

WP2: Review of advances in toxicogenomics with relevance for the reduction of uncertainties in EMF related cancer risk assessment

Bottom-up approach: State of the art review of current toxicogenomics applications with respect to EMF research

PD Dr. Ludger Klein-Hitpass
BioChip-Labor
Institute for Cell Biology
University of Duisburg-Essen
Virchowstr. 173
D-45122 Essen

1. Introduction and aim of evaluation

Toxicogenomics is a relatively new scientific subdiscipline that combines the technologies of toxicology, genomics and bioinformatics to identify and characterize mechanisms of action of known and suspected toxicants. Currently, the most widely used toxicogenomic tool is clearly the DNA microarray, which is commonly known as GeneChip[®] (a registered trademark of Affymetrix, Inc.) or genome chip, DNA chip, biochip, or gene array. The terms microarray and macroarray are used interchangeably, but the term macroarray is frequently used to differentiate arrays with relatively large spot sizes and fewer spots or gene probes on the support. Microarrays have been used successfully in a wide range of applications, such as sequencing, detection of single nucleotide polymorphism (SNP), and identification of chromosomal aberrations in tumors. However, the main application of microarrays is to measure the steady-state mRNA levels of nearly all well-characterized genes in cells, normal and diseased tissues or whole organs from a number of species, including human, mouse and rat. By analyzing the effects of known toxic and potentially toxic compounds on the expression of nearly the whole genome, researchers aim at identifying toxicant or toxicant-group-specific gene signatures or patterns, which indicate exposure, initiation of toxic events and mechanisms involved.

In this evaluation, I will introduce the whole-genome gene expression analysis technologies, gene expression studies and critically evaluate peer-reviewed publications using microarrays in EMF research.

1.1 Microarray systems

Microarray systems consist of three key components: the microarray, labelled samples and a detection system that quantifies the hybridization signal. The first component, the microarray is in its conventional form the result of continued development and miniaturization of previously used macroarray platforms, such as filter-based dot blots or slot blots with nitrocellulose or nylon membranes. Such microarrays consist of a collection of nucleic acid sequences immobilized onto a solid support so that each feature forms a tiny spot or square. There are different types of nucleic acids (cDNA fragments amplified by polymerase chain reaction or synthetic oligonucleotides) and different ways for depositing the nucleic acids onto the support. While home-made cDNA arrays printed in few specialized arraying facilities on nylon membranes or glass slides enabled already some years ago researchers to perform first nearly genome-wide gene expression studies, today oligonucleotide arrays produced by commercial manufacturers are the preferred type of array, as large cDNA clone collections and cost-expensive arraying facilities are no longer required, thereby allowing even smaller institutes/facilities to perform microarray analyses. While the relative large cDNA fragments or purified synthetic oligonucleotides have to be spotted or printed onto the solid support by robots with some distance in between the spots, the most dense packing of very small features can be reached using in situ synthesis of short single-stranded oligonucleotides (25-mers) by a photolithographic procedure. The latter method was invented and patented by Affymetrix Inc., the current leader in the microarray market, and allows to array up to 6×10^6 features each containing millions of DNA molecules on an area of less than 2.5 square centimeter. These oligonucleotide microarrays, which are produced at an industrial scale, show much less batch-to-batch variation than self-printed or –spotted cDNA arrays and have become an attractive, cost-effective and reliable alternative. Compared to cDNA arrays, oligonucleotide arrays have additional benefits as different parts of the same gene can be represented on the array, enabling a more robust design and analysis of the signals. Moreover, using carefully designed oligonucleotide probes, alternatively spliced transcript variants, different alleles and closely related members of gene families can be distinguished. Current whole-genome array generations for human samples offered by Affymetrix contain so-called probesets interrogating some 54,000 transcripts derived from about 27,000 genes, thereby covering a large part of the protein-coding genes of the genome. Each probeset contains at least 11 perfectly matching oligonucleotide probes and a corresponding number of mismatch oligonucleotides, which serve as controls.

The second key component of microarray analysis is the labelled sample or samples that are hybridized with the microarray. The samples to be analyzed in a gene expression study are either enriched polyadenylated mRNA preparations or total RNA preparations. For analysis, the mRNA has to be converted into a labelled population of nucleic acids, the 'target'. While in earlier days, the targets were labelled by radioisotopes, such as P^{32} or P^{33} , most current microarray analysis protocols have adopted fluorescent dyes as the predominant label. In the most widely used labelling procedure, first strand cDNA synthesis by Reverse Transcriptase is primed with a modified oligo-d(T)₂₁ primer, that includes a T7 RNA Polymerase promoter sequence. Thus, after synthesis of the complementary second cDNA strand, all double-stranded cDNA molecules carry a functional T7 RNA polymerase binding site at their 3'-end, thereby allowing incorporation of biotin-modified nucleotides into the copyRNA (cRNA) generated by in vitro transcription with T7 RNA polymerase. Since the polymerase transcribes each cDNA molecule many times, about 100- to 1000-fold linear amplification of the original mRNA is achieved, while the representation of the different mRNA species is maintained.

The labelled targets can then be hybridized to the array. During hybridization, the labelled nucleic acid fragments in the probe form duplexes with the immobilized complementary sequences on the array. To avoid binding of non-complementary fragments, hybridization conditions have to be tightly controlled and sequences to be deposited on the array have to be selected carefully to represent unique sequences. During hybridization single-stranded labelled fragments in the sample of interest are expected to form duplexes with their immobilized complementary sequences on the array. As long as the amount of immobilized nucleic acid is in excess and not limiting the kinetics of hybridization, the number of duplexes formed reflects the relative number of each specific fragment in the sample. Thus, by measuring the amount of label associated with each spot, the relative abundance of specific sequences in each of the samples can be determined and compared to other samples.

The third key component of microarray analysis is the detection system that quantitates the hybridization signal. For detection of gene expression by macroarrays using radioactively labelled RNA samples, bound nucleic acids hybridizing to the immobilized DNA fragments can be identified by exposure to X-ray films, followed by densitometric quantification. Since x-ray films have only a limited dynamic range, several different exposures have to be taken and analyzed in order to obtain a set of films in which abundant and rare transcripts can be both analyzed with some accuracy. Alternatively, bound radioactive label can be quantified

more conveniently by phosphor-imaging. However, due to the hazards of radioactive isotopes, relative low sensitivity, quite long exposure periods and the difficulties associated with the densitometric analysis of X-ray films, radioactive labeling has been overcome by labeling with fluorescent dyes. Moreover, by labelling control and experimental samples with two different dyes, e.g. Cy3 and Cy5, fluorescence labelling allows simultaneous hybridization of both samples to a single array. In the case of Affymetrix GeneChip arrays, after stringent washing steps eliminating unbound cRNA fragments, bound biotin-modified cRNA molecules have to be visualized through a post-hybridization staining step with a streptavidin-conjugate carrying a fluorescent dye. To measure the amount of fluorescence associated with the spots on the array, a microarray laser scanner is required, in which the bound dye is excited at a specific wavelength and the emitted light is quantitated with the help of a high-resolution confocal microscope and a photo-multiplier tube. This method allows to monitor 1000-fold differences in expression levels and is sufficiently sensitive to detect a single copy of mRNA per 100,000 RNA molecules (Affymetrix GeneChip system). Depending on the cells or tissues analyzed, this level of sensitivity allows detection of transcripts from 40-55% of the genome. Chemiluminescence detection appears to yield a similar, if not superior sensitivity. Here, a digoxigenin-labelled nucleotide is incorporated into cDNA or cRNA, which can be stained by an antibody-enzyme conjugate, that will produce light upon addition of a chemiluminescence substrate. Both fluorescence and chemiluminescence labelling techniques have the advantage that more samples can be analyzed in parallel, increasing the throughput considerably.

The raw images acquired by densitometry, phosphorimaging or laser scanning have to undergo a number of quality control and processing steps in order to obtain a final result. These steps include e.g. visual inspection of the images to eliminate arrays showing damaged areas or high background, spot finding or grid alignment, background determination and subtraction, spot quantification and normalization procedures, which compensate for differences due to unequal loading of the chip and other variables of non-biological origin. Finally, the goal of image processing is to compute a unique value that hopefully is directly proportional to the quantity of mRNA present in the solution hybridized to the chip.

On Affymetrix GeneChips for expression analysis, genes are represented by a so-called probesets consisting of 11 to 20 pairs of perfect match (PM) and mismatch (MM) oligonucleotides interrogating some 300 to 600 bp within the 3'-end of the known sequence of the gene. The PM probe is the exact complement of a 25 bp subsequence of the target gene

and is supposed to bind the labelled RNA derived from the mRNA in the sample. The MM probe is identical to the corresponding PM probe, except that the central base is changed to decrease the affinity to the gene's labelled cRNA. The signal measured on the MM probes is supposed to largely represent non-specific hybridization, which is subtracted from the PM signal during the signal summarization step in the Affymetrix image analysis software (MAS5 algorithm). Next, the signal differences of the 11-20 probe pairs are summarized by a one-step Tukey biweight statistic to yield the final quantification value that is proportional to the level of expression of the gene (Hubbell et al, 2002). Besides the desired quantitative information, the analysis of Affymetrix GeneChip images delivers also a qualitative information, which is captured in a set of "detection calls". A high proportion of probe pairs showing positive signal differences favours a "Present" detection call indicating faithful detection of the gene, whereas probesets showing low and intermediate proportions of positive probepair differences tend to be tagged with "Marginal" or "Absent" calls. In many studies, this qualitative information is taken into account during the data mining process leading to the identification of differentially expressed genes.

Another whole genome expression analysis system, that is completely different from the above described array technologies, is serial analysis of gene expression (SAGE), which was first described by Velculescu et al. (1995). In principle, through a series of steps mediated by a variety of enzymes, short (10-14 bp) sequence tags are obtained from the transcripts. This short sequence contains sufficient information to identify a transcript, provided that the tag is derived from a unique sequence position within that transcript. Many sequence tags are then linked together to form long serial tag concatemers which can be cloned into a plasmid vector and sequenced. The gene expression level is then obtained by quantitation of the number of times a particular unique tag is found in the cloned tag library. While the SAGE principle is quite simple, optimization of each step is a very demanding task and requires excellent laboratory skills. Since SAGE allows the quantitation of expressed sequences not previously known, it is particularly suited for studies in organism for which no complete genome and hence no array available. However, since generation of SAGE libraries is clearly not a high-throughput technique, it has been established successfully only in very few specialized laboratories. Due to the large number of genes and the wide range of expression levels, it is obvious that large clone libraries have to be sequenced in order to obtain a faithful representation of the transcribed genome and to detect low-abundance transcripts. Moreover, identification of differentially expressed genes by SAGE requires tag libraries of approximately equal size and specifically developed statistical methods (ref). Differential

expression analysis by SAGE is hampered by the fact, that SAGE libraries can show large intra- and inter-individual variations (Lu et al., 2005).

2. Key challenges and solutions in using microarrays in gene expression studies

2.1 Identification of regulated genes

One of the greatest challenges that researchers performing microarray studies have to face is noise or experimental variation, which introduces unpredictable fluctuations in the measurements resulting in statistical errors. Noise can be introduced through subtle variations at all steps of the analysis and through all reagents, including the sampling process (cell sorting, tissue or organ preparation), cell culture conditions, RNA preparation, target labeling, hybridization, scanning, arrays and reagent batches, pipetting errors, as well as through variations of the experimental treatment itself. In whole animal studies, additional sources of variation can arise from hormonal fluctuations, unidentified infections, stress and circadian rhythms. In EMF research, inhomogeneities of the field or inaccurate dosimetry may represent another source of variation. One way to limit noise is the development and strict use of standard operating procedures (SOP), which also ease comparison of datasets generated with the same microarray system in different laboratories. However, even when the same target preparation resulting from a single RNA preparation is hybridized to two microarrays and processed using highly standardized operating procedures, many genes will show different quantification values, particularly those showing signals close to background level. The challenge appears, when comparing control cells with treated/exposed cells, since it is per se not clear, whether a measured signal difference is due to a genuine change of the mRNA level resulting from the treatment/exposure or simply the result of noise. When designing microarray studies, it is most important to consider this problem appropriately by planning sufficiently high replicate numbers, as replication will allow an estimation of the degree of experimental variability, which is evaluated in suitable statistical tests. Moreover, an increased number of replicates also reduces the probability of missing true positives (type II error).

The fact that by using microarrays the expression of tens of thousands of genes can be analyzed in parallel in many samples at a time is fascinating and has led to the wide adoption of this technology in many areas of the life sciences. However, the large number of variables

(genes) measured represents a major challenge, as appropriate data mining strategies and new statistical methods had to be developed, to identify truly regulated genes. Most importantly, classical statistical techniques (e.g. t-test, Mann-Whitney or Wilcoxon rank test) designed to answer the crucial question whether expression of genes differs significantly between the groups considered, can not be applied directly because in microarray experiments the number of variables (several ten thousands of genes/probesets) is far greater than the number of experiments performed. This multiple testing problem is best illustrated by analyzing the result of a data table representing the signal values obtained from six replicate analyses originating from a single RNA sample. If the 6 samples are split into two experimental groups and subjected to the t-test, the test will yield approximately 2,500 (5%) variables with significantly different mean signals at $p \leq 0.05$, although no true differences can exist (type I error). If the experimental arrays exhibit a similar level of variability, a comparable fraction of false positives has to be expected along with the genes which represent true positives, when experimental arrays are analyzed against the control arrays. Thus, a list of genes showing significance ($p < 0.05$) for about 5% of all genes subjected to significance testing does not necessarily indicate that the experimental parameter has indeed caused gene expression changes, a fact that has been overseen in a number of microarray studies, including EMF research (Mayo et al., 2006).

To deal with the multiple testing error problem, several more general multiple testing adjustment methods (Bonferroni correction, Westfall-Young step-down approach, Holm-Bonferroni method a.o.) have been developed. However, with 2-3 replicates and multiple testing correction, traditional statistical methods often fail to find any significant changes, even when known truly regulated genes are present in the dataset.

A widely accepted alternative approach, which provides a better balancing of both error rates than classical t-tests with multiple testing correction in microarray studies, is the Significance Analysis of Microarrays (SAM) developed by Tibshirani and coworkers (Tusher et al., 2001). In a first step, SAM computes a statistic for each gene. Then, the statistic is compared to the distribution of statistical testing of random permutation of the sample categories. For each permutation test, a certain proportion of all genes in the permutation set (control set) will be found to be 'significant' by chance and this parameter is then used to calculate a "False Discovery Rate" (FDR). This is presented as a q value for each gene in the final list of significant genes, which defines the expected percentage of false positives among the significant tests and, thus, provides a means to discriminate truly significant genes from

genes, which are likely to represent false significant candidates. As demonstrated by analysis of model datasets, the SAM method is much less conservative than classical t-tests with multiple testing correction and identifies expected networks of regulated genes in model datasets. However, applying pre-analysis exclusion criteria based on detection call thresholds as well as post-hoc FDR thresholds and fold change filtering can have an impact on the outcome and the biological interpretation of SAM results, as obtained by gene ontology overrepresentation analysis (Larsson et al., 2005). While this is not a problem unique to SAM results, it is recommended to reanalyze datasets by SAM using a range of thresholds and exclusion criteria. The SAM tool is relatively easy to use and freely available for academics. Importantly, it should be noted that FDR estimation by SAM requires at least 5 arrays in each class, since the estimation of variance is dependent on a large number of truly different permutations of the datasets. A similar permutation approach to correct for multiple testing is used in Microarray Analysis of Variance (MAANOVA, Kerr et al., 2000), which has, however, not reached the broad acceptance as SAM. MAANOVA uses a modified ANOVA method to provide estimates of changes in gene expression and provides normalization methods, which can adjust potential confounding effects, such as array batch effects.

A widely used approach to classify genes as differentially expressed in studies involving few replicates is the application of an arbitrarily chosen constant ratio, commonly a 1.5- or 2-fold change, often in combination with a p-value threshold. While this method has led to successful identification of truly regulated genes in many studies, it does not effectively remove false positives associated with low signals (Mariani et al. 2003) and eliminates all potentially interesting candidates showing changes below the threshold value. In studies using Affymetrix GeneChip, it has been recommended to make use of the qualitative information obtained with the detection call. Here, exclusion of probesets called “Absent” in the majority of the arrays of a study is supposed to reduce the fraction of false positives in the list of significant genes considerably. The effectiveness of such exclusion strategies can be estimated by applying them to datasets obtained from intragroup comparisons of experimental and/or control arrays. However, whatever action is taken, it is not possible to determine precisely, how effectively a chosen exclusion strategy removes false positives in the intergroup comparisons, in which experimental arrays are compared with controls. In general, any kind of exclusion strategy or threshold has to be designed very cautiously, to avoid introduction of a bias, severe loss of sensitivity and statistical power and possible impact on the biological interpretation.

Affymetrix microarray studies involving few replicates frequently analyze the data by using the pair-wise comparison analysis algorithm implemented in the Affymetrix MicroArray Suite v5 (MAS5). In such a pair-wise comparison of experimental versus baseline (control) array, the underlying algorithm performs a Wilcoxon test using the 11 PM-MM signal differences of a given probeset on the two arrays to be compared. The resulting p-value is compared against empirically determined threshold p-values to yield Increase (I), Marginal Increase (MI), Decrease (D), Marginal Decrease (MD) or No Change (NC) calls. Studies involving two or three arrays each in the experimental and control group can perform pair-wise cross-comparisons yielding 4 or 9 comparisons, respectively, to identify those probesets showing concordant Increased or Decrease Change calls. Due to the multiple testing situation, the Wilcoxon test will produce a considerable number of false positives. However, the probability to pick up a gene consistently in 100% of all cross-comparisons by chance is relatively low and can be estimated by multiplying the percentages of genes recovered in each individual comparison analysis. Thus, identification of concordantly regulated genes is a practicable data mining strategy that can produce lists of genes, which contain candidates most likely representing true positives. However, if 'concordance' is defined conservatively, the risk to miss truly regulated genes (type II error) is increased. Finally, the reliability of this method depends on the number of cross comparisons performed.

2.2 Recommended experimental design and validation strategies

Due to the type I error problem, all microarray projects, including those that aim at identifying effects of exposure to hazardous compounds, ionizing or non-ionizing radiation, which involve low replicate numbers, are very likely to end up with an apparently positive result, i.e. list of genes which are apparently regulated by the treatment. Depending on the technical quality of the arrays, the data mining strategy and applied thresholds, these candidate lists contain a variable fraction of true positives, but in the worst case may consist of false positives only. This worst case scenario is particularly unsatisfying, because it may lead to completely wrong conclusion about the potential health risk associated with a particular compound or type and dose of radiation.

The best way to handle this problem is to increase the number of replicates of independent samples to six or more, because this will allow the use a permutation based statistics, such as

SAM, which handles the type I error problem in an elegant way and can evidently produce the expected negative results, if e.g. 12 identical biological replicates are split into two groups. In this way, a genome-wide microarray experiment involving six or more replicates, in which no further signal or fold change thresholds are being applied pre- or post-SAM analysis, represents a most comprehensive and sensitive analysis which produces target gene lists with low rates of both error types. However, even when this “Gold Standard” is met, it is advisable to validate a number of candidate genes identified by SAM analysis by alternative methods. Various methods with different sensitivity and dynamic range can be used to validate and extend the initial observation made in microarray studies, including RNA quantification methods, e.g. Northern Blotting, semi-quantitative RT-PCR, real-time RT-PCR, and various protein detection methods, e.g. Western Blotting and immunocytological techniques. The method of quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) has been proven as the gold-standard in determining transcript levels, because it displays a much broader dynamic range than microarrays, is clearly more sensitive and allows parallel analysis of quite a number of genes and samples at an affordable price.

Importantly, by subjecting an increased sample number of experimental and control samples to qRT-PCR analysis, a more robust statistical analysis proving or disproving the initial observation in the microarray experiment can be obtained. Moreover, by choosing a considerable number of potential candidate genes for validation, a more precise estimation of the fraction of false positives observed in the microarray target gene list is obtained. Clearly, independent samples should be analyzed, because by using the same samples hybridized to arrays one validates only, that the microarrays “worked”. However, what one wants to validate is that the genes are differentially expressed in the “population”, which can only be done by use of independent samples. If appropriate statistical analysis of the PCR data confirms differential expression of the candidate gene, this should be taken as conclusive evidence for an effect on gene regulation.

The costs of a microarray experiment using commercially available arrays are somewhere in the range of 500 to 1000 € per sample. Thus, following the “Gold Standard” approach, a single EMF exposure experiment involving six controls and six experimental replicate samples may cost from 6000 to 12000 €. Considering the fact that a thorough analysis of EMF effects will require a great number of experimental variables (cell type, SAR, frequency, modulation forms, exposure time, recovery time) and combinations thereof to be studied, such a systematic investigation will require large funds. An alternative approach may thus involve

only three or less independent microarray analyses, which can be analyzed by pair-wise cross comparison analysis or classical statistics, in which the application of multiple testing correction methods should be carefully studied. If no signal or fold change thresholds and no multiple testing correction methods are being applied, this analysis will retain a high level of sensitivity, but produce a relative high rate of false positives, which may obscure the recognition of a certain gene pattern and clues to the mechanisms involved. However, in a second step a broad validation approach should be undertaken, by analyzing a larger number number (>10) of independent samples in quantitative PCR experiments. Moreover, validation should involve a relatively high number of candidate genes obtained from the microarray experiment, as this will allow an estimation of the false positives rate in the microarray and of the probability to find a true positive target in the remaining list of unvalidated candidate genes.

2.3 RNA and target quality assessment

A quite important step in microarray analysis is the preparation of intact RNA samples devoid of DNA, proteins and other compounds. RNA is relatively unstable and can be degraded either physically, chemically or enzymatically by ribonucleases, which are present in all cells and be introduced easily back into purified RNA samples from skin or hair. Thus, care has to be taken to limit the degradation of the RNA during cell or tissue harvesting, to inactivate these enzymes most rapidly during homogenization of the sample, to avoid contamination of the purified RNA samples with ribonucleases from external sources by using ribonuclease-free solutions and plastic ware, and to limit non-enzymatical degradation through proper storage and handling. As RNA degradation occurring in the sample prior to and after RNA extraction can greatly affect the data, it is highly recommended to verify the quality of each RNA preparation prior to target preparation either by agarose gel electrophoresis or on Agilent Bioanalyzer chips. In intact total RNA samples separated on agarose gels stained with ethidium bromide, the ratio of the two predominant RNA species, the large and small ribosomal RNAs, should be close to 2. Bioanalyzer chips, which represent a more sensitive and quantitative alternative, allow a more sophisticated estimation of a ribosomal RNA integrity number (RIN), which should be between 8 and 10. Due to target labeling and amplification procedures, which mostly involve reverse transcription primed with a poly(dT)₁₈₋₂₁ primer binding to the polyA tail at the 3'-end of the mRNAs, more 5' located

mRNA sequences will be underrepresented in partially degraded RNA samples (3'-bias). Thus, if hybridized to an array and compared with an intact sample, the partially degraded sample will artificially show many down-regulated genes. As shown by Klein-Hitpass and Möröy (2005) with a GeneChip harboring 22283 probesets, an increase of the 3'/5'-ratio of the GAPDH mRNA introduced by controlled heat fragmentation of an intact RNA from 0.8 to 0.99 led to the identification of more than 1600 apparently regulated genes showing average fold changes ranging from 1.5- to 10-fold.

Comparably high quality of RNA preparations is a prerequisite for a good microarray study, however, it may not guarantee unbiased results, since incomplete first strand cDNA synthesis due to the presence of inhibitors of the Reverse Transcriptase may also lead to 3'-biased target preparations. Thus, it is also recommended to assess the quality of the labelled targets prior to hybridization by analyzing their size distribution by gel electrophoresis or Bioanalyzer chip. Moreover, the addition of several external spike-in transcripts is recommended, which allow to control the efficiency of the reverse transcription step, the sensitivity and linearity of the assay and the success of the data mining step in identifying a transcript spiked-in at known different concentrations. It is worth mentioning that the popular Affymetrix GeneChip system offers built-in control checkpoints, as it allows post-hybridization target quality evaluation by estimating the representation of 3'-, middle and 5'-sequences of two housekeeping genes and a set of external spike-in controls. By comparing the 3'/5' signal ratios of different samples, poor targets can be identified and the corresponding arrays can be excluded from further analysis. Moreover, a possible quality bias in control or treated groups can be identified by subjecting the 3'/5'-ratios to statistical analysis. If those 3'/5'-ratios were reported explicitly in microarray studies using the Affymetrix system, the quality of the hybridized sample could be judged immediately.

2.4 Biological interpretation of the data

Due to the great number of genes affected by treatment exposure and the possible presence of false positive genes, the interpretation of gene expression changes observed in microarray studies represents another major challenge. In many cases, interpretation is also hampered by the fact that many microarray studies are not driven by a hypothesis of what the outcome is likely to be, but rather by the simple question of whether a certain compound or treatment may alter gene expression or not. By using clustering methods, class prediction and other

techniques, it is possible to correlate expression profiles induced by unknown toxicants with profiles obtained by known toxicants to identify common pattern or gene signatures characterizing certain groups of compounds. However, if such reference expression profiles are not available or a novel type of toxicants is analyzed, one of the first steps towards an understanding of the toxicant mechanism is an analysis of the molecular function, cellular localization and biological processes associated with the genes affected by the toxicant under investigation. In this context, the Gene-Ontology database (www.geneontology.org) provides a useful tool to annotate and analyze large numbers of genes. Analysis of gene lists is also conveniently achieved by a number of software tools, such as Gostat (Beissbarth and Speed, 2004), GoMiner (Zeeberg et al. 2003), GSEA (Subramanian et al., 2005) or GeneTrail (Keller et al., 2007). Importantly, these tools also provide statistical algorithms (hypergeometric distribution, Fisher's exact test) including multiple testing correction methods to analyze whether a list of identified candidate genes is significantly enriched for genes associated with a particular GO term (e.g. apoptosis, cell cycle, cellular defense). The identification of significantly overrepresented GO terms within target gene lists provides strong hints towards biological processes potentially affected by the treatment/exposure, due to the activity changes resulting from the observed changes at the mRNA levels. In contrast, the mere sorting of genes according to most frequently observed GO terms is of limited value only, since any list of genes picked by chance might yield similar results. Moreover, the sorting approach is prone to bias, if only GO terms of interest being analyzed.

GO enrichment analysis can be further supplemented by so-called pathway analysis which focuses on the identification of regulated genes encoding members of signaling, metabolic or disease pathways. Differential expression of pathway constituents may alter its activity and therefore affect the processes regulated by the pathway. Most pathway analysis programs provide statistical approaches to determine whether there is a significant enrichment of regulated pathway constituents within the target gene list analyzed. Similarly, Gene Set Enrichment Analysis provides statistical means to demonstrate significant enrichment of gene lists for specific sets of genes, e.g. genes known to be regulated by a specific transcription factor. A gene list enriched for P53-regulated genes would indicate that the experimental treatment most likely led to activation of P53, thereby providing some understanding of the mechanism responsible for observed gene expression changes.

2.5 Microarray data accessibility

Since each microarray experiment produces ten thousands of data points, it is impossible to provide the entire information along with the paper in scientific journals. Thus, another challenge in microarray studies is the improvement of the accessibility of microarray data, since a fundamental principle guiding the publication of scientific results is that the data must be made fully available to the research community in a form that allows the basic conclusions to be reevaluated independently. In line with this, many high impact journals require the submission of whole genome expression analyses, as obtained by microarrays or SAGE, to public data repositories as part of the publication process and as a condition of acceptance. Public data repositories, such as Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo/), ArrayExpress (www.ebi.ac.uk) or CIBEX (<http://cibex.nig.ac.jp>), provide a solution to make data sets permanently available to the scientific community. These repositories allow deposition of microarray and SAGE data in standardized formats following MIAME (Minimum Information About a Microarray Experiment) guidelines, including final processed (normalized) gene expression signals for all genes present on the array, experimental factors (treatment, dose, compound), experimental design (number and type of replicates), array platform and annotation, the essential laboratory and data processing protocol and the raw data files for each hybridization, e.g. CEL (Affymetrix GeneChip images) or GPR (GenePix images) files. As indicated on the GEO homepage, this largest database stores more than 170000 array experiments ordered in more than 6700 series, which represents an extremely valuable resource for gene expression data browsing and query. Retrieval of the raw images permits detailed reevaluation of array image and hybridization target quality, data processing steps and statistical analysis as well as studying the impact of alternative signal summarization, normalization steps or statistical methods on the biological interpretation of the experiment. It is highly recommended that all studies involving microarray studies make their raw data accessible to the public through submission to the databases.

3. Evaluation of studies using microarrays in EMF research

For this review, only original papers published up to May 2008 in peer-review journals were considered. Papers were searched through Pubmed (<http://www.ncbi.nlm.nih.gov>) and the database provided by the IMBA project using various keywords, including EMF,

electromagnetic field, and radiofrequency alone or in combination with the terms microarray, array, chip, whole genome analysis, and gene expression. In the first part (chapter 3.1), papers were evaluated based on some experimental and technical aspects, e.g. cell type used, exposure conditions and array system used. The second part (chapter 3.2) contains a critical evaluation of how these studies deal with the major challenges in microarray studies as defined in chapter 2.

3.1 Models and microarray systems used in EMF research

As shown in Table 1 and 2, 15 microarray studies in EMF research were evaluated, published in the period between 2003 and 2008, with the bulk of papers (13/15) appearing in the years 2006 and 2007. *Proteomics* (4 studies, impact factor 5.735), *Radiation Research* (3 studies, impact factor 2.602) and *Bioelectromagnetics* (4 studies, impact factor 1.514) were the most frequently chosen journals for publishing.

Table 1 and 2 summarize a number of experimental details the EMF studies identified, including sample type, exposure details and microarray system used. The majority of studies (11/15) used cell lines originating from a very broad range of tumors, such as breast cancer, leukaemia, glioblastoma, neuroblastoma or immortalized cells, which can be easily maintained in the lab and provide a convenient source for experiments. In 4 studies, primary neuronal cultures from rat or mouse and human T-lymphocytes from peripheral blood were exposed to EMF ex-vivo. Two studies study (Belayaev et al, 2006; Paparini et al., 2008) analysed an intact tissue (cerebellum or whole brain) isolated from rats exposed to EMF in-vivo. In the Belayaev study the in-vivo exposure setup may represent a source of variation because the animals were not restrained in the exposure device. Moreover, the cerebellum preparation may have caused additional variation, since individual cerebellum samples might contain variable fractions of contaminating cells or tissues. The relative frequent use of primary cells or cell lines derived from brain may reflect the fact that specific brain regions may receive relatively high field exposures during mobile phone usage, however, a rationale for choosing other cell systems is rarely presented.

Similarly, a wide range of exposition parameters is evident from Table 1. While most studies used SARs up to 2 W/kg, which represents the recommended safety limit based on ICNIRP guidelines, some studies investigated SARs above the safety limit only (Lee et al., 2005;

Nylund and Leszczynski, 2006; Whitehead et al., 2006). The most frequently selected field frequencies were close to 900 or 1800 MHz, reflecting the frequencies used in mobile phone communication. However, these major field frequencies were combined with a number of different modulation forms supposed to mimic phone conversations more closely, a large range of exposure time periods from 1 to 72 h and recovery periods of 2 or 6 h after exposure (Gurisek et al., 2006; Chauhan et al., 2007), resulting in a large variety of unique exposition conditions. One study (Zhao TY et al., 2007), which simply used a mobile phone as exposure device, did not report any SAR, but it seems unlikely that the safety limit was exceeded in this case. All but two of the cell culture studies (Zhao TY et al., 1997; Gurisek et al., 2006) report explicitly, that the temperatures of the samples were kept within $37\pm 0.5^{\circ}\text{C}$, making thermal effects on gene expression unlikely.

To demonstrate that the microarray technique used and the data mining strategy applied to EMF exposed cells can pick up expected expression changes, it is reasonable to consider some kind of positive control in the experimental design. Four EMF studies (Qutop et al., 2006; Hirose et al., 2006; Hirose et al., 2007; Chauhan et al., 2007), chose heat-treated cells as a positive control, while Whitehead et al. (2006) selected X-ray treated cells. In these studies, heat and X-ray treatment resulted in identification of induced heatshock protein genes and some genes previously reported to be induced by X-ray in other studies, respectively. However, in light of the multiple testing problem, those controls would be even more convincing, if by subjecting the entire lists of heat or X-ray targets to GO term enrichment analysis, a significant overrepresentation of genes involved in stress or DNA damage response could be confirmed.

Concerning the labeling and detection system used, 11/15 microarray studies relied on the more sensitive fluorescence- or chemiluminescence-based detection method ensuring relative high sensitivity, as indicated by a high fraction of probesets (39%) showing Present detection calls in the study of Hirose et al. (2007). Three studies (Port et al., 2003, Nylund and Leszczynski, 2006; Remondini et al., 2006) used radioisotope labelling in combination with densitometric analysis of X-ray films or phosphorimaging for their analysis. As exemplified by the study of Nylund and Leszczynski (2006), who detected only approximately 100 genes per cell line analyzed, this method shows relatively poor sensitivity and limits the analysis to high abundance transcripts. The sensitivity of the study of Remondini et al. (2006) cannot be estimated, since it is not stated how many genes were discarded by the pre-processing steps (background removal, exclusion of poor quality spots).

Among the array system used, the Affymetrix GeneChip system is clearly the most frequently used (8/14 studies), reflecting the broad distribution and acceptance of this technology. However, the various chip types used represent different chip generations and types with a genome coverage ranging from 1200 (Zhao R et al., 2007) up to 39000 genes (Hirose et al., 2006, 2007). Two studies (Qutop et al., 2006; Chauhan et al., 2007) used spotted oligonucleotide arrays from Agilent, which also interrogate a large number (18000) of genes. The study of Remondini et al. (2006) used a nylon macroarray, which contained as many as 75000 different cDNA fragments spotted onto 3 subarrays each 22x22 cm in size and therefore also covered a large part of the genome. In contrast, Zhao TY et al. (2007) used an array, that covered only 96 apoptosis-related genes. Quite obviously, this theme-specific array does not cover a significant fraction of the mouse genome and does not allow an unbiased analysis of the processes potentially altered upon EMF exposure. Moreover, the studies using the Clontech Atlas array covering only 1176 genes (Port et al., 2003; Nylund and Leszczynski, 2006), the Rat Neurobiology array (~1200 genes; Zhao R et al., 2006) can also hardly be regarded as whole genome analysis, since only 5-10% of the expressed genome were covered.

3.2 How do published paper in EMF research meet the major challenges of microarray studies?

As discussed in chapter 2.1 to 2.5, there are some major challenges in performing and analyzing microarray studies. Based on the considerations presented above, I will evaluate how each publication presenting microarray data in EMF research deals with these challenges. Special attention will be paid to the most important aspects, i.e. experimental design and identification of regulated genes as well as experimental validation of candidate genes.

Port et al. (2003) performed two independent experiments with one replicate each at 4 different time points following EMF exposure. Though not documented, RNA samples as well the labelled hybridization target had been subjected to appropriate quality checks. Based on tests in which the reproducibility at different signal thresholds was analyzed, all genes showing an intensity value below 10000 were adjusted to a value of 10000, however, it is not mentioned, how many genes were excluded by this strategy. The ratios of adjusted intensities were calculated and differentially expressed genes were identified by applying a change

threshold of 2-fold. Depending on the time point analyzed, only 0 to 6 candidate genes were identified by this approach. Since not a single candidate gene appeared to be regulated concordantly in both independent experiments performed, the authors conclude that no reproducible differential gene expression could be demonstrated. As indicated by the small number of regulated genes showing fold changes greater than 2-fold, the signal adjustment strategy in combination with the relatively insensitive technology used (radioactively labelled targets, array representing 1176 genes, densitometric analysis of X-ray films), may have excluded many genes from analysis. Thus, this study hardly represents a sensitive whole genome approach. Due to the negative outcome, experimental validation and a statistical approach to data interpretation were unnecessary. The microarray data has not been deposited in a public database.

Using the SAGE method, which is known to show significant within and between tag library variation (Lu et al., 2005), the study of Lee et al. (2005) identified 221 and 759 genes with fold changes >4 at 2 h and 6 h by EMF. While these numbers were obtained from single SAGE analyses, 84 genes showed concordant changes at both time points. Since the number of common genes is significantly larger than the number expected by chance, these genes might represent at least in part truly regulated genes. Genes identified were apparently subjected to a GO term enrichment analysis and various processes were found, including apoptosis and cell cycle regulation. Remarkably, the data has been deposited in the NCBI GEO database. Unfortunately, the study does not demonstrate how reproducible their SAGE analysis is and no effort is made to validate candidate genes by a different technique. Moreover, since the SAR applied to the cells (10 W/kg) was far above the existing threshold values, the relevance of this study for EMF risk assessment remains questionable.

Nylund and Leszczynski (2006) performed a microarray study, in which three controls and exposed arrays were analyzed for two different cell lines. RNA sample and hybridization target quality checks are not mentioned. Due to the limitations of the technology used (radioactively labelled targets, array representing 1176 genes, densitometric analysis of X-ray films), only 136 genes out of 1167 analyzed remained after exclusion of genes showing signal intensities below a threshold of 3000 in both cell lines. While the average fold change exceeded 2-fold for 93 and 63 genes, only 2 and 12 were significant in a t-test without multiple testing correction at $p \leq 0.05$ in EA.hy929 and EA.hy929v1, respectively. Clearly, the number of candidates identified was lower (EA.hy929) or only 2-fold higher (EA.hy929v1) than the expected number of type I errors. Although EA.hy929 and

EA.hy929v1 cells represent the same line obtained from two different sources, not a single gene proved to be significant in both cell lines. A statistical approach to data interpretation was obviously not feasible due to the limited number of significant genes. The dataset has not been fully published or deposited in a public database and, therefore, cannot be reanalyzed independently. Since no detailed information on the reproducibility of the array analyses is given and no experimental validation is presented at all, it cannot be ruled out that the list of regulated genes contains only few, if any, true positives. This study, which comes to the conclusion that EMF causes broad changes in gene expression, ignores the multiple testing problem completely and, therefore, represents a typical example of inappropriate data analysis and interpretation.

Whitehead et al. (2006) also analyzed 3 replicates and controls each for two differently modulated EMF fields (CDMA and FDMA). Since the unexposed controls for the CDMA and FDMA experiment should be essentially identical, the authors used these 6 control replicates to estimate their false positives rate by performing t-tests on permuted datasets. As recommended by the Affymetrix data analysis fundamentals manual, the authors excluded all probesets, which did not show 3 Present calls in either controls or exposed arrays, to reduce the noise associated with probesets showing Absence calls only. Analysis of CDMA and FDMA exposed samples revealed even slightly less significant probesets with a fold change >1.3 than the average false positives rate estimated by analysis of permuted control arrays. Thus, the authors came to the correct conclusion that EMF exposure had no significant effect on gene expression. In contrast, X-ray exposure led to number of significant genes, which clearly exceeded the estimated false positives rate and identified a high number of genes known to respond to X-rays. While in this Affymetrix analysis the hybridization target quality, as indicated by the 3²/5²-ratio, is not reported and the data is not fully accessible, experimental design as well as the strategy to estimate the false positives rate is very well thought.

The paper of Remondini et al. (2006) reports microarray data obtained within the so-called REFLEX study, a multinational study supported by the EU. In this study six different cell lines were exposed to 900 and 1800 MHz at SAR between 1.5 and 2.5 W/kg for different times and various modulation forms, yielding 9 different experimental settings. RNA samples were analyzed on Bioanalyser chips, while no quality control is reported for labeled cDNA targets. Unamplified P-32-labelled samples were hybridized to cDNA arrays containing about 75000 cDNA clones spotted in duplicates onto 3 nylon membranes of 22x22 cm at the RZPD

facility. As indicated in Table 1 of the paper, depending on the experimental setting, 2, 3 or 5 independent exposure experiments were performed; however, in the results section it is clearly noted that for each experimental condition only a single pooled RNA sample was analyzed twice by array analysis, to yield two technical replicates. Data processing and analysis were performed separately on each part of the array to avoid possible biases related to different behavior of the arrays during hybridization or scanning. Only spots representing 4 “good measurements” in the two replicate analyses resulting for both control and experiment were further processed, but no information regarding number of genes retained and of the reproducibility of the technical replicates is given. Importantly, according to Material and Methods, target genes were identified as the intersection of genes identified by two different methods including SAM, which provides an FDR correction term. As the array analyses comprise only 2 technical replicates for controls and experiment each, it remains obscure why this obviously inappropriate approach has been taken and how it might lead to a reasonable FDR estimate. Anyway, by this approach in three out of 9 experimental settings up to 34 differentially expressed genes were identified, which mostly represented genes associated with metabolism. In contrast, comparison of RNA from sham-exposed sample with incubator controls did not show any regulated gene corresponding to the significance criteria applied. A statistical approach to data interpretation has not been taken and no consistent EMF signature could be detected. The data has not been deposited into a public data repository. Altogether, due to the obscurities associated with data analysis and the lack of independent validation, this dataset has to be regarded as highly preliminary.

Qutob et al. (2006) performed 5 independent experiments with single replicates of unexposed controls and samples exposed at 0.1, 1.0 and 10 W/kg using an Agilent oligonucleotide chip (18000 genes, dual color experiment). The quality of the RNA samples, but not of the labelled cRNA, was checked by analysis on Bioanalyzer chips. Since a preliminary analysis indicated that the day of hybridization was the greatest source of variation, the data were subjected to complex normalization procedures that removed the effect. The normalized data were analyzed by MAANOVA, which includes a multiple testing adjustment specifically designed for two-color microarray experiments, as well as by SAM. According to the authors, the MAANOVA method did not identify any significant gene. Similarly, “low-stringency” SAM analysis did not identify any significantly modulated gene. Since the authors did not specifically state what p-value and FDR q-value thresholds were applied to their MAANOVA and SAM analysis, respectively, one can not rule out, that the negative outcome might be the result of too conservative thresholds. However, the analysis of heat-treated controls led to the

identification of 99 significant genes, including a number of heatshock genes and other stress-response genes. Six heat-induced genes selected were all successfully validated by semi-quantitative RT-PCR, suggesting that the data analysis method used is suitable to pick up truly regulated genes while producing a low number of false positives. Again, a reanalysis of the data is not possible, since the data has not been submitted to a public database.

Belayaev et al. (2006) exposed rats in vivo and analyzed gene expression in cerebellum by using an Affymetrix chip with 8800 probesets representing approximately 8000 different genes. RNA quality was analyzed using bioanalyzer chips. Three exposed and unexposed replicates were analysed by cross-comparison analysis using the Affymetrix comparison algorithm. Eleven genes were found to be concordantly up-regulated (Increased) and 2 down-regulated (Decreased) in all 9 cross-comparisons. It can be estimated, that this number of concordantly regulated genes could occur by chance only, if each individual comparison produced Increased or Decreased calls on more than 50% of all probesets on the chip, which appears to be rather unlikely. However, the reproducibility observed in this study is not documented and could have been compromised, because of the use of animals, in vivo exposure and cerebellum dissection, which might represent sources for additional variability. Again, since the dataset has not been deposited in a public database, the actual level of variability can not be determined. As no effort is presented to validate the candidate genes, this study fails to provide substantial hints for an EMF effect on gene expression.

Zeng et al. (2006) performed two independent microarray experiments on Affymetrix HG-U133A chips (~22000 probesets) involving only single replicates of unexposed MCF-7 cell cultures and cells exposed at 2 or 3.5 W/kg. The quality of RNA samples and labeled cRNAs were checked by agarose gel electrophoresis and test chip hybridization, but the observed 3'/5'-ratios are not documented. Arrays were analyzed by Affymetrix cross comparison analysis yielding four comparison data sets for each SAR applied. Only genes showing concordant Increased or Decreased calls in all 4 comparisons were considered as potentially regulated genes. In cells exposed to a SAR of 2 W/kg no concordantly regulated gene could be identified, while at 3.5 W/kg five less than 2-fold up-regulated genes with 100% consistency were identified. Quantitative RT-PCR failed to confirm regulation of all five candidate genes, implying that the effects had occurred by chance. Thus, this study, which is based on a very limiting number of replicates and therefore may have missed some truly regulated genes, does not provide evidence for EMF induced gene expression changes. The microarray data has not been submitted to a public database.

Gurisek et al. (2006) exposed SK-N-SH cells to a SAR of 0.2 W/kg and analyzed gene expression on Affymetrix Human Focus chips representing ~8400 genes. Information on RNA or cRNA quality is not indicated. Only a single pair of exposed and unexposed samples was processed and evaluated by Affymetrix comparison analysis, which according to the paper identified 6 decreased genes with less than 2-fold changes. The unusual low fraction (0.07%) of changed genes identified by Affymetrix comparison analysis appears to indicate an unusually high level of reproducibility; however, re-evaluation of the data is not possible, since the full data set is not published or deposited in a database. RT-PCR analysis failed to produce any PCR product in four cases. Another two genes, for which RT-PCR could be established, were not confirmed in a set of three independent samples. Thus, it is possible that all candidates represent false positive observations.

The study by Hirose et al. (2006) presents Affymetrix HG-U133Plus_2.0 (~54000 probesets) microarray analyses of the A172 cell line exposed to SAR of 80 or 800 mW/kg for 24 h. In two independent experiments, exposed and unexposed controls were analyzed in triplicates, yielding as much as six replicates for each condition. Regarding RNA or cRNA quality, no information is given. Obviously, the two independent experiments were analyzed separately by using a modified t-test (Welch's t-test) without a correction for multiple testing. The results show, that none of 21 selected p53-related genes tested significant in both independent experiments. Moreover, RT-PCR analysis of four p53-related genes failed to show significant differences under the experimental conditions employed for the microarray analysis, thereby confirming the negative microarray data. Strangely, since the authors strictly focus on the p53-related genes, it remains completely open whether there were any non-p53-related genes, which tested significant in both independent experiments. Moreover, it remains unclear, why the authors did not exploit the statistical power associated with six replicates arrays, e.g. by SAM analysis or related methods, but rather preferred to separately analyze the two experiments by a much less powerful t-test variant combined with a relatively conservative correction method.

A second paper published by the same group (Hirose et al., 2007) presents an extended study combining two cell lines (A172 and IMR-90) with different SARs (80, 250, 800 mW/kg), exposure times (2, 24, and 48 h) and two different field modulation forms. Again, in two independent experiments, exposed and unexposed controls were analyzed in triplicates, yielding as much as six replicates for each condition. In addition, for each cell line heat-shock controls were employed. Regarding RNA or cRNA quality, no information is given. Analysis

of the Affymetrix HG-U133Plus_2.0 dataset (~54000 probesets) was this time obviously restricted to genes called Present or Marginal, thereby introducing a systematic bias through exclusion of potentially most interesting genes switching from Absent to Present or vice versa upon exposure. Since the description of the data mining procedure is essentially identical to the one given in Hirose et al. (2006), it is again not completely clear how the data is analyzed. However, in Table 2, the authors show that the expression ratios (exposed/control) of four mRNAs encoding hsp27, hsp40, hsp70 and hsp105/100 did not vary by more than 2-fold under any of the 12 different conditions employed. In contrast, heat-shock treatment led to up to 23-fold induction of heat-shock genes. The authors state in the result section that “*No significant differences were observed between any of the RF field exposure groups and the sham-exposed controls in the gene expression profile, including hsp-related genes in either experiments of exposure to W-CDMA or CW signal for 2-48 h (Table 2)*”. While this statement could be understood, as if not a single gene of the approximately 20000 probesets subjected to statistical analysis was found to be significant in both independent experiments, this important finding is not at all picked up in the discussion. Re-analysis of the dataset, which is one of the largest in EMF research, by a more appropriate method under relaxed exclusion criteria is highly advisable, but not possible, since the data has not been deposited in a public database. My repeated requests to provide the full dataset for confidential re-analysis remained unanswered by the corresponding author. Interestingly, both studies presented by Hirose et al. (2006, 2007) were supported by a Japanese mobile phone company.

Zhao R et al. (2007) analyzed primary rat neuronal cultures exposed to an SAR of 2 W/kg for 24 h using an Affymetrix chip (RN U34, ~1200 genes) specifically designed for neurobiological research. The integrity of the RNA and cRNA was analyzed, but is not documented. The number of replicates analyzed by microarray is not clearly indicated; however, from the data analysis procedure it can be inferred that only single replicates have been analyzed. Regulated genes were identified by Affymetrix comparison analysis, yielding 34 (2.8%) genes called Increased or Decreased, all with changes below 2-fold. Although this percentage of regulated candidate genes is within the range of false positives observed by this analysis method, the authors report that by quantitative RT-PCR 23 out of 25 candidates could be confirmed. How many RNA samples were analyzed by RT-PCR and whether these RNAs represented independent samples, remains unclear. Regulated genes were shown to belong to a variety of functional categories, but no statistical approach to data interpretation was taken. The dataset has not been deposited into a public database. Due to the uncertainties of the experimental design and the apparent discrepancy between expected and

experimentally observed false positives rate, final prove that EMF may alter gene expression in primary neuronal cells requires confirmation by an independent study.

Zhao TY et al. (2007) analyzed gene expression in primary neuron cultures from mouse exposed to a rather unconventional EMF device, namely a mobile phone. No SAR is reported, but one can assume that the average SAR most likely did not exceed 2 W/kg. No information regarding the RNA or the labeled sample is given. Two independent sets of single replicate samples were analyzed on a special array representing only 96 apoptosis-related genes, which hardly represents a whole genome expression analysis. The fold change of gene expression level was calculated for each independent pair of samples and 9 genes showing an increase or decrease $\geq 35\%$ in both sets were defined as candidates and further examined. By quantitative RT-PCR, expression of 3 out of 5 genes was confirmed to be regulated by exposure to the mobile phone, both in the standby and on modes. However, it remains unclear how many RNA samples were analyzed and whether they represented independent samples. Whether under the given experimental conditions exposure indeed led to induction of apoptosis, was not investigated by a more direct method. Also due to the unconventional exposure device and associated uncertainties regarding SAR, possible field inhomogeneity and thermal effects, this study can hardly be regarded as a prove that EMF, as generated by a mobile phone, regulates gene expression in neuronal cells.

Chauhan et al. (2007) used an Agilent oligonucleotide array representing ~18000 genes to analyze gene expression in U87MG and MM6 cells exposed to SARs up to 10 W/kg for 6 and 24 hours. Integrity of RNA samples, but not the Cy3-labelled amplified samples, was checked on Bioanalyzer chips. The entire dataset obtained for each cell line was first variance-stabilized and normalized in R, then renormalized in SAS and analyzed by MAANOVA, including an FDR adjustment method. Spearman correlation analysis of normalized intensities confirmed that the 5 independent array replicates in each group showed quite some variability, while median correlations for exposed cells were acceptable, but not superior (0.831-0.939). Statistical analysis revealed neither in U87MG nor in MM6 cells any significant gene ($p \leq 0.05$) even prior to any filtering based upon minimum fold change. In contrast, identical analysis of RNA samples isolated 6 and 18 hours after heat-treatment (1h, 43°C), yielded between 439 and 2420 significant genes. Genes regulated by heatshock were reported to include stress related genes and others, but a statistical analysis proving significant enrichment for specific GO terms is not presented. By semi-quantitative RT-PCR, some HSP genes were confirmed to be regulated in heat-treated cells, but not in EMF treated cells. Most

likely, for RT-PCR analysis the very same set of samples subjected array analysis was used. The successful identification of large numbers of heat-induced gene expression changes, some of which were successfully validated, proves that the data mining strategy was capable of identifying truly differentially expressed genes. Thus, this paper provides some hints that at least in the given experimental settings no EMF induced gene expression changes can be observed.

Paparini et al. (2008) used an Affymetrix MOE430A array (22000 transcripts/genes) to analyze the effect of EMF (1800 MHz) on gene expression in mice whole brain tissue. Here, the restrained animals were exposed for 1 h at a whole body SAR of 1.1 W/kg (average SAR in the brain at 0.2 W/kg). RNA was prepared from 15 sham controls and 15 exposed brains and RNA quality checked by gel electrophoresis. Three pools of RNA each consisting of equal parts of 5 RNA preparations were prepared from controls and exposed samples and subjected to array analysis to yield 3 replicate arrays in each group. It is stated that “*no variability was observed within either the sham or exposed group, and the correlation coefficients between replicates were very high indeed*”, however, no supporting data is given. Data obtained by Affymetrix analysis “*was filtered for present calls to yield 13152 probesets*” and subjected to statistical analysis. How the presence call filtering was exactly done remains unclear and it appears possible that it may have introduced a severe bias. Omitting a multiple testing correction, the analysis of the preprocessed list of probesets yielded 75 significant ($p < 0.05$) candidate genes in a Welch’s t-test. 20 up- and 10 down-regulated candidate genes were subjected to validation in 30 RNA preparations. While this is an expressing number of replicates, it appears likely that these samples correspond to the 30 individual samples used to prepare the RNA pools for microarray analysis and, thus, are strictly no independent samples. Anyway, according to the authors “*there was no significant correlation ... between gene modulations obtained from microarray and real-time PCR*”, while a statistical evaluation of the PCR data of individual genes is not presented. While the data may support the authors conclusion that there is no evidence of major transcriptional changes in the brain of exposed mice, the possible bias introduced in data pre-processing as well as the sparse analysis of the PCR data raises some concern about the study quality.

4. Conclusions

Analyzing the experimental design and the statistical approach to identify EMF regulated genes, it is striking, that all four studies, which analyzed replicate numbers of 5 or more, did not provide evidence for EMF induced gene regulation. Two of these studies (Chauhan et al., 2007; Qutob et al., 2006) analyzed their data by SAM or MMANOVA, which provide a suitable balancing of both error rates. The studies published by Hirose et al (2006, 2007), which produced 6 replicates under each condition, for unknown reasons chose to analyze the two independent sets of experiments with 3 replicates each separately by applying Welch's T-test without multiple testing correction (Hirose et al., 2006) or a simple fold change ratio approach (Hirose et al., 2007). However, since the authors disclosed the negative results of a small number of selected genes only and do not clearly state or discuss, whether there were any genes significant in both sets, it remains questionable whether these two studies can indeed be counted as negative studies. Similarly, the data pre-processing steps and the lack of a statistical analysis of the PCR validation experiments raises some concern about the "negative" study of Papparini et al. (2008). Among the studies showing negative results, the study of Whitehead et al. (2006) exhibits a well thought experimental design that shows that even with a limited budget a reasonable estimation of the expected false positives rate can be achieved. Based on the estimated false positives rate, the authors concluded that the candidates testing significance in EMF exposed samples, most likely represent false positives. In the study of (Zeng et al., 2006), which analyzed two replicates each, a low number of candidate targets observed could not be validated by RT-PCR. Similarly, another study (Port et al., 2003) with two independent replicates also failed to identify regulated genes, which could be confirmed by RT-PCR. In both studies, the data mining strategy may have resulted in a low type I error rate; however, due to rather conservative approaches (concordance analysis; fold change threshold combined with signal intensity threshold), those studies are suspected to exhibit low sensitivity and could have produced high type II errors (false negatives).

Among the studies, which claimed to identify EMF effects on gene expression, one study (Nylund and Leszczynski, 2006), which consists of 3 replicates, ignores the multiple testing problem completely and in the absence of validation experiments, cannot be regarded as a substantial hint for EMF induced gene regulation. The study of Lee et al. (2005), which relied on the analysis of single SAGE libraries, did not provide any validation experiments and used a SAR, which might not be relevant for risk assessment of electromagnetic fields up to the

existing safety limit. The study of Remondini et al. (2006) used a rather strange data analysis strategy that can hardly include a reasonable form of FDR adjustment and, due to the absence of validation efforts, can at best be regarded as a preliminary. Belayaev et al. (2006) analyzed 3 replicates by cross-comparison analysis and identified 13 concordantly regulated genes in rat cerebellum. Since this number of concordantly regulated is unlikely to occur by chance by this rather conservative approach, this study provides a hint for EMF regulated genes in the brain of rat exposed to EMF. However, the experimental procedure exhibits some weak points and confirmation in independent samples is lacking. The study of Zhao R et al (2007), which studies neuronal cells, identified 34 EMF regulated candidate genes with diverse functions. Since 23 out of 25 candidate genes were successfully confirmed by RT-PCR, this study has to be regarded as a hint for EMF regulated in neuronal cells. However, the significance of this study is very difficult to judge, since the number of array replicates analyzed as well as the data mining strategy and the nature of the samples used for validation remain unclear. Finally, Zhao TY et al. (2007) presented hints for EMF induced genes in primary neurons from mice by analysis of two pairs of samples on an array representing 96 apoptosis-related genes. While 3 of the candidate genes identified by array analysis could be successfully confirmed by quantitative RT-PCR, again the significance of this study can be challenged, as it is not clear whether independent samples were analyzed.

Taken together, it is striking that most studies exhibiting a more advanced experimental design and data analysis fail to produce hints for EMF induced gene expression changes. However, only a limited number of negative studies exhibit a reasonable design and in some cases the negative result might be due to too conservative approaches which might have created high type II error rates. In contrast, among the studies presenting EMF regulated genes, there is none which used more than 3 replicates and more advanced statistical analysis methods, such as SAM. Remarkably, among the apparently positive studies also presenting successfully validated genes, there is none which clearly described the samples used for validation as independent or used an increased number of samples. Therefore, there is not a single positive study which meets two most important rules in microarray studies, i.e. a experimental design allowing a suitable statistical approach to data analysis and confirmation of regulated genes in a sufficiently large set of independent samples.

Looking at the candidate gene lists presented by the “positive” studies, no common EMF gene signature can be identified. Since different cell systems and exposure conditions were used in

most studies, this could mean that different cell types, different SARs and modulation forms engage different signaling pathways. Except for the study using SAGE analysis, there is not a single case, which by a statistical approach, such as GO term or gene set enrichment or pathway analysis, provides any hint for the involvement of a specific biological process, known signaling pathway or involved transcriptional factors. This could indicate that EMF affects gene expression and biological processes through currently unknown signaling pathways and regulators. However, in light of the many shortcomings of the available studies, it appears also possible, that the genes identified may not represent true positive findings. Clearly, independent confirmation by other groups would be required to substantiate the observations indicating EMF regulated gene expression. Ideally, confirmatory studies should be reconciled with statisticