Propionibacterium acnes – disease causing agent or common contaminant? Detection in diverse patient samples by next generation sequencing

Sarah Mollerup1*, Jens Friis-Nielsen2, Lasse Vinner1, Thomas Arn Hansen1, Stine Raith Richter1, Helena Fridholm1, Jose Alejandro Romero Herrera2, Ole Lund2, Søren Brunak2,3, Jose M. G. Izarzugaza2, Tobias Mourier1§, Lars Peter Nielsen4§, Anders Johannes Hansen1*§

1 Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Denmark.
2 Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kgs. Lyngby, Denmark.
3 Disease Systems Biology Program, Panum Institute, University of Copenhagen, Denmark.
4 Department of Autoimmunology and Biomarkers, Statens Serum Institut, Copenhagen S, and Aalborg University, Health Sciences, Denmark.
§These authors contributed equally to this work.

*Corresponding authors
Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1350 Copenhagen, Denmark, sarah.mollerup@snm.ku.dk, ajhansen@snm.ku.dk

Running title: Detection of P. acnes in diverse patient samples

Keywords: Propionibacterium acnes, next generation sequencing, clinical samples, shotgun sequencing, microbe enrichment, human skin flora, suggested pathogen, contaminant
Abstract

Propionibacterium acnes (P. acnes) is the most abundant bacterium on human skin, particularly in sebaceous areas. P. acnes is suggested to be an opportunistic pathogen involved in the development of diverse medical conditions, but is also a proven contaminant of human samples and surgical wounds. Its significance as a pathogen is consequently a matter of debate.

In the present study we investigated the presence of P. acnes DNA in 250 next generation sequencing datasets generated from 180 samples of 20 different sample types, mostly of cancerous origin. The samples were either subjected to microbial enrichment, involving nuclease treatment to reduce the amount of host nucleic acids, or shotgun-sequenced.

We detected high proportions of P. acnes in enriched samples, particularly skin derived and other tissue samples, with levels being higher in enriched compared to shotgun-sequenced samples. P. acnes reads were detected in most samples analysed, though the proportions in most shotgun-sequenced samples were low.

Our results show that P. acnes can be detected in practically all sample types when employing molecular methods such as next generation sequencing. The possibility of contamination from the patient or other sources, including laboratory reagents or environment, should therefore always be considered carefully when P. acnes is detected in clinical samples. We advocate that detection of P. acnes is always accompanied by experiments validating the association between this bacterium and any clinical condition.
Introduction

Propionibacterium acnes (P. acnes) is a facultative anaerobic gram positive bacterium present on human skin as part of the normal flora, as well as in the oral cavity, large intestine, conjunctiva, and external ear canal (1). P. acnes is the most prevalent bacterium in sebaceous areas of the skin (2, 3), but is also abundant in dry areas (3). P. acnes predominates in the pilosebaceous follicles of the skin (4).

P. acnes is proposed to play a role in the development of acne vulgaris (5) and is considered an opportunistic pathogen causing postoperative infections (6). It has been isolated from patients with endocarditis (7), SAPHO (synovitis, acne, pustulosis, hyperostosis, osteitis) syndrome (8), and sarcoidosis (9), among other syndromes. Some studies also identify P. acnes as a contaminant of blood products, tissue cultures, and surgical wounds (10-13), and the significance of this widely abundant skin commensal is consequently debated.

The postulated role of P. acnes in various conditions is often highly speculative and based on the mere detection of the bacterium (9, 14-20). To expand the knowledge of the abundance of P. acnes, we investigated the proportions of P. acnes sequences in samples of a wide variety of cancer types, as well as blood samples. Next generation sequencing was performed on DNA extracted either after enrichment for intact microbes or directly from the samples. This study is the first large scale analysis investigating the abundance of P. acnes DNA in such diverse tissue types by next generation sequencing, and we show that P. acnes can be readily detected in literally all sample types investigated, particularly when samples are subjected to microbial enrichment.
Materials and Methods

Ethics statement

Human sample collection, handling, and analysis were performed under the ethical protocols H-2-2012-FSP2 (Regional Committee on Health Research Ethics) and case no. 1304226 (National Committee on Health Research Ethics). In accordance with National legislation (Sundhedsloven), all human samples were processed anonymously.

