Accepted: 6 September 2016

DOI: 10.1111/exd.13210

METHODS REVIEW (INVITED)

WILEY Experimental Dermatology

Experimental metagenomics and ribosomal profiling of the human skin microbiome

Pamela Ferretti[1] | **Stefania Farina**[2] | **Mario Cristofolini**[2] | **Giampiero Girolomoni**[3] | **Adrian Tett**[1] | **Nicola Segata**[1]

[1]Centre for Integrative Biology, University of Trento, Trento, Italy

[2]Istituto G. B. Mattei, Trento, Italy

[3]Section of Dermatology, Department of Medicine, University of Verona, Verona, Italy

Correspondence
Adrian Tett and Nicola Segata, Centre for Integrative Biology, University of Trento, Trento, Italy.
Emails: adrianjames.tett@unitn.it (AT) and nicola.segata@unitn.it (NS)

Funding information
Istituto G.B. Mattei; Fondazione Caritro; European Union's Seventh Framework Programme, Grant/Award Number: PCIG13-GA-2013-618833; Centre for Integrative Biology (University of Trento); MIUR "Futuro in Ricerca", Grant/Award Number: E68C13000500001

Abstract

The skin is the largest organ in the human body, and it is populated by a large diversity of microbes, most of which are co-evolved with the host and live in symbiotic harmony. There is increasing evidence that the skin microbiome plays a crucial role in the defense against pathogens, immune system training and homoeostasis, and microbiome perturbations have been associated with pathological skin conditions. Studying the skin resident microbial community is thus essential to better understand the microbiome-host crosstalk and to associate its specific configurations with cutaneous diseases. Several community profiling approaches have proved successful in unravelling the composition of the skin microbiome and overcome the limitations of cultivation-based assays, but these tools remain largely inaccessible to the clinical and medical dermatology communities. The study of the skin microbiome is also characterized by specific technical challenges, such as the low amount of microbial biomass and the extensive human DNA contamination. Here, we review the available community profiling approaches to study the skin microbiome, specifically focusing on the practical experimental and analytical tools necessary to generate and analyse skin microbiome data. We describe all the steps from the initial samples collection to the final data interpretation, with the goal of enabling clinicians and researchers who are not familiar with the microbiome field to perform skin profiling experiments.

KEYWORDS

metagenomics, next generation sequencing, profiling, ribosomal, skin microbiome, skin microbiome, metagenomics, ribosomal profiling

1 | INTRODUCTION

The human body is inhabited by trillions of microbes,[1] with up to one million bacteria per square centimetre colonizing the skin.[2] The importance of this skin-associated diversity—the skin microbiome—is well recognized but many members of the skin microbiome remain elusive if studied using traditional cultivation-based approaches. The recent advent of next-generation sequencing platforms has revolutionized the field by enabling the ecology of the (skin) microbiome to be studied without the need of an in vitro cultivation step. Here, we summarize and review the available approaches to profile the skin microbiome.

2 | THE SKIN MICROBIOME: THE IMPORTANCE OF STUDYING OUR MICROBIAL INTERFACE

The skin is the largest organ in the human body providing an outer protective layer against physicochemical damages and microbial infections. Rather than simply a static shield, the skin is a dynamic and complex ecosystem characterized by different environmental niches, which are inhabited by heterogeneous microbial communities. The healthy skin microbiome includes hundreds of species,[3–5] and its composition is dependent on the microenvironment characteristics, including humidity

and temperature, and the presence of sebaceous and sweat glands.[6–8] However, far from being a passive resident of the skin, the host and the microbiome interact at the systems level. Members of the microbial community can model the host local activity by releasing factors[6,9] and *Staphylococcus epidermidis* strains can drive small-scale environmental changes[10] as well as perform specific functions useful to the host.[9–11] The skin microbiome is also involved in the immune system training,[12] enhancement of innate barrier immunity[13] and homoeostasis,[10,14] by triggering inflammatory cell recruitment and cytokines release.[15] The skin displays a large variability in terms of composition and species abundances between individuals (interpersonal variability), but it has also been shown that there are strong underlying individual specific signatures (intra-personal variability).[4,6,16,17] The site-specific variability of the skin microbiome discourages the comparison of results provided by different body areas. Furthermore, age and gender can be confounding factors that need to be considered and properly addressed.[18–20] For these reasons, it is crucial to have properly powered cohorts and *ad hoc* longitudinal study designs.[21–24]

Propionibacterium acnes and *S. epidermidis* are two prevalent members of the skin microbiome, but many other bacteria from the Actinobacteria, Firmicutes, Bacteroidetes and Proteobacteria phyla are frequently found.[3,23] The micro-eukaryotic component of the microbiome also comprises many organisms (most notably *Malassezia* spp.[7,25]). Viruses are also present on the skin: bacteriophages, in particular, play a crucial role in shaping bacterial communities,[26] but the study of microbiome–virome relationships is still in its infancy.[27–29] Although several microbiome studies are unravelling the complementary aspects of the skin microbiome,[28,30] further research is required to truly understand the complex interplay between the host and the resident microbiome.

3 | LOW- AND HIGH-THROUGHPUT APPROACHES FOR STUDYING THE SKIN MICROBIOME

Cultivation-based approaches rely on isolating and growing microbes in vitro. While there are distinct advantages to cultivating organisms, the vast majority of microorganisms are refractory to cultivation or are unable to grow under the selected conditions,[31] and thus, this approach dramatically underestimates the complexity of the sample.[32] Cultivation-free methods, based on collecting and analysing total DNA from the community under study, have revolutionized our understanding of the host's resident microbiome. This total DNA can be sequenced directly, as in shotgun metagenomics, or specific microbial regions are first amplified and subsequently sequenced. In this second case, the targeted region is usually contained in the rRNA genes; herein, this technique is referred to as ribosomal community profiling.

The most common high-throughput ribosomal community profiling approach utilizes the 16S rRNA gene that is universally conserved in all bacteria and archaea. In this review, we focus on 16S rRNA profiling although, for eukaryotic microbes including fungi, the equivalent 18S rRNA can be targeted. Other ribosomal regions and rRNA genes include ITS (Internal Transcribed Spacer regions) 28S and 5.8S rRNA.[7,33] The 16S rRNA gene contains conserved sequences and several hypervariable regions, which can be used to infer the taxonomic composition of the community.[34] The hypervariable regions are targeted by primers designed on the conserved regions and are amplified by PCR. As human DNA does not contain 16S rRNAs, sequencing of human DNA contained in the sample is avoided. After PCR, the amplicons are sequenced and clustered according to their sequence similarities. Sequences that have more than 97% of identity are frequently assumed to belong to the same or closely related organisms and are therefore clustered together, in so-called operational taxonomic units (OTUs)[35] which are the basis of the quantitative and taxonomic characterization (Figure 1). The 16S rRNA sequencing method is cost-effective, thanks to the relatively small size of the gene and the short amplicon size. However, this method does not provide any functional information, it has a limited taxonomic resolution compared to metagenomics,[36,37] and as there is no equivalent 16S or 18S rRNA universal genes in viruses/phages, this approach cannot assess the viral component of the microbiome.

