



**HAL**  
open science

## Multi-omics analysis to decipher the molecular link between chronic exposure to pollution and human skin dysfunction

Namita Misra, Cécile Clavaud, Florent Guinot, Nasrine Bourokba, Stephanie Nouveau, Sakina Mezzache, Paul Palazzi, Brice M R Appenzeller, Arthur Tenenhaus, Marcus H y Leung, et al.

### ► To cite this version:

Namita Misra, Cécile Clavaud, Florent Guinot, Nasrine Bourokba, Stephanie Nouveau, et al.. Multi-omics analysis to decipher the molecular link between chronic exposure to pollution and human skin dysfunction. *Scientific Reports*, 2021, 11 (1), 10.1038/s41598-021-97572-1 . hal-03369683

**HAL Id: hal-03369683**

<https://hal.sorbonne-universite.fr/hal-03369683v1>

Submitted on 7 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPEN

## Multi-omics analysis to decipher the molecular link between chronic exposure to pollution and human skin dysfunction

Namita Misra<sup>1✉</sup>, Cécile Clavaud<sup>1</sup>, Florent Guinot<sup>1</sup>, Nasrine Bourokba<sup>1</sup>, Stephanie Nouveau<sup>1</sup>, Sakina Mezzache<sup>1</sup>, Paul Palazzi<sup>2</sup>, Brice M. R. Appenzeller<sup>2</sup>, Arthur Tenenhaus<sup>3,4</sup>, Marcus H. Y. Leung<sup>5</sup>, Patrick K. H. Lee<sup>5</sup>, Philippe Bastien<sup>1</sup>, Luc Aguilar<sup>1</sup> & Nükhet Cavusoglu<sup>1</sup>

Environmental pollution is composed of several factors, namely particulate matter (PM<sub>2.5</sub>, PM<sub>10</sub>), ozone and Ultra Violet (UV) rays among others and first and the most exposed tissue to these substances is the skin epidermis. It has been established that several skin disorders such as eczema, acne, lentigines and wrinkles are aggravated by exposure to atmospheric pollution. While pollutants can interact with skin surface, contamination of deep skin by ultrafine particles or Polycyclic aromatic hydrocarbons (PAH) might be explained by their presence in blood and hair cortex. Molecular mechanisms leading to skin dysfunction due to pollution exposure have been poorly explored in humans. In addition to various host skin components, cutaneous microbiome is another target of these environment aggressors and can actively contribute to visible clinical manifestation such as wrinkles and aging. The present study aimed to investigate the association between pollution exposure, skin microbiota, metabolites and skin clinical signs in women from two cities with different pollution levels. Untargeted metabolomics and targeted proteins were analyzed from D-Square samples from healthy women (n = 67 per city), aged 25–45 years and living for at least 15 years in the Chinese cities of Baoding (used as a model of polluted area) and Dalian (control area with lower level of pollution). Additional samples by swabs were collected from the cheeks from the same population and microbiome was analysed using bacterial 16S rRNA as well as fungal ITS1 amplicon sequencing and metagenomics analysis. The level of exposure to pollution was assessed individually by the analysis of polycyclic aromatic hydrocarbons (PAH) and their metabolites in hair samples collected from each participant. All the participants of the study were assessed for the skin clinical parameters (acne, wrinkles, pigmented spots etc.). Women from the two cities (polluted and less polluted) showed distinct metabolic profiles and alterations in skin microbiome. Profiling data from 350 identified metabolites, 143 microbes and 39 PAH served to characterize biochemical events that correlate with pollution exposure. Finally, using multiblock data analysis methods, we obtained a potential molecular map consisting of multi-omics signatures that correlated with the presence of skin pigmentation dysfunction in individuals living in a polluted environment. Overall, these signatures point towards macromolecular alterations by pollution that could manifest as clinical sign of early skin pigmentation and/or other imperfections.

Pollution remains the world's largest environmental threat to human health, responsible in 2017 for 15% of all deaths globally, and 275 million Disability-Adjusted Life Years, reported by *The Lancet Commission on Pollution and Health, 2017* and ambient air pollution kills more people around the globe than any other form of pollution (<https://gahp.net/pollution-and-health-metrics/>). Epidemiological and clinical studies have shown that short- and long-term exposure to particulate matter (PM<sub>2.5</sub>: Particle matter with aerodynamic diameter

<sup>1</sup>Research and Innovation, L'Oréal SA, Aulnay Sous Bois, France. <sup>2</sup>Human Biomonitoring Research Unit, Luxembourg Institute of Health, Strassen, Luxembourg. <sup>3</sup>CentraleSupélec Laboratoire des Signaux et Systèmes, Université Paris-Saclay, CNRS, Gif-sur-Yvette, France. <sup>4</sup>Brain and Spine Institute, Paris, France. <sup>5</sup>School of Energy and Environment, City University of Hong Kong, Kowloon, Hong Kong SAR, China. ✉email: namita.misra@rd.loreal.com

less than 2.5 microns, and PM<sub>10</sub>: Particle matter with aerodynamic diameter less than 10 microns) and ozone increase respiratory and cardiovascular morbidity and lead to the development of certain cancers<sup>1–3</sup>. Among all organs, skin is the most visible and there is accumulating scientific evidence that air pollution plays an important role in extrinsic aging<sup>4,5</sup>. Skin is a multi-layered tissue composed of a top layer, the stratum corneum, which is in direct contact with atmospheric pollutants, while the deeper layers, such as dermis, are exposed to diverse xenobiotics through systemic route. A major mechanism by which PM exert their detrimental effects is through the generation of oxidative stress<sup>6,7</sup> which is an important contributor to extrinsic skin aging<sup>8,9</sup>. PM<sub>2.5</sub> is often associated with toxic chemicals such as heavy metals or polycyclic aromatic hydrocarbons (PAHs), and some photo-reactive PAHs can induce strong oxidative stress under UVA exposure<sup>10</sup>. Clinical studies performed in China and elsewhere indicate towards premature skin aging phenotype such as pigmented spots and wrinkles<sup>11,12</sup>. In a previous study using in-vitro skin models and untargeted proteomics, we have shown that exposure to PAH(s) leads to disruption of several cellular processes<sup>6</sup>. However, the molecular mechanisms that link exposure and its impact on clinical manifestations remain to be deciphered.

The present work is part of a multi-parametric study, in which non-invasive facial skin samples (D-Squame) were collected from hundred and thirty-four healthy women living in two Chinese cities with different levels of exposure to pollution<sup>13</sup>. Exposure was determined by quantifying 39 parent PAHs and their mono-hydroxy metabolites in hair samples collected from the same individuals, according to a validated method providing information on the average internal dose of chemicals<sup>13</sup>. Each individual was also evaluated for a set of dermatological skin parameters. Microbiome profiling from all the subjects from both cities gave us the first insight into microbiome alteration resulting from pollution exposure<sup>14</sup>. To explore the mechanistic link between pollution exposure and its clinical manifestation, we considered several molecular characterizations. We first applied an untargeted metabolomic analysis on skin samples to characterize pollution-dependent biochemical events followed by a targeted proteomic analysis to complement the biochemical profiles generated through metabolomics. Finally, we complemented the multi-omics data by 16S and ITS amplicon sequencing. PAH quantification along with clinical evaluation of the same individuals completed this global characterization. These datasets were computed in a block structured multivariate analysis to propose a molecular map of human skin that links pollution exposure to skin dysfunction.

## Results and discussion

**Facial skin pigmentary disorders.** To assess the effect of chronic exposure to pollution on exposed area of skin, we performed dermatological clinical assessment across 35 facial parameters on 67 women living in the city of Baoding (the most polluted area) and 67 women living in the city of Dalian (less polluted area), in China. The women had been living in respective cities for at least 15 years and were aged between 25 to 45 years. The clinical parameters analyzed on these women are presented in Suppl. Table 1. Briefly, each individual was evaluated for 10 major facial clinical themes/clusters that included 35 sub-conditions. Pigmentary disorders and wrinkles are the two clinical themes that were most significantly increased on facial skin of individuals from the polluted area. Shiny and dull skin were also modulated in a second instance (Suppl. Table 1). Previously, we have reported that individuals from the two cities differ significantly in PAH levels measured in hair shafts, suggesting different level of exposure to outdoor pollution<sup>13</sup>. These initial clinical evaluations were done on a larger population of 204 women and this difference remained significant for the study sub-population of 134 (Suppl. Fig. 1). Therefore, we calculated a PAH score (described in “Materials and methods” section) as a complementary factor to be analyzed together with the other datasets (clinical, metabolites, microbiome etc.) presented in this paper. As shown in Table 1, spread macules on forehead and cheeks and are significantly more prevalent (25.4% vs 6% and 56.7% vs 35.8% respectively) in women from the most polluted city of Baoding versus Dalian. A similar pattern is observed in the prevalence of spread macules on forehead and cheeks for women with PAH score higher than the median of the study population with regard to women with PAH score lower than the median (22.4% vs 9% and 53.7% vs 38.8% respectively). Previous studies have shown that the occurrence of some disorders of facial hyperpigmentation such as melasma, spots and lentigenes are highest in individuals of skin type III-IV living in India and Southeast Asia<sup>15</sup>, these are also the geographical regions reporting heavy environmental PAH burden (<https://gahp.net/pollution-and-health-metrics/>). Results from our study show a similar trend. Spread macule (SP), a facial pigmentary condition is less known and has been described previously in South Asian population<sup>16</sup> living in large metropolitan cities in India that are often the most polluted ones.

**Skin metabolite profile.** Untargeted metabolomic analysis on facial skin tape strips allowed the identification of a total of 350 metabolites. We first conducted a pathway enrichment analysis to explore if chronic exposure to pollution induced accumulation or perturbations in specific metabolic pathways. At a higher level, increased enrichment was observed in samples from Baoding that included amino acid and fatty acid metabolism (Fig. 1A and Suppl. Table 2). Xenobiotics of dietary origin appeared to be decreased compared to Dalian. Further analysis revealed elevated levels of N-acetyl amino-acids (Fig. 1B), gamma-glutamyl amino-acids (Fig. 1C), and urea cycle intermediates. N-acetyl amino acids are derived from proteins that have undergone post-translational acetylation or from free amino acids reacting with acetyl groups. Gamma-glutamyl transferase (GGT) is an enzyme that transfers the gamma-glutamyl moiety of glutathione to an acceptor that may be an amino acid or peptide<sup>17</sup>. The GGT system thus plays an important role in transporting amino acids and dipeptides into the cells as well as regulating the exchange of intra- and extracellular glutathione. Higher levels of gamma-glutamyl amino acids in the polluted group suggest an upregulated GGT system, perhaps to provide cells pyrrolidone carboxylic acid (PCA or 5-oxoproline), a natural moisturizing factor (NMF) released from GGCT enzyme following GGT activation<sup>18</sup>. NMF are the amino acids or their derivatives (PCA and urocanic acid, together with lactic acid, urea, citrate and sugars) that are produced due to flaggrin proteolysis and are

Clinical signs	Intensity	Dalian	Baoding	<i>p</i> value	PAH < MED	PAH > MED	<i>p</i> value
<b>Pigmentary disorders</b>							
Spread macules on forehead	No	94.03	74.63		91.04	77.61	
	Weak	2.99	17.91		7.46	13.43	
	Moderate	2.99	2.99	0.0029	1.49	4.48	0.0292
	Severe		4.48			4.48	
Spread macules on cheeks	No	64.18	43.28		61.19	46.27	
	Weak	32.84	32.84		34.33	31.34	
	Moderate	2.99	19.40	0.0024	2.99	19.40	0.0212
	Severe		4.48		1.49	2.99	
Hyperpigmented spots on forehead	No	26.87	20.90		25.37	22.39	
	Weak	73.13	76.12		74.63	74.63	
	Moderate		2.99	0.2844		2.99	0.4938
Simplex lentigo on cheek	No	25.37	13.43	0.0805	25.37	13.43	0.0805
	Yes	74.63	86.57		74.63	86.57	
Actinic lentiginos on cheek	No	77.61	76.12	0.8377	82.09	71.64	0.1516
	Yes	22.39	23.88		17.91	28.36	
Melasma or melasma like on forehead	No	97.01	91.04	0.1447	97.01	91.04	0.1447
	Yes	2.99	8.96		2.99	8.96	
Melasma or melasma like on cheek	No	80.60	85.07	0.4919	80.60	85.07	0.4919
	Yes	19.40	14.93		19.40	14.93	

