral to the change and hence are scored lowest (scenario 1). Contrary to the NESH score, the Jaccard index of the neighbors fails to correctly quantify these changes in most cases and is also unable to differentiate several other changes (scenario 2–7 and 9–4). In a nutshell, the NESH score is minimum when the associated partners of a node are same in both 'case' and 'control' followed by scenarios when there is only a subset of the associated partners present in the 'case' and maximum when a completely new set of associated partners appear in the 'case' network.

Case studies with real world microbiome data

Two studies on gut and oral (saliva and plaque) microbiome samples in allergy (asthma) and HIV, respectively, were chosen for as case studies [53, 54]. Both datasets had a set of healthy ('control') and diseased ('case') samples and reported no clear differences between the 'case' and 'control'. Hence, the NetShift analysis route was chosen to check whether any additional insights can be obtained from these datasets.

Analyzing gut microbiome for Allergy dataset

Several recent reports have established the emerging relationship between allergy (asthma, atopy, etc.) and gut microbiome [55]. One of the hypothesized mechanisms relating to this connection pertains to alteration in gut microbial composition, which in turn influences allergic immune responses [56]. However, these changes may not always be apparent from the taxonomic composition statistics or abundance based clustering. Analyzing microbial

association networks may thus complement in increasing our understanding in allergic diseases. We selected a dataset for analysis where the authors studied 21 adult patients suffering from allergic asthma and compared their fecal microbiota with those from 22 healthy controls [53]. The authors did not find any clustering or differentially abundant taxa specific to the allergy samples.

We generated the microbial association network (as described in the Methods section) for the allergy ('case') and healthy ('control') sets and applied the NetShift workflow using the implemented web-server. The observed low Jaccard edge index (0.17) pointed towards sufficient rewiring between the two networks. Prominent changes in the community structure were clearly evident from the community shuffling plots (Fig. 2a, b). High density and lower average path length in the allergy network indicates higher information transport which might suggest a colonization activity. A closer look into the community shuffling network (Fig. 2a) revealed two distinct 'core hub community', one in the 'control' ('control_C1') and other in the 'case' network ('case_C2'). While the 'control_C1 core hub' consisted of genera namely *Dehalobacterium*, *Eubacterium*, *Desulfovibrio*, and *Akkermansia*, the 'case_C2 core hub' was composed of genera *Fusibacter*, *SMB53*, *Alkaliphilus*, and *Turicibacter*. The member genera in the 'control core hub' are mostly well known gut commensals like *Eubacterium* which is a butyrate producer [57]. *Dehalobacterium* is known to produce acetate from dichloromethane [58] which is utilized mainly by other butyrate-producing bacteria found in the human colon [59]. *Desulfovibrio* is a prominent Sulfate-reducing bacteria (SRB) which helps to maintain redox balance in the gut [60] and *Akkermansia* is the most abundant mucus degrading bacteria found in healthy individuals [61]. On the other hand, genera in the 'case core hub' had well known gut pathogens like *Turicibacter* which are known to be strongly associated with immune function and bowel disease [62]. Additionally, the members of the 'case core hub' were seen to have lower coreness values in the 'control' network as compared to the 'control core hub' (Fig. 2a) which might be due to the commensals effort in subduing pathogen colonization. However, in the disease ('case' network), the pathogenic 'case core hub' is seen to clearly colonize while commensals cease to be the core members. Interestingly, a closer look into the 'driver genera' between the 'case' and 'control' (Supplementary material S1 Figure 4), revealed *Ruminococcus*, *Clostridium*, *Granulicatella*, *Desulfohalobium* and *Dorea* as the top five critical nodes which had a high NESH score and an increased betweenness (quantified by bigger red nodes). When investigated only for high NESH score nodes (i.e., nodes undergoing high rewiring irrespective of increase in importance), the same taxa were observed except *Dorea* which was replaced by *Turicibacter*.

