PNAS Nexus, 2025, 4, pgaf127
https://doi.org/10.1093/pnasnexus/pgaf127
Advance access publication 13 May 2025
Research Report

# 5G-exposed human skin cells do not respond with altered gene expression and methylation profiles

Jyoti Jyoti [a,1], Isabel Gronau [a,1], Eda Cakir [a,1], Marc-Thorsten Hütt [a], Alexander Lerchl [a] and Vivian Meyer [a,b,*]

[a]School of Science, Constructor University, Bremen 28759, Germany
[b]Department of Biology and Environmental Sciences, Carl von Ossietzky Universität Oldenburg, Oldenburg 26129, Germany
*To whom correspondence should be addressed: Email: vivian.meyer2@uol.de
[1]J.J., I.G., and E.C. contributed equally to this work.
**Edited By** Shibu Yooseph

## Abstract

Due to the ever-increasing wirelessly transmitted data, the development of new transmission standards and higher frequencies in the 5G band is required. Despite basic biophysical considerations that argue against health effects, there is public concern about this technology. Because the skin penetration depth at these frequencies is only 1 mm or less, we exposed fibroblasts and keratinocytes to electromagnetic fields up to ten times the permissible limits, for 2 and 48 h in a fully blinded experimental design. Sham-exposed cells served as negative, and UV-exposed cells as positive controls. Differences in gene expression and methylation due to exposure were small and not higher than expected by chance. These data strongly support the assessment that there is no evidence for exposure-induced damage to human skin cells.

**Keywords:** 5G radiofrequency, electromagnetic fields, in vitro, systems biology

## Significance Statement

Widespread adoption of 5G wireless technology and the imminent introduction of higher frequencies have intensified public concerns regarding its potential health effects, in particular cancer risk due to (epi-)genetic alterations. Previous studies have faced criticism for methodological shortcomings, including lack of blinding, temperature control, and transparent statistical methods. Our results show with great clarity that in human skin cells, even under worst-case conditions, no significant changes in gene expression or methylation patterns are observed after exposure. These results will contribute to counteracting the uncertainties with well-founded facts. Beyond this, the statistical toolbox designed here can be applied to a range of other biological and medical scenarios where the absence of an effect needs to be confirmed.

## Introduction

The 5G standard in mobile communications technology is being introduced because the amount of transmitted data is increasing. Currently, the roll-out of 5G takes place and frequencies of the 5G New Radio Frequency Range 1 (5G NR FR1) below 6 GHz are used. In the upcoming years, 5G NR FR2 will add on with frequencies of 24.3–27.5 GHz and 39.5–43.3 GHz. The biological effects of electromagnetic fields (EMF) are strongly dependent on the frequency. In the range up to 3 GHz the penetration depths into the skin are about 10 mm, in the range from 10 GHz it is 1 mm and less (1). The safety reference level for the general population in the range 2–300 GHz is 10 $W/m^2$ (1 $mW/cm^2$), and for occupationally exposed persons, 50 $W/m^2$ (5 $mW/cm^2$) (2).

Although there are scientific studies on biological effects for frequencies above 6 GHz, they are heterogeneous in terms of power flux density, the biological model, the frequencies and the biological endpoints (3, 4). Since the absorption of electromagnetic energy takes place in the upper layers of the skin, thermal effects are particularly critical. At these frequencies and with very high power flux densities far beyond the limits, therefore, pain and also burns may occur. Thermal effects also include those associated with protein denaturation and can cause gene or chromosome damage (5–7). Overall, the study results so far do not indicate any nonthermal damage caused by exposure to radiofrequency EMF in the frequency range of 3–100 GHz. The few exceptions point to experiments in which the methods show numerous shortcomings.

Here, we present the results of a strictly blinded, temperature-controlled transcriptomics and methylation study in human keratinocytes and human dermal fibroblasts exposed to 5G EMF at different frequencies (27 GHz and 40.5 GHz), power flux densities (1 $mW/cm^2$ and 10 $mW/cm^2$), and exposure times (2 h and 48 h). To our knowledge, this is the first study that uses state-of-the-art

**Competing Interest:** The authors declare no competing interests.
**Received:** July 4, 2024. **Accepted:** March 31, 2025

methods like whole-genome RNA-Seq and methylation array to analyze the genetic and epigenetic effects of 5G NR FR2 frequencies on human material. These methods are complemented by a newly developed combinatorial technique to examine if the signals were higher than expected by chance. This may serve as a future standard for lack-of-signal confirmation in "omics" data.

## Experimental pipeline

The experiments were conducted in an exposure facility, which was characterized for cell monolayer dosimetry, and allows for blinded exposure and temperature monitoring (Fig. 1A, (8)). Due to the heterogeneities in previous studies (3, 4), we devised a comprehensive experimental setup and design (Fig. 1B and C) that includes variations in cell type, power flux density, frequency, and time of exposure to 5G EMF under compensation of temperature increase. There are total 96 samples, with 3 technical replicates per experiment for exposed as well as sham-exposed cells ($2^5 \times 3$). Additionally, 24 samples (12 HaCat, 12 HDF) were considered for UV, and 5G exposure without temperature compensation treatment, respectively. Differential RNA expression (DE) analysis and differential DNA methylation (DM) analysis were performed on exposure category variable (exposed/sham) for each treatment combination. Both analyses followed the standard pipeline (Fig. 1D), i.e. quality control, data normalization, batch detection and correction, hypothesis testing to find differentially expressed genes (DEGs), and differentially methylated probes (DMPs). Then, we performed a network analysis (Fig. 1E) by mapping the resulting genes on two gene-centric metabolic networks (Recon2 (9) and Recon3D (10)) and two gene level protein–protein interaction networks (String (11) and Biogrid (12)). The connectivity of the subnetworks generated by DEGs and DMPs was compared to the