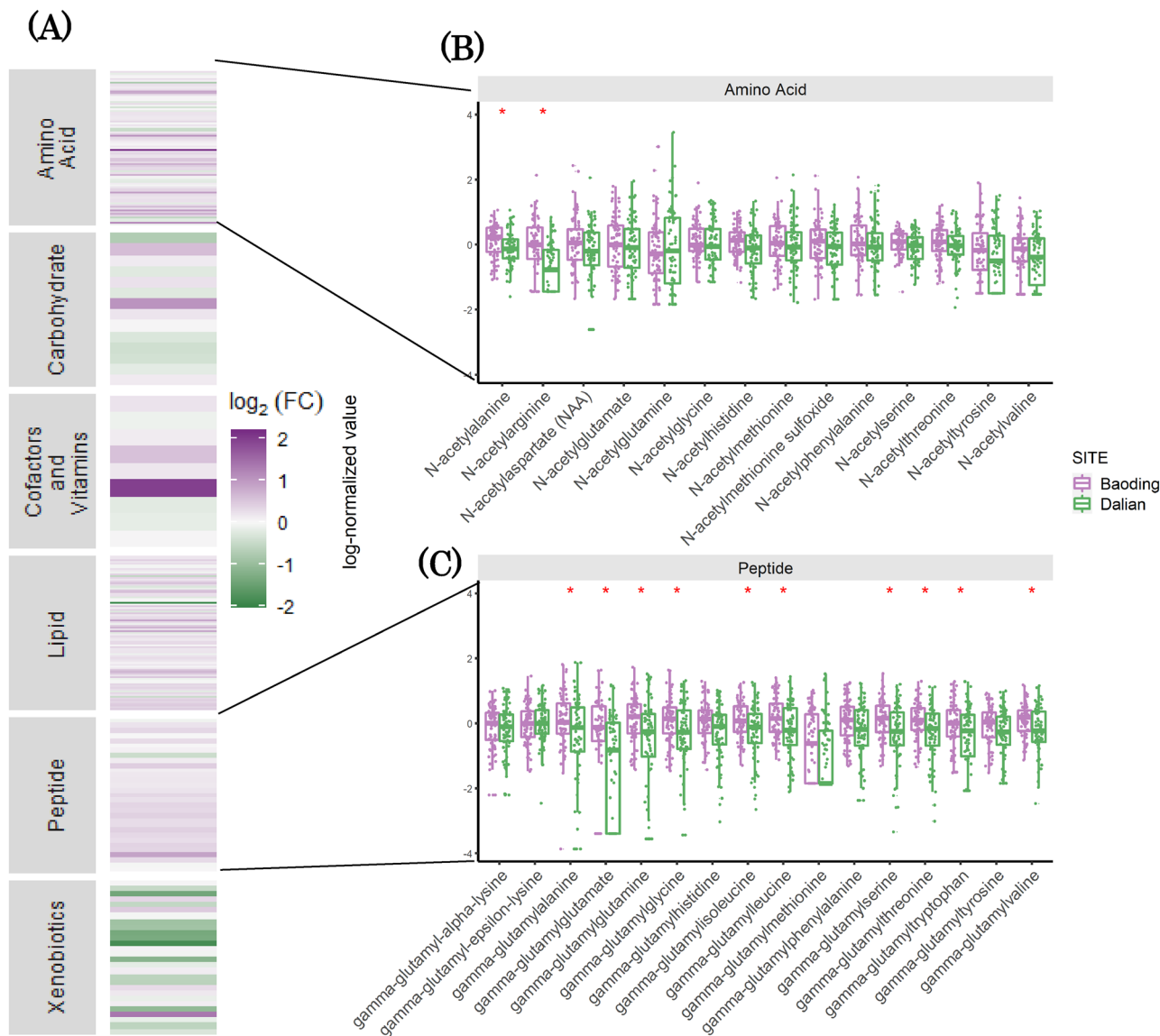
**Table 1.** Clinical facial signs for pigmentary disorders, their evaluation across Cities and PAH levels. PAH median corresponds to the median of the PAH score, the first component of a principal components analysis (PCA) on log-normalized PAH measurements. *p* values were calculated using either unpaired two-samples Wilcoxon or  $\chi^2$  tests depending on whether the clinical scores were ordinal or binary. The table was created using SAS version 9.4, SAS Institute Inc., Cary, NC, USA.

found exclusively in stratum corneum (SC). Free amino acids (Fig. 2A), 5-oxoproline and lactic acid levels were elevated in Baoding versus Dalian (Fig. 2B) suggesting an upregulation of the filaggrin proteolysis pathway. To further explore the latter, we performed targeted protein analysis on proteins involved in skin barrier function. Surprisingly we observed decreased levels of CASP14, PADI1, TGM3 and GGCT. Latter, the gamma-glutamyl transferase enzyme is implicated in the last step of PCA synthesis, and since PCA levels were increased, this observation would suggest another mechanism that could be responsible for PCA increase. Interestingly, KLK7 a major enzyme in desquamation was decreased by 2 folds, suggesting a slow-down in surface renewal that could as a consequence lead to PCA and other amino acids accumulation. KLK5, another desquamation activating enzyme could be considered unchanged with a 1.2-fold increase. To further support this accumulation hypothesis, Filaggrin and Filaggrin2 which are the principal NMF providers, remained unaltered in skin samples from Baoding (Suppl\_Fig. 2). Taken together, protein and metabolite analysis show a perturbation of molecular entities that contribute towards the skin barrier functioning and an enhanced repair/compensation in skin samples from Baoding. It appears that chronic exposure to pollution induces skin to be in a dynamic state to manage damage and repair simultaneously. Pollution induced effects on the skin barrier function have been reported by us previously using in-vitro models and untargeted proteomics<sup>19</sup>.

As a next step and in order to quantify the differences between the metabolomics profile between two cities more precisely and irrespective of classification into sub or super pathway, the log<sub>2</sub>-Fold change of all metabolites with Baoding as the reference population were calculated and represented as a cumulative distribution in Suppl\_Fig. 3. Furthermore, to test the significance of the fold change, we performed a *t*-test for each of them with a volcano plot visualized in Fig. 3A.

To further investigate the relationship between the metabolite profiles and the city membership, we also conducted a multivariate analysis using a random forest classification model (see “Methods” section for more details). The variable importance values are shown in Suppl\_Fig. 4A and further details on the RF model are given in Suppl\_Fig. 4B,C.

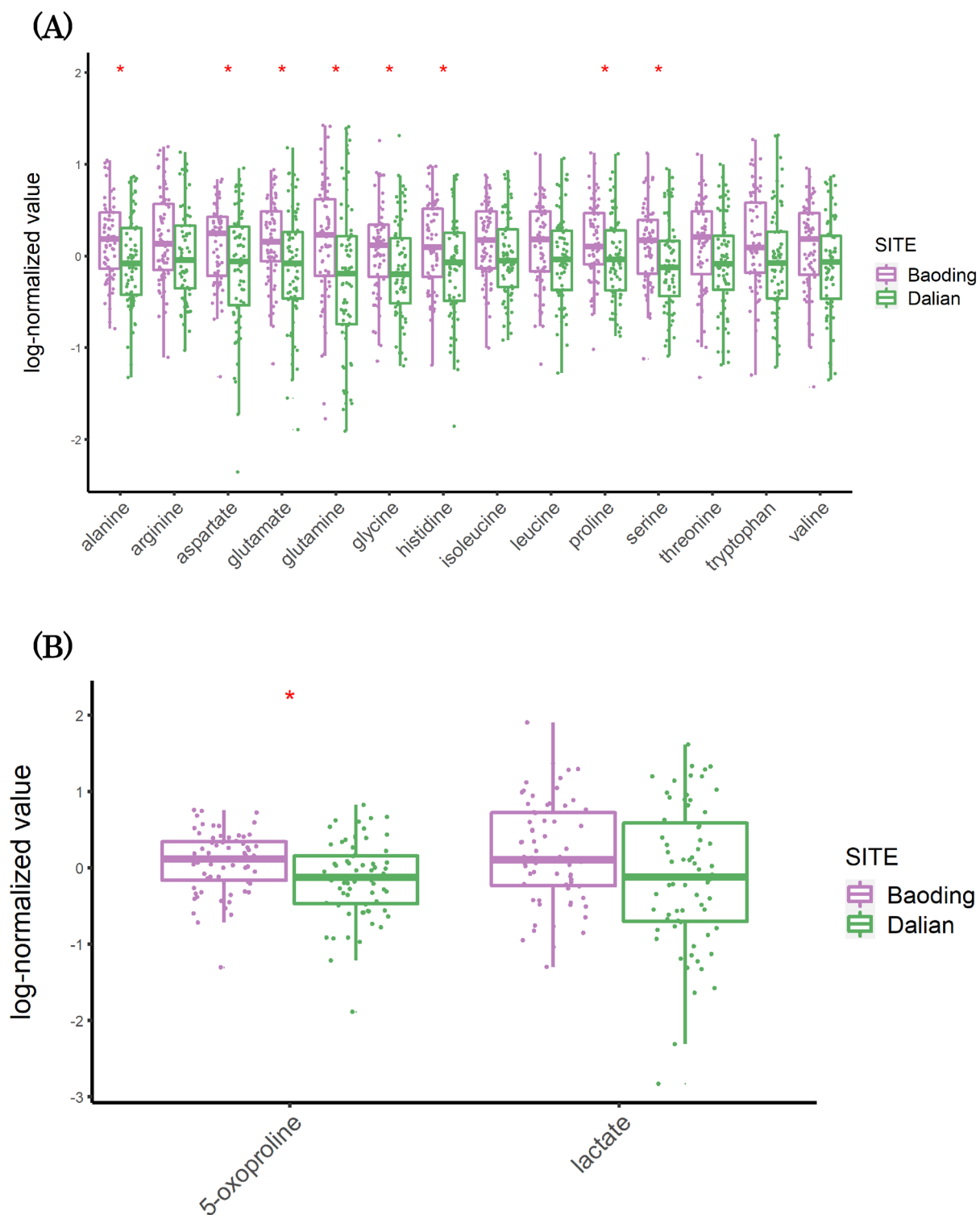
When re-grouped by pathway (Fig. 3B), we find the most significant members of the different class of metabolites presented above and therefore allows for a deeper insight into the mechanisms that are most significantly associated with pollution exposure. Among lipids group, oleamide, oleoyl and linoleoyl ethanolamides are the fatty acid amides and are bioactive lipid signaling molecules. These are reported to be involved in a wide range of physiological responses in various tissues including skin conditions such as atopic and contact dermatitis<sup>20</sup>. Particulate matter leads to secretion of proinflammatory cytokines, e.g., tumor necrosis factor (TNF)- $\alpha$ , IL-1 $\alpha$ , IL-8, and upregulation of matrix metalloproteinases 1, 2, and 9<sup>21,22</sup>, it is likely that bioactive fatty acids is the response against pollution induced inflammatory stress. Among other metabolites, 4-imidazole acetate, N-acetyl arginine and unknown X-13737 are significantly upregulated while p-cresol sulfate is downregulated (Fig. 3B). Detection of elevated levels of 4 imidazole-acetate indicates perturbation of Histidine metabolism and is discussed in the next section. Elevated levels of kynurenate indicates perturbation of Tryptophan metabolism and a focused