Patient samples

Samples collected in Denmark: Malignant melanoma biopsies and cells sorted by fluorescence activated cell sorting (FACS) from acute myeloid leukemia (AML), B-cell chronic lymphocytic leukemia (B-CLL), and chronic myelogenous leukemia (CML) were obtained from Aarhus University Hospital. T-lineage acute lymphoblastic leukemia (T-ALL) samples were obtained from either Aarhus University Hospital (FACS sorted cells; n=9) or Rigshospitalet (Copenhagen University Hospital) (bone marrow samples; n=2). B-cell precursor acute lymphoblastic leukemia (BCP-ALL) samples (bone marrow), oropharyngeal head and neck cancer and testicular cancer biopsies were obtained from Rigshospitalet. Biopsies from basal cell carcinoma and mycosis fungoides (cutaneous T-cell lymphoma) were obtained from Bispebjerg Hospital. Bladder, breast, and colon cancer biopsies, blood samples from colon cancer patients and ascites fluid from colon, ovarian, and pancreatic cancer patients were obtained from or collected in collaboration with the Danish Cancer Biobank, Herlev Hospital. Ascites fluid samples were subjected to low speed centrifugation upon reception to pellet intact host cells and cellular debris. The supernatant was used for microbial enrichment, whereas pelleted cells were used for shotgun sequencing. Cryopreserved, fully transformed B-cell lymphoma cell lines (OCI Ly3_M, OCI-Ly7_M, OCI-Ly8,
SU-DHL-4, SU-DHL-5, and U698M) and multiple myeloma cell lines (KMS-12-BM, KMS-12-PE, MOLP-2, MOLP-8, RPMI-8226, and U266) were obtained from Aalborg University Hospital. Samples collected outside Denmark: Vulva cancer biopsies were obtained from the National Institute of Oncology, Budapest, Hungary. All samples are listed in Table 1.

Laboratory facilities
All sample processing was performed under conditions designed to diminish contamination, through a unidirectional workflow. Separate laboratories were used for reagent preparation, sample extraction and sequencing library preparation, and library amplification. All laboratory work was conducted wearing lab coat and gloves, as well as sleeves during extraction and library preparation. All sample processing was carried out in laminar flow cabinets equipped with UV lamps. All bench tops were cleaned with 5% bleach and 70% ethanol before and after any lab work. Only sterile, disposable lab utensils were used. Filter tips were used for all pipetting.

Microbial enrichment, nucleic acid extraction, and library preparation
Samples used for enrichment were fresh frozen after collection with no addition of nucleic acid preservatives. Each batch of processed samples included a non-template control containing only phosphate buffered saline (PBS) ran in parallel with the patient samples (n=19). Samples were kept on ice unless otherwise stated. Biopsies were thawed and transferred to 400 μl cold PBS. For the leukaemia samples, 150-400 μl were mixed with cold PBS to a total volume of 400 μl. Two stainless steel beads of 2-3 mm diameter were added to the samples, which were subsequently homogenised using the TissueLyser II (Qiagen, Hilden, Germany) for 6 minutes at 30 Hz. All tissue homogenates, leukaemia samples, and ascites fluids (1 ml) devoid of pelleted cells were centrifuged for 2 minutes at 800 × g to
remove tissue debris and the supernatants were subsequently filtered through 5 μm centrifuge filters (Millipore, Darmstadt, Germany). The filtrates were nuclease digested to remove unprotected nucleic acids while retaining nucleic acids within bacterial cells or viral particles using 14 μl TURBO DNase (2U/μl) (Ambion, Thermo Fisher Scientific, Waltham, MA, USA), 12 μl Baseline-ZERO DNase (1U/μl) (Epicentre, Madison, WI, USA), 16 μl RNase Cocktail Enzyme Mix (Ambion), and 40 μl 10× TURBO DNase buffer in a total volume of 400 μl, and incubated at 37°C for two hours. Nucleic acids from the enriched samples were extracted using the High Pure Viral RNA Kit (Roche, Basel, Switzerland) according to the manufacturers instructions, with the addition of 10 μg linear acrylamide carrier (Applied Biosystems, Foster City, CA, USA).