Ruminococcus, *Clostridium*, *Desulfotomaculum*, and *Dorea* are mostly reported as commensals and are known to digest resistant starches thereby contributing to maintenance of gut homeostasis [63, 64] or produce acetates to help other butyrate producing commensals [59, 65]. These genera can hence be considered as the commensal 'drivers'. On the other hand, *Granulicatella*, also observed as one of the 'drivers', is reported to have pathogenic roles [66] and hence can be considered as a pathogenic 'driver'. The 'case' specific associations (displayed as red edges) of *Granulicatella* as seen in the common sub-network (Supplementary material S1 Figure 4) point toward other pathogens like *Staphylococcus* and *Veillonella* [67] also having high NESH scores thereby strengthening its role in pathogen colonization.

The presence of a particular combination and relative abundances of commensal microbes are known to generate distinct immune environments and immune responses in the host [68] and prevent pathogen colonization. However, some pathogens are able to disrupt the resistance and subsequently infect the host, especially during disturbed immune response in an infection [69]. The associations of *Turibacter* and *Granulicatella* may be an indicator of two independent pathogens trying to colonize in the 'case' network but only one being able to become a driver. On the other hand, the existence of prominent commensal 'driver' genera might be an effort by them to prevent the pathogen colonization.

Oral and lung microbiome analysis in HIV patients

Recent investigations have indicated the relationship between the altered microbiome and HIV [70]. In pursuit of understanding the role of microbiome in HIV infection, we selected another publicly available study on oral (saliva and plaque) microbiome in HIV individuals. The original study reported the saliva and plaque microbiome to be distinctly different (in both HIV positive and negative individuals). However, neither the oral nor the plaque microbiome was reported to have any abundance level differences between the HIV positive and negative groups. For community analysis, the authors in the original study [54] used Jaccard and thetaYC distance matrices to quantify differences between the 'case' and 'control' groups in the two datasets. However, these matrices may be limited to reproduce only abundance level variations without any weightage to inter microbial associations. Hence, NetShift was used to investigate whether any community level differences are evident which may complement the original report. The NetShift plots showed lesser community shuffling in the saliva dataset (Supplementary material S1 Fig. 3) as compared to the plaque (Supplementary material S1 Fig. 2). The network plot also showed more conserved core members in the

saliva dataset. The Jaccard edge index indicated high rewiring in the plaque (0.228) and a fair amount in the saliva (0.581). Also low average path length and high network density in plaque 'case' network provided some signs of colonization. While saliva dataset had four distinct drivers (Supplementary material S1 Figure 6), namely, *Porphyromonas*, *TG5*, *Oribacterium*, and *Leucobacter*, the plaque dataset showed *Selenomonas*, *Fusobacterium*, *Catonella*, *Capnocytophaga*, *Streptococcus*, and *Granulicatella* as prominent drivers (Supplementary material S1 Figure 5). One of the drivers identified in the saliva dataset, *Porphyromonas*, is known to be a major etiologic agent of chronic periodontitis [71] and also reported to play important role in dysimmunity and inflammation in oral lichen planus [72] which is related to HIV infection [73]. On the other hand, all the 'drivers' identified in plaque dataset consisting of genera namely *Selenomonas*, *Fusobacterium*, *Catonella*, and *Granulicatella*, are well reported periodontal pathogens [74–77]. Interestingly, periodontal diseases are often reported to be first clinical sign of human HIV infection [78, 79] which is caused by the coordinated action of a complex microbial community resulting in inflammation of tissues supporting the teeth. Thus, the importance of understanding the bacterial component of HIV especially in the plaque microbiome, their changes in community structure upon HIV invasion and associated differences across diseased and healthy can help in early diagnostics as well as in treatment.