**Figure 1.** Metabolite pathways that are differentially modulated in skin samples of individuals from polluted city of Baoding. **(A)** Heatmap of log<sub>2</sub> of Fold Change between Baoding (More Polluted) and Dalian (Less Polluted), positive values indicate a higher amount in Baoding and negative values a higher amount in Dalian. **(B)** Focus on N-acetyl amino acids with boxplots of log normalized values, colored by site. **(C)** Focus on gamma-glutamyl amino acids with boxplots of log normalized values, colored by site. A red star indicates, for a given metabolite, that the difference in mean between the 2 cities is significant (q-value < 0.05, t-test with correction for multiple testing using Benjamini–Hochberg method). Figures were created using the R software: R Core Team (2017), R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

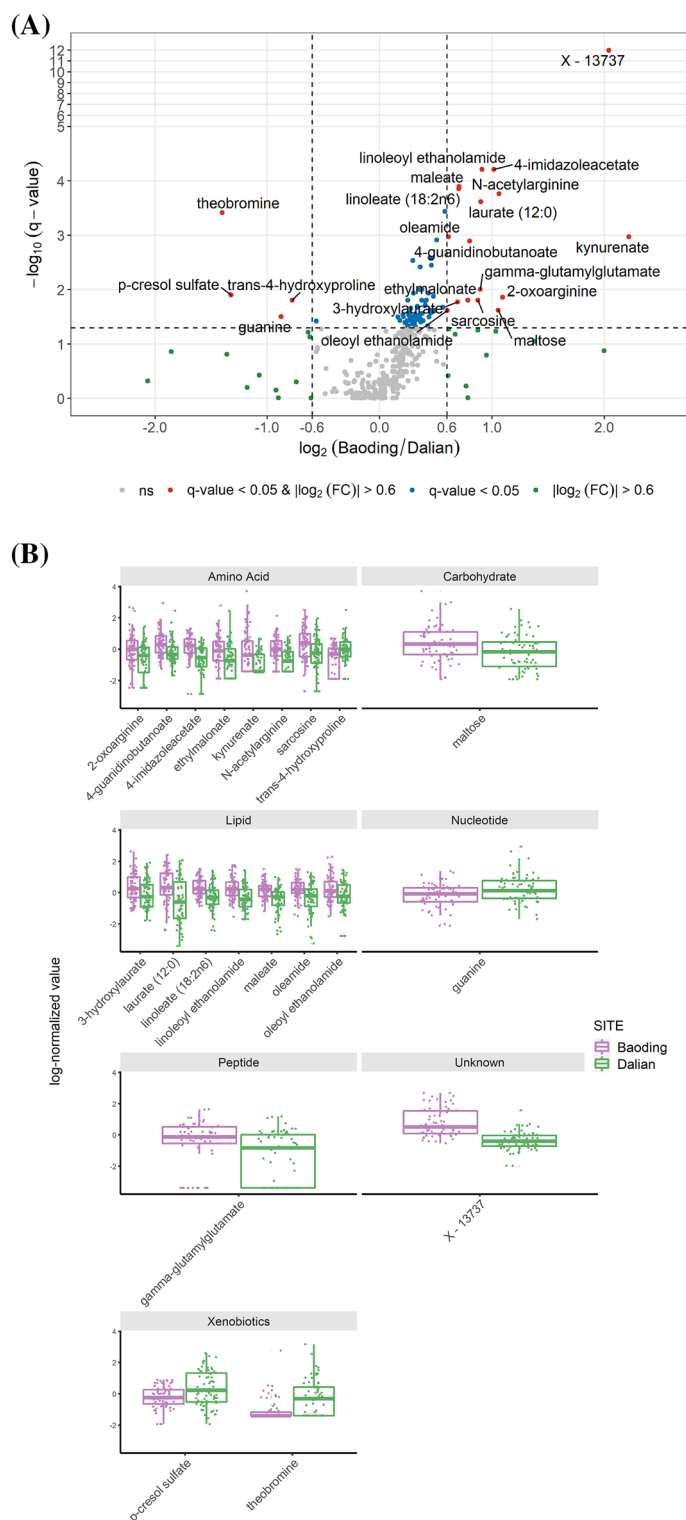
analysis showed reduced indoxyl-3-sulfate in individuals in Baoding (Suppl\_Fig. 5). The kynurenine pathway (KP) of tryptophan metabolism accounts for most of the tryptophan that is not involved in protein synthesis and includes compounds active in the nervous and immune systems. Kynurenine acts on the aryl hydrocarbon receptor (AhR), affecting metabolism of xenobiotics and affecting carcinogenesis. Limited studies have focused on the effects of acute ultraviolet exposure and the induction of the KP in human skin derived fibroblasts and keratinocytes. UV exposure is known to elicit an inflammatory component in skin cells, it is probable that the KP may be induced in these cells in response to UV exposure and this KP metabolites could be the mediators of inflammatory and anti-inflammatory responses<sup>23</sup>. We observed reduced indoxyl-3-sulfate in individuals living in the most polluted area. Several studies have identified tryptophan metabolite indole as a major extracellular metabolite produced by gut bacteria such as *E. coli* and interestingly, indoxyl-3-sulfate has been characterized as AhR agonist<sup>24</sup>. P-cresol sulfate was found to be reduced in skin samples from Baoding (Fig. 2A). P-cresol has been previously reported as a biomarker for healthy aging<sup>25</sup>.

Another top marker in the volcano plot is the unknown metabolite “X-13737” (Fig. 3A,B). Upon further investigation, it was determined that the MS/MS spectrum for “X-13737” actually contained a mixture of ions from two metabolites, one being (S)- $\alpha$ -amino- $\omega$ -caprolactam and the other being a still unidentified molecule



**Figure 2.** (A) Boxplots of log-transformed values of free amino acids (B) Boxplots of log-transformed values of NMF. A red star indicates, for a given metabolite, that the difference in mean between the 2 cities, Baoding (More Polluted) and Dalian (Less Polluted), is significant ( $q$ -value  $< 0.05$ ,  $t$ -test with correction for multiple testing using Benjamini–Hochberg method). Figures were created using the R software: R Core Team (2017), R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

(Suppl\_Fig. 6). Although the precise identity and contribution of this additional co-eluting unknown to the “X-13737” signal has not been fully verified, data suggest that it is one of the major actors that differentiates metabolomics profile of individuals living in one city from the other (Suppl\_Fig. 4). The novelty of this finding



**Figure 3.** (A) Volcano plot representing the  $\log_2(\text{FC})$  in x-axis and  $-\log_{10}(\text{q-value})$  in y-axis. A threshold of 0.6 for the  $\log_2(\text{FC})$  is added as a dotted vertical black line and a threshold of  $-\log_{10}(0.05)$  as a dotted horizontal black line for the q-value. For representation purpose, the y-axis scale has been squashed by a factor of 5 between 5 and 12. A metabolite is considered significantly modulated if its  $|\log_2(\text{FC})| > 0.6$  and  $\text{q-value} < 0.05$  (in red and labelled on the figure). (B) Boxplots of significantly modulated metabolites, gathered by pathway, Baoding (More Polluted) and Dalian (Less Polluted). Figures were created using the R software: R Core Team (2017), R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

is that caprolactam itself is a moderately toxic irritant, which has previously been regarded as a hazardous air pollutant. Although it is not known how (S)- $\alpha$ -amino- $\omega$ -caprolactam can be formed from caprolactam (or if it can be at all), the fact that caprolactam is a known air pollutant, it is noteworthy that the metabolomic data yielded an increase in signal for a spectrum which contains (S)- $\alpha$ -amino- $\omega$ -caprolactam in the skin of the women with higher exposure to pollutants as compared to the other ones.

**Skin metabolites and PAH exposure.** From the analysis presented above, we saw significant differences in metabolomic profiles of skin from individuals living in polluted and non-polluted environments. To go one step further and to make a direct correlation between exposure and its molecular impact on skin, we carried out an analysis between PAHs concentrations, measured in individual hair samples, and skin SC metabolites using a sparse Canonical Correlation Analysis (sCCA) to identify sets of PAHs and metabolites that are correlated together. Hair analysis is increasingly used for the assessment of exposure, and several studies demonstrated that pollutant concentration in hair is representative of the body burden<sup>26–28</sup>. Hair analysis provides integrated information on chronic exposure covering up to several months (considering an average growth of one centimeter per month) and allows for the detection of both parent pollutants and their metabolites contrary to biological fluids<sup>26</sup>. Among the different correlations observed between PAH and skin metabolites, a positive correlation with N-acetyl amino acids was interesting (Fig. 4A). Acetylation is post-translational protein modifications that has multiple effects on cellular proteins and metabolites. Acetyl-coenzyme A donates acetyl group that can be post translationally attached to either the alpha amino group of N-terminus of proteins or to epsilon amino group of lysine residue. The reaction is catalyzed by N-acetyltransferase (NATs) and they are involved in a multitude of signaling pathways impacting diverse cellular functions<sup>29</sup>. The correlation between N-acetyl aa and PAH is intriguing. Previous studies have shown a link between exposure to particulate matter and histone acetylation<sup>30</sup>. Other studies show an association of NAT and cotinine levels from exposure to secondhand smoke<sup>31</sup>. Interestingly, we observed strong correlations of skin metabolites with 2OH Naphthalene, a PAH metabolite associated with smoking<sup>32</sup>. In conclusion, association between N-acetyl amino acids and PAH needs further investigation. Another pathway of interest was histidine metabolism since a correlation was observed between Cis-urocanate and 4 imidazole-acetate and PAH (Fig. 4A). Previous studies have shown that Benzo[a]pyrene perturbs histidine metabolism in human lung epithelial cells<sup>33</sup>. We observed elevated levels of Cis-urocanate in individuals from the polluted city (Fig. 4B). Trans-urocanate levels remained unchanged between the two cities resulting in elevated Cis/trans ratio in polluted city of Baoding (Fig. 4B). Histidine is deaminated by histidase to form trans-urocanate. While trans-urocanate acts as photo-protectant, cis-urocanate is implicated to suppress the immune response, leading to UV-induced immunosuppression.