DNA libraries were built using the Nextera or Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA, USA), according to the manufacturer’s instructions. Nextera libraries were amplified by 20 cycles of PCR, Nextera XT libraries by 12 cycles. The amplified and indexed libraries were purified using the Agencourt AMPure XP PCR purification system (Beckman Coulter, Indianapolis, IN, USA). In case of insufficient amplification, libraries were re-amplified using AccuPrime Pfx DNA polymerase (Life Technologies, Carlsbad, CA, USA) and P5 and P7 sequence primers.

Total DNA extraction and library preparation

Total DNA was extracted from samples using the QIAamp DNA Mini kit (Qiagen) following the manufacturer’s instructions. DNA libraries were prepared from 1 μg of DNA using either the TruSeq DNA protocol (PE-940-2001) (Illumina, San Diego, CA, USA) or an in-house protocol (21) using the NEBnext reagents (E6070) (New England Biolabs, Ipswich, MA, USA).
Sequencing

2 ×100 bp paired-end sequencing was performed on the Illumina HiSeq 2000 platform at BGI-Europe (Copenhagen N, Denmark).

Analysis of sequence data

Pre-processing and human depletion: Paired-end sequencing reads were trimmed of adapter sequences and overlapping read pairs merged using AdapterRemoval (22) (version 1.5.3). Reads shorter than 30 nucleotides after trimming were excluded from further analysis. For digital subtraction of human reads, the remaining reads were mapped to the human genome (hg38) using the mem algorithm implemented in BWA (23) (version 0.7.7). Reads of a pair were evaluated independently. Reads containing 25 bp or more of low complexity regions were filtered out using the DustMasker algorithm (24) (version 1.0.0).

Analysis of reads mapping to P. acnes: Genomic sequences from 97 P. acnes strains representing complete genomes, scaffolds, or contigs were downloaded from the NCBI Genome resource (http://www.ncbi.nlm.nih.gov/genome/genomes/1140) (Table S1). Reads filtered of human and low complexity sequences were mapped to the P. acnes sequences using the aln algorithm in BWA (23) after duplicate reads had been marked using SAMBLASTER (25). Only reads with a perfect match were kept for further analysis. The proportion of P. acnes reads was calculated as parts per million (ppm) unique P. acnes reads relative to the total number of demultiplexed reads.

Analysis of the P. acnes strains present: The presence of the different P. acnes strains was assessed on the basis of the mapping to the P. acnes genomic sequences (above). Only reads mapping exclusively to a single strain were considered for this analysis. Information regarding the phylogenetic type of the different strains was obtained from the P. acnes multilocus sequence typing (MLST) database (http://pubmlst.org/pacnes/).
Analysis of the coverage of the *P. acnes* genome: To visualise the genomic coverage of *P. acnes*, all non-human reads were mapped to the 12 full *P. acnes* genomes currently present in GenBank using the mem algorithm implemented in BWA (23) (version 0.7.7). The resulting alignments were represented as Circos plots (26).

Availability of supporting data

Sequencing reads mapping to *P. acnes* will be uploaded to the Sequence Read Archive once the article is accepted for publication.
Results

P. acnes in microbe-enriched and shotgun-sequenced samples

The relative number of P. acnes sequencing reads was analysed in data from 180 samples of 20 different sample types (Table 1). Ascites fluid samples and lymphoma/myeloma cell lines were analysed as each being one common sample category. 143 samples were processed to enrich for intact microbes before extracting DNA and preparing sequencing libraries. Sequencing reads subtracted of human reads were mapped to a custom database containing the 97 P. acnes genomes currently present in GenBank as full genomes or genome assemblies. P. acnes levels were assessed as parts per million (ppm) unique P. acnes reads relative to the total number of demultiplexed reads (see materials and methods). Using this approach, we detected relative levels of P. acnes reads exceeding 300 ppm in basal cell carcinoma (median of 576 ppm), mycosis fungoides (480 ppm), malignant melanoma (446 ppm), and breast cancer (373 ppm) (Fig. 1, black bars, and Table S2). Proportions of P. acnes reads below 100 ppm were detected in the remaining enriched samples (0.8-89 ppm). P. acnes reads were also detected in non-template control samples containing PBS run in parallel with the enriched samples (n=19). In these samples the relative proportion of P. acnes were high (median of 10,377 ppm) (see discussion).

Total DNA extracted from 107 samples, was subjected to shotgun sequencing (Table 1). Only basal cell carcinoma (median of 3.9 ppm), malignant melanoma (1.2 ppm), and breast cancer (1.1 ppm) contained relative P. acnes levels above 1 ppm (Fig. 1, grey bars, and Table S2). The fold increase in the median P. acnes ppm in enriched relative to shotgun-sequenced samples ranged from 35 in colon cancer biopsies (12 vs. 0.3 ppm) to >1000 in mycosis fungoides (480 vs. 0.5 ppm). The blood samples contained the lowest relative P. acnes levels of all samples included in this study (0.05 ppm).
Distribution of *P. acnes* strains

To assess the *P. acnes* strains and phylogenetic groups present in our samples we analysed the distribution of reads mapping to the 97 *P. acnes* genomes used in our analysis. Only reads mapping unambiguously to one *P. acnes* strain were included in the analysis. Using this criterion, an average of 8.6% of the *P. acnes* reads were included in the analysis (Table S3), however, with this approach we ensure high stringency when assigning the reads. Fig. 2 shows the relative abundance of strains for each sample. The phylogenetic type is indicated next to the individual strain names. For 20 of the strains included the phylogenetic type is not determined, these are designated unknown (Unkn).

Overall, two strains predominated across samples for both enriched and shotgun-sequenced samples (Table 2 and Table S4), namely strain 119_PAVI and SK182B-JCVI (both of unknown phylogenetic type). Strain HL025PA1 and HL050PA2 (type IA1 and II, respectively), were also among the ten most abundant strains for both methods. Certain strains, such as JCM 18920 (unknown type), HL001PA1 (type II), and C1 (type IA1), were mainly detected in enriched samples, while other strains, such as HL110PA4 (type II) and HL411PA1 (unknown type), were mainly detected in shotgun-sequenced samples.

Generally, larger strain diversity was seen in enriched tissue samples (Fig. 2 left, indicated by the striated pattern of the horizontal lines, and Fig. S1), ranging from an average of 10 different strains in testicular cancer to an average of 29 different strains in mycosis fungoides. The remaining samples contained an average of 2 to 9 different strains.