In contrast to ribosomal profiling, shotgun metagenomics does not target specific DNA regions, but instead total community DNA is sequenced directly. This offers several advantages, including the taxonomic assessment of the microbiome (eukaryotic, prokaryotic and viral/phage) at species and even strain-level resolution,[38] and permits the investigation of the functional potential (ie which genes and pathways are present) of the microbiome. However, this approach is more expensive, the downstream analysis is more challenging, and any contaminating human DNA in the sample is also sequenced in conjugation with the microbial DNA.

For both metagenomic and ribosomal community profiling, the computational analysis can be characterized as primary and secondary (Figure 1). The primary analysis for ribosomal community profiling, generally involves OTU clustering and taxonomic assignment, whereas for shotgun metagenomics, the first steps are genome reconstruction

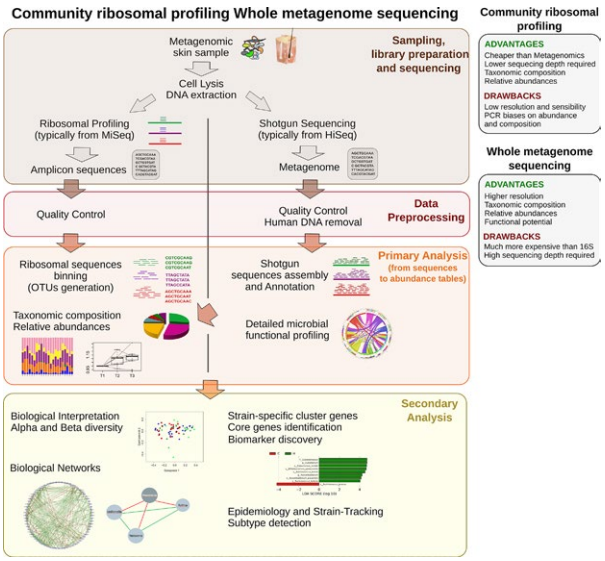

FIGURE 1 Overview on complementary approaches for studying the skin microbiome

and taxonomic and functional potential profiling. The secondary analysis includes tasks such as biomarker detection, ecological inference and statistical postprocessing.

4 | SAMPLE COLLECTION AND DNA EXTRACTION PROCEDURES

Sample collection and DNA extraction are particularly relevant in studies targeting the microbiome of the skin, because of specific technical challenges related to the low microbial biomass and high human DNA contamination.

Skin sampling can be performed using non-invasive swabs,[17,39–42] biopsies[43,44] or scrapes.[42,45] Swabs are moistened with a buffer solution and are firmly rubbed back and forth for approximately 30 seconds over the skin area,[46] and the procedure can be repeated twice to maximize the amount of collected biomass. Invasive techniques requiring medical expertise include punch biopsies, where analysis of deeper layers of the dermis is required, and skin scrapes.[47] Because each method will result in different profiles due to slightly different but overlapping environments being sampled, it is extremely important that the sampling procedure is standardized within a study.

Although the sampling method can influence the amount of collected microbial biomass,[47] generally a low amount of DNA is recovered. Importantly, the amount of extracted DNA is sometimes insufficient to apply enrichment methods (eg QIAamp DNA Microbiome Kit (Qiagen GmbH, Hilden, Germany), NEBNext® (New England Bio- Labs, Inc., Ipswich, MA) Microbiome DNA Enrichment Kit, MolYsis™ (MolZym GmbH & Co. KG, Bremen, Germany), LOOXSTER® (Analytik Jena AG, Germany) Enrichment Kit) to remove contaminating human genetic material that have been instead extensively validated for other human microbiome samples such as saliva and blood.[48–50] Recently, some microbial enrichment methods specifically for skin samples have been proposed[48,51] but are not yet commonly adopted. Besides the problem posed by the human DNA, also external bacterial contamination has to be considered before defining a microbial-disease causation.[52,53]

The choice of the DNA extraction method is similarly important. Frequently, commercially available kits are used with popular choices being the Qiagen DNA Extraction Kit (Qiagen GmbH, Hilden, Germany),[12,47] the MO BIO PowerSoil DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA)[5,17,54] or the MO BIO Ultraclean Microbial DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA).[55] Extraction methods typically combine chemical (sometimes with additional enzymatic treatment) with some form of physical disruption such as heat and/or mechanical bead-beating, to lyse cells and free their DNA (Table S1). As the cellular walls of some microbes are easier to lyse than others, this can result in biases in the sample composition.[56] If the lysis is too gentle, the more recalcitrant microbes will be under-represented, on the other hand if too excessive, it can lead to DNA fragmentation of the microbes more amenable to lysis.

Excess DNA fragmentation can increase the DNA loss during library preparation and compromise the sequencing read length, reducing the quality of the reconstructed genomes. Therefore, a compromise inevitably needs to be made and caution taken when comparing studies that employ different extraction techniques.[53]

5 | PREPARATION AND SEQUENCING AMPLICON-BASED ASSAYS

When considering a ribosomal community profile study, the investigator needs to choose which hypervariable region of the 16S rRNA gene (or the 18S rRNA for micro-eukaryotic organisms) to target and which primers to use. Both variables strongly impact on what members of the microbial community are accurately sampled.[41,56–59] Full-length 16S rRNA[60] or subset of the nine specific regions (from V1 to V9) have been targeted for skin studies: V1-V3,[41,43,57,58,60,61] V3-V5,[41,57,58,61] V6-V9,[57,58,61] V2[62] and V4.[62–64] Some regions provide more accurate taxonomic classification than others,[65] but no region is able to distinguish all bacteria and archaea.

As archaea could constitute up to 4% of the skin microbiome,[66] an ideal primer should be able to amplify both bacterial and archaeal sequences and degenerated primers are frequently used to increase the number detectable organisms.[57,67,68] For the primer set 515F/R806, targeting V4 region, a well-established Illumina protocol is available[69,70] with a recent improved version.[71] However, a recent study [163] revealed that the V4 region poorly amplifies the most common skin-associated bacteria and that the V1-V3 regions targeted with primer set 27F/534R are a more suitable choice. Nevertheless no primers are truly universal and primer choice therefore inevitably leads to some preferential PCR amplification, biasing the taxonomic interpretation.[56,57] Ultimately, the decision of which hypervariable region to target and which primer to use might be driven by the set of microbes the researcher is most interested in.

The length of the amplified fragments (amplicons) depends on the experimental goal. Although longer amplicons provide higher accuracy in the taxonomical identification,[35,37,72] there is an upper limit to the length achievable through PCR amplification[73] and read length is limited by the sequencing technology chosen. Popular choices for ribosomal profiling studies have included IonTorrent PGM, Roche 454 (now discontinued) and Illumina MiSeq, but the Illumina platforms have become the most popular choice.[74] Although ribosomal community profiling can generate chimeric amplicons[75] up to 8% of the total sequenced reads[76] and its taxonomic resolution is rather low, it is still widely used as it is a very efficient and cost-effective approach.

6 | SHOTGUN METAGENOMIC SEQUENCING

Shotgun metagenomics has the potential to reconstruct the genomes of many of the skin's resident microbes including bacteria, viruses and micro-eukaryotes. For common skin organisms, reconstructing the genome is crucial to elucidate strain- and sample-specific variations, whereas for more elusive microorganisms this approach can provide the first insight into their characteristics, taxonomy and functional potential.