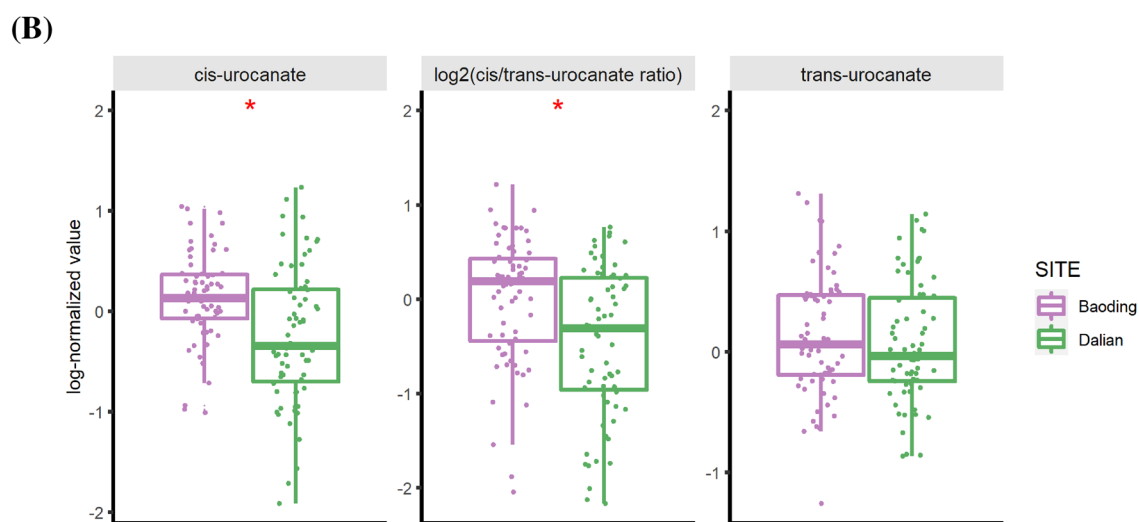
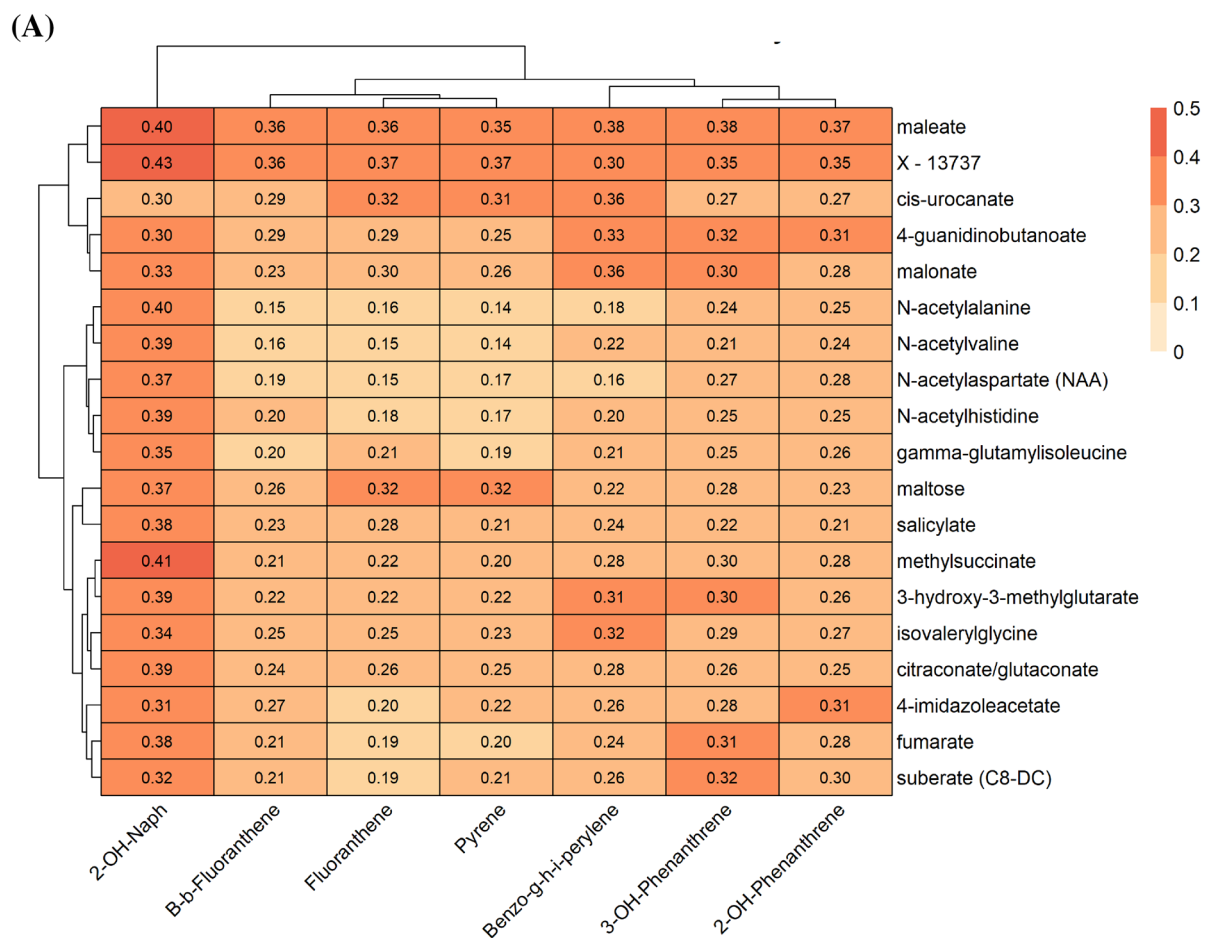
**Impact of skin microbiome on metabolites and vice-versa.** On the skin surface, there is a complex interplay between the host metabolism and its microbiota. It cannot be ruled out that a subset of metabolites presented above are either directly produced or are altered by the skin microflora. Indeed, skin microbiome plays an important role in maintaining cutaneous health and the skin microflora is constantly adapting in response to intrinsic and extrinsic factors. Environment and therefore pollution exposure has the potential to influence the skin microflora and bacterial isolated from the human skin have been shown to degrade PAHs and related xenobiotic compounds<sup>33</sup>.

In a previous study based on the same cohort, we have shown that an increase in Shannon diversity was correlated to PAH levels score. This observation was further supported by identification of pathways that point towards host-microbe interaction and degradation of aromatic compounds in a metagenomics analysis on a limited number of individuals ( $n = 32$ ) who presented with very high and very low levels of PAH in their hair samples<sup>14</sup>. Here, to explore the influence of microbiome on skin metabolites and vice-versa, we used PLS prediction model and Pearson's correlations between metabolites and microbiome on all the 134 individuals. In previous studies performed on gut, fecal metabolome has been reported to account for 68% of the variance in inter individual gut microbiome composition<sup>34</sup>, such comparisons are largely unexplored for skin. There has been a focus on persistence of the metabolites derived from skin care products<sup>35</sup> or chemicals associated with urbanization<sup>36</sup> or a higher level 3D modeling of the skin surface components<sup>37</sup>. In our study, for the first time, we have simultaneously analyzed endogenous metabolites and microbial diversity.

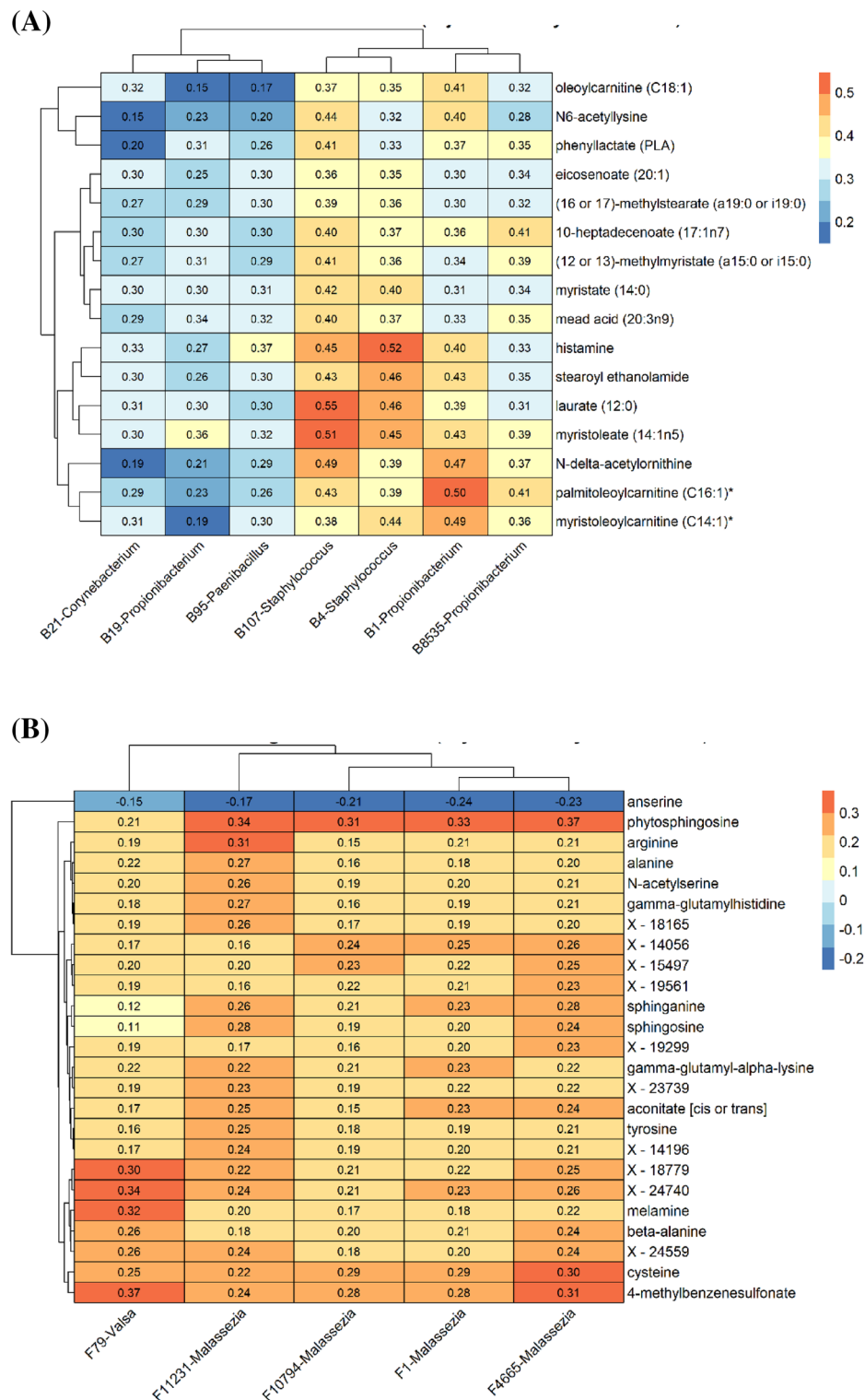
Using the prediction model, metabolome of the skin samples from the two cities accounted for 35% of the variance of bacterial diversity. Using the prediction model, metabolome of the skin samples accounted for only 14% of the variance of fungal skin diversity suggesting a higher contribution of the bacterial diversity to skin metabolites composition compared to fungi. Pearson's correlations highlighted 42 metabolites significantly correlated to bacterial diversity, mainly amino acids ( $n = 19$ ; including kynurenate), peptides ( $n = 10$ ; including gamma-glutamylglutamate) and lipids ( $n = 11$ ; including maleate and oleamide) (Suppl\_Table 3). Two acetylated amino acids, N6-acetyllysine and N-delta-acetylornithine, were identified as the strongest correlated metabolites with a decrease in bacterial Shannon diversity. For fungal diversity, Pearson's correlations were very weak ( $< 0.25$ ), except for three metabolites: trans-4-hydroxyproline; linoleate (18:2n6) and nicotine (Suppl\_Table 3). Our observations indicate a higher contribution of the bacterial diversity to metabolite composition compared to fungi.

To dig deeper into this analysis, we conducted sparse CCA ("Methods" section). As shown in Fig. 5A, mainly commensal bacterial taxa (*Propionibacterium* B1, B19 and B8535; *Staphylococcus* B107 and B4; *Corynebacterium* B21) were retrieved by sCCA with an exception of *Paenibacillus* B52, an environmental bacterium. A positive correlation was observed mainly with lipid metabolites (6/11 metabolites); laurate and myristoleate being the strongest correlated metabolites with *Staphylococcus* B107. Presence of free fatty acids could mirror triglyceride hydrolysis by microbial lipases as reported earlier<sup>37,38</sup>. A second group of metabolites was positively correlated with OTUs, gathering three acyl-carnitine (C18:1; C14:1; C16:1), that are not synthesized by bacteria but can





**Figure 4.** (A) Heatmap of cross-correlation between metabolites and PAHs selected by sparse Canonical Correlation Analysis (sCCA). The values in the heatmap correspond to Pearson's correlation coefficient. (B) Focus on cis- and trans-urocanate metabolites with boxplots of the log-normalized values. We also represent the cis/trans urocanate ratio. A red star indicates that the difference in mean between the 2 cities, Baoding (More Polluted) and Dalian (Less Polluted), is significant ( $q$ -value  $< 0.05$ ,  $t$ -test with correction for multiple testing using Benjamini–Hochberg method). Figures were created using the R software: R Core Team (2017), R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.