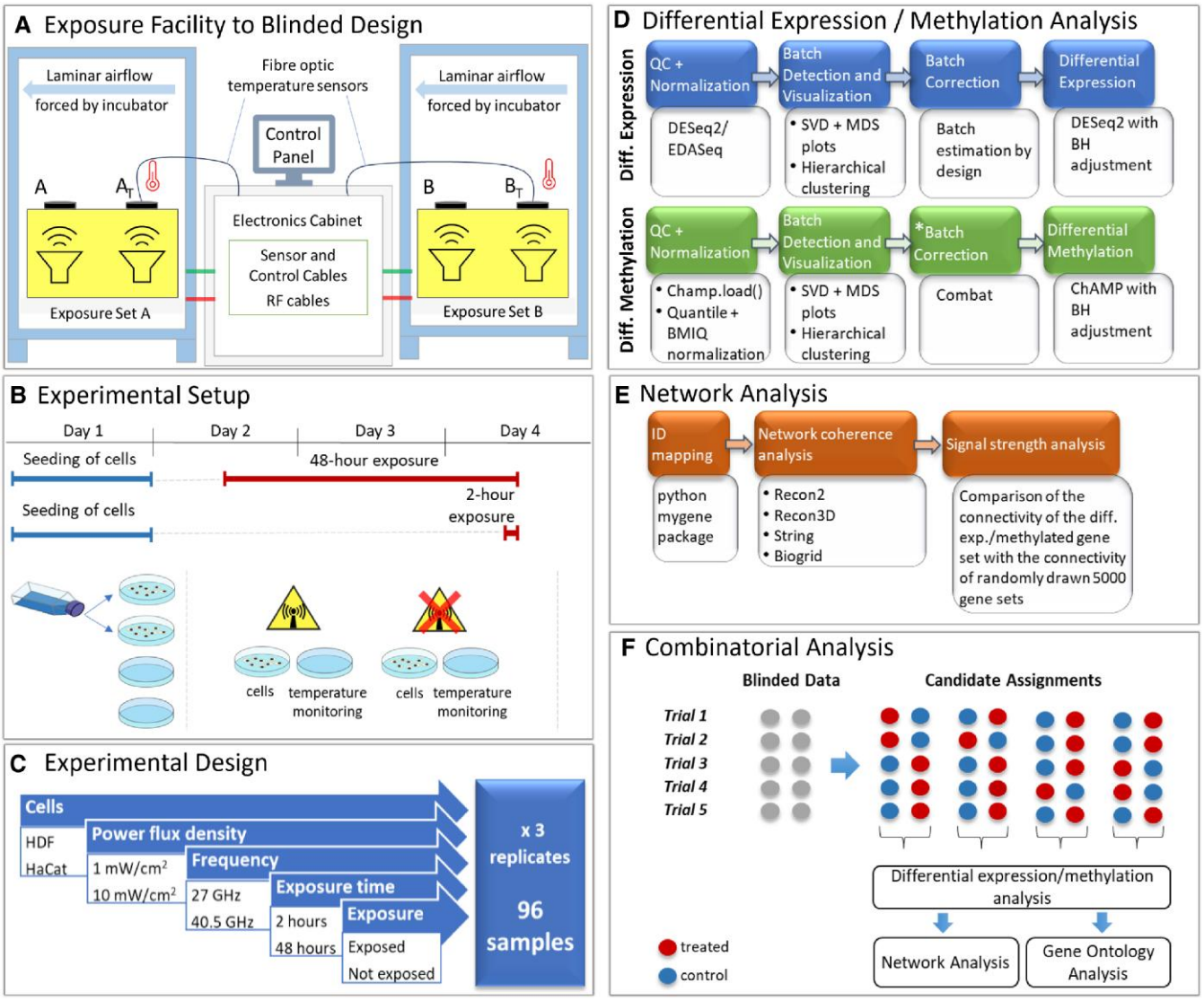


**Fig. 1.** Overview of the experiment. A) Exposure facility to blinded design: Our novel exposure facility was characterized for cell monolayer dosimetry in 60 mm petri dishes and allows for standard in vitro incubation (37 °C, 5% $CO_2$) during exposure; redrawn from (8). It enables a randomized and blinded application of parallel sham exposure and exposure with temperature monitoring during experiments using fiber optic temperature probes. B) Experimental setup: One day (48 h experiments) or 3 days (2 h experiments) after seeding of the cells, experiments were started using one petri dish with cells and one with medium only to monitor the temperature in each incubator. C) Experimental design: Our study on transcriptomics and methylation under temperature-controlled conditions involves the investigation of two cell types exposed to 5G electromagnetic fields at different frequencies, power flux densities and exposure times. D) The summary of the pipeline used for conducting differential expression and differential methylation analysis; *in supplementary material. E) The summary of the pipeline employed for conducting network analysis. F) Combinatorial analysis: The method aims to detect the real query group separation (sham/exposed) of samples from other randomly generated sample groups.

coherence of 5,000 randomly drawn gene sets to evaluate the significance of the network signal. Gene ontology (GO) enrichment analysis was performed to evaluate the functional importance of the cluster of genes that exhibited differential expression or methylation. The evaluation of the analysis results was accomplished through a comparison of the quantity of significant GO terms that were identified. Finally, we performed a combinatorial analysis (Fig. 1F), i.e. under blinded conditions, we executed the four analyses (DE, DM, network, and GO analysis) for all possible query group separation of samples (called combinations). We hypothesized retrieval of a significant signal when the strongest signal out of all combinations is the original query group separation, given the original separation indeed has significant differences in expression.

## Results

The results of the various analyses, i.e. differential gene expression (DE), DM, network coherence, and GO enrichment analysis, differed only slightly. Gene expression and DNA methylation remained almost unchanged after exposure. Overall, the results of the different power flux densities, frequencies and exposure times, were very similar. Therefore, we here show a subset of the results, which is complemented by the supplementary material.

### RNA-Seq analysis

RNA-Seq analysis of HaCat cells under 5G exposure with a power flux density of 10 mW/cm$^2$ showed no clear clusters based on the treatment category (exposed/sham) in a multidimensional scaling (MDS) plot (Fig. 2A). Although some clusters based on the frequency and time of exposure can be seen, these clusters mainly reflect the segregation of samples based on nonbiological variables (batches) in the experiment. In RNA-Seq analysis, batch effects were compensated by incorporating the batch variable in the design while hypothesis testing (batches indicated in Dataset S1). In a heatmap representation (Fig. 2B) of gene expressions, no clustering of samples based on the treatment category of genes is seen. The DE analysis resulted in a handful of DEGs given in the table (Fig. 2C). Quantitative Real-Time PCR (qRT-PCR)

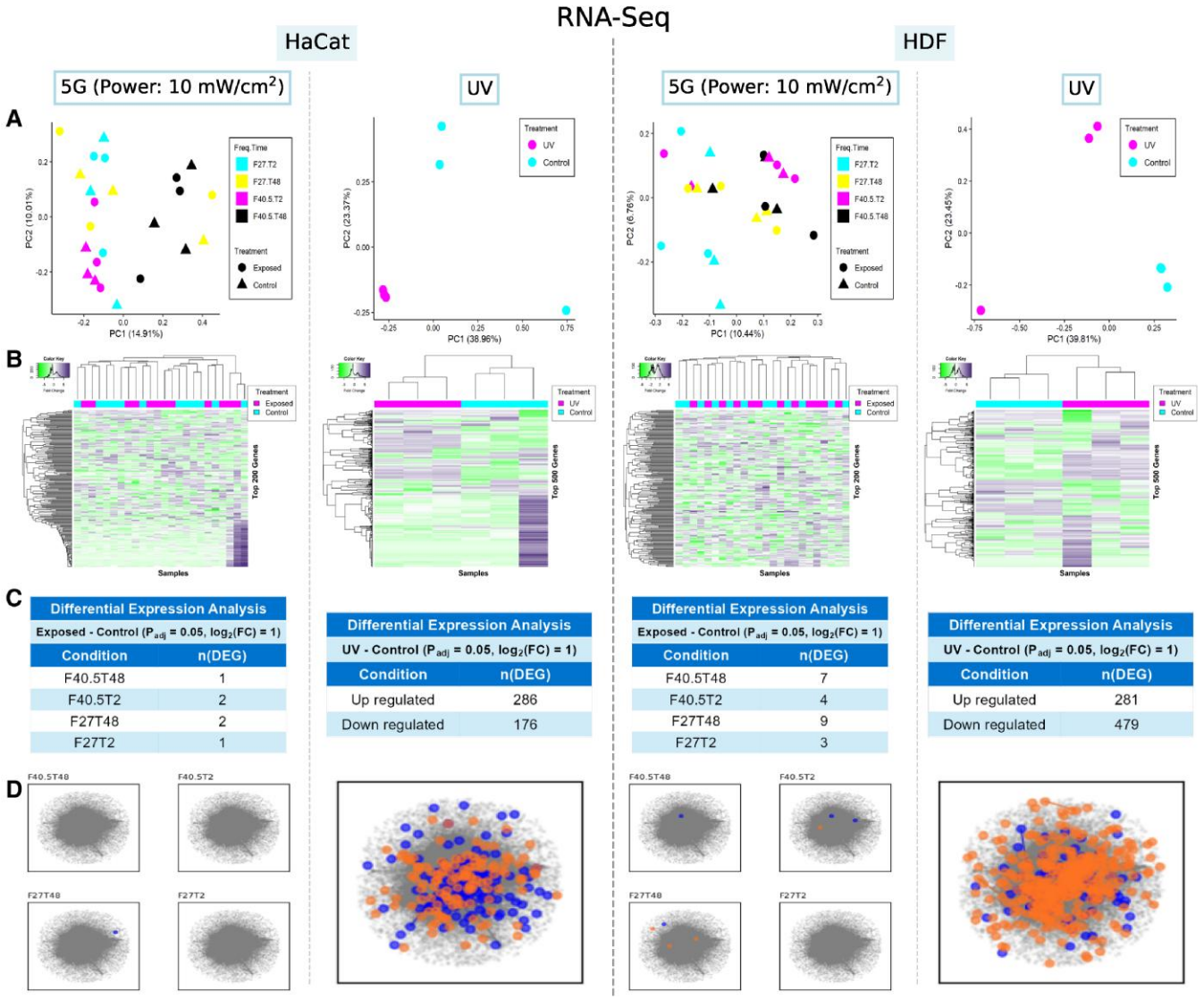


**Fig. 2.** Visualization of results from differential expression (RNA-Seq) and its network analysis. HaCat and HDF skin cells, exposed to 5G with power 10 mW/cm$^2$ and UV radiation (positive controls). Figure parts from top to bottom: A) Multidimensional scaling plots; B) heatmaps; C) tables with the number of significant DEGs, with Benjamini–Hochberg (13) adjusted *P*-value ≤0.05, $|\log_2(\text{foldchange})| \geq 1$ for Wald hypothesis testing; D) network plots, DEGs were mapped to a gene-level protein–protein interaction network derived from String.

validation was done for all 10 coding DEGs of the main experiments (Tables S1 and S2. A Ct range ≥29 was measured in 7 out of the 10 validated genes. In 21 of these 42 measurements with a high Ct value (seven genes in six samples), no Ct value could be determined at all as no amplicon was present. For the remaining three gene measurements, *MMP23B, SULT1A3, and SULT1A4*, the Ct values were 21 to 28. The gene expression of the exposed and sham-exposed samples was significantly different in only one of the 10 measured genes (Student's *t*-test; *SULT1A3*: $P = 0.027$).

Projecting the DEGs onto networks resulted in empty subgraphs due to the extremely small number of DEGs that are not present in the network (Fig. 2D), so network analysis could not be performed for these cases. On the contrary, UV treatment showed clear sample clustering based on the treatment in the MDS plots as well as the heatmaps (Fig. 2, HaCat-UV: A and B). We found 286 upregulated and 176 downregulated genes. Collective DEG results are shown in Fig. S1A. The network coherence analysis results in a z-score >2, i.e. coherence of the subgraph is higher than expected at random (Fig. S1B). The representation of DEGs projected onto different networks can be seen in Figs. S15 and S16.