As the amount of nucleic material recovered from skin samples is small, when constructing a metagenomic library, it is usual for the

community DNA to be amplified. This may be performed prior to or as part of library preparation.[77,78] Several amplification methods are available,[79] such as multiple displacement amplification[80] or limited-cycle PCR which is used in the popular Illumina NexteraXT Kit (Illumina®, San Diego,CA).[81] Regarding the choice of the sequencing platform, Illumina has become the *de facto* standard for metagenomic sequencing mainly because of its high throughput.[82] Metagenomic libraries are usually multiplexed on the Illumina platform.[83–85] Because the number and the size of the organisms in a microbiome are not known *a priori*, the sequencing throughput is usually set based on previous experience with similar experiments and the available budget. High depths provide additional information to reconstruct the low-abundance organisms in the sample and improve the detection of base-calling errors. For skin samples, there is the additional issue of the presence of human DNA whose fraction is hard to be estimated prior to sequencing, unless qPCR is used to preliminary assess human contamination.[86] Altogether, from our experience and published works, the recommended sequencing throughput of a skin metagenome is ~5 Gb which is in line with that of more validated gut metagenomes, because the relatively low species diversity of the skin is counterbalanced by the more dramatic presence of human DNA. To accurately test and validate each sequencing library, it is possible to multiplex multiple samples and sequence them on a low-throughput platform, for example Illumina MiSeq, before committing the metagenomes to a high-throughput run (eg Illumina HiSeq).

7 | PRIMARY COMPUTATIONAL ANALYSIS

A crucial step of both ribosomal community profiling and shotgun metagenomics is the computational analysis.[87] The high number of sequencing reads produced (up to many millions of reads per sample), the sequencing noise, and the incompleteness of available reference genomes, means analysis can be challenging. In the primary computational analysis, the focus is on processing the sequencing information to obtain manageable numerical and annotated data (eg which microbes are present and at what abundances in each sample). This phase is followed by a downstream statistical and cross-sample comparative analysis summarized in the next section.

After the read preprocessing performed internally by the sequencing platform, the first computational step for both 16S and metagenomic studies is the quality control of the sequencing output. This includes removing or trimming low-quality reads, discarding partial leftover adapter sequences and collapsing identical reads that are likely artifacts.[88,89]

In ribosomal community profiling, similar sequences are clustered to form OTUs as mentioned above, and a taxonomic label is assigned to each OTU by exploiting the sequenced hypervariable regions. Typically, for this approach, the taxonomic resolution is limited to the genus level, but it is not rare that many OTUs in a sample do not have assigned taxonomy and they are thus thought to represent organisms still uncharacterized in the databases.

As a representative example of a 16S rRNA skin study, Grice *et al*.[23] sampled 20 skin sites from 10 healthy individuals detecting 19 bacterial phyla (with 51% of the OTUs assigned to Actinobacteria). The volar forearm was found to be the site with the greatest number of median OTUs,[44] while the retroauricular crease showed the lowest OTU richness (only 15 OTUs). For ribosomal community profiling, there are computational pipelines which can perform a complete analysis from initial reads to taxonomic assignment such as QIIME[90] and mothur[91] with documented advantages and drawbacks.[92–94] For shotgun metagenomic studies, the standard quality control procedure includes the removal of human DNA reads, as on the skin they can represent more than 90% of the reads.[61] Human DNA can be removed from the metagenomic sample using DeconSeq,[95] CS-SCORE[96] or by mapping the reads against the human genome BWA[97] or Bowtie2.[98]

Once the preprocessing is completed, the metagenomic reads can be assembled with the goal of reconstructing the genomes of the organisms in the sample. Reference-based assembly methods align the sample's reads against a known (draft) genome stored in a database, while de-novo assembly is applied when no reference genome is available.[99] For a review of different assemblers and their respective performance, the reader can refer to several studies.[100–102] Assembly based methods are however typically very computationally and time demanding.[103] Taxonomy characterization of metagenomes can also be based on the assembly free classification of reads,[104] with several methods focusing on mapping hits,[78,105] or on using probabilistic models,[106] machine learning[107] and tetra-nucleotide frequencies.[108] A recent alternative approach infers the taxonomic classification analysing the discrete distribution of the total oligonucleotide composition.[104] Other state-of-the-art taxonomic profiling tools for shotgun metagenomics include those based on markers,[109,110] both universal and species specific,[78,111] with extension to strain-level profiling.[112]

A comprehensive metagenomic analysis should not only address the composition of the community, but also its functional potential. Functional profiling aims to catalogue genes, generally through homology-based methods such as BLAST[113] or HMMER,[114] by mapping the gene sequences against databases of functional genes.[115] The analysis of the community functional potential infers the function of genes, reveals the presence or absence of biological pathways and provides insights into the evolution and the survival strategies of the community microorganisms.[116] Pathways analysis relies on annotated database such as KEGG,[117] PFAM, UNIREF,[118] MetaCyc[119] or eggNOG[120] which integrate genomic, chemical and systemic functional information. Several computational pipelines for functional analysis are currently available.[121–123]

8 | STATISTICAL AND COMPARATIVE STUDIES

The richness and the diversity of microbial communities have historically been described using macroecology variables such as alpha- and beta-diversity defined as "the community's richness in species" and "the extent of differentiation of communities along habitat gradients," respectively.[124] Applied to microbial ecology, the alpha-diversity describes the local community richness or within-community diversity,

while beta-diversity describes the between-community diversity. These two estimators provide insightful information about the structure of microbial communities.

Alpha-diversity can be calculated on the OTUs possibly integrating the information about their phylogenetic relation.[125,126] It can be quantified by computing the number of observed species and evenness, Shannon index,[127] Chao1 index[46] and phylogenetic diversity whole tree,[128] among others. On the other hand, beta-diversity is generally computed using estimators such as Bray–Curtis dissimilarity[129] or the UniFrac distance.[130] A global view of the beta-diversity structure of a set of microbiomes is frequently obtained using principal coordinate analysis (PCoA) by embedding the distances in low-dimensional spaces where they can be visualized with the associated metadata. Beta- and alpha-diversity values can also be the input of advanced statistical approaches including network analysis, clustering and biomarker discovery as summarized in Table 1.