**Figure 5.** Heatmaps of the cross-correlation between OTU abundance and metabolites selected by sparse Canonical Correlation Analysis (sCCA), adjusted for city confounder. The values in the heatmap correspond to Pearson's correlation coefficient. **(A)** bacterial OTU. **(B)** fungal OTU. Figures were created using the R software: R Core Team (2017), R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

be a source of carnitine, a known nutrient and osmoprotectant, for bacteria<sup>39</sup>, suggesting that these acylcarnitine/carnitine favor skin commensals on the skin. This results contrasts with the gut metabolome<sup>40</sup>, where acylcarnitines were enriched in the case of gut dysbiosis. Only two amino acids were positively correlated with *Propionibacterium* and *Staphylococcus* OTUs that corresponded to the same metabolites associated to bacterial diversity: N6-acetyllysine and the arginine metabolite N-delta-acetylornithine. Although skin bacteria such as *Staphylococcus* have the capacity to produce these amino acids, to our knowledge there are no existing data supporting their contribution to the skin metabolome content, nor link with skin dysbiosis. Furthermore, two metabolites were positively correlated with bacterial taxa: histamine and phenyllactate<sup>30</sup>. Histamine is a well-known mediator of the allergic reaction<sup>41</sup>. Histamine is not only synthesized in mast cells but is also produced by commensal microorganisms in the gut under physiological conditions<sup>42,43</sup>. We can postulate a similar case on the skin, since *Propionibacterium* and *Staphylococcus* have the capacity to secrete histamine too<sup>44</sup>. PLA is a biopreservative produced by propionic acid bacteria such as *Propionibacterium*, lactic acid bacteria (LAB)<sup>45</sup>, and *Paenibacillus*. Given the symbiotic association of *Paenibacillus* with a stingless bee, associated with antimicrobial activity of PLA, our observation raise the possibility that PLA produced by skin commensals protect the skin from external pathogens<sup>46</sup>. Skin metabolome analysis presented in previous sections showed tryptophan pathway to be modulated in individuals living in polluted city of Baoding (Suppl. Fig. 5). Specifically, we observed reduced indoxyl-3-sulfate, an aryl hydrocarbon receptor (AhR) agonist, in individuals living in more polluted city and is reported to be significantly lower in the skin of AD patients<sup>47</sup>. The present analysis did not show significant correlation between indoxyl-3-sulfate and bacterial taxa. Additional analysis at the functional level using metagenomics is underway to investigate the link between tryptophan metabolism and the microbiome.

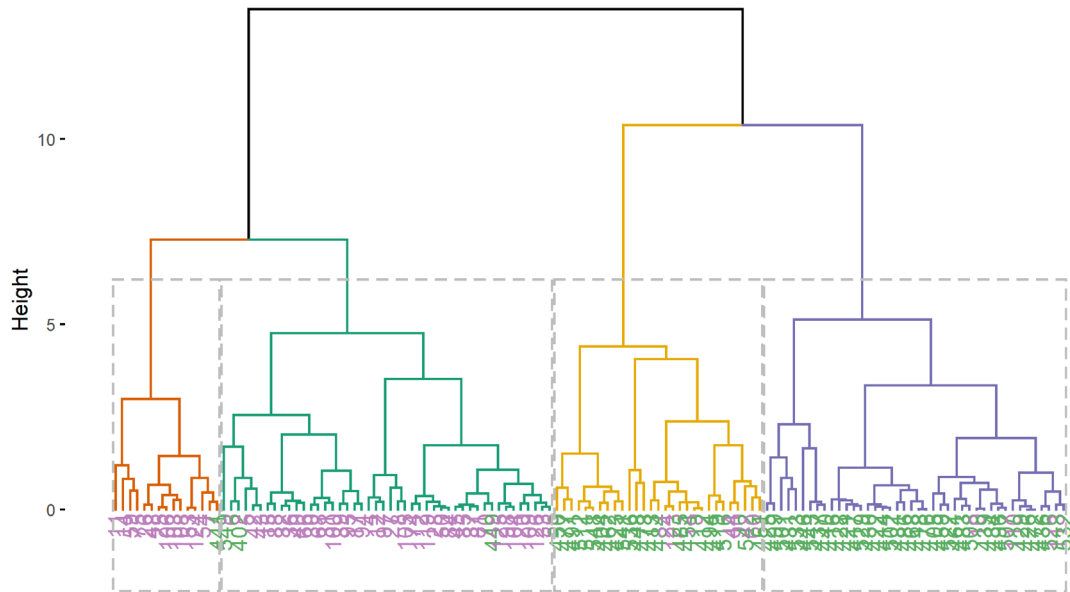
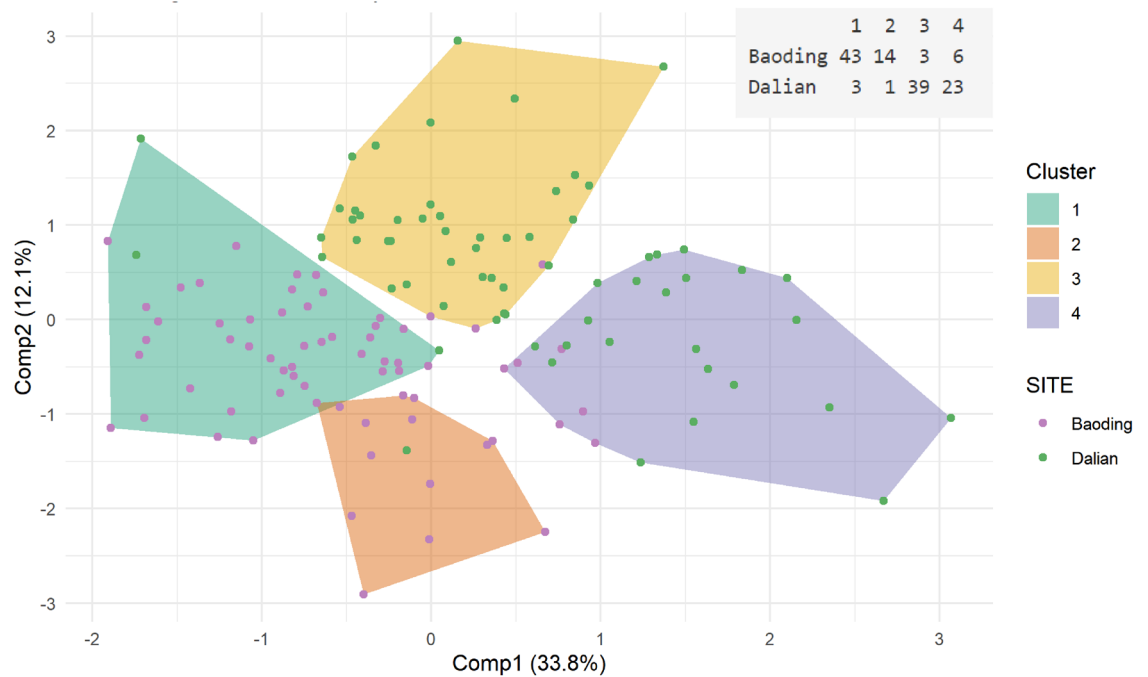
The majority of fungal taxa selected with sCCA corresponded to *Malassezia* OTU (Fig. 5B). They have distinct correlations profiles compared to bacteria with either amino acids/peptides (arginine, alanine, gamma-glutamylhistidine), the long-chain sphingoid bases phytosphingosine, sphinganine and sphingosine and unknown metabolites. Phytosphingosine, sphingosine and sphinganine are main constituents of the stratum corneum ceramides. In addition to their role in barrier function, these molecules are reported to favor the growth of *Malassezia* sp. while inhibiting the growth of opportunistic pathogens such as *Candida albicans*<sup>48</sup>; which could explain the correlation observed in the present study.

Comparing skin microbiota and skin metabolome highlighted a strong link between sebum degradation and bacterial taxa (*Propionibacterium* and *Staphylococcus*), this also revealed a potential link between these taxa and the presence of carnitine, histamine and PLA on skin, which could represent new factors involved in the commensal—host homeostasis. As the correlation between bacterial diversity and metabolites was stronger, we decided to focus on bacterial taxa for further analysis.

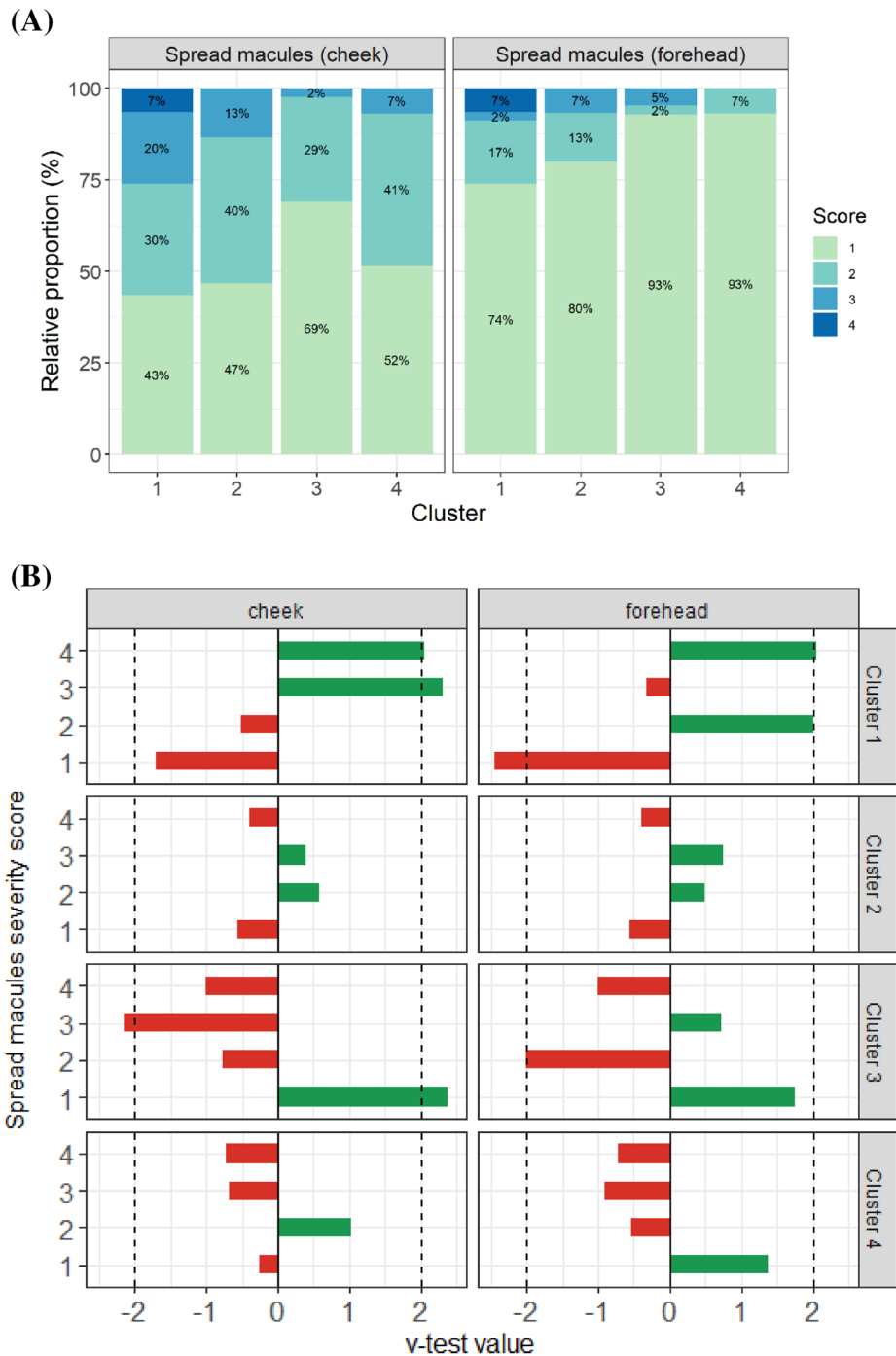
**Multi-omics signature of pollution exposure induced clinical signs.** The results presented so far allowed us to characterize individual parameters such as clinical, metabolites, microbiome and their correlation to pollution exposure. This analysis revealed certain mechanisms linked to barrier function as illustrated by metabolites and proteins and/or sebum degradation from microbiome analysis. To investigate the possible link between these diverse biological parameters, exposure and a visible perceptible clinical skin dysfunction, we conducted a multiblock data analysis (see “Methods” section for details) that concerns the analysis of several sets of variables (blobs) observed on same individual. The need to analyze conjointly the different data sets to discover their main sources of covariation requires the use of very specific computational methods like the one recently proposed by Tenenhaus et al.<sup>49</sup> in the framework of generalized canonical correlation analysis. This approach makes it possible to obtain a unique representation of individuals and descriptors in a consensus space, an essential step in the search for specific profiles/clusters based on classification<sup>49,50</sup>.

By following the methodology described in the method section, the entire population was partitioned into 4 distinct clusters as shown in Fig. 6. Cluster 1 and Cluster 3 are the largest and consist of 93.4% (n = 46) of Baoding and 92.8% (n = 42) of Dalian population respectively. Cluster 3 (n = 15) and Cluster 4 (n = 29) are smaller and have a mixed population. Each cluster has its unique molecular profile that consists of a list of metabolites, microbes and PAH and their derivatives. Each of these clusters was characterized by the variables differentially expressed in the clusters with regard to the whole study population, using a v-test (see “Methods” section for more details). We then obtained a subset of meaningful/relevant clinical parameter(s) for each of the clusters. It was interesting to note that among the 46 individuals of Cluster 1, 26 (57%) and 11 (25%) of them presented with spread macules on cheek and forehead respectively and that this cluster also had the highest scores for the severity for this pigmentary disorder (Fig. 7).

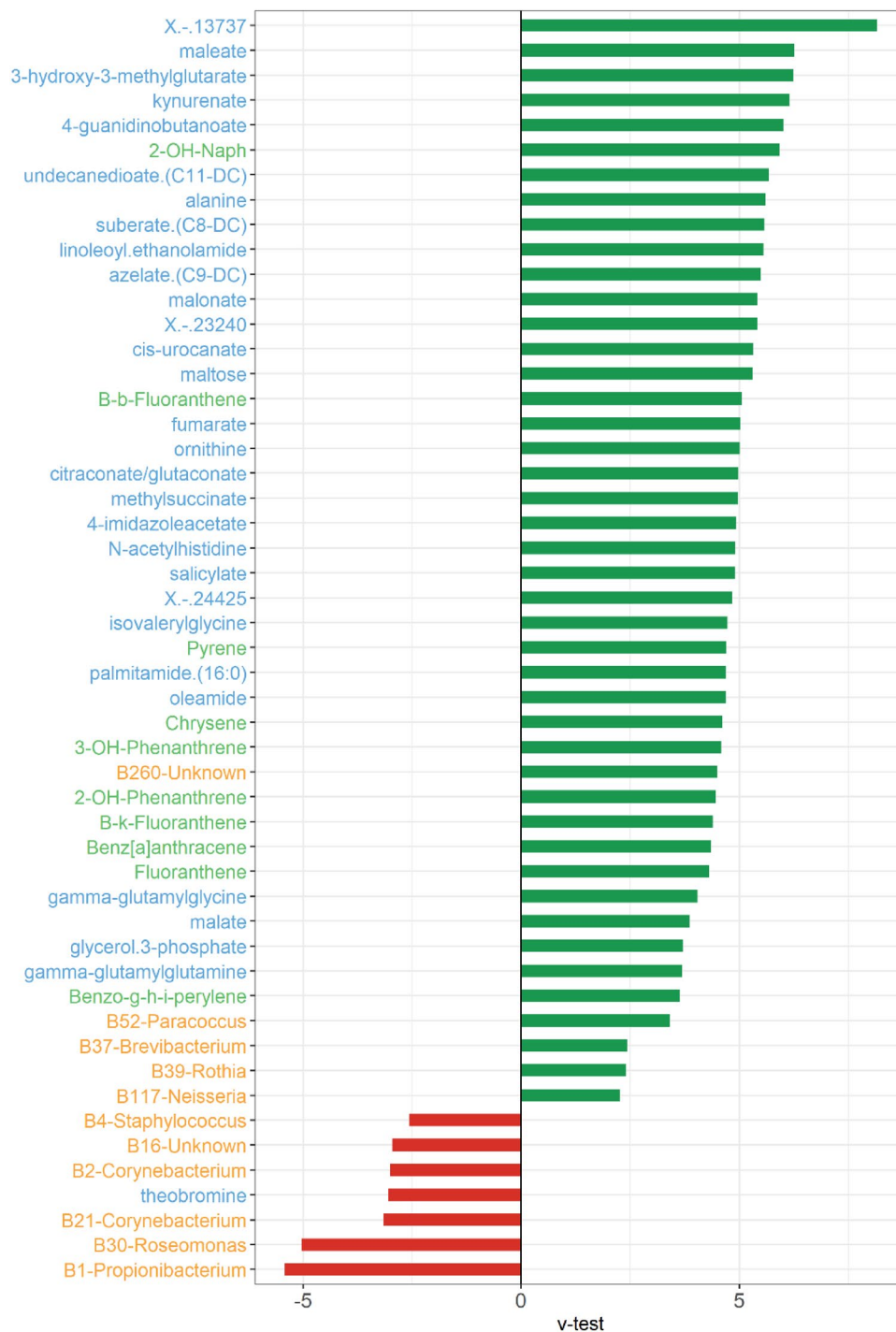
Cluster 2, 3 and 4, on the other hand, did not show spread macule as a dominant condition. In fact, these clusters showed mixed clinical characters (data not shown) and therefore, it was not possible to classify them into a precise clinical parameter category. Figure 8 represents the v-test values of the variables which are significantly modulated (under- or over-expressed) in the Cluster 1 and consists of 30 metabolites, 14 bacteria and 9 PAH and their derivative. We observed a good overlap of metabolites that were described earlier in this manuscript and were the most significantly modulated between polluted and less polluted cities. These include metabolites of tryptophan and histidine pathways (kynurenate, cis-urocanate and 4-imidazoleacetate), peptide (gamma-glutamyl glycine) and fatty acids (oleamide). Cluster 1 was characterized by a negative association with commensal bacteria taxa (*Propionibacterium* B1; *Staphylococcus* B4 and *Corynebacterium* B2 and B21) and positive associations were observed with two genera previously observed on aged skin (*Neisseria* B119; *Rothia* B39)<sup>51</sup> and two genera linked to PAH (*Brevibacterium* B37<sup>14</sup> and *Paracoccus* B52<sup>22</sup>). This analysis has allowed us to come up with a potential molecular ID or a road-map to investigate the association between chronic exposure to pollution and appearance of the pigmentary disorders. It is too early to speculate on the mechanisms and role of different

**(A)** Cluster Dendrogram**(B)**

**Figure 6.** (A) Cluster dendrogram of the hierarchical clustering performed on the consensus space constructed by MAXVAR-A model from the PAH, metabolites and Bacteria OTUs selected by sCCA. Individuals are colored by cities. Based on the height of the gap between 2 consecutive level of the dendrogram, we chose to construct 4 clusters of individuals. (B) Representation of hierarchical clustering on the MAXVAR-A consensus space, with colored convex hull around cluster. The two first component of the consensus space represent 45.9% of the total variability. The table in the upper-right of the figure represents the repartition of individuals between the 2 cities, Baoding (More Polluted) and Dalian (Less Polluted), for each cluster. Figures were created using the R software: R Core Team (2017), R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.



**Figure 7.** (A) Relative proportion of Spread Macules severity score in the consensus clusters. A score of 1 correspond to “no spread macules”, a score of 2 to “weak spread macules”, a score of 3 to “moderate spread macules” and a score of 4 to “severe spread macules”. (B) Characterization of consensus clusters by Spread Macules severity score using v-test statistic. In row we represent the consensus clusters and in column the location of the spread macule score on the face (cheek or forehead). A positive v-test value (in green) indicates that the severity score is in higher proportion in the cluster compare to the global population. A negative v-test value (in red) indicates that the severity score is in higher proportion in the cluster compare to the global population. The v-test value is considered significant if it reaches a threshold of 2 (black dotted vertical line on the plot). Figures were created using the R software: R Core Team (2017), R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.



**Figure 8.** Focus on the characterization of the consensus cluster 1 by PAHs, metabolites and bacteria OTUs using v-test statistic. A positive v-test value (in green) indicates that the variable has mean value in the cluster higher than the global population. A negative v-test value (in red) indicates that the variable has an average lower in the cluster higher than the global population. Only the variables with a significant v-test value are represented. The y-axis text is colored by type of variable (blue for metabolites, green for PAHs and yellow for bacterial OTUs). Figures were created using the R software: R Core Team (2017), R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

entities (metabolites, Microbiome and PAH) that lead to this clinical condition. These are some of the questions that we are asking in our follow up studies through extensive metagenomics analysis and other clinical studies.

In conclusion, our study gives first insight into potential molecular perturbations and phenotype changes of skin due to chronic exposure to PAH, it also provides a roadmap of biological and clinical measurements and computational tools that could be applied to other similar studies including larger cohorts.

## Materials and methods

**Subject recruitment and sample collection.** This multi-parametric clinical study was performed by the Sino-German Cosmetics Institute (Beijing) in agreement with the recommendations of the Declaration of Helsinki and was approved by the local ethics committee (Study n° 2015-033-DY-024, CAIQ Cosmetics Tech Center, Beijing, China, 19.08.2015). Informed written consent was obtained from all participants prior to any study-related procedure. Details of subject recruitment were described extensively elsewhere<sup>13,14,28</sup>. Briefly, study volunteers lived in two Chinese cities selected for being similar as regards UV exposure, geographical location and population lifestyle, but having different pollution levels, measured by monitoring stations as Air Quality Index (Suppl\_Fig. 9). The selection was based on the data collected by the Chinese national air reporting system over several months, that showed different air pollution profiles of the two cities but that are located on comparable latitude and elevation and also have similar climate conditions<sup>53</sup>. Sixty-seven women were recruited in Dalian, the less polluted city, and 67 women in the more polluted city of Baoding. Age of subjects was limited to 25–45 years, to avoid major effects of hormonal variations during either puberty or menopause. Exclusion criteria included pregnancy, use of medications (e.g. antibiotics or antifungals), skin pathologies and current smoking. All participants provided information regarding health status, medical history, and daily habits. Each subject filled a self-administered questionnaire on skin care habits and time spent outdoors. Information on demographics and life-style of individuals is presented in Suppl\_Table 4. Participants also underwent a clinical assessment by dermatologists with skin measurements, including scoring of skin signs and phenotypes. For microbiome analysis, cheek skin sampling was performed in a climate-controlled room at 22 °C and 60% humidity, as described previously<sup>14</sup>. Sterile cotton-tipped dry swabs were rubbed firmly on the cheek for 60 s to cover a surface area of 2 cm<sup>2</sup>. Cotton swabs were then placed into coded microfuge tubes, immediately flash frozen in liquid nitrogen, and stored at –80 °C prior to further analyses. For metabolomics and proteomics D-Squame samples were collected from the cheek area as per manufacturer's instructions. D-SQUAME is the registered name of a circular strip tape used to sample stratum corneum, Produced by CuDerm Corporation (Dallas, TX, U.S.A). For PAH analysis, a sample of hair was cut with stainless scissors from the occipital region of each subject, as described in<sup>13,28</sup>. Each sample comprised only the first 12 cm of hair starting from scalp, corresponding to around a year of hair growth before sampling. Hair samples were transferred into aluminium-foil paper and kept at room temperature until analysis.