The RNA-Seq analysis for HDF cells after 5G (power flux density: 10 mW/cm$^2$) and UV exposure gave similar results (Figs. 2 and S1A) with the number of DEGs slightly larger compared to HaCat, but not enough to assume substantial differences in gene expression for 5G exposure. For UV, it resulted in a total of 760 DEGs, which also have a significantly higher network coherence than expected at random (Fig. S1B). The volcano plots for the aforementioned cases are presented in Fig. S2. The results of the remaining experiments, 5G exposure with a power flux density of 1 mW/cm$^2$ and 5G exposure without temperature compensation are shown in Figs. S1, S3, and S4. The RNA-Seq analysis of these experiments resulted in very low number of DEGs for all cases, therefore, the network coherence analysis could not be performed for them, except for HaCat cells exposed to 5G EMF without temperature compensation (Fig. S1).

The outcome of the GO enrichment analysis conducted by using GOstat aligns with these findings. Concerning biological processes, molecular functions, and cellular components, we observed higher functional significance of the DEGs for HaCat after UV exposure and 5G exposure without temperature compensation and for HDF, only after UV exposure (Fig. S12). The lists of 10 most significant GO terms can be seen in Fig. S14. The main significant GO term lists derived from the fgsea analysis are presented in Fig. S18.

## Methylation analysis

Similar to RNA-Seq analysis, there are no visible clusters of samples based on the treatment category (exposed/sham) in MDS plots as well as the heatmaps for 5G exposure with power flux density of 10 mW/cm$^2$ for both HaCaT and HDF cells (Fig. 3A and B), and the DM analysis gave almost negligible number of DMPs (Figs. 3, S1A, and S5). Unlike RNA-Seq analysis, the control samples exposed to UV also showed no clear difference between exposure and sham exposure for both cell types. The remaining DM analysis results (5G with power flux density 1 mW/cm$^2$, 5G without temperature compensation) showed no difference in exposed and sham exposed samples (Figs. S6 and S7). The network or gene ontology analysis for any of these cases could not be realized because of the extremely small number of DMPs (<5) found. Since batch correction in methylation analysis, with currently available methods, leads to spurious results by either inflating the false positives or deflating the true positives (14–16), the results mentioned above were obtained by performing DM analysis on samples without correcting for batch effects. Nevertheless, we also present the results for which batch effects were corrected before performing the DM analysis in supplementary materials (see Fig. S8). Here, an inconsistent inflation in number of DMPs can be seen. In network analysis, where the power flux density is 1 mW/cm$^2$, a slightly enhanced signal was apparent for HDF cells but this enhancement was not observed across two distinct networks (Fig. S8). These inflations are later nullified by the combinatorial analysis (see Fig. S9 for String analysis results, see Fig. S17 for Biogrid). The outcome of GO analysis (numbers of significant GO terms) is consistent with the result of DM analysis (Fig. S12). The significant GO terms are listed in Fig. S14. We also conducted a parallel enrichment analysis to simultaneously evaluate the enrichment of RNA-Seq and methylation data. The 10 most significant GO terms are listed in Fig. S19.

## Combinatorial analysis

So far, we can conclude that the gene expression and DNA methylation remained almost unchanged after 5G exposure with temperature compensation, since the number of DEGs and DMPs is very low for all the treatments. To further solidify these findings and remove any biases from the earlier analyses, we performed a combinatorial analysis. A detailed explanation of this technique with examples is given in Fig. 4. If the original combination has differences in expression, the circled dot would appear significantly higher than other combinations. Such a signal is only noticed in three of the positive controls (HaCat and HDF after UV exposure, HaCat after 5G EMF without temperature compensation) for RNA-Seq, confirming the results obtained above (Fig. 4C and D). Since the number of DMPs obtained from DM analysis without batch correction is extremely low (<5), combinatorial analysis was not executed for it. Rather, we performed the combinatorial analysis on DMPs obtained from DM analysis with batch correction. Again, we see that the original combinations do not appear significantly higher than the rest for any of the experiments, and in contrast to the RNA-Seq combinatorial results, the control samples without temperature compensation and the UV samples do not stand out so strongly in methylation analysis (Fig. S9). The GO combinatorial analysis findings are consistent with the results obtained (Fig. S13).

The results of the combinatorial analysis were quantified by finding the z-scores of the original combination against all other combinations. The z-scores of DEGs, DMPs (Fig. S10A) and their respective network coherences (Fig. S10B) are below the threshold value for significant differences of $z = 2$ for all tested experimental conditions, confirming that the actual assignment of sham-exposed and exposed samples does not show a stronger signal than expected by chance.

# Discussion

Overall, the data show no indication that the gene expression and methylation of human skin cells were altered by the exposure conditions selected here. Using the combinatorial analysis for these low-signal data provides the chance to manifest a clear statement.

## 5G EMF exposure did not alter gene expression or DNA methylation profiles

### *RNA-Seq*

The results indicate that exposure did not lead to any changes, particularly in gene expression. The few coding genes that were

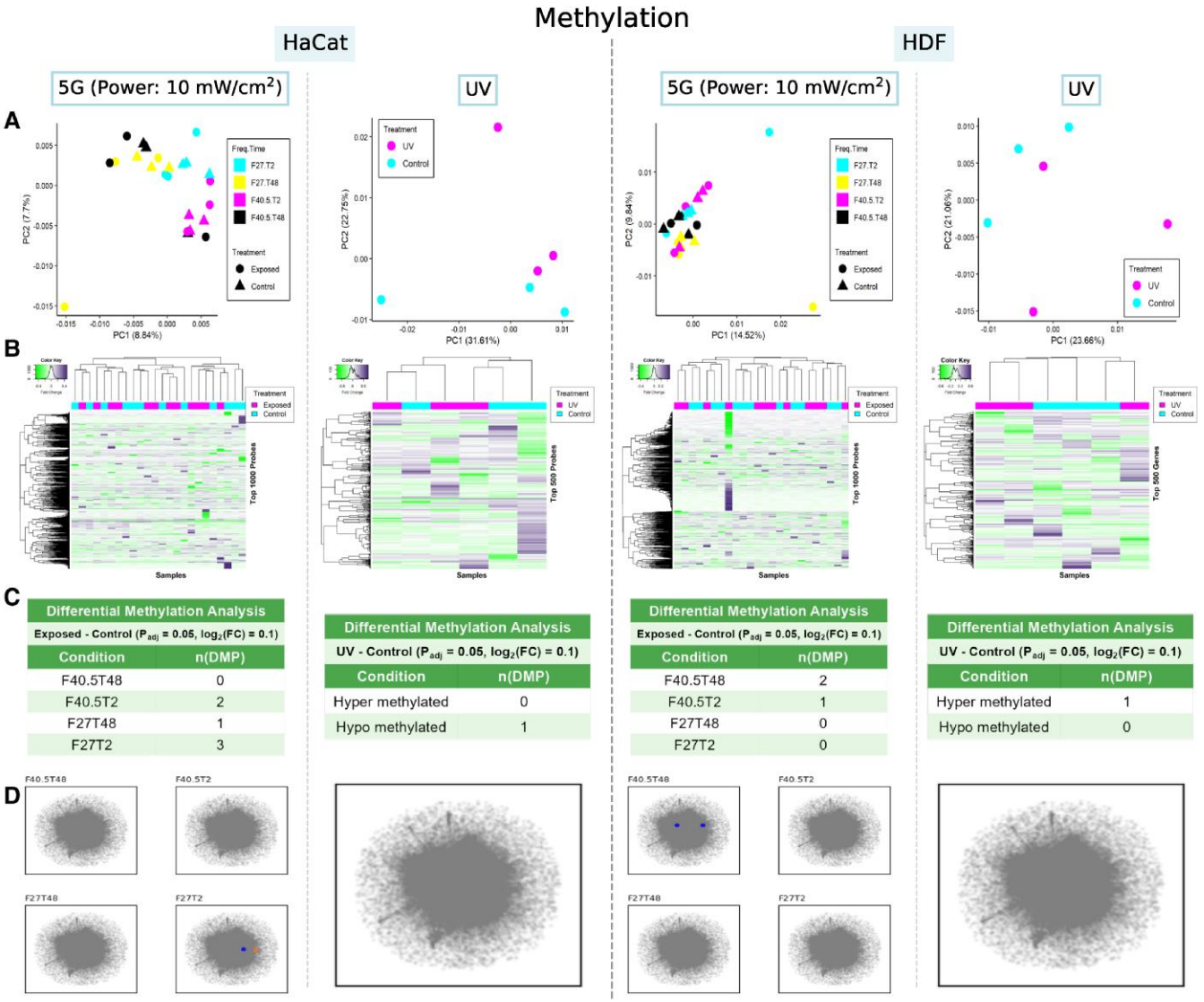


**Fig. 3.** Visualization of results from differential methylation and its network analysis. HaCat and HDF skin cells, exposed to 5G with power 10 mW/cm$^2$ and UV radiation (Positive controls). Figure parts from top to bottom: A) Multidimensional scaling plots; B) heatmaps; C) tables with the number of significant DMPs; D) network plots, DMPs were mapped to a gene-level protein–protein interaction network derived from String.

found differentially expressed after RNA-Seq analysis could mostly not be confirmed by qRT-PCR validation. As the expression of seven of the ten genes was very low or could not be measured at all using qRT-PCR, it is reasonable to assume that these DEGs were stochastic artifacts from the RNA-Seq analysis.