In order to perform meaningful analysis on a 'case-control' study, it is important to appropriately choose the 'control' set. For example, to understand a disease ('case') population, a comparable 'control' set can be selected by matching characteristics like age, geography, food habit, environmental conditions, drug intake, etc. of the individuals belonging to the healthy ('control'). Additionally, it is advisable to use similar methods and reagents for sampling, extraction as well as sequencing of DNA followed by similar computational methods to obtain the microbial abundance information for the samples belonging to each of the sets. It is preferable to select 'case' matched 'control' group over using 'normal' individuals from an altogether different study [80]. In time series experiments pertaining to an intervention study, individuals can be treated as their own control using the collected baseline samples (i.e., pre-intervention). However, in order to generate statistically significant microbial association networks, it is essential to have sufficient sample time points for the pre and post treatment states [81]. When time series data is not available, closely matched control subjects (like monozygotic twins) may serve as better alternatives. For a well designed study, the corresponding 'case' and 'control' microbial association networks are inherently expected to be comparable and can be used as an input for 'NetShift'.

Conclusion

The resident microbiome in the human body is known to harbor an ecosystem that fosters host immune development as well as protect it from pathogen colonization, often referred to as colonization resistance [82]. However, the mechanism used by pathogens to overcome these barriers of colonization resistance is still poorly understood [69]. The “keystone pathogen” hypothesis is one such which says that certain low abundant microbial pathogen can orchestrate inflammatory disease by remodeling a normal microbiota into a dysbiotic one [28]. The observations in the allergy microbiome in the case study show some evidences of such pathogen colonization and subsequent disempowerment of the commensals. Similarly, several pathogens identified in the HIV dataset as ‘drivers’ may serve as early indicators of a pathogen invasion. Studying microbial co-occurrence networks serve as a valuable means for understanding changes in association patterns, which in many cases cannot be inferred based on differential abundances. In this communication, we present a special analysis methodology, called NetShift, that can quantify these salient changes and utilize the same to identify taxonomic groups that may act as ‘drivers’ for a disease. The Net-shift web tool is expected to help researchers in easily applying the methodology for analyzing ‘case-control’ microbiome studies as well as complement other existing methods.

Acknowledgements We acknowledge Mr. Pramod Nanadikar and Mr. Govind Gopal for their help in deploying the web application. We also acknowledge Dr. Anirban Dutta and Dr. Swadha Anand for their help in preparing the figures and proofreading the manuscript. Kuntal Kumar Bhusan is an industry sponsored PhD student at Chemical Engineering & Process Development Division, CSIR-National Chemical Laboratory (NCL), Pune 411008 (India) and would like to acknowledge the Academy of Scientific and Innovative Research (AcSIR) and NCL for its support.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Eloe-Fadrosh EA, Rasko DA. The human microbiome: from symbiosis to pathogenesis. *Annu Rev Med*. 2013;64:145–163.
- Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, Reyes-Gavilán GC, Salazar N. et al. Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Front Microbiol*. 2016;7:185
- Das A, Srinivasan M, Ghosh TS, Mande SS. Xenobiotic metabolism and gut microbiomes. *PLoS ONE*. 2016;11:e0163099
- Brown J, de Vos WM, DiStefano PS, Doré J, Huttenhower C, Knight R, et al. Translating the human microbiome. *Nat Biotechnol*. 2013;31:304–308.
- Garrett WS, Gordon JI, Glimcher LH. Homeostasis and inflammation in the intestine. *Cell*. 2010;140:859–870.
- Huttenhower C, Kostic AD, Xavier RJ. Inflammatory bowel disease as a model for translating the microbiome. *Immunity*. 2014;40:843–854.
- Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. *J Physiol*. 2009;587:4153–4158.
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490:55–60.
- Knights D, Lassen KG, Xavier RJ. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. *Gut*. 2013;62:1505–1510.
- Li B, Selmi C, Tang R, Gershwin ME, Ma X. The microbiome and autoimmunity: a paradigm from the gut-liver axis. *Cell Mol Immunol*. 2018;15:595–609.
- Vieira SM, Pagovich OE, Kriegel MA. Diet, microbiota and autoimmune diseases. *Lupus*. 2014;23:518–526.
- Mayer EA, Tillisch K, Gupta A. Gut/brain axis and the microbiota. *J Clin Invest*. 2015;125:926–938.
- Rogers GB, Keating DJ, Young RL, Wong M-L, Licinio J, Wesselingh S. From gut dysbiosis to altered brain function and mental illness: mechanisms and pathways. *Mol Psychiatry*. 2016;21:738–748.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature*. 2007;449:804–810.
- Cho I, Blaser MJ. The Human Microbiome: at the interface of health and disease. *Nat Rev Genet*. 2012;13:260–270.
- D’Argenio V, Salvatore F. The role of the gut microbiome in the healthy adult status. *Clin Chim Acta*. 2015;451:97–102.
- Ghosh TS, Gupta SS, Bhattacharya T, Yadav D, Barik A, Chowdhury A, et al. Gut microbiomes of Indian children of varying nutritional status. *PLoS ONE*. 2014;9:e95547.
- Gupta SS, Mohammed MH, Ghosh TS, Kanungo S, Nair GB, Mande SS. Metagenome of the gut of a malnourished child. *Gut Pathog*. 2011;3:7.
- Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. *Curr Opin Gastroenterol*. 2015;31:69–75.
- Faust K, Lima-Mendez G, Lerat J-S, Sathirapongsasuti JF, Knight R, Huttenhower C, et al. Cross-biome comparison of microbial association networks. *Front Microbiol*. 2015;6:1200.
- Yadav D, Ghosh TS, Mande SS. Global investigation of composition and interaction networks in gut microbiomes of individuals belonging to diverse geographies and age-groups. *Gut Pathog*. 2016;8:17.
- Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. *Science*. 2012;336:1262–1267.
- Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol*. 2012;10:538–550.
- Belenguer A, Duncan SH, Calder AG, Holtrop G, Louis P, Lobley GE, et al. Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. *Appl Environ Microbiol*. 2006;72:3593–3599.
- Flint HJ, Duncan SH, Scott KP, Louis P. Interactions and competition within the microbial community of the human colon: links between diet and health. *Environ Microbiol*. 2007;9:1101–1111.
- Round JL, Mazmanian SK. The gut microbiome shapes intestinal immune responses during health and disease. *Nat Rev Immunol*. 2009;9:313–323.
- Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasi-kala M, Reddy DN. Role of the normal gut microbiota. *World J Gastroenterol*. 2015;21:8787–8803.
- Hajishengallis G, Darveau RP, Curtis MA. The keystone pathogen hypothesis. *Nat Rev Microbiol*. 2012;10:717–725.

29. Sears CL, Pardoll DM. Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. *J Infect Dis*. 2011;203:306–311.
30. Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, et al. Using graph theory to analyze biological networks. *BioData Min*. 2011;4:10.
31. Yaveroğlu ÖN, Milenković T, Pržulj N. Proper evaluation of alignment-free network comparison methods. *Bioinformatics*. 2015;31:2697–2704.
32. Di Lena P, Wu G, Martelli PL, Casadio R, Nardini C. MIMO: an efficient tool for molecular interaction maps overlap. *BMC Bioinformatics*. 2013;14:159.
33. Mandakovic D, Rojas C, Maldonado J, Latorre M, Travisany D, Delage E, et al. Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. *Sci Rep*. 2018;8:5875.
34. Kuntal BK, Dutta A, Mande SS. CompNet: a GUI based tool for comparison of multiple biological interaction networks. *BMC Bioinformatics*. 2016;17:185.
35. Zhou X, Nardini C. A method for automated pathogenic content estimation with application to rheumatoid arthritis. *BMC Syst Biol*. 2016;10:107.
36. Peura S, Bertilsson S, Jones RI, Eiler A. Resistant microbial cooccurrence patterns inferred by network topology. *Appl Environ Microbiol*. 2015;81:2090–2097.
37. Wang H, Wei Z, Mei L, Gu J, Yin S, Faust K, et al. Combined use of network inference tools identifies ecologically meaningful bacterial associations in a paddy soil. *Soil Biol Biochem*. 2017;105:227–235.
38. CCREPE. Bioconductor. <http://bioconductor.org/packages/ccrepe/>. Accessed on 24 March 2018.
39. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLOS Comput Biol*. 2012;8:e1002687.
40. Deng Y, Jiang Y-H, Yang Y, He Z, Luo F, Zhou J. Molecular ecological network analyses. *BMC Bioinformatics*. 2012;13:113.
41. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. *PLOS Comput Biol*. 2012;8:e1002606.
42. Lü L, Zhou T, Zhang Q-M, Stanley HE. The H-index of a network node and its relation to degree and coreness. *Nat Commun*. 2016;7:10168.
43. Dani A. Colonization and infection. *Cent European J Urol*. 2014;67:86–87.
44. Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci USA*. 2006;103:8577–8582.
45. Lai D, Lu H, Nardini C. Enhanced modularity-based community detection by random walk network preprocessing. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2010;81:066118.
46. Kuntal BK, Ghosh TS, Mande SS. Community-analyzer: a platform for visualizing and comparing microbial community structure across microbiomes. *Genomics*. 2013;102:409–418.
47. Handelsman J. *Encyclopedia of life sciences*. Chichester, UK: John Wiley & Sons, Ltd; 2007.
48. Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc Natl Acad Sci USA*. 2013;110:12804–12809.
49. Yang Z, Algesheimer R, Tessone CJ. A comparative analysis of community detection algorithms on artificial networks. *Sci Rep*. 2016;6:30750.
50. Seidman SB. Network structure and minimum degree. *Social Netw*. 1983;5:269–287.
51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–2504.
52. Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, et al. EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res*. 2018;46:D726–D735.
53. Hevia A, Milani C, López P, Donado CD, Cuervo A, González S, et al. Allergic patients with long-term asthma display low levels of bifidobacterium adolescentis. *PLoS One*. 2016;11:e0147809–e0147809.
54. Kistler JO, Arirachakaran P, Poovorawan Y, Dahlén G, Wade WG. The oral microbiome in human immunodeficiency virus (HIV)-positive individuals. *J Med Microbiol*. 2015;64:1094–1101.
55. Fujimura KE, Lynch SV. Microbiota in allergy and asthma and the emerging relationship with the gut microbiome. *Cell Host Microbe*. 2015;17:592–602.
56. Fujimura KE, Demoor T, Rauch M, Faruqi AA, Jang S, Johnson CC, et al. House dust exposure mediates gut microbiome Lactobacillus enrichment and airway immune defense against allergens and virus infection. *Proc Natl Acad Sci USA*. 2014;111:805–810.
57. Rivièrè A, Selak M, Lantin D, Leroy F, De Vuyst L. Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. *Front Microbiol*. 2016;7:979.
58. Chen G, Murdoch RW, Mack EE, Seger ES, Löffler FE. Complete genome sequence of dehalobacterium formicoaceticum strain dmc, a strictly anaerobic dichloromethane-degrading bacterium. *Genome Announc* 2017;5:e00897–17.
59. Duncan SH, Holtrop G, Lobley GE, Calder AG, Stewart CS, Flint HJ. Contribution of acetate to butyrate formation by human faecal bacteria. *Br J Nutr*. 2004;91:915–923.
60. Rey FE, Gonzalez MD, Cheng J, Wu M, Ahern PP, Gordon JI. Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proc. Natl Acad. Sci. USA*. 2013;110:13582–13587.
61. Collado MC, Derrien M, Isolauri E, de Vos WM, Salminen S. Intestinal integrity and akkermansia muciniphila, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl Environ Microbiol*. 2007;73:7767–7770.
62. Clark A, Mach N. Exercise-induced stress behavior, gut-microbiota-brain axis and diet: a systematic review for athletes. *J Int Soc Sports Nutr*. 2016;13:43.
63. Ze X, Duncan SH, Louis P, Flint HJ. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J*. 2012;6:1535–1543.
64. Lopetuso LR, Scaldaferri F, Petito V, Gasbarrini A. Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog*. 2013;5:23.
65. Blaut M, Collins MD, Taras D, Dorea. *Bergey's Manual of Systematics of Archaea and Bacteria*. 2015. American Cancer Society, pp 1–7.
66. Cargill JS, Scott KS, Gascoyne-Binzi D, Sandoe JAT. Granulocytocidal infection: diagnosis and management. *J Med Microbiol*. 2012;61:755–761.
67. Rigsbee L, Agans R, Shankar V, Kenche H, Khamis HJ, Michail S, et al. Quantitative profiling of gut microbiota of children with diarrhea-predominant irritable bowel syndrome. *Am J Gastroenterol*. 2012;107:1740.
68. Ivanov II, Honda K. Intestinal commensal microbes as immune modulators. *Cell Host Microbe*. 2012;12:496–508.
69. Stecher B, Hardt W-D. The role of microbiota in infectious disease. *Trends Microbiol*. 2008;16:107–114.
70. Saxena D, Li Y, Yang L, Pei Z, Poles M, Abrams WR, et al. Human microbiome and HIV/AIDS. *Curr HIV/AIDS Rep*. 2012;9:44–51.

71. How KY, Song KP, Chan KG. *Porphyromonas gingivalis*: an overview of periodontopathic pathogen below the Gum Line. *Front Microbiol.* 2016;7:53.
72. Wang K, Lu W, Tu Q, Ge Y, He J, Zhou Y et al. Preliminary analysis of salivary microbiome and their potential roles in oral lichen planus. *Sci Rep.* 2016;6:22943.
73. Ficarra G, Flaitz CM, Gaglioti D, Piluso S, Milo D, Adler-Storthz K, et al. White lichenoid lesions of the buccal mucosa in patients with HIV infection. *Oral Surg Oral Med Oral Pathol.* 1993;76:460–466.
74. Gonçalves LFH, Fermiano D, Feres M, Figueiredo LC, Teles FRP, Mayer MPA, et al. Levels of *Selenomonas* species in generalized aggressive periodontitis. *J Periodontal Res.* 2012;47:711–718.
75. Han YW. *Fusobacterium nucleatum*: a commensal-turned pathogen. *Curr Opin Microbiol.* 2015;0:141–147.
76. Kumar PS, Griffen AL, Moeschberger ML, Leys EJ. Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *J Clin Microbiol.* 2005;43:3944–3955.
77. Karched M, Bhardwaj RG, Asikainen SE. Coaggregation and biofilm growth of *Granulicatella* spp. with *Fusobacterium nucleatum* and *Aggregatibacter actinomycetemcomitans*. *BMC Microbiol.* 2015;15:114.
78. Tenenbaum HC, Mock D, Simor AE. Periodontitis as an early presentation of HIV infection. *CMAJ.* 1991;144:1265–1269.
79. Ryder MI, Nittayananta W, Coogan M, Greenspan D, Greenspan JS. Periodontal disease in HIV/AIDS. *Periodontol* 2000. 2012;60:78–97.
80. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, et al. Conducting a microbiome study. *Cell.* 2014;158:250–262.
81. Baksi KD, Kuntal BK, Mande SS. ‘TIME’: a web application for obtaining insights into microbial ecology using longitudinal microbiome data. *Front Microbiol.* 2018;9:36.
82. Buffie CG, Pamer EG. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat Rev Immunol.* 2013;13:790–801.