**Metabolomic analysis of skin samples.** The untargeted metabolomic analysis was performed at Metabolon, Inc (Morrisville, NC). Method of sample preparation and analysis described in this section was performed at Metabolon, Inc (Morrisville, NC) and has been described elsewhere before<sup>54</sup>. All samples were stored at –80 °C until processed. Sample Preparation: Samples were prepared using the automated MicroLab STAR system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. All methods utilized a Waters ACQUITY (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyser operated at 35,000 mass resolution. The MS analysis alternated between MS and data-dependent MS scans using dynamic exclusion. The scan range varied slightly between methods but covered 70–1000 m/z. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities based on retention time, molecular weight, preferred adducts and in-source fragments, as well as associated MS spectra and curated by visual inspection for quality control using software developed by Metabolon. Proteomic analysis by targeted MRM/SRM: Protein extraction, digestion, and mass spectrometry analyses were performed by the proteomics platform of the CHU de Quebec research center, Quebec, Qc, Canada. The samples (one D-Squame per subject) were resuspended in 550µl of extraction buffer (50 mM ammonium bicarbonate (ABC), 0.5% sodium deoxycholate, 50 mM DTT, protease inhibitor cocktail and 1µM pepstatin). After successive filtration and centrifugation, pellets were solubilized in 50µL of a 50 mM ABC/1% DOC solution for protein quantification by Bradford protein assay.

**Tryptic digestion:** Approximately 10µg of each sample was heated at 95 °C for 5 min and proteins were reduced with 0.2 mM DTT at 37 °C 30 min and alkylated with 0.8 mM IAA (iodoacetamide) for 30 min at RT in the dark. The proteins were then digested with trypsin (1 µg) and incubated at 37 °C overnight.

**SRM optimisation:** Three peptides/protein of interest were selected by the three most intense transitions with an optimal collision energy. Selected peptides for SRM assay were synthesized with a heavy isotope on the C-ter amino acid ([13C6]-Lys and [13C6]-Arg) by ThermoFisher Scientific (Germany). Just before injection, 2.8 µL of sample are mixed with 2.8 µL of the stock heavy labeled synthetic peptide standard solution.

1 µg of each sample were analyzed on an Eksigent NanoLC 400 chromatography system (Sciex, Concord, Ontario) coupled online to a 6500QTRAPTM (ABSciex, Concord, Ontario) mass spectrometer with a nanospray

ion source. The peptides were eluted with a linear gradient from 5 to 40% solvent B (A: 0.1% FA, B: ACN, 0.1% FA) in 30 min then 40 to 95% B in 10 min at 300 nL/min. Samples were injected randomly. Raw files were imported in Skyline v3.6 software for peak integration.

Results from Skyline were treated with excel, briefly the two more intense transitions for two peptides of each protein were used for quantification. The peak areas were normalized based on the heavy-labeled synthetic peptides. A normalisation factor was calculated and applied to the endogenous peptides. For each group, the mean of the sum of the normalized areas were calculated for each condition, then a ratio and a t-test were applied.

**Analysis of PAH and their metabolites in hair samples.** Details on the methodology used for the analysis of hair PAH and metabolites have been previously reported<sup>13</sup>, here we describe briefly only the main steps. First, hair samples were washed and decontaminated, to remove any compounds on the hair surface without removing those incorporated in the bulk matrix via biological mechanisms and representative of the dose present in body. Then, hair samples were pulverized, hydrolyzed, extracted and analyzed according to well-established methods based on gas and liquid chromatography coupled with tandem mass spectrometry (GC-MS/MS and LC-MS/MS). PAH, PAH metabolites and nicotine/cotinine were analyzed separately, and quantified against stable isotope labeled analogue. The method allowed for the analysis of 15 parent PAH (based on US-EPA priority list), 56 PAH monohydroxy-metabolites (all the commercially available standard), nicotine and cotinine. Limits of detection were also assessed.

**Analysis of skin microbiota.** Detailed methodology for the analysis of skin microbiome has been reported earlier<sup>14</sup>. Briefly, gDNA was extracted from skin samples using the PowerSoil DNA isolation kit (MO BIO Laboratories, Carlsbad, CA, USA) following the manufacturer's instruction with modifications as described previously, and each gDNA sample underwent triplicate PCR by primers targeting the bacterial 16S rRNA gene V1-3 region and the fungal ITS1 region. Following amplicon purification and indexing -PCR the library was prepared, and bacterial and fungal paired -end sequencing was performed on Illumina Miseq platform by Seq-Matic LLC (Fremont, CA, USA). 16S rRNA gene and ITS sequences underwent processing and bioinformatics analysis using USEARCH (v9.2.64) and QIIME (v1.9). Bacterial OTUs was obtained for the representative sequences after clustering at 97% sequence identity using the USEARCH algorithm within USEARCH, against the SILVA database (128 release) and fungal OUT were interrogated against previously curated fungal database designed for skin microbiome analysis<sup>55</sup>. Following quality controls and after removing undesirable reads, a total of 9,656,916 and 14,649,172 bacterial and fungal reads were retained, respectively. Within-sample Shannon diversity (or alpha diversity) was estimated using the breakaway (v4.0) package in R v3.5.1 as reported earlier<sup>14</sup>. Correlations between metabolites and bacterial and fungal taxa were subsequently identified using a sCCA approach.

**Statistical methods.** The article is structured into several parts, each with specific statistical approaches which are described below.

**Association of clinical signs with city and PAH:** Descriptive statistics for clinical signs have been tabulated by city and PAH groups. Clinical scores between cities or PHA groups were compared using a Wilcoxon test. The PAH score is the best univariate summary of all 39 PAHs measurements. It is the first principal component of a principal component analysis (PCA) based on the log-normalized PAH measurements.

The study being observational, *p* values must be interpreted more in an exploratory than confirmatory way, associations being potentially associated with confounding factors.

**Metabolites and City:** A pathway enrichment analysis has been performed. Metabolite pathways that are differentially modulated in skin samples of individuals from polluted versus less polluted cities are displayed using both heatmap of  $-\log_{10}$  *q*-values (correction of *p* values using Benjamini-Hochberg multiple testing adjustment method) and boxplots by city of selected log normalized descriptors.

In order to evaluate the differences between the metabolomics profile between the two cities, irrespective of classification into sub or super pathway, a cumulative distribution of Log<sub>2</sub>-Fold change was calculated with the most polluted city, Baoding, as the reference population (Suppl\_Fig. 3). The Fold changes have been calculated after removing the outliers having a value superior or equal to 3 times the interquartile range. Significance of these fold changes was tested using a t-test and the results are represented using a volcano plot which represents  $\log_2(\text{FC})$  in x-axis and  $-\log_{10}(\text{q-value})$  in y-axis. This bivariate representation allows to highlight the most relevant metabolites, i.e. those having a *q*-value < 0.05 and a  $|\log_2(\text{FC})| > 0.6$ .

A complementary multivariate approach using a random Forest classification model was also used to confirm the robustness of the previous results (Suppl\_Fig. 4A). The random forest allows to model the city membership as a non-linear multivariate function of the metabolites and to extract variable importance, i.e. the metabolites which contribute the most to the predictive accuracy of the model. We use the *rfsrc* function from the Random Forest SRC R package to perform the random forest classification model, more details on the model are given in Suppl\_Fig. 4B. We also evaluate the predictive performance of the model on the Out-Of-Bag (OOB) samples, the accuracy results are given in Suppl\_Fig. 4C. To compute the variable importance (VIMP) values we use a prediction error approach involving "noising-up" each variable in turn. VIMP for a variable  $X_j$  is the difference between prediction error when  $X_j$  is noised up by randomly permuting its values, compared to prediction error under the observed values<sup>21</sup>. Since VIMP is the difference between OOB prediction error before and after permutation, a large VIMP value indicates that misspecification detracts from the variable predictive accuracy in the forest. VIMP close to zero indicates the variable contributes nothing to predictive accuracy, and negative values indicate the predictive accuracy improves when the variable is misspecified. For the sake of clarity, we only represent the 20 metabolites with highest VIMP values in Suppl\_Fig. 4A.



**Metabolites and PAH:** To make a direct comparison between exposure and its molecular impact on skin, a sparse Canonical Correlation Analysis (sCCA) has been carried out to identify sets of PAHs and metabolites that covary together. sCCA is a regularized version of canonical correlation analysis (CCA) used to study the relationship between two datasets while selecting only significant correlations<sup>17</sup>. If we have two datasets X and Y of random variables, and there are correlations among the variables, then the sCCA will find sparse linear combinations (components with null weights for non-relevant variable) of X and Y which have maximum correlation with each other. The sparsity parameters of sCCA were estimated using a permutation scheme (nperms = 500) with the MultiCCA.permute function from the PMA R package<sup>17,18</sup>. The permutation procedure gave a *p* value of 0.004 associated with a correlation coefficient of 0.539 between the 1st sparse components of the two datasets. The 1st sparse components have 20 and 7 variables with non-null weights for metabolomics and PAH dataset respectively. The cross-correlation between the selected metabolites and PAHs was visualized using a heatmap representation.

**Metabolites and skin microbiome:** Correlation between the Shannon index of alpha bacterial or fungi diversity and the metabolites were estimated using Pearson correlation coefficient and ranked according to significance.

Prediction of Shannon index for bacterial or fungi diversity using the metabolites as independent variables was performed using a partial least squares regression with SIMCA software version 16.0, Umetrics, Umea, Sweden. sCCA was then performed to explore global associations between relative abundances of bacterial and fungal OTUs and metabolites (Fig. 5A,B). The sparsity parameters of sCCA were estimated using a permutation scheme (nperms = 500) with the MultiCCA.permute function from the PMA R package. The permutation procedure of the sCCA between the bacterial OTUs and the metabolites gave a *p* value < 0.001 associated with a correlation coefficient of 0.618 between the 1st sparse components of the two datasets. The 1st sparse components have 16 and 7 variables with non-null weights for the metabolomics and bacterial OTU datasets respectively. The permutation procedure of the sCCA between the fungal OTUs and the metabolites gave a *p* value of 0.206 associated with a correlation coefficient of 0.436 between the 1st sparse components of the two datasets. The 1st sparse components have 25 and 5 variables with non-null weights for the metabolomics and fungal OTU datasets respectively. The cross-correlation between the selected metabolites and OTUs was visualized using a heatmap representation.

**Multiblock analysis:** The need to analyze conjointly the different data sets to discover their main sources of covariation requires to generalize the sCCA approach to more than two groups. In that context, we use a sparse version of Regularized Generalized Canonical Correlation Analysis [ref] called Sparse Generalized Canonical Correlation Analysis (SGCCA) [ref]. S/RGCCA is a general component-based framework for a component-based framework for the integrative exploration of multimodal and high-dimensional data sets. S/RGCCA subsumes as special cases many multiblock component methods as special cases including MAXVAR-A (see Tenenhaus et al.<sup>49</sup> for details). MAXVAR-A allows visualizing in a single space (the so-called consensus space), the relationships between variables belonging to the different blocks. We thus combine the metabolomics, PAH and bacterial OTU datasets using MAXVAR-A. Based on the 2 first components of the consensus space, we performed a hierarchical clustering to segment the population into 4 clusters (see Fig. 6A,B). The choice of 4 clusters was based on the visual evaluation of the dendrogram in Fig. 6A, more specifically by looking at the height of the gap between 2 consecutive level of the hierarchy. Each cluster was then characterized by clinical signs, metabolites, PAHs and bacterial OTUs using a *v*-test<sup>20</sup>, a procedure available from the catdes function of the Factominer R package. The *v*-test values allow determining if a variable is significantly over-represented or under-represented in the subgroup compared to the total population. For a continuous variable we test if the mean in a particular subgroup is different for the mean of the total population. For discrete/qualitative variables, we test if the proportion of a modality in a particular subgroup is over-expressed or under-expressed in the subgroup compared to the whole population. The results of *v*-test are shown in Fig. 7B for the spread macules clinical sign and in Fig. 8 for metabolites, PAHs and bacterial OTUs which are significantly modulated in cluster 1.

Received: 24 March 2021; Accepted: 3 August 2021

Published online: 15 September 2021

## References

- Gonzales, M. C., Yu, P. & Shiao, S. P. MTHFR gene polymorphism-mutations and air pollution as risk factors for breast cancer: A metaprediction study. *Nurs. Res.* **66**(2), 152–163 (2017).
- Heroux, M. E. et al. Quantifying the health impacts of ambient air pollutants: Recommendations of a WHO/Europe project. *Int. J. Public Health* **60**(5), 619–627 (2015).
- Masiol, M., Agostinelli, C., Formenton, G., Tarabotti, E. & Pavoni, B. Thirteen years of air pollution hourly monitoring in a large city: Potential sources, trends, cycles and effects of car-free days. *Sci. Total Environ.* **494–495**, 84–96 (2014).
- Araviiskaia, E. et al. The impact of airborne pollution on skin. *J. Eur. Acad. Dermatol. Venerol.* **33**(8), 1496–1505 (2019).
- Schikowski, T. & Huls, A. Air pollution and skin aging. *Curr. Environ. Health Rep.* **7**(1), 58–64 (2020).
- Rajagopalan, P. et al. How does chronic cigarette smoke exposure affect human skin? A global proteomics study in primary human keratinocytes. *OMICS* **20**(11), 615–626 (2016).
- Vattanasit, U. et al. Oxidative DNA damage and inflammatory responses in cultured human cells and in humans exposed to traffic-related particles. *Int. J. Hyg. Environ. Health* **217**(1), 23–33 (2014).
- Vierkotter, A. et al. Airborne particle exposure and extrinsic skin aging. *J. Invest. Dermatol.* **130**(12), 2719–2726 (2010).
- Krutmann, J. Premature skin aging by ultraviolet radiation and other environmental hazards. The molecular basis. *Hautarzt* **54**(9), 809–817 (2003).
- Marrot, L. Pollution and sun exposure: A deleterious synergy. Mechanisms and opportunities for skin protection. *Curr. Med. Chem.* **25**(40), 5469–5486 (2018).
- Lefebvre, M. A. et al. Consequences of urban pollution upon skin status. A controlled study in Shanghai area. *Int. J. Cosmet. Sci.* **38**(3), 217–223 (2016).

12. Lefebvre, M. A. *et al.* Evaluation of the impact of urban pollution on the quality of skin: A multicentre study in Mexico. *Int. J. Cosmet. Sci.* **37**(3), 329–338 (2015).
13. Palazzi, P. *et al.* Exposure to polycyclic aromatic hydrocarbons in women living in the Chinese cities of BaoDing and Dalian revealed by hair analysis. *Environ. Int.* **121**(Pt 2), 1341–1354 (2018).
14. Leung, M. H. Y. *et al.* Changes of the human skin microbiota upon chronic exposure to polycyclic aromatic hydrocarbon pollutants. *Microbiome* **8**(1), 100 (2020).
15. KrupaShankar, D. S. *et al.* A cross-sectional, multicentric clinico-epidemiological study of melasma in India. *Dermatol. Ther. (Heidelb.)* **4**(1), 71–81 (2014).
16. Hourblin, V., Nouveau, S., Roy, N. & de Lacharriere, O. Skin complexion and pigmentary disorders in facial skin of 1204 women in 4 Indian cities. *Indian J. Dermatol. Venereol. Leprol.* **80**(5), 395–401 (2014).
17. Whitfield, J. B. Gamma glutamyl transferase. *Crit. Rev. Clin. Lab. Sci.* **38**(4), 263–355 (2001).
18. Rawlings, A. V. & Harding, C. R. Moisturization and skin barrier function. *Dermatol. Ther.* **17**(Suppl 1), 43–48 (2004).
19. Rajagopalan, P. *et al.* Proteome-wide changes in primary skin keratinocytes exposed to diesel particulate extract—A role for antioxidants in skin health. *J. Dermatol. Sci.* **91**(3), 239–249 (2018).
20. Biro, T., Toth, B. I., Hasko, G., Paus, R. & Pacher, P. The endocannabinoid system of the skin in health and disease: Novel perspectives and therapeutic opportunities. *Trends Pharmacol. Sci.* **30**(8), 411–420 (2009).
21. Kim, K. E., Cho, D. & Park, H. J. Air pollution and skin diseases: Adverse effects of airborne particulate matter on various skin diseases. *Life Sci.* **152**, 126–134 (2016).
22. Mancebo, S. E. & Wang, S. Q. Recognizing the impact of ambient air pollution on skin health. *J. Eur. Acad. Dermatol. Venereol.* **29**(12), 2326–2332 (2015).
23. Stone, T. W., Stoy, N. & Darlington, L. G. An expanding range of targets for kynurenine metabolites of tryptophan. *Trends Pharmacol. Sci.* **34**(2), 136–143 (2013).
24. Hubbard, T. D. *et al.* Adaptation of the human aryl hydrocarbon receptor to sense microbiota-derived indoles. *Sci. Rep.* **5**, 12689 (2015).
25. Tenenhaus, A., & Tenenhaus, M. Regularized generalized canonical correlation analysis. *Psychometrika* **76**(2), 257 (2011).
26. Appenzeller, B. M. R. *et al.* Hair analysis for the biomonitoring of pesticide exposure—Comparison with blood and urine in a rat model. *Arch. Toxicol.* **91**(8), 2813–2825 (2017).
27. Grova, N., Hardy, E. M., Fays, F., Duca, R. C. & Appenzeller, B. M. R. Hair analysis for the biomonitoring of Polycyclic Aromatic Hydrocarbon exposure: Comparison with urinary metabolites and DNA-adducts in a rat model. *Arch. Toxicol.* (2018) (**in press**).
28. Naudin, G. *et al.* Human pollution exposure correlates with accelerated ultrastructural degradation of hair fibers. *Proc. Natl. Acad. Sci. U. S. A.* **116**(37), 18410–18415 (2019).
29. Aksnes, H., Ree, R. & Arnesen, T. Co-translational, post-translational, and non-catalytic roles of N-terminal acetyltransferases. *Mol. Cell* **73**(6), 1097–1114 (2019).
30. Leclercq, B. *et al.* Genetic and epigenetic alterations in normal and sensitive COPD-diseased human bronchial epithelial cells repeatedly exposed to air pollution-derived PM<sub>2.5</sub>. *Environ. Pollut.* **230**, 163–177 (2017).
31. LeMasters, G. K. *et al.* N-acetyltransferase 1 polymorphism increases cotinine levels in Caucasian children exposed to secondhand smoke: The CCAAPS birth cohort. *Pharmacogenom. J.* **15**(2), 189–195 (2015).
32. Hagedorn, H. W. *et al.* Urinary excretion of phenolic polycyclic aromatic hydrocarbons (OH-PAH) in nonsmokers and in smokers of cigarettes with different ISO tar yields. *J. Anal. Toxicol.* **33**(6), 301–309 (2009).
33. Smith, M. R. *et al.* Benzo[a]pyrene perturbs mitochondrial and amino acid metabolism in lung epithelial cells and has similar correlations with metabolic changes in human serum. *J. Occup. Environ. Med.* **61**(Suppl 12), S73–S81 (2019).
34. Zierer, J. *et al.* The fecal metabolome as a functional readout of the gut microbiome. *Nat. Genet.* **50**(6), 790–795 (2018).
35. Bouslimani, A. *et al.* The impact of skin care products on skin chemistry and microbiome dynamics. *BMC Biol.* **17**(1), 47 (2019).
36. McCall, L. I. *et al.* Home chemical and microbial transitions across urbanization. *Nat. Microbiol.* **5**(1), 108–115 (2020).
37. Bouslimani, A. *et al.* Molecular cartography of the human skin surface in 3D. *Proc. Natl. Acad. Sci. U. S. A.* **112**(17), E2120–2129 (2015).
38. Xie, W., Khosasih, V., Suwanto, A. & Kim, H. K. Characterization of lipases from *Staphylococcus aureus* and *Staphylococcus epidermidis* isolated from human facial sebaceous skin. *J. Microbiol. Biotechnol.* **22**(1), 84–91 (2012).
39. Meadows, J. A. & Wargo, M. J. Carnitine in bacterial physiology and metabolism. *Microbiology (Reading)* **161**(6), 1161–1174 (2015).
40. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**(7758), 655–662 (2019).
41. Thurmond, R. L., Gelfand, E. W. & Dunford, P. J. The role of histamine H1 and H4 receptors in allergic inflammation: The search for new antihistamines. *Nat. Rev. Drug Discov.* **7**(1), 41–53 (2008).
42. Barcik, W. *et al.* Histamine-secreting microbes are increased in the gut of adult asthma patients. *J. Allergy Clin. Immunol.* **138**(5), 1491–1494 (2016).
43. Sudo, N. Biogenic amines: Signals between commensal microbiota and gut physiology. *Front. Endocrinol. (Lausanne)* **10**, 504 (2019).
44. Allaker, R. P., Greenman, J. & Osborne, R. H. The production of inflammatory compounds by *Propionibacterium acnes* and other skin organisms. *Br. J. Dermatol.* **117**(2), 175–183 (1987).
45. Zhang, X., Zhang, S., Shi, Y., Shen, F. & Wang, H. A new high phenyl lactic acid-yielding *Lactobacillus plantarum* IMAU10124 and a comparative analysis of lactate dehydrogenase gene. *FEMS Microbiol. Lett.* **356**(1), 89–96 (2014).
46. Menegatti, C. *et al.* *Paenibacillus polymyxa* associated with the stingless bee *Melipona scutellaris* produces antimicrobial compounds against entomopathogens. *J. Chem. Ecol.* **44**(12), 1158–1169 (2018).
47. Yu J. *et al.* A tryptophan metabolite of the skin microbiota attenuates inflammation in patients with atopic dermatitis through the aryl hydrocarbon receptor. *J. Allergy Clin. Immunol.* **143**(6), 2108–2119.e12. <https://doi.org/10.1016/j.jaci.2018.11.036> (2019).
48. Nenoff, P. & Haustein, U. F. In vitro activity of phytosphingosines against *Malassezia furfur* and *Candida albicans*. *Acta Derm. Venereol.* **82**(3), 170–173 (2002).
49. Tenenhaus A. *et al.* Variable Selection for Generalized Canonical Correlation Analysis. *Biostatistics* **15**(3), 569–583 (2014).
50. Garali, I. *et al.* A strategy for multimodal data integration: Application to biomarkers identification in spinocerebellar ataxia. *Brief. Bioinform.* **19**(6), 1356–1369 (2018).
51. Shibagaki, N. *et al.* Aging-related changes in the diversity of women's skin microbiomes associated with oral bacteria. *Sci. Rep.* **7**(1), 10567 (2017).
52. Sowada, J. *et al.* Toxication of polycyclic aromatic hydrocarbons by commensal bacteria from human skin. *Arch. Toxicol.* **91**(6), 2331–2341 (2017).
53. Rohde, R. A. & Muller, R. A. Air pollution in China: Mapping of concentrations and sources. *PLoS ONE* **10**(8), e0135749 (2015).
54. Lind, L. *et al.* The plasma metabolomic profile is differently associated with liver fat, visceral adipose tissue, and pancreatic fat. *J. Clin. Endocrinol. Metab.* **106**(1), e118–e129 (2021).
55. Findley, K. *et al.* Topographic diversity of fungal and bacterial communities in human skin. *Nature* **498**(7454), 367–370 (2013).

## Acknowledgements

Eric Arbey, L'Oréal Research and Innovation, France and Brian Keppler, Metabolon Inc, USA are acknowledged for additional analysis to identify the unknown metabolites mentioned in this study.

## Author contributions

N.M., N.C. and L.A. conceived and designed the study. N.B. and S.N. designed the clinical protocol and executed the study. P.P., B.M.R.A. and S.M. conducted the experiments to measure P.A.H. quantity in hair fibre. N.M., C.C. conducted metabolomics and proteomics analysis. M.H.Y.L., P.K.H.L. and C.C. conducted microbiome and metagenomics analysis. A.T., P.B. and F.G. developed computational strategy for this study and performed statistical analysis on data. N.M., N.C., C.C., F.G. and P.B. wrote the manuscript and all authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97572-1>.

**Correspondence** and requests for materials should be addressed to N.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021