In addition to the *MMP23B* gene, whose altered expression could not be confirmed, there was also an overexpression of the *SULT1A4* gene (also not confirmed) and an underexpression of *SULT1A3* in the same sample set. Both *SULT1A* genes originate from the same gene family of sulfotransferase 1A. These genes code for enzymes that catalyze the sulphate conjugation of various substances (including hormones and neurotransmitters) (17). The number and length of exons is similar in all members of this gene family, *SULT1A3* and *SULT1A4* are more than 99% homologous and code for identical proteins (18). Thus, overexpression of *SULT1A4* may compensate for the underexpression of *SULT1A3* and would have no biological effect.

The gene expression results are consistent with literature data from studies on the effect of millimeter waves on the transcriptome of human skin cells. Habauzit et al. (19) found no differentially expressed genes when HaCaT cells were exposed to a frequency of 60.4 GHz, a power flux density of 20 mW/cm$^2$ for 3 h and temperature compensation. The same study group recently stated that the transcriptional landscape of HaCaT and other keratinocytes is not altered under athermal conditions (20). A similar study found clear effects of exposure on gene expression of primary dermal fibroblasts at 60 GHz and 2.6 mW/cm$^2$ hourly for several days (21). Therefore, it would be interesting to repeat this experimental design with the setup shown here.

It might be argued that the likelihood ratio test (LRT) is more suitable for the number of samples presented here, therefore, we performed the same analysis with LRT. The results of the two tests do not differ much (File S1, Fig. S11).

### *Methylation*

The DM results do not suggest any alteration in the DNA methylation profiles. The methylation data have been challenging to analyze because of the increased variability within the respective samples due to the influence of confounding factors and

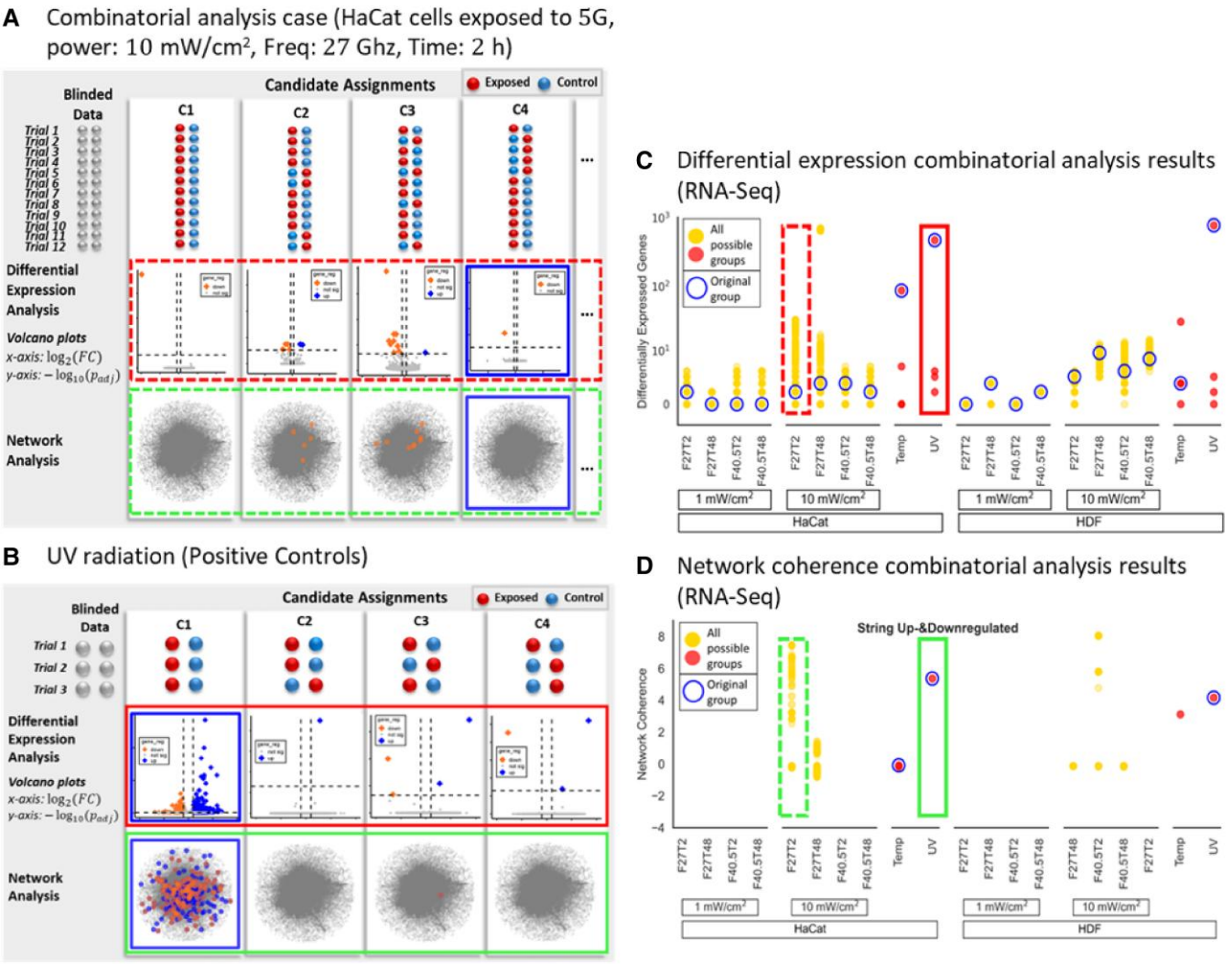


**Fig. 4.** Visualization of combinatorial (RNA-Seq) analysis results. A) There are 12 trials in an experiment, containing one exposed and one sham-exposed sample each. Assuming blinded conditions, we created all possible combinations of samples ($2^{12}/2$ (excluding symmetric combinations)), with samples in their original trials. Each candidate assignment (C1, C2, …) is one combination of exposed and sham-exposed samples. DE- and network analysis were performed on all these query groups. Volcano plots show the number of DEGs, which are then mapped to a network (String). Here, C4 is the original combination, i.e. the unblinded distribution of exposed and sham-exposed samples. B) UV experiment, C1: original combination. C) Each dot represents the number of DEGs from each combination; one vertical block of dots represents one experiment; circled dot: original combination; dashed block: number of DEGs from all combinations in (A); solid block: number of DEGs from all combinations from (B), here original combination is highest. D) Each vertical block represents the network coherence of DEGs mapped to the network; dashed and solid block: network coherence for all combinations in (A) and (B), respectively; circled dot: original combination. Temp: 5G exposure without temperature compensation.

inconsistencies in the available analysis tools. Since batch correction is a highly debatable topic, especially in methylation analysis (14–16), we opted to do the analysis with as well as without batch correction. For both cases, we did not find many significant genes, but an inconsistent inflation of DMPs from analysis with batch correction. The results with batch correction seem very contradictory. Because of the high number of DMPs at a power flux density of 1 mW/cm² and almost negligible DMPs at 10 mW/cm², no dose–response relationship could be observed.

However, it must be said here that the number of DMPs of the UV control also does not stand out. The reasons for this are difficult to pinpoint. Altered methylation in promoter regions can lead to altered gene expression (22). Due to the comparatively high number of DEGs found in the RNA-Seq analysis, it was therefore reasonable to expect that there must also be a certain number of DMPs, although altered gene expression can also occur through other mechanisms, such as histone modifications or RNA-based mechanisms (23). The analysis of DNA methylation in this project is extremely sensitive and easily influenced by external factors. Side effects should therefore be monitored as standard in future studies and included in the analysis. A further recommendation would be to increase the number of replicates to compensate for the high variability of the samples.

The relationship between exposure to EMF and DNA methylation is still poorly understood. In the range of frequencies investigated here, there are no known in vitro studies to date. Recently, two independent studies on the effect of exposure to 900 MHz on DNA methylation of keratinocytes (24) and two cancer cell lines (25) were published. In both studies, clear effects were found, but the penetration depth of the waves into the cells or the medium at 900 MHz is much higher (approx. 12 mm) than at the frequencies investigated here (~ 0.3–0.5 mm).

### DEGs and DMPs did not form a coherent signal in functional interpretations

The network analysis provides a functional interpretation of the DEGs and DMPs by studying their network coherences in biological networks via their reaction and metabolites or via interactions. The network coherence of DEGs/DMPs from exposure experiments either did not show any significant signal or could not be realized due to the small number of significant genes. In RNA-Seq analysis, a clear statement can only be made about the controls. The significantly increased network coherence of the HaCaT temperature control and the samples after UV radiation indicates that various signaling pathways are stimulated, particularly after UV radiation. In case of DMPs from analysis with batch correction, network coherence of few inflated DMPs appeared significantly higher than by chance, but this enhancement was not observed across two distinct networks and is later nullified by the combinatorial analysis.