Coverage of the *P. acnes* genome

To evaluate the genomic coverage of *P. acnes*, non-human sequencing reads mapping to the 12 full genomes were presented with Circos (Fig. S2). The plots show evenly distributed sequencing reads leading to almost full coverage of the *P. acnes* genome in some samples, whereas more sporadic coverage was seen in samples containing lower *P. acnes* levels. A disproportionately high
sequencing depth was seen in two or three regions, depending on which reference genome was used for mapping, corresponding to the genes encoding the rRNAs. The high depth of sequencing could be a result of reads originating from bacteria other than *P. acnes* mapping to the rRNA regions as a result of high sequence similarity between different bacterial rRNAs, as the algorithm used for generating the plots allows for some mismatches as opposed to the more stringent mapping used for quantification. To investigate if an excess of reads in the rRNA regions was impacting the quantitative proportions of *P. acnes*, we reanalysed this data excluding reads mapping to the rRNA regions. For the majority of the samples the reanalysed values of the median relative levels of unique *P. acnes* reads were <1% lower than the results presented above, except some sample types containing very low numbers of *P. acnes* reads (Table S2). These results indicate that an excess of reads mapping to rRNA regions generally was not found in our quantification analysis.
Discussion

During metagenomic analysis of next generation sequencing data from patient samples we identified considerable amounts of genetic material from the bacterium *P. acnes*. *P. acnes* is known to contaminate human samples, and this initial discovery therefore prompted us to investigate the abundance of *P. acnes* in metagenomic data from a range of different human samples. In the present study we investigated the proportions of *P. acnes* DNA in data from samples subjected to enrichment for intact microbes, and from samples subjected to shotgun sequencing of the total pool of DNA. We detected the highest relative levels of *P. acnes* in samples enriched for intact microbes compared to shotgun sequencing of total DNA, with relative *P. acnes* levels being highest in skin derived samples and breast cancer (Fig. 1). Basal cell carcinoma samples contained the highest *P. acnes* levels of the shotgun-sequenced samples, while most other shotgun-sequenced samples as well as enriched BCP-ALL samples contained *P. acnes* levels below 1 ppm (Table S2).

Prior to surgical procedures or penetration of the skin for sampling of blood the skin is routinely sterilised using antiseptics. However, antiseptic preparation of the skin does not kill all bacteria or affects bacteria present in the deeper layers of the skin (27, 28), and bacteria from surrounding areas of the skin may rapidly recolonize the antiseptically treated areas (29, 30). Moreover, antiseptics will not remove extracellular bacterial DNA present on the skin. Therefore, detection of *P. acnes* in skin samples is not unexpected. As the breast tissue lies just beneath the subcutaneous adipose tissue in an area of the skin rich in *P. acnes*, contamination from the skin during surgery is conceivable.

In our study, microbial enrichment was achieved by reducing the amount of host endogenous DNA by centrifugation and filtration followed by nuclease treatment of fluid samples or tissue homogenates. This procedure is widely used for enrichment of intact viral particles in virus discovery studies (31-34), and results in an increase in intact viral nucleic acids relative to host...
nucleic acids (35). We did not culture bacteria from the samples, and therefore we do not know if
the detected bacterial DNA originates from live bacteria capable of forming colonies. As the DNA
of live bacteria is protected from degradation by DNases (36, 37), the up to >1000-fold increase in
the median proportion of P. acnes achieved upon enrichment could suggest that P. acnes DNA
detected in our samples originates, at least partially, from live bacteria.

P. acnes has over the years been detected in diverse disease samples, which has led to its proposed
association with several conditions. Among these, P. acnes has been implicated in the pathogenesis
of acne vulgaris, although its role is still debated (38). Other diseases for which the bacterium is
proposed to play a role is sarcoidosis (9), diverse spine conditions (15-17), SAPHO syndrome (8),
aortic aneurysms (18), and prostate cancer (19). Common to these studies is, that detection is based
on culturing, PCR, or a combination of the two. In most of the studies, roughly half of the samples
are positive, and the identified bacterium is P. acnes in about half of the cases. Besides the
proposed role in certain diseases and syndromes, P. acnes is considered a contaminant of blood
products (11, 39), surgical wounds (12, 40), and heart valve cultures used in the diagnosis of
infection during valve replacement surgery (10). The fact that P. acnes is often detected with
similar frequencies in surgical wounds and disease samples could imply that the detection of P.
acnes is incidental as a result of its widespread presence. Studies investigating the role of P. acnes
or other ubiquitous bacteria in infections or diseases should therefore be designed to allow
researchers to distinguish between contamination and true infection. Approaches suggested to help
distinguish include performing multiple cultures (11), histopathological analysis (7), distinguishing
single bacterial cells from biofilm-related infections by immunofluorescence microscopy (12), and
investigating the ability of P. acnes isolates to form biofilm (41). The invasiveness of P. acnes
isolates may also hint to a possible disease causing role of P. acnes in some cases (42). MLST
schemes have revealed three major phylogenetic groups of P. acnes; I, II, and III (43, 44), with
group I being further divided into IA1, IA2, IB, and IC. Clustering analysis found acne vulgaris to be
predominantly associated with certain clonal complexes within type IA1, while type IB and II were more frequently recovered from soft tissue and retrieved medical devices. Certain P. acnes sequence types from group II and III were only recovered from healthy skin (45).

In our samples, recurring P. acnes strains predominated across sample types, including in the non-template controls (Table S4). Two of these strains were also ubiquitously detected in skin microbiomes (3). These findings could suggest that some strains are widespread among individuals or that they arise from reagent or lab contamination, as P. acnes or Propionibacteria are well known contaminants of next generation sequencing data (46-48). Our non-template controls containing PBS run in parallel with the enriched samples contained high relative levels of P. acnes. This could occur as a result of cross-contamination during sample processing, contamination from lab personnel or reagents used, or of bleeding over from libraries with high P. acnes levels sequenced on the same lanes of the flow cell (35, 49). With the exception of two, the non-template control libraries contained DNA levels below the detection limit of both the Qubit and Bioanalyzer (data not shown), but were nevertheless sequenced. Comparing the non-template control datasets to the other enriched samples the controls contain an average of 12% human reads, while the biological samples contain an average of 91% human reads (Table S2). This dissimilar composition of the datasets suggests that the sequence reads in the controls, including the P. acnes reads, do not mainly originate from cross-contamination or bleeding over. Studies investigating contamination in next generation sequencing data show that more dilute samples contain higher fractions of contaminating reads (47, 48) and that negative controls contain a variety of bacterial genera, including Propionibacteria (48). This could be explained by the more extensive PCR amplification of contaminant templates from samples containing nearly no DNA compared to the same amount of contaminant templates present in a complex sample containing human DNA as well, competing for the same quantity of primers and other PCR reagents. Considering this skewed amplification and the circumstance that the biological datasets are “diluted” with a large amount of endogenous...
human reads, the relative level of *P. acnes* reads in the controls becomes less astounding. The
detection of *P. acnes* in the controls can thus be interpreted as further evidence of how easily
samples are contaminated with this ubiquitous bacterium, regardless of the source.

We believe that the high relative *P. acnes* levels detected particularly in enriched skin and breast
cancer samples reflect patient-derived contamination. Similar patterns showing traces of *P. acnes* in
most sample types and higher levels in *P. acnes* rich body sites are seen in other microbiome
studies (50). Our hypothesis is supported by the high diversity of *P. acnes* strains seen in these
samples, compared to the low diversity in other samples, including the controls (Fig. S1). The low
strain diversity in the controls supports the hypothesis of over amplification occurring in these
samples. More adequate negative controls for *P. acnes* levels in our study are the blood samples
from colon cancer patients, as these are in principle sterile samples containing high levels of
endogenous DNA. In these samples we detect practically no *P. acnes* reads (median of 5 in absolute
numbers or 0.05 ppm, Table S2).

Considering our findings it is possible that contamination, deriving from the patient or the hospital
or lab personnel, or arising during sample processing and analysis, contributes to the presence of *P.
acnes* DNA in samples processed in clinical and molecular laboratories. The risk of contamination
should therefore be considered carefully when interpreting molecular and sequencing data showing
the presence of *P. acnes*, and negative controls should always be included. Typing of *P. acnes* can
be a relevant tool when tracing the origin of *P. acnes*, but does not necessarily provide clinically
relevant information, as presumed cutaneous *P. acnes* strains are readily detected in diverse sample
types. Efforts should be made to exclude that *P. acnes* originates from contamination and,
importantly, to prove the causal relationship between this bacterium and the disease in question.

**Funding information**
The work was supported by the Innovation Fund Denmark (The GenomeDenmark platform, grant no. 019-2011-2), Danish National Research Foundation (grant no. DNRF94), and the Lundbeck Foundation. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgements

We thank professor Eske Willerslev for support and use of laboratories. We thank Robert Gniadecki (Bispebjerg Hospital), Christian von Buchwald (Rigshospitalet), Ewa Rajpert-De Meyts (Rigshospitalet), Line Groth-Pedersen (Rigshospitalet), Torben Steinike (Aarhus University Hospital), Hans Johnsen and Karen Dybkaer (Aalborg University Hospital), Estrid Høgdal (The Danish Cancer Biobank), Peter Hokland (Aarhus University Hospital), Ulrik Baandrup (Aalborg University Hospital), Jacob Rosenberg (Herlev Hospital), Zoltan Szallasi (Technical University of Denmark), Ildikó Vereczkey (National Institute of Oncology, Budapest, Hungary), Imre Pete (National Institute of Oncology, Department of Gynecology, Budapest, Hungary), and Zsolt Baranyai (1st Department of Surgery, Semmelweis University, Budapest, Hungary) for providing clinical samples. We thank BGI Europe for sequencing of the samples.

Author contributions

Conceived concept and designed experiments: LPN, AJH, SM, TM, LVI, EW, SB. Performed the experiments: SM, LVI, SRR, HF. Conducted computational analysis: TM, JFN, JMG1, TAH. Analyzed the computational data: SM, TM, AJH, LVI, JFN, JMG1, JARH. Prepared the manuscript: SM. Critical revision and approval of manuscript: All.

Conflicts of interest

The authors declare no conflicts of interest.
References


16. Albert HB, Lambert P, Rollason J, Sorensen JS, Worthington T, Pedersen MB,


Daly GM, Bexfield N, Heaney J, Stubbs S, Mayer AP, Palser A, Kellam P, Drou N,


**Figure legends**

Fig. 1 Relative *P. acnes* levels in enriched (black bars) and shotgun sequenced (grey bars) samples having a median *P. acnes* proportion >1 ppm. The proportions of unique *P. acnes* reads in ppm organised according to tissue types are shown on a logarithmic scale. The error bars indicate one standard deviation. Subcut: subcutaneous.

Fig. 2 Relative distribution of *P. acnes* strains in the individual samples. The y-axis shows the percentage accounted for by each strain, and each vertical line represents a sample. The bottom horizontal lines indicate enriched or shotgun-sequenced samples, and the bottom vertical ticks delimit the sample categories. The phylogenetic type of the *P. acnes* strains are indicated next to the strain name in the figure legend. The strains mentioned in the main text are highlighted in bold.

NTC: non-template control, Unkn: unknown.
Table 1. Samples included in the study. The diseases, sample types, and number of samples included are shown.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Abbreviation used</th>
<th>Sample type</th>
<th>Microbial enrichment</th>
<th>Shotgun sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute myeloid leukaemia</td>
<td>AML</td>
<td>Bone marrow (sorted cells)</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>B-cell chronic lymphocytic leukaemia</td>
<td>B-CLL</td>
<td>Blood/bone marrow (sorted cells)</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>B-cell precursor acute lymphoblastic</td>
<td>BCP-ALL</td>
<td>Bone marrow</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>leukaemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basal cell carcinoma</td>
<td>BCC</td>
<td>Tumour biopsies</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast cancer (ductal/lobular)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic myelogenous leukaemia</td>
<td>CML</td>
<td>Bone marrow (sorted cells)</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Colon cancer</td>
<td></td>
<td></td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Lymphoma/myeloma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malignant melanoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycosis fungoides (cutaneous T-cell lymphoma)</td>
<td>Mycosis</td>
<td>Tumour biopsies</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Oropharyngeal cancer</td>
<td>Oral</td>
<td>Tumour biopsies</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-lineage acute lymphoblastic leukaemia</td>
<td>T-ALL</td>
<td>Bone marrow (non-sorted/sorted cells)</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Testicular cancer (sem/non-sem)</td>
<td>Testis</td>
<td>Tumour biopsies</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>Vulva cancer</td>
<td>Valva</td>
<td>Tumour biopsies</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Total number of samples and sample types 20 143 107
Table 2. Prevalence (%) of the 20 most frequently detected *P. acnes* strains in samples investigated by microbial enrichment, shotgun sequencing, and all in samples combined.

<table>
<thead>
<tr>
<th>Strain</th>
<th>All samples</th>
<th>Microbial enrichment</th>
<th>Shotgun sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>119_PAV1</td>
<td>82</td>
<td>83</td>
<td>64</td>
</tr>
<tr>
<td>SK182B-JCVI</td>
<td>78</td>
<td>81</td>
<td>60</td>
</tr>
<tr>
<td>HL025PA1</td>
<td>56</td>
<td>62</td>
<td>36</td>
</tr>
<tr>
<td>JCM 18920</td>
<td>45</td>
<td>66</td>
<td>6</td>
</tr>
<tr>
<td>HL050PA2</td>
<td>44</td>
<td>52</td>
<td>23</td>
</tr>
<tr>
<td>HL001PA1</td>
<td>34</td>
<td>51</td>
<td>4</td>
</tr>
<tr>
<td>HL411PA1</td>
<td>32</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>C1</td>
<td>31</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td>HL110PA4</td>
<td>29</td>
<td>14</td>
<td>46</td>
</tr>
<tr>
<td>ATCC 6919</td>
<td>27</td>
<td>32</td>
<td>14</td>
</tr>
<tr>
<td>JCM 18909</td>
<td>26</td>
<td>37</td>
<td>4</td>
</tr>
<tr>
<td>HL082PA2</td>
<td>24</td>
<td>35</td>
<td>-</td>
</tr>
<tr>
<td>J165</td>
<td>24</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>HL201PA1</td>
<td>20</td>
<td>29</td>
<td>-</td>
</tr>
<tr>
<td>HL110PA1</td>
<td>19</td>
<td>-</td>
<td>27</td>
</tr>
<tr>
<td>HL030PA1</td>
<td>16</td>
<td>-</td>
<td>21</td>
</tr>
<tr>
<td>SK187</td>
<td>15</td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>HL099PA1</td>
<td>15</td>
<td>21</td>
<td>-</td>
</tr>
<tr>
<td>HL078PA1</td>
<td>12</td>
<td>18</td>
<td>-</td>
</tr>
<tr>
<td>J139</td>
<td>11</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>HL087PA2</td>
<td>-</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>KPA171202</td>
<td>-</td>
<td>14</td>
<td>-</td>
</tr>
<tr>
<td>HL056PA1</td>
<td>-</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>HL110PA3</td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>HL025PA2</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>HL046PA2</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>HL074PA1</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
</tbody>
</table>
Proportion of unique *P. acnes* reads (ppm)

- Skin
  - BCC
  - Melanoma
  - Breast
  - Vulva
  - Oral
  - Bladder
  - Testis
  - Colon
  - B-CLL
- Mucosa
  - AML
  - T-ALL
  - CML
  - Ascites
  - BCC
  - Melanoma
  - Breast

Bars represent the proportion of unique *P. acnes* reads in various tissues and conditions. The x-axis represents the number of reads per million (ppm), ranging from 0 to 10,000.