9 | CASE STUDY: THE SKIN MICROBIOME OF THE ELBOW AS SEEN BY SHOTGUN METAGENOMICS

To illustrate the full workflow from sample collection to computational analysis, we provide a step-by-step guide to the analysis of

TABLE 1 Tools grouped according to their main functionality for 16S, 18S and ITS data analysis

Category	Name	Description	Reference
Complete Pipelines	QIIME	Microbiome analysis from raw DNA sequencing data	[90]
	Mothur	Expandable software for microbial ecology analysis	[91]
Quality Control—Preprocessing	FASTQC	Quality control for high-throughput sequence data	[134]
	BayesHammer	Bayesian clustering for read error correction	[135]
	PEAR	Illumina paired-end read merger	[136]
	PANDAseq	Paired-end assembler for Illumina sequences	[137]
	Sickle	Windowed adaptive trimming tool for FASTQ files	[138]
	UCHIME	Algorithm for detecting chimeric sequences	[139]
	ChimaeraSlayer	Chimera detection for Sanger and 454-FLX sequences	[58]
	SolexaQA	Statistics and quality control for Illumina reads	[140]
	StreamingTrim	Trimming of amplicon library data	[141]
Databases of annotated ribosomal sequences	SILVA	Quality checked and aligned rRNA sequence data	[142]
	GreenGenes	Extensive 16S rRNA gene database	[143]
	RDP	Bacterial, archaeal and fungal rRNA sequences	[144]
OTUs Generation	UCLUST/USEARCH	Algorithms for local and global sequence	[145]
	UPARSE	Tool for highly accurate OTU sequences	[146]
	CROP	Clustering 16S rRNA for OTU Prediction	[147]
	SLP	Single-linkage preclustering	[148]
	CD-HIT	Clustering and comparison of nucleotide sequences	[149]
OTUs Classification	RDP Classifier	Classification of bacterial 16S rRNA sequences	[63]
	RTAX	Taxonomic classification of 16S sequences	[150]
	Simrank	Comparison between database strings and query strings	[151]
Statistical Postprocessing	Bioconductor	Analysis of high-throughput genomic data	[152]
	phyloseq	R package for interactive analysis of microbiome data	[153]
	SourceTracker	Contaminants/source identification and quantification	[154]
	LEfSe	Algorithm for high-dimensional biomarker discovery	[155]
	STAMP	Comparative analysis of taxonomic/functional profiles	[156]
	Metastats	Analysis of differentially abundant features	[157]
Other analysis	Oligotyping	Identification of information-rich nucleotide positions	[158]
	GraPhlAn	Compact visualizations of microbial (meta)genomes	[159]
	Cytoscape	Platform for networks and attribute data visualization	[160]
	MEGAN	General metagenomic data analysis	[105]
	UniFrac	Phylogeny-based microbial communities comparison	[161]
	PICRUSt	Inferring community metagenomic potential	[162]

TABLE 2 Summary of the reconstructed contigs and their annotation for the four analysed shotgun metagenomic samples. For each human elbow sample, we report basic statistics on the genome assembly as obtained by MEGAHIT and genome annotation as provided by Prokka

	Sample 1	Sample 2	Sample 3	Sample 4
MEGAHIT results				
SRA ID	SRR1569207	SRR1571292	SRR1571295	SRR1571296
# Contigs	6710	44 185	16 218	15 643
# Contigs >1 kb	987	6239	1628	3856
N50	3587	2032	1687	1934
Avg contig length (bp)	2779.3	1956.7	1808.7	1885.7
Total length (bp)	~2.7M	~12.2M	~2.9M	~7.3M
PROKKA results				
# tRNA	84	360	127	138
# rRNA	14	41	14	14
# CDS	5073	34 212	10 783	15 381
# CDS without known function	2700	21 827	7423	10 630

four shotgun metagenomic skin samples. Samples were collected by swabbing the external elbow area and sequenced on the Illumina HiSeq-2000 platform.[78] The sequences are available for download (BioProject n.260277, accession number PRJNA260277) and as these are already been preprocessed samples, no quality control or human DNA removal is required.

Our proposed analysis includes metagenomic assembly using MEGAHIT,[131] genome annotation with Prokka,[132] taxonomic profiling with MetaPhlAn2[78] and functional profiling with HUMAnN2.[122] All tools used in the analysis are free and open source, and a tutorial practically detailing the performed pipeline is available in Supporting information.

MEGAHIT[131] successfully reconstructed up to 6000 genome fragments (contigs) longer than 1000 nucleotides. These contigs can be the basis of further genetic analysis such as the taxonomical characterization of the most abundant organisms (by mapping the obtained contigs against known genomes) or genome annotation. Basic statistics about the reconstructed contigs and their annotation results are reported in Table 2.

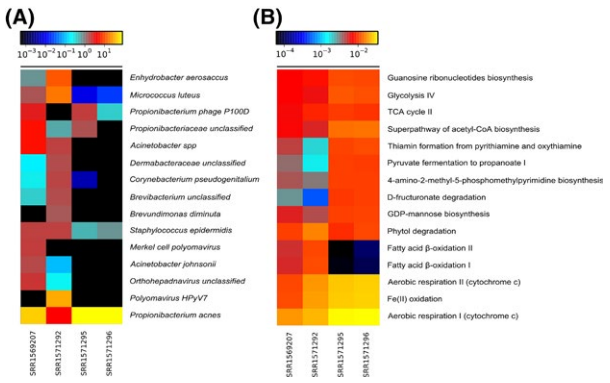

FIGURE 2 Shotgun sequencing and analysis of four human elbow samples. (A) Heatmap of taxonomic abundances generated with MetaPhlAn2. (B) Heatmap of functionalities encoded in the genome generated with HUMAnN2. We report here only the 15 most abundant species and functions (full tables available as Tables S3-S6)

The quantitative taxonomic composition was assessed with MetaPhlAn2,[78] and the output files of each sample were merged to generate a table of species abundances for all samples (Figure 2A). *Propionibacterium acnes*, *Micrococcus luteus* and *Straphylococcus epidermidis* are the most abundant species. *Merkel cell polyomavirus* and *Polyomavirus HPyV7*, recognized as common constituent of the human skin microbiome,[133] are also present.

HUMAnN2[122] was used to perform functional profiling. The most abundant metabolic pathways (Figure 2B) include guanosine and acetyl-CoA biosynthesis, metabolism-related processes (such as glycolysis, TCA cycle and aerobic respiration I and II), iron (II) oxidation and aerobic phytol degradation, but these are only a subset of the many functions encoded by the microorganisms in the metagenomic samples.

10 | CONCLUSIONS AND PERSPECTIVES

The recent advances in high-throughput methodologies and the continuous reduction in the cost of sequencing have opened new frontiers in microbial ecology by dramatically broadening the discovery potential of skin-related research studies. In this context, the metagenomic approach provides unprecedented means to understand and describe in a comprehensive way the complex interaction networks underlying the host-skin microbiome symbiosis and dysbiosis. We described here the basic experimental and analysis procedures to perform skin microbiome studies summarizing the current state of the art of a subfield of dermatology that we foresee will be of crucial relevance in the next few years.

ACKNOWLEDGEMENTS

This work has been supported by "Istituto G.B. Mattei" and Terme di Comano to N.S. and by Fondazione Caritro to A.T. The work as also been partially supported by the European Union's Seventh

Framework Programme (FP7/2007-2013) under REA grant agreement no PCIG13-GA-2013-618833, start-up funds from the Centre for Integrative Biology (University of Trento), MIUR "Futuro in Ricerca" E68C13000500001, and by from Fondazione Caritro to N.S.

CONFLICT OF INTERESTS

The authors have declared no conflicting interests.

AUTHOR CONTRIBUTION

P.F., A.T. and N.S. designed the review and drafted the manuscript, and S.F., M.C. and G.G. provided critical feedback and contributed to specific sections. All authors contributed and approved the final text.

REFERENCES

[1] R. Sender, S. Fuchs, R. Milo, *PLoS Biol* **2016**, *14*, e1002533.
[2] J. J. Leyden, K. J. McGinley, K. M. Nordstrom, G. F. Webster, *Dermatology* **1987**, *88*, 65.
[3] Z. Gao, C. H. Tseng, Z. Pei, M. J. Blaser *Proc. Natl Acad. Sci. USA* **2007**, *104*, 2927.
[4] E. K. Costello, C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, R. Knight, *Science* **2009**, *326*, 1694.
[5] The Human Microbiome Project Consortium, *Nature* **2012**, *486*, 207.
[6] H. H. Kong, J. A. Segre, *J. Invest. Dermatol.* **2012**, *132*, 933.
[7] E. A. Grice, J. A. Segre, *Microbiology* **2011**, *9*, 244.
[8] E. A. Grice, *Semin. Cutan. Med. Surg.* **2014**, *33*, 98.
[9] A. L. Cogen, K. Yamasaki, K. M. Sanchez, R. A. Dorschner, Y. Lai, D. T. MacLeod, J. W. Torpey, M. Otto, V. Nizet, J. E. Kim, R. L. Gallo *J. Invest. Dermatol.* **2010**, *130*, 192.
[10] Y. Lai, A. L. Cogen, K. A. Radek, H. J. Park, D. T. Macleod, A. Leichtle, A. F. Ryan, A. Di Nardo, R. L. Gallo *J. Invest. Dermatol.* **2010**, *130*, 2211.
[11] T. Iwase, Y. Uehara, H. Shinji, A. Tajima, H. Seo, K. Takada, T. Agata, Y. Mizunoe, *Nature* **2010**, *465*, 346.
[12] K. A. Capone, S. E. Dowd, G. N. Stamatas, J. Nikolovski, *J. Invest. Dermatol.* **2011**, *131*, 2026.
[13] S. Naik, N. Bouladoux, J. L. Linehan, S. J. Han, O. J. Harrison, C. Wilhelm, S. Conlan, S. Himmelfarb, A. L. Byrd, C. Deming, M. Quinones, J. M. Brenchley, H. H. Kong, R. Tussiwand, K. M. Murphy, M. Merad, J. A. Segre, Y. Belkaid, *Nature* **2015**, *520*, 104.
[14] Y. Lai, A. Di Nardo, T. Nakatsuji, A. Leichtle, Y. Yang, A. L. Cogen, Z. R. Wu, L. V. Hooper, R. R. Schmidt, von Aulock S., K. A. Radek, C. M. Huang, A. F. Ryan, R. L. Gallo, *Nat. Med.* **2009**, *15*, 1377.
[15] R. L. Gallo, T. Nakatsuji, *J. Invest. Dermatol.* **2011**, *131*, 1974.
[16] N. O. Verhulst, W. Takken, M. Dicke, G. Schraa, R. C. Smallegange, *FEMS Microbiol. Ecol.* **2010**, *74*, 1.
[17] N. Fierer, C. L. Lauber, N. Zhou, D. McDonald, E. K. Costello, R. Knight, *Proc. Natl Acad. Sci. USA* **2010**, *107*, 6477.
[18] S. Ying, D. N. Zeng, L. Chi, Y. Tan, C. Galzote, C. Cardona, S. Lax, J. Gilbert, Z. X. Quan, *PLoS One* **2015**, *10*, e0141842.
[19] N. Fierer, M. Hamady, C. L. Lauber, R. Knight, *Proc. Natl Acad. Sci. USA* **2008**, *105*, 17994.
[20] A. SanMiguel, E. A. Grice, *Cell. Mol. Life Sci.* **2015**, *72*, 1499.
[21] R. Knight, J. Jansson, D. Field, N. Fierer, N. Desai, J. A. Fuhrman, P. Hugenholtz, der van Lelie D., F. Meyer, R. Stevens, M. J. Bailey, J. I. Gordon, G. A. Kowalchuk, J. A. Gilbert, *Nat. Biotechnol.* **2012**, *30*, 513.
[22] T. L. Tickle, N. Segata, L. Waldron, U. Weingart, C. Huttenhower, *ISME J.* **2013**, *7*, 2330.
[23] E. A. Grice, H. H. Kong, S. Conlan, C. B. Deming, J. Davis, A. C. Young; NISC Comparative Sequencing Program, G. G. Bouffard, R. W. Blakesley, P. R. Murray, E. D. Green, M. L. Turner, J. A. Segre, *Science* **2009**, *324*, 1190.
[24] G. E. Flores, J. G. Caporaso, J. B. Henley, J. R. Rideout, D. Domogala, J. Chase, J. W. Leff, Y. Vázquez-Baeza, A. Gonzalez, R. Knight, R. R. Dunn, N. Fierer, *Genome Biol.* **2014**, *15*, 531.
[25] G. Wu, H. Zhao, C. Li, M. P. Rajapakse, W. C. Wong, J. Xu, C. W. Saunders, N. L. Reeder, R. A. Reilman, A. Scheynius, S. Sun, B. R. Billmyre, W. Li, A. F. Averette, P. Mieczkowski, J. Heitman, B. Theelen, der Schrö M. S., P. F. De Sessions, G. Butler, S. Maurer-Stroh, T. Boekhout, N. Nagarajan, T. L. . Dawson Jr, *PLoS Genet.* **2015**, *11*, e1005614.
[26] J. Liu, R. Yan, Q. Zhong, S. Ngo, N. J. Bangayan, L. Nguyen, T. Lui, M. Liu, M. C. Erfe, N. Craft, S. Tomida, H. Li, *ISME J.* **2015**, *9*, 2078.
[27] L. A. Ogilvie, B. V. Jones, *Front. Microbiol.* **2015**, *6*, 918.
[28] G. D. Hannigan, J. S. Meisel, A. S. Tyldsley, Q. Zheng, B. P. Hodkinson, A. J. SanMiguel, S. Minot, F. D. Bushman, E. A. Grice, *mBio* **2015**, *6*, e01578.
[29] K. Cadwell, *Immunity* **2015**, *42*, 805.
[30] E. A. Grice, *Genome Res.* **2015**, *25*, 1514.
[31] E. J. Stewart, *J. Bacteriol.* **2012**, *194*, 4151.
[32] H. Al-Awadhi, N. Dashti, M. Khanafer, D. Al-Mailem, N. Ali, S. Radwan, *Springerplus* **2013**, *2*, 369.
[33] K. Findley, J. Oh, J. Yang, S. Conlan, C. Deming, J. A. Meyer, D. Schoenfeld, E. Nomicos, M. Park; NIH Intramural Sequencing Center Comparative Sequencing Program, H. H. Kong, J. A. Segre, *Nature* **2013**, *498*, 367.
[34] C. R. Woese, O. Kandler, M. L. Wheelis, *Proc. Natl Acad. Sci. USA* **1990**, *87*, 4576.
[35] M. Hamady, R. Knight, *Genome Res.* **2009**, *19*, 1141.
[36] R. Poretsky, L. M. Rodriguez-R, C. Luo, D. Tsementzi, K. T. Konstantinidis, *PLoS One* **2014**, *9*, e93827.
[37] M. J. Claesson, O. O'Sullivan, Q. Wang, J. Nikkilä, J. R. Marchesi, H. Smidt, de Vos W. M., R. P. Ross, P. W. O'Toole, *PLoS One* **2009**, *4*, e6669.
[38] T. J. Sharpton, *Front. Plant Sci.* **2014**, *5*, 209.
[39] H. H. Kong, J. Oh, C. Deming, S. Conlan, E. A. Grice, M. A. Beatson, E. Nomicos, E. C. Polley, H. D. Komarow; NISC Comparative Sequence Program, P. R. Murray, M. L. Turner, J. A. Segre, *Genome Res.* **2012**, *22*, 850.
[40] M. J. Blaser, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, I. Estrada, Z. Gao, J. C. Clemente, E. K. Costello, R. Knight, *ISME J.* **2013**, *7*, 85.
[41] S. M. Huse, Y. Ye, Y. Zhou, A. A. Fodor, *PLoS One* **2012**, *7*, e34242.
[42] J. Oh, S. Conlan, E. C. Polley, J. A. Segre, H. H. Kong, *Genome Med.* **2012**, *4*, 77.
[43] T. Nakatsuji, H. I. Chiang, S. B. Jiang, H. Nagarajan, K. Zengler, R. L. Gallo, *Nat. Commun.* **2013**, *4*, 1431.
[44] A. Fahlén, L. Engstrand, B. S. Baker, A. Powles, L. Fry, *Arch. Dermatol. Res.* **2012**, *304*, 15.
[45] T. Staudinger, A. Pipal, B. Redl, *J. Appl. Microbiol.* **2011**, *110*, 1381.
[46] A. Chao, *Scand. J. Stat.*, **1948**, *11*, 265.
[47] E. A. Grice, H. H. Kong, G. Renaud, A. C. Young, NISC Comparative Sequencing Program, G. G. Bouffard, R. W. Blakesley, T. G. Wolfsberg, M. L. Turner, J. A. Segre, *Genome Res.* **2008**, *18*, 1043.
[48] G. R. Feehery, E. Yigit, S. O. Oyola, B. W. Langhorst, V. T. Schmidt, F. J. Stewart, E. T. Dimalanta, L. A. Amaral-Zettler, T. Davis, M. A. Quail, S. Pradhan, *PLoS One* **2013**, *8*, e76096.
[49] H. P. Horz, S. Scheer, F. Huenger, M. E. Vianna, G. Conrads, *J. Microbiol. Methods* **2008**, *72*, 98.
[50] L. Zhou, A. J. Pollard, *BMC Infect. Dis.* **2012**, *12*, 164.
[51] M. Garcia-Garcerà, K. Garcia-Etxebarria, M. Coscollà, A. Latorre, F. Calafell, *PLoS One* **2013**, *8*, e74914.
[52] S. Mollerup, J. Friis-Nielsen, L. Vinner, T. A. Hansen, S. R. Richter, H. Fridholm, J. A. R. Herrera, O. Lund, S. Brunak, J. M. G. Izarzugaza,

T. Mourier, L. P. Nielsen, A. J. Hansen, *J. Clin. Microbiol.* **2016**, *54*, 980.
[53] S. J. Salter, M. J. Cox, E. M. Turek, S. T. Calus, W. O. Cookson, M. F. Moffatt, P. Turner, J. Parkhill, N. J. Loman, A. W. Walker, *BMC Biol.* **2014**, *12*, 87.
[54] C. L. Lauber, N. Zhou, J. I. Gordon, R. Knight, N. Fierer, *FEMS Microbiol. Lett.* **2010**, *307*, 80.
[55] P. L. J. M. Zeeuwen, J. Boekhorst, van den Bogaard E. H., de Koning H. D., de van Kerkhof P. M. C., D. M. Saulnier, van Swam I. I., van Hijum S. A. F. T., M. Kleerebezem, J. Schalkwijk, H. M. Timmerman, *Genome Biol.* **2012**, *13*, R101.
[56] J. Kuczynski, C. L. Lauber, W. A. Walters, L. W. Parfrey, J. C. Clemente, D. Gevers, R. Knight, *Nat. Rev. Genet.* **2012**, *13*, 47.
[57] Jumpstart Consortium Human Microbiome Project Data Generation Working Group, *PLoS One* **2012**, *7*, e39315.
[58] B. J. Haas, D. Gevers, A. M. Earl, M. Feldgarden, D. V. Ward, G. Giannoukos, D. Ciulla, D. Tabbaa, S. K. Highlander, E. Sodergren, B. Methé, T. Z DeSantis; Human Microbiome Consortium, J. F. Petrosino, R. Knight, B. W. Birren, *Genome Res.* **2011**, *21*, 494.
[59] E. A. Eloe-Fadrosh, N. N. Ivanova, T. Woyke, N. C. Kyrpides, *Nat. Microbiol.* **2016**, *1*, 15032.
[60] J. Oh, A. F. Freeman, N. I. S. C. Comparative Sequencing Program, M. Park, R. Sokolic, F. Candotti, S. M. Holland, J. A. Segre, H. H. Kong, *Genome Res.* **2013**, *23*, 2103.
[61] The Human Microbiome Project Consortium, *Nature* **2012**, *486*, 215.
[62] Z. Liu, T. Z. De Santis, G. L. Andersen, R. Knight, *Nucleic Acids Res.* **2008**, *36*, e120.
[63] Q. Wang, G. M. Garrity, J. M. Tiedje, J. R. Cole, *Appl. Environ. Microbiol.* **2007**, *73*, 5261.
[64] S. Lax, D. P. Smith, J. Hampton-Marcell, S. M. Owens, K. M. Handley, N. M. Scott, S. M. Gibbons, P. Larsen, B. D. Shogan, S. Weiss, J. L. Metcalf, L. K. Ursell, Y. Vázquez-Baeza, W. Van Treuren, N. A. Hasan, M. K. Gibson, R. Colwell, G. Dantas, R. Knight, J. A. Gilbert, *Science* **2014**, *345*, 1048.
[65] J. S. Meisel, G. D. Hannigan, A. S. Tyldsley, A. J. SanMiguel, B. P. Hodkinson, Q. Zheng, E. A. Grice, *J. Invest. Dermatol.* **2016**, *136*, 947.
[66] A. J. Probst, A. K. Auerbach, C. Moissl-Eichinger, *PLoS One* **2013**, *8*, e65388.
[67] G. C. Baker, J. J. Smith, D. A. Cowan, *J. Microbiol. Methods* **2003**, *55*, 541.
[68] Y. Wang, P. Y. Qian, *PLoS One* **2009**, *4*, e7401.
[69] J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, R. Knight, *Proc. Natl Acad. Sci. USA* **2011**, *108*(Suppl 1), 4516.
[70] J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith, R. Knight, *ISME J.* **2012**, *6*, 1621.
[71] E. R. H. William Walters, D. Berg-Lyons, G. Ackermann, G. Humphrey, A. Parada, J. A Gilbert, J. K. Jansson, J. Gregory Caporaso, J. A. Fuhrman, A. Apprill, R. Knight. *mSystems* **2015**, *1*, e00009.
[72] W. J. Sul, J. R. Cole, C. Jesus Eda, Q. Wang, R. J. Farris, J. A. Fish, J. M. Tiedje, *Proc. Natl Acad. Sci. USA* **2011**, *108*, 14637.
[73] L. Mamanova, A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, J. Shendure, D. J. Turner, *Nat. Methods* **2010**, *7*, 111.
[74] S. J. Salipante, T. Kawashima, C. Rosenthal, D. R. Hoogestraat, L. A. Cummings, D. J. Sengupta, T. T. Harkins, B. T. Cookson, N. G. Hoffman, *Appl. Environ. Microbiol.* **2014**, *80*, 7583.
[75] D. J. Lahr, L. A. Katz, *Biotechniques* **2009**, *47*, 857.
[76] P. D. Schloss, D. Gevers, S. L. Westcott, *PLoS One* **2011**, *6*, e27310.
[77] J. Oh, A. L. Byrd, C. Deming, S. Conlan; NISC Comparative Sequencing Program, H. H. Kong, J. A. Segre, *Nature* **2014**, *514*, 59.
[78] D. T. Truong, E. A. Franzosa, T. L. Tickle, M. Scholz, G. Weingart, E. Pasolli, A. Tett, C. Huttenhower, N. Segata, *Nat. Methods* **2015**, *12*, 902.
[79] M. Fakruddin, K. S. Bin Mannan, A. Chowdhury, R. M. Mazumdar, M. N. Hossain, S. Islam, M. A. Chowdhury, *J. Pharm. Bioallied Sci.* **2013**, *5*, 245.
[80] C. Spits, C. Le Caignec, M. De Rycke, L. Van Haute, A. Van Steirteghem, I. Liebaers, K. Sermon, *Nat. Protoc.* **2006**, *1*, 1965.
[81] S. Lamble, E. Batty, M. Attar, D. Buck, R. Bowden, G. Lunter, D. Crook, B. El-Fahmawi, P. Piazza, *BMC Biotechnol.* **2013**, *13*, 104.
[82] M. A. Quail, M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, Y. Gu, *BMC Genom.* **2012**, *13*, 341.
[83] K. H. Wong, Y. Jin, Z. Moqtaderi. *Curr. Protoc. Mol. Biol.* **2013**, *Chapter 7*, Unit 7.11.
[84] P. D. Hebert, A. Cywinska, S. L. Ball, J. R. deWaard, *Proc. Biol. Sci.* **2003**, *270*, 313.
[85] P. Parameswaran, R. Jalili, L. Tao, S. Shokralla, B. Gharizadeh, M. Ronaghi, A. Z. Fire, *Nucleic Acids Res.* **2007**, *35*, e130.
[86] Z. Gao, G. I. Perez-Perez, Y. Chen, M. J. Blaser, *J. Clin. Microbiol.* **2010**, *48*, 3575.
[87] N. Segata, D. Boernigen, T. L. Tickle, X. C. Morgan, W. S. Garrett, C. Huttenhower, *Mol. Syst. Biol.* **2013**, *9*, 666.
[88] Q. Zhou, X. Su, A. Wang, J. Xu, K. Ning, *PLoS One* **2013**, *8*, e60234.
[89] R. K. Patel, M. Jain, *PLoS One* **2012**, *7*, e30619.
[90] J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, *Nat. Methods* **2010**, *7*, 335.
[91] P. D. Schloss, S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, C. F. Weber, *Appl. Environ. Microbiol.* **2009**, *75*, 7537.
[92] J. T. Erica Plummer, D. M. Bulach, S. M. Garland, S. N. Tabrizi. *J. Prot. Bioinform.* **2015**, *8*, 283.
[93] E. Kopylova, J. A. Navas-Molina, C. Mercier, Z. Z. Xu, F. Mahé, Y. He, H. W. Zhou, T. Rognes, J. Gregory Caporaso, R. Knight, *mSystems* **2016**, *1*, e00003.
[94] P. D. Schloss. *mSystems* **2016**, *1*, e00027.
[95] R. Schmieder, R. Edwards, *PLoS One* **2011**, *6*, e17288.
[96] M. M. Haque, T. Bose, A. Dutta, C. V. Reddy, S. S. Mande, *Genomics* **2015**, *106*, 116.
[97] H. Li, R. Durbin, *Bioinformatics* **2009**, *25*, 1754.
[98] B. Langmead, S. L. Salzberg, *Nat. Methods* **2012**, *9*, 357.
[99] T. Thomas, J. Gilbert, F. Meyer, *Microb. Inform. Exp.* **2012**, *2*, 3.
[100] T. Magoc, S. Pabinger, S. Canzar, X. Liu, Q. Su, D. Puiu, L. J. Tallon, S. L. Salzberg, *Bioinformatics* **2013**, *29*, 1718.
[101] W. Zhang, J. Chen, Y. Yang, Y. Tang, J. Shang, B. Shen, *PLoS One* **2011**, *6*, e17915.
[102] M. Hunt, C. Newbold, M. Berriman, T. D. Otto, *Genome Biol.* **2014**, *15*, R42.
[103] M. Comin, M. Schimd, *BMC Bioinformatics* **2014**, *15*(Suppl 9), S1.
[104] P. Meinicke, K. P. Asshauer, T. Lingner, *Bioinformatics* **2011**, *27*, 1618.
[105] D. H. Huson, A. F. Auch, J. Qi, S. C. Schuster, *Genome Res.* **2007**, *17*, 377.
[106] A. Brady, S. L. Salzberg, *Nat. Methods* **2009**, *6*, 673.
[107] N. N. Diaz, L. Krause, A. Goesmann, K. Niehaus, T. W. Nattkemper, *BMC Bioinformatics* **2009**, *10*, 56.
[108] H. Teeling, A. Meyerdierks, M. Bauer, R. Amann, F. O. Glöckner, *Environ. Microbiol.* **2004**, *6*, 938.
[109] S. Sunagawa, D. R. Mende, G. Zeller, F. Izquierdo-Carrasco, S. A. Berger, J. R. Kultima, L. P. Coelho, M. Arumugam, J. Tap, H. B. Nielsen, S. Rasmussen, S. Brunak, O. Pedersen, F. Guarner, de Vos W. M., J. Wang, J. Li, J. Doré, S. D. Ehrlich, A. Stamatakis, P. Bork, *Nat. Methods* **2013**, *10*, 1196.
[110] N. Segata, D. Börnigen, X. C. Morgan, C. Huttenhower *Nat. Commun.* **2013**, *4*, 2304.

[111] N. Segata, L. Waldron, A. Ballarini, V. Narasimhan, O. Jousson, C. Huttenhower, *Nat. Methods* **2012**, *9*, 811.
[112] M. Scholz, D. V. Ward, E. Pasolli, T. Tolio, M. Zolfo, F. Asnicar, D. T. Truong, A. Tett, A. L. Morrow, N. Segata, *Nat. Methods* **2016**, *13*, 435.
[113] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403.
[114] R. D. Finn, J. Clements, S. R. Eddy, *Nucleic Acids Res.* **2011**, *39*, W29.
[115] T. Prakash, T. D. Taylor, *Brief. Bioinform.* **2012**, *13*, 711.
[116] G. W. Tyson, J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, J. F. Banfield, *Nature* **2004**, *428*, 37.
[117] A. Masoudi-Nejad, S. Goto, T. R. Endo, M. Kanehisa, *Methods Mol. Biol.* **2007**, *406*, 437.
[118] B. E. Suzek, H. Huang, P. McGarvey, R. Mazumder, C. H. Wu, *Bioinformatics* **2007**, *23*, 1282.
[119] C. J. Krieger, P. Zhang, L. A. Mueller, A. Wang, S. Paley, M. Arnaud, J. Pick, S. Y. Rhee, P. D. Karp, *Nucleic Acids Res.* **2004**, *32*, D438.
[120] J. Huerta-Cepas, D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M. C. Walter, T. Rattei, D. R. Mende, S. Sunagawa, M. Kuhn, L. J. Jensen, von Mering C., P. Bork, *Nucleic Acids Res.* **2016**, *44*, D286.
[121] S. Nayfach, P. H. Bradley, S. K. Wyman, T. J. Laurent, A. Williams, J. A. Eisen, K. S. Pollard, T. J. Sharpton, *PLoS Comput. Biol.* **2015**, *11*, e1004573.
[122] S. Abubucker, N. Segata, J. Goll, A. M. Schubert, J. Izard, B. L. Cantarel, B. Rodriguez-Mueller, J. Zucker, M. Thiagarajan, B. Henrissat, O. White, S. T. Kelley, B. Methé, P. D. Schloss, D. Gevers, M. Mitreva, C. Huttenhower, *PLoS Comput. Biol.* **2012**, *8*, e1002358.
[123] J. R. Kultima, S. Sunagawa, J. Li, W. Chen, H. Chen, D. R. Mende, M. Arumugam, Q. Pan, B. Liu, J. Qin, J. Wang, P. Bork, *PLoS One* **2012**, *7*, e47656.
[124] R. H. Whittaker. *Taxon*. **1972**, *21*, 213.
[125] X. C. Morgan, C. Huttenhower, *PLoS Comput. Biol.* **2012**, *8*, e1002808.
[126] S. W. Kembel, J. A. Eisen, K. S. Pollard, J. L. Green, *PLoS One* **2011**, *6*, e23214.
[127] C. E. Shannon, *MD Comput*. **1997**, *14*, 306.
[128] D. P. Faith, A. M. Baker, *Evol. Bioinf. Online* **2006**, *2*, 121.
[129] J. R. Bray, J. T. Curtis. *Ecological Monographs* **1957**, *27*, 326.
[130] C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh, R. Knight, *ISME J.* **2011**, *5*, 169.
[131] D. Li, C. M. Liu, R. Luo, K. Sadakane, T. W. Lam, *Bioinformatics* **2015**, *31*, 1674.
[132] T. Seemann, *Bioinformatics* **2014**, *30*, 2068.
[133] R. M. Schowalter, D. V. Pastrana, K. A. Pumphrey, A. L. Moyer, C. B. Buck, *Cell Host Microbe* **2010**, *7*, 509.
[134] S. Andrews. **2010**, Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
[135] S. I. Nikolenko, A. I. Korobeynikov, M. A. Alekseyev, *BMC Genom.* **2013**, *14*(Suppl 1), S7.
[136] J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, *Bioinformatics* **2014**, *30*, 614.
[137] A. P. Masella, A. K. Bartram, J. M. Truszkowski, D. G. Brown, J. D. Neufeld, *BMC Bioinformatics* **2012**, *13*, 31.
[138] N. A. Joshi, J. N. F. Sickle. **2011**, Available online at: https://github.com/najoshi/sickle.
[139] R. C. Edgar, B. J. Haas, J. C. Clemente, C. Quince, R. Knight, *Bioinformatics* **2011**, *27*, 2194.
[140] M. P. Cox, D. A. Peterson, P. J. Biggs, *BMC Bioinformatics* **2010**, *11*, 485.
[141] G. Bacci, M. Bazzicalupo, A. Benedetti, A. Mengoni, *Mol. Ecol. Resour.* **2014**, *14*, 426.
[142] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, *Nucleic Acids Res.* **2013**, *41*, D590.
[143] T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, G. L. Andersen, *Appl. Environ. Microbiol.* **2006**, *72*, 5069.
[144] J. R. Cole, Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske, J. M. Tiedje, *Nucleic Acids Res.* **2014**, *42*, D633.
[145] R. C. Edgar, *Bioinformatics* **2010**, *26*, 2460.
[146] R. C. Edgar, *Nat. Methods* **2013**, *10*, 996.
[147] X. Hao, R. Jiang, T. Chen, *Bioinformatics* **2011**, *27*, 611.
[148] S. M. Huse, D. M. Welch, H. G. Morrison, M. L. Sogin, *Environ. Microbiol.* **2010**, *12*, 1889.
[149] W. Li, A. Godzik, *Bioinformatics* **2006**, *22*, 1658.
[150] D. A. Soergel, N. Dey, R. Knight, S. E. Brenner, *ISME J.* **2012**, *6*, 1440.
[151] T. Z. DeSantis, K. Keller, U. Karaoz, A. V. Alekseyenko, N. N. Singh, E. L. Brodie, Z. Pei, G. L. Andersen, N. Larsen, *BMC Ecol.* **2011**, *11*, 11.
[152] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, J. Zhang, *Genome Biol.* **2004**, *5*, R80.
[153] P. J. McMurdie, S. Holmes, *PLoS One* **2013**, *8*, e61217.
[154] D. Knights, J. Kuczynski, E. S. Charlson, J. Zaneveld, M. C. Mozer, R. G. Collman, F. D. Bushman, R. Knight, S. T. Kelley, *Nat. Methods* **2011**, *8*, 761.
[155] N. Segata, J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett, C. Huttenhower, *Genome Biol.* **2011**, *12*, R60.
[156] D. H. Parks, G. W. Tyson, P. Hugenholtz, R. G. Beiko, *Bioinformatics* **2014**, *30*, 3123.
[157] J. R. White, N. Nagarajan, M. Pop, *PLoS Comput. Biol.* **2009**, *5*, e1000352.
[158] A. M. Eren, L. Maignien, W. J. Sul, L. G. Murphy, S. L. Grim, H. G. Morrison, M. L. Sogin. *Methods Ecol. Evol.* **2013**: *4*, 1111.
[159] F. Asnicar, G. Weingart, T. L. Tickle, C. Huttenhower, N. Segata, *PeerJ* **2015**, *3*, e1029.
[160] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res.* **2003**, *13*, 2498.
[161] C. Lozupone, R. Knight, *Appl. Environ. Microbiol.* **2005**, *71*, 8228.
[162] M. G. Langille, J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. Vega Thurber, R. Knight, R. G. Beiko, C. Huttenhower, *Nat. Biotechnol.* **2013**, *31*, 814.
[163] J. S. Meisel, G. D. Hannigan, A. S. Tyldsley, A. J. SanMiguel, B. P. Hodkinson, Q. Zheng, E. A. Grice, *J. Invest. Dermatol.* **2016**, *136*, 94.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Table S1 Advantages and drawbacks of different cell disruption methods

Table S2 Summary of the parameters for the four analysed shotgun metagenomic samples

Table S3 Full table of taxonomic abundancies for the four analysed skin metagenomes as estimated by MetaPhlAn2

Table S4 Full table of gene family abundancies for the four analysed skin metagenomes as estimated by HUMAnN2

Table S5 Full table of pathway abundancies for the four analysed skin metagenomes as estimated by HUMAnN2

Table S6 Full table of pathway coverages for the four analysed skin metagenomes as estimated by HUMAnN2

Data S1 Tuitorial Form