It is striking that the HaCaT cells show significant values in both the metabolic and protein interaction networks, while the HDF cells only show significant values in the protein network. This could be due to the fact that both cell types react differently to UV radiation. This has already been shown by other studies in which fibroblasts synthesize the enzyme matrix metalloproteinase-1 (MMP-1) after UV irradiation, which is thought to be involved in photoaging processes (26). In contrast, keratinocytes did not increase their MMP-1 production but instead secreted other proteins (interleukin-1$\alpha$ and interleukin-6), which indirectly stimulated the production of MMP-1 in neighboring fibroblasts.

GO enrichment analysis functions as a valuable instrument in evaluating the functional importance of a specific set of genes, such as those exhibiting differential expression or methylation, with respect to biological processes, molecular functions, and cellular components. In line with our aforementioned findings, we observe higher functional significance of the DEGs for HaCat after UV exposure and 5G exposure without temperature compensation and for HDF, only after UV exposure. For the experiments, the analysis with DEGs cannot be performed due to the small number.

We also performed a parallel enrichment analysis employing the mitch library (27) to be able to jointly analyze the differential expression/methylation analysis results of RNA-Seq and methylation data. The number of significant GO terms identified via mitch was quite volatile with no visible trend across the parameters of our study. Shuffling labels ("exposed"—"sham-exposed") and recomputing significant GO terms for these null model datasets gave similar numbers of significant GO terms. Furthermore, when grouping experiments according to cell types and computing the intersection of significant GO terms, these were small and did not point to a common biological signal like DNA damage and other cellular stress pathways. This analysis therefore confirms the absence of a 5G-driven signal. The cell-type-level intersections of significant GO terms can be found in Fig. S19.

### DEGs and DMPs did not outperform the bulk of data separations during combinatorial analysis

The combinatorial analysis of the gene expression data also categorizes the few DEGs found in the RNA-Seq analysis in the stochastic probability range. In comparison to the control after UV radiation, it is very clear that the number of DEGs found does not stand out from the number of DEGs found by chance after variable sample combination. The combinatorial analysis of the batch-corrected methylation results provides further evidence that the number of DMPs found does not stand out from the number of DMPs found by chance after variable sample combination.

Although the DMPs of two exposure conditions showed significant network coherence, this could not be distinguished from other randomly generated sample combinations. This suggests that the methylation status of the cells was not comprehensively changed by the exposure. The combinatorial analysis outcomes of the methylation GO results show that the original combinations do not show any stronger than other combinations.

## Conclusion

The almost complete absence of effects of exposure of human skin cells on the tested parameters, even at 10 times the exposure limits, are on the one hand very well in line with biophysical facts: at the frequencies tested here, the quantum energies are far too low to have photochemical or even ionizing effects. On the other hand, it shows how important it is to precisely control, document and, if necessary, compensate for the temperature effects caused by the exposure. In spite of assessments from the WHO in 2010 (28) and the National Toxicology Program (NTP) (29), this topic is still prevalent in the media, in public opinion and in the political sphere. This is in part due to a few isolated scientific studies providing opposing evidence. Due to our strong emphasis on highly controlled experimental conditions and our combinatorial analysis, we hope to close this debate and in particular cast fundamental doubt on the existence of possible nonthermal biological effects of exposure.

## Materials and methods

### Cell culture

Human dermal fibroblasts (HDF, Cell Applications, San Diego, USA) and HaCaT cells (CLS Cell Lines Service, Eppelheim, Germany) were obtained as a primary culture and in passage 31, respectively. Cells were cultured in DMEM, high glucose, GlutaMAX™ (Thermo Fisher Scientific, Waltham, USA) supplemented with 10% FBS (Thermo Fisher Scientific, Waltham, USA) and 1% Penicillin Streptomycin (Thermo Fisher Scientific, Waltham, USA) in an incubator with 37 °C, 5% $CO_2$ and saturated humidity. For each exposure experiment 120,000 HDF or 200,000 HaCaT cells were seeded onto a 60 mm CELLBIND® Surface cell culture dish (Corning, New York, USA). Cells were used in passages 6–7 (HDF) and 35 (HaCaT).

### Exposure system and experimental design

The exposure system is described in detail by Schmid et al. (8). Briefly, the novel exposure system has been characterized for cell monolayer dosimetry in 60 mm petri dishes and allows for standard in vitro incubation (37 °C, 5% $CO_2$, saturated humidity) during exposure. The system consisted of two similar incubators in which the exposure setups were integrated. The software allowed for random and blinded allocation of exposure and sham-exposure (Fig. 1A).

One day (48 h experiments) or 3 days (2 h experiments) after seeding of the cells, the dishes were transferred to the exposure incubators and experiments were started (Fig. 1B). All cells stem from one stock that was initially frozen within aliquots. The different times for culture (1 or 3 days after seeding) were chosen to have all cells at the same age after thawing an aliquot/ seeding and be comparable at the end of exposure. Cells were exposed or

sham-exposed for 2 h or 48 h, with a frequency of 27 GHz or 40.5 GHz, and a power flux density of 1 mW/cm$^2$ or 10 mW/cm$^2$ (Fig. 1C). The exposure times of 48 and 2 h were based on the known kinetics of changing gene expression already after 2 h of exposure to changing environments. DNA methylation is incorporated during the cell cycle which takes ∼1–2 days. The highest level of DMPs would therefore be expected after 48 h of exposure. In all experiments, the temperature was measured continuously using a fiber optic sensor.

The preliminary tests showed that the temperature deviations between the exposed and sham-exposed cell culture dishes occurring during the experiments with a power flux density of 1 mW/cm$^2$ were tolerable. In the experiments with a power flux density of 10 mW/cm$^2$, the temperature increase measured in the center of the dish was up to 1.2 °C at 27 GHz and up to 1.5 °C at 40.5 GHz. To minimize potential temperature effects, the internal temperature of the incubator in which the exposure takes place was reduced by manually controlling the incubator temperature so that an average temperature of 37 °C could be maintained in the petri dishes. To set up the incubator temperature, a single person was unblinded for the experiments at 10 mW/cm$^2$ who was not involved in any way in processing the samples, performing the experiments, or analyzing the data. In these experiments, a lid heater was used in all four petri dishes of the two experimental setups to prevent condensation in the exposed petri dish and maintain blinding conditions. The lid heater consisted of a circular copper plate (thickness 2 mm) to which two high-load resistors were attached. A low, adjustable direct current (<0.3 A) caused the series-connected resistors to heat up and transfer the heat to the plate, heating the lid and preventing condensation. Three technical replicates per treatment combination were processed.

The transmitting power flux density or SAR showed a relatively high inhomogeneity within the cell monolayer (8). To keep the homogeneity as high as possible, it was decided to exclude cells outside a defined area during harvesting. As this was a compromise between the highest possible homogeneity and the largest possible amount of sample material obtained, the area range was adjusted accordingly based on the calculations of Schmid et al. (8). After exposure, the medium of the cell dishes was discarded and a template made with a 3D printer (at 27 GHz an ellipse with a radius of 24 mm/23 mm; at 40.5 GHz a capped circle with a radius of 24 mm and a distance of the capped area to the center of 19 mm) was applied to the outside of the bottom of the dish and the areas to be omitted were marked. The cells were removed according to the marking with a cell scraper and the cell monolayer was then rinsed twice with PBS (Gibco, Waltham, USA). As a result, a low deviation in the SAR value of only 3.66 dB with a cell loss of 12% was achieved for the 27 GHz tests and only 3.35 dB with 13% cell loss for the 40.5 GHz tests. Since RNA yield of HDF cells was only sufficient for RNA-Seq analyses but not qPCR validation, HDF experiments were repeated once.

Two different approaches were pursued for further controls in addition to sham exposure. Firstly, temperature controls were created. Three replicates each of HaCaT and HDF cells were exposed or sham-exposed at 40.5 GHz and 10 mW/cm$^2$ for 48 h without compensating for the temperature increase. The starting temperature of the medium was 37.0 °C in these experiments and increased to an average of 38.3 °C in the exposed samples. In addition, no lid heaters were used, so that the heating of the medium in the exposed samples led to medium condensation. Secondly, cells were treated with UV radiation. Three replicates each of HaCaT and HDF cells were exposed or sham-exposed with a MinUVIS analysis lamp (DESAGA, Wiesloch, Germany) at a wavelength of 254 nm and an illuminance of 1 mW/cm$^2$. The medium was removed and collected, the cells were rinsed twice with PBS (Gibco, Waltham, USA) and thinly coated with 600$\mu$l PBS per dish. Exposure was performed without the dish lid for 20 s, for sham exposure the dishes were stored in the dark at room temperature. The PBS was then aspirated, and the cells were layered with the previously collected medium. After 2 h incubation in the incubator, the medium was discarded, and the cells were harvested and analyzed as in the main experiments.

## DNA and RNA extraction

Following exposure, the cells were immediately harvested by washing the remaining cell monolayer once with PBS (Thermo Fisher Scientific, Waltham, USA) and transferring 350$\mu$l RLT Plus lysis buffer (Qiagen, Hilden, Germany) to each of the dishes. In the case of the HDF repeated experiments, 350$\mu$l RLT lysis buffer (Qiagen, Hilden, Germany) was used. The cell lysates were collected using a cell scraper and transferring them into reaction tubes. The lysates were frozen at −80 °C until further processing.

DNA and RNA extraction was done using the DNA/RNA Allprep Mini Kit (Qiagen, Hilden, Germany) or the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Briefly, DNA, RNA, and waste materials were separated using silica membrane columns. DNase digestion was done by integrating the RNAse-Free DNase Set (Qiagen, Hilden, Germany) into the extraction procedure. After separating and washing the nucleic acids with several buffers and centrifugation steps, DNA was eluted using EB buffer (Qiagen, Hilden, Germany) at 70 °C which was incubated on the column for 2 min. RNA was eluted using RNase-free water which was incubated on the column for 10 min to increase RNA amount.

## RNA-Seq

A total of 96 RNA samples from the main experiments and 24 RNA samples from the controls were sent to IMGM Laboratories (Martinsried, Germany) for analysis of gene expression. The samples were first tested for integrity using the 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany) with RNA Nano LabChip kits (Agilent Technologies, Waldbronn, Germany). If the integrity was sufficiently high (RIN value >7.0), the samples were used, otherwise the experiment was repeated and the new samples were sent in.

The transcriptome library was generated using TruSeq® Stranded mRNA technology according to the manufacturer's protocol (Illumina, San Diego, USA). It included fragmentation, poly-T-oligo pulldown, and sequencing adapter ligation. The NovaSeq® 6,000 Next Generation Sequencing System (Illumina, San Diego, USA) was used for RNA sequencing. After excluding reads from low quality clusters which did not pass quality criteria, the total number per sample was 14.1 +/− 3.9 million reads. The resulting reads were trimmed to 75 bp and mapped against the human reference genome using CLC Genomics Workbench 20.0.4 (CLC bio, Qiagen, Hilden, Germany). A proportion of 13.9 +/− 3.9 million reads could be mapped to the reference genome which is 99.05 +/− 0.29 %. Values are given as means +/− standard deviation.

## DE analysis

We extracted total exon reads for each sample from the read counts data received from IMGM. Then, we adjusted the raw counts to enable direct comparison of samples by normalization. In RNA-Seq analysis, normalization of data is an important step

which needs careful assessment (30, 31). Main assumptions for the choice of normalization by distribution are (32): (a) Most genes are non-DEGs (differentially expressed genes). (b) Technical variation is the same for DEGs and non-DEGs. (c) The differential expression is symmetric across conditions and there is no global shift. These assumptions are in line with our experimental expectations; therefore, we choose normalization by distribution (DESeq2 normalization (33, 34)) which equilibrates expression levels for non-DEGs. We visualize the results by density plots created using geneplotter::multidensity function in R. When the sample distributions do not overlap, we perform within lane normalization to remove the GC content bias of the genes using EDASeq (35). After normalization, we check for the presence of batch effects in the data. Batch effects are nonbiological variables in the experiments that, if not adjusted/corrected, can result in spurious results, e.g. the effects encountered in (36, 37) which were later addressed in (38, 39). To detect the batches, we calculate correlations between variation in the data (principal components) and the covariates using DEGreport::degCovariates function, and visualize it using SVD plots using ggplot2. Batch effect correction is one of the most debatable steps in DE analysis. Two primary ways of dealing with batch effects are (14): (a) Adding batch as a covariate in the statistical model design of the analysis. (b) Removing batch effects from the read counts and then performing the statistical analysis. For our study, we choose the first approach for the following reasons: (i) Authors of DESeq2 and other similar tools suggest using the first approach for statistical analysis and second to visualize batch corrected data (in DESeq2 package vignettes and Bioconductor support: https://support.bioconductor.org/p/121408/). (ii) DESeq2 requires integer counts as input, while most batch correction tools return noninteger values except Combat-seq from SVA, and some studies (14) have suggested that correcting for batch effects by Combat-seq can induce false positives by inflating the F-statistics of query analysis. To visualize clustering of samples, we use MDS plots (limma::plotMDS) and heatmaps with samples and genes clustered hierarchically according to average linkage (stats::as.dist, stats::hclust). Next, we perform the statistical analysis to find differentially expressed genes between two study groups. We use the robust and powerful tool DESeq2 (34) for the statistical inference which uses the negative binomial distribution while accounting for inherent variability of RNA-Seq data. Parameters for genes to be significantly differentially expressed are: adjusted $P$-value $\leq 0.05$, $|\log_2(\text{foldchange})| \geq 1$ for Wald hypothesis testing. DEGs are finally visualized using volcano plots using ggplot2.

## qRT-PCR validation

The differentially expressed coding genes which were identified by RNA-Seq analysis were validated by qRT-PCR, read-through transcripts and other noncoding sequences were excluded from the analysis. In addition, the amplification of three reference genes was measured in each sample to normalize the data accordingly. The selection of reference genes was based on known genes that were found to be suitable for both HaCaT and HDF cells (40, 41).

RNA samples were reverse transcribed using the QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Briefly, 1 $\mu$g RNA was treated with gDNA Wipeout Buffer and then reverse transcribed using the kit's oligo-dT and random primer mix. The cDNA was diluted 1:5 with RNase-free water and stored at −80 °C.

The qRT-PCR was performed using the QIAquant 96 5plex qPCR Cycler (Qiagen, Hilden, Germany) and the QuantiNova LNA PCR Kit (Qiagen, Hilden, Germany). The efficiency of the 13 QuantiNova LNA PCR assays used was determined by measuring standard curves (Table S1). Seven of the ten target genes were only slightly expressed (Ct value ≤29), so that the standard curves were partly generated with the amplification products of qPCR (reamplification). According to the manufacturer's protocol, the QuantiNova SYBR Green PCR Master Mix (Qiagen, Hilden, Germany) was mixed with the QuantiNova LNA PCR assay to be measured (Qiagen, Hilden, Germany) and RNase-free water and distributed in white PCR plates for qPCR (Brand, Wertheim, Germany). 2 $\mu$l of each cDNA sample was added and the measurement was performed with the qPCR cycler according to the following program: 2 min at 95 °C, 45 cycles of 5 s at 95 °C and 10s at 60 °C. Following the last cycle, a melting curve of the amplification product was generated to detect any unspecific products.

The qRT-PCR validation data, corrected for assay efficiency and normalized to the three reference genes, were determined using the ΔΔCt method (42). The gene expression of the samples was tested for significant differences between exposure and sham exposure/treatment using Student's $t$-test.

## Methylation

A total of 96 DNA samples from the experiments and 24 DNA samples from the controls were sent to IMGM Laboratories (Martinried, Germany) for analysis of methylation. To ensure quality, the concentration of the DNA was determined using the Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, USA) in comparison to a standard contained in the kit. The genomic DNA was then modified by bisulfite conversion. For this purpose, the EZ DNA Methylation™ Direct Kit (Zymo Research, Freiburg, Germany) was used according to the manufacturer's instructions. The modified DNA was denatured and neutralized using the Infinium HD Assay for Methylation (Illumina, San Diego, USA) according to the manufacturer's protocol. It was amplified and precipitated with isopropanol after enzymatic fragmentation. The DNA was resuspended and hybridized to the Infinium Methylation EPIC arrays (Illumina, San Diego, USA). The primers on the BeadChips were extended with labeled nucleotides complementary to the DNA sample. The hybridized DNA was removed, and the labeled, extended primers were stained and dried. Fluorescence was measured using the iScan™ system (Illumina, San Diego, USA).

## DM analysis

For methylation analysis, we opt for a comprehensive package ChAMP (43) in R for an integrated methylation analysis. The raw intensity files were imported with minfi method (44, 45) using champ.load() function. Preprocessing of the data was done by filtering out probes with detection pvalue >0.01, bead count <3 in at least 5% samples, overlapping with SNP sites (46), overlapping with multiple locations on human genome or aligned to X/Y chromosome. The intensities are imported to beta values ranging from 0 to 1. After filtering the probes, we normalize the beta values by using Quantile (47) + BMIQ (Beta-Mixture Quantile)(48) normalization. QN+BMIQ normalization has been found highly reliable for microarray data (49–51) including DNA methylation protocols by Illumina (52). It focuses on transforming the distribution of Type II probes to be similar to Type I probes. After normalization, batches are detected by visualizing correlations between principal components and covariates using singular value decomposition (SVD) plots (53). The detected batches are removed using champ.combat() which uses combat method from SVA package (54, 55).

The findings in various studies (14–16) suggest that batch correction by combat, while preserving the query group differences, can result in inflation of F-statistics and false positives, therefore we filter out artifacts as strong isolated signals not discernible without batch corrections. Batch corrected data is then statistically analyzed to find DMPs using limma (56, 57) with multiple testing correction (Benjamini–Hochberg (1995) (13)) threshold: adjusted pvalue $\leq 0.05$, and $|\log_2(\text{foldchange})| \geq 0.1$.

## Network analysis

To investigate the network coherence of the differentially expressed/methylated genes in the context of biological networks, we employed two gene-centric metabolic networks (Recon2 (9) and Recon3D (10)) and two gene-level protein–protein interaction networks (String (11) and Biogrid (12)). Our focus was on the effective subnetworks, which represent the projections of differentially expressed/methylated genes onto the employed gene-centric metabolic network or gene-level protein–protein interaction network. We conducted a comprehensive analysis of network coherence on these subnetworks, by thoroughly assessing whether the network coherence happens to be greater or lower than expected at random.

### *Gene-centric metabolic network construction*

We extracted gene-centric metabolic models (GCMNs) from the Recon2 (9) and Recon3D (10) human metabolic network models following the methods described in Refs. (58–61). In GCMNs, nodes represent genes and edges represent the associations of genes via metabolic reactions. Stated differently, a connection between two genes is established if the metabolic reactions associated with these genes share a common metabolite.

The primary exchange metabolites, namely ATP, ADP, CO2, H, NAD, NADH, among others, are the most highly connected metabolic species that are unlikely to establish links between genes with similar metabolic functions, leading to an artificially denser metabolic network (62–64). In order to mitigate this effect, we opt to remove a subset of metabolites that represent the top 2% of the most highly connected metabolites, prior to network construction.

The resulting gene-centric metabolic network obtained through the utilization of the Recon2 human metabolic network model has 1,806 nodes and 31,699 edges. Similarly, the one extracted from Recon3D has 3,449 nodes and 233,235 edges.

### *Gene-level protein–protein interaction network construction*

We perform our analysis using two different protein–protein interaction networks on gene-level (GPINs) derived from String (11) and Biogrid (12) protein–protein interaction databases.

To construct GPIN derived from the String database, we selectively included protein interactions associated with the human organism, and only the ones with a score exceeding 850 to include solely direct interactions. We then cross-referenced the protein IDs with associated gene IDs using the Ensembl database (65). The GPIN is depicted as a graph, with genes serving as nodes and the edges representing the associations between the genes through protein interactions. Notably, the graph has 12,468 nodes and 139,560 edges.

The GPIN derived from the Biogrid database is constructed following the methodology explained above without any filtration of interaction scores. The graph consists of a total of 21,797 nodes and 1,018,531 edges.

### *ID mapping*

We employed the Python mygene package to map gene names to both Entrez IDs and Ensembl IDs, for the purpose of performing GO and network analyses, respectively. In instances where multiple hits are identified, meaning that a singular gene name may correspond to multiple IDs, all of the mapped IDs were included.

### *Network coherence*

The differentially expressed/methylated genes were projected onto the network employed to extract the effective subnetwork. The coherence value of a subnetwork was determined by the ratio of nonisolated genes to the total number of genes that exist within the network. The analysis was only performed if the number of differentially expressed/methylated genes was greater than or equal to 5.

To obtain the null distribution of coherence values, 5,000 gene sets, each being equivalent in size to the effective subnetwork, are randomly drawn from the employed network, and the coherence values of these randomly drawn subnetworks were subsequently calculated. The z-score of the coherence of the effective subnetwork was then determined utilizing this distribution.

## GO enrichment

GO enrichment analysis serves as a valuable tool to assess the functional significance of a particular collection of genes, such as those that are differentially expressed or methylated, in relation to biological processes, molecular functions, and cellular components. In order to determine statistical significance, a hypergeometric test is conducted to evaluate whether any observed enrichment of GO terms in the gene list is statistically significant. This test determines whether any GO terms are over-represented or if the number of selected genes associated with a particular term is greater than expected.

We conducted the GO enrichment analysis using GOstats (66), a package available in R through Bioconductor, when the number of differentially expressed or methylated genes is greater than or equal to 5. The org.Hs.eg.db database (67), which serves as an organism-level package utilizing a central gene identifier, namely the Entrez gene ID, and encompasses mappings between this identifier and various other types of identifiers, specifically GO terms, was utilized as the background gene list for this analysis. The terms associated with fewer than five genes were discarded. This database comprises 20,692 genes and 18,348 GO terms, with the filtered database yielding 7,349 GO terms when the size threshold was established at a minimum of five genes: 5,029 Biological Process (BP), 920 Cellular Component (CC), and 1,400 Molecular Function (MF).

We used the Benjamini–Hochberg method (13) for multiple test corrections, i.e. to calculate adjusted *P*-values, since the analysis output consists of only *P*-values. We assess the analysis outcomes by comparing the number of significant GO terms identified with an adjusted *P*-value threshold of 0.05.

Since functional class sorting was reported to be more sensitive than over-representation analysis (68), we also conducted a fast preranked gene set enrichment analysis using the fgsea library (69) for the RNA-Seq data. Gene ontology biological process gene sets, obtained from the Molecular Signatures Database (MSigDB) (70), were used as background gene set with a minimum gene set size threshold of five. The ranking metric was the *stat* value, corresponding to the t-statistic provided in the output of the DESeq2 analysis.

Additionally, we executed a parallel enrichment analysis of the RNA-Seq and methylation data employing the mitch library (27). The same background gene set was utilized once more, with the minimum set size threshold maintained at 5. For ranking, t-statistics provided in the output of the ChAMP analysis was used.

The combinatorial analysis was performed using the GOstat package.

### Combinatorial analysis

Considering the experiment to be strictly blinded (meaning that the data analysts were blinded until writing this manuscript), we devised a method that aims to detect the real query group separation (sham/exposed) of samples from other randomly generated sample groups by combinatorics. We have 24 samples in each experiment, divided into 12 trials containing one sham and one exposed sample. Without knowing which sample is sham or exposed, we create all possible combinations ($2^{12}$) of samples such that each trial consists of one sham and one exposed sample. Then, we perform the differential expression/methylation analysis by putting each possible sample-grouping as the model design in the statistical analysis. This yields a compilation of lists of differentially expressed/methylated genes for all groupings.

This study investigated two signals that were derived by conducting combinatorics: (i) the number of differentially expressed/methylated genes and whether the original grouping of sham and exposed samples display a stronger signal than expected at random, i.e. the z-score of the number of differentially expressed/methylated genes and (ii) the network coherence of those differentially expressed/methylated genes (the z-score) and the strength of this signal with respect to other randomly generated sample groupings (the z-score of z-scores).

The resulting signal strengths are investigated for all possible groupings where the original grouping is expected to display the strongest signal under the assumption that the original grouping has significant differences in expression/methylation.

## Acknowledgments

We thank Karen Drees for her technical assistance and her valuable solution-oriented thinking. We are grateful to Gernot Schmid and Seibersdorf Laboratories for comprehensive and promptly engineering support.

## Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

## Funding

This study was funded by the Bundesamt für Strahlenschutz (FKZ 3619S82470).

## Author Contributions

Conceptualization: I.G. and A.L.; Data curation: J.J. and E.C.; Formal analysis: J.J., and E.C.; Methodology: all authors; Funding acquisition: I.G. and A.L.; Investigation: J.J., I.G., E.C., M.H., and V.M.; Project administration: I.G. and V.M.; Resources: I.G., A.L., and V.M.; Supervision: M.H. and A.L.; Visualization: J.J., I.G., E.C., M.H., and V.M.; Writing—original draft: J.J., E.C., and V.M.; Writing—review & editing: all authors.

## Data Availability

All raw data and result tables have been submitted to the NCBI Gene Expression Omnibus and are accessible using the accession numbers GSE271547 and GSE271704. HaCaT cell line was kindly provided by Deutsches Krebsforschungszentrum (DKFZ) and originally created by Profs Dr Petra Boukamp and Dr Norbert Fusenig. The code developed for generating the results, along with session information, is available in the GitHub repository (https://github.com/Computational-Systems-Biology/5G.git).

## References

1 Alanen E, Lahtinen T, Nuutinen J. 1999. Penetration of electromagnetic fields of an open-ended coaxial probe between 1 MHz and 1 GHz in dielectric skin measurements. *Phys Med Biol*. 44(7): N169.
2 International Commission on Non-Ionizing Radiation Protection (ICNIRP). 2020. Guidelines for limiting exposure to electromagnetic fields (100 kHz to 300 GHz). *Health Phys*. 118(5):483–524.
3 Karipidis K, Mate R, Urban D, Tinker R, Wood A. 2021. 5G mobile networks and health—a state-of-the-science review of the research into low-level RF fields above 6 GHz. *J Expo Sci Environ Epidemiol*. 31(4):585–605.
4 Simkó M, Mattsson M-O. 2019. 5G wireless communication and health effects—a pragmatic review based on available studies regarding 6 to 100 GHz. *Int J Environ Res Public Health*. 16(18):3406.
5 Pakhomov AG, Akyel Y, Pakhomova ON, Stuck BE, Murphy MR. 1998. Current state and implications of research on biological effects of millimeter waves: a review of the literature. *Bioelectromagnetics*. 19(7):393–413.
6 Vijayalaxmi, Prihoda TJ. 2008. Genetic damage in mammalian somatic cells exposed to radiofrequency radiation: a meta-analysis of data from 63 publications (1990–2005). *Radiat Res*. 169(5):561–574.
7 Ziskin MC. 2013. Millimeter waves: acoustic and electromagnetic. *Bioelectromagnetics*. 34(1):3–14.
8 Schmid G, *et al.* 2022. Design and dosimetric characterization of a broadband exposure facility for in vitro experiments in the frequency range 18–40.5 GHz. *Bioelectromagnetics*. 43(1):25–39.
9 Thiele I, *et al.* 2013. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*. 31(5):419–425. https://www.nature.com/articles/nbt.2488.
10 Brunk E, *et al.* 2018. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*. 36(3):272–281.
11 Szklarczyk D, *et al.* 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 49(D1):D605–D612.
12 Oughtred R, *et al.* 2021. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*. 30(1):187–200.
13 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B: Methodol*. 57(November 1995):289–300.
14 Nygaard V, Rødland EA, Hovig E. 2016. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics*. 17(1): 29–39.
15 Zindler T, Frieling H, Neyazi A, Bleich S, Friedel E. 2020. Simulating ComBat: how batch correction can lead to the systematic introduction of false positive results in DNA methylation microarray studies. *BMC Bioinformatics*. 21(1):1–15.

16 Price EM, Robinson WP. 2018. Adjusting for batch effects in DNA methylation microarray data, a lesson learned. *Front Genet*. 9: 338551.
17 Riches Z, Stanley EL, Bloomer JC, Coughtrie MWH. 2009. Quantitative evaluation of the expression and activity of five major sulfotransferases (SULTs) in human tissues: the SULT "pie". *Drug Metab Dispos*. 37(11):2255–2261.
18 Butcher NJ, *et al.* 2018. Sulfotransferase 1A3/4 copy number variation is associated with neurodegenerative disease. *Pharmacogenomics J*. 18(2):209–214.
19 Habauzit D, *et al.* 2014. Transcriptome analysis reveals the contribution of thermal and the specific effects in cellular response to millimeter wave exposure. *PLoS One*. 9(10):e109435.
20 Martin C, *et al.* 2024. Transcriptional landscape of human keratinocyte models exposed to 60-GHz millimeter-waves. *Toxicol Vitro*. 97:105808.
21 Lawler NB, *et al.* 2022. Millimeter waves alter DNA secondary structures and modulate the transcriptome in human fibroblasts. *Biomed Opt Express*. 13(5):3131–3144.
22 Ehrlich M, Lacey M. 2013. DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. *Epigenomics*. 5(5):553–568.
23 Zhang L, Lu Q, Chang C. 2020. Epigenetics in health and disease. *Epigenetics Allergy Autoimmunity*. 1253(5):3–55.
24 Cantu JC, Butterworth JW, Peralta XG, Payne JA, Echchgadda I. 2023. Analysis of global DNA methylation changes in human keratinocytes immediately following exposure to a 900 MHz radiofrequency field. *Bioelectromagnetics*. 44(3–4): 77–89.
25 Ravaioli F, *et al.* 2023. Evaluation of DNA methylation profiles of LINE-1, Alu and ribosomal DNA repeats in human cell lines exposed to radiofrequency radiation. *Int J Mol Sci*. 24(11): 9380.
26 Wang X-Y, Bl Z-G. 2006. UVB-irradiated human keratinocytes and interleukin-1$\alpha$ indirectly increase MAP kinase/AP-1 activation and MMP-1 production in UVA-irradiated dermal fibroblasts. *Chin Med J (Engl)*. 119(10):827–831.
27 Kaspi A, Ziemann M. 2020. Mitch: multi-contrast pathway enrichment for multi-omics and single-cell profiling data. *BMC Genomics*. 21(1):1–17.
28 World Health Organization. 2010. WHO research agenda for radiofrequency fields [accessed 2024 Apr 10]. https://iris.who.int/handle/10665/44396.
29 National Toxicology Program. 2024. Cellphone radio frequency radiation studies fact sheet [accessed 2024 Apr 10]. https://www.niehs.nih.gov/sites/default/files/NTP_cell_phone_factsheet_jan_2024_508.pdf.
30 Bullard JH, Purdom E, Hansen KD, Dudoit S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 11(1):1–13.
31 Hansen KD, Irizarry RA, Wu Z. 2012. Removing technical variability in RNA-Seq data using conditional quantile normalization. *Biostatistics*. 13(2):204–216.
32 Evans C, Hardin J, Stoebel DM. 2018. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief Bioinform*. 19(5):776–792.
33 Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Nat Precedings*. 11:1–1.
34 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15(12):1–21.
35 Risso D, Schwartz K, Sherlock G, Dudoit S. 2011. GC-content normalization for RNA-Seq data. *BMC Bioinformatics*. 12(1):1–17.
36 Spielman RS, *et al.* 2007. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet*. 39(2):226–231.
37 Petricoin EF, *et al.* 2002. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*. 359(9306):572–577.
38 Akey JM, Biswas S, Leek JT, Storey JD. 2007. On the design and analysis of gene expression studies in human populations. *Nat Genet*. 39(7):807–808.
39 Baggerly KA, Edmonson SR, Morris JS, Coombes KR. 2004. High-resolution serum proteomic patterns for ovarian cancer detection. *Endocr Relat Cancer*. 11(4):583–584.
40 Allen D, Winters E, Kenna PF, Humphries P, Farrar GJ. 2008. Reference gene selection for real-time rtPCR in human epidermal keratinocytes. *J Dermatol Sci*. 49(3):217–25.
41 Brug'e F, Venditti E, Tiano L, Littarru GP, Damiani E. 2011. Reference gene validation for qPCR on normoxia- and hypoxia-cultured human dermal fibroblasts exposed to UVA: is $\beta$-actin a reliable normalizer for photoaging studies? *J Biotechnol*. 156(3):153–62.
42 Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 29(9):e45.
43 Morris TJ, *et al.* 2014. ChAMP: 450 k chip analysis methylation pipeline. *Bioinformatics*. 30(3):428–430.
44 Aryee MJ, *et al.* 2014. Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. *Bioinformatics*. 30(10):1363–1369.
45 Fortin J-P, Triche Jr TJ, Hansen KD. 2017. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 33(4):558–560.
46 Zhou W, Laird PW, Shen H. 2017. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. 45(4):e22–e22.
47 Touleimat N, Tost J. 2012. Complete pipeline for Infinium® Human Methylation 450 K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 4(3):325–341.
48 Teschendorff AE, *et al.* 2013. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 29(2):189–196.
49 Wang Z, Wu XL, Wang Y. 2018. A framework for analyzing DNA methylation data from Illumina Infinium HumanMethylation450 BeadChip. *BMC Bioinformatics*. 19(1):15–22.
50 Wang T, *et al.* 2015. A systematic study of normalization methods for Infinium 450 K methylation data using whole-genome bisulfite sequencing data. *Epigenetics*. 10(7):662–669.
51 Marabita F, *et al.* 2013. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics*. 8(3):333–346.
52 Wu MC, Kuan P-F. 2018. A guide to Illumina BeadChip data analysis. *DNA Methylation Protoc*. 1708(4):303–330.
53 Teschendorff AE, *et al.* 2009. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One*. 4(12):e8274.
54 Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 8(1):118–127.
55 Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 28(6):882–883.
56 Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 3(1):Article 3.
57 Wettenhall JM, Smyth GK. 2004. limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics*. 20(18):3705–3706.

58 Ø Palsson B. *Systems biology: properties of reconstructed networks* Cambridge University Press, 2006.
59 Sonnenschein N, Geertz M, Muskhelishvili G, Hütt M-T. 2011. Analog regulation of metabolic demand. *BMC Syst Biol*. 5(1):1–13.
60 Sonnenschein N, *et al*. 2012. A network perspective on metabolic inconsistency. *BMC Syst Biol*. 6(1):1–13.
61 Knecht C, Fretter C, Rosenstiel P, Krawczak M, Hütt M-T. 2016. Distinct metabolic network states manifest in the gene expression profiles of pediatric inflammatory bowel disease patients and controls. *Sci Rep*. 6(1):1–11.
62 Ma H-W, Zeng A-P. 2003. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics*. 19(11):1423–1430.
63 Ma H, Zeng A-P. 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*. 19(2):270–277.
64 Kharchenko P, Church GM, Vitkup D. 2005. Expression dynamics of a cellular metabolic network. *Mol Syst Biol*. 1(1): 2005–0016.
65 Cunningham F, *et al*. 2022. Ensembl 2022. *Nucleic Acids Res*. 50(D1): D988–D995.
66 Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics*. 23(2):257–258.
67 Carlson M, Falcon S, Pages H, Li N. 2019. org. hs. eg. db: genome wide annotation for human. R package version, 3(2):3.
68 Ziemann M, Schroeter B, Bora A. 2024. Two subtle problems with overrepresentation analysis. *Bioinform Adv*. 4(1):vbae159.
69 Korotkevich G, *et al*. 2021. Fast gene set enrichment analysis. bioRxiv 060012. https://doi.org/10.1101/060012, preprint: not peer reviewed.
70 Liberzon A, *et al*. 2015. The molecular signatures database hallmark gene set collection. *Cell Syst*. 1(6):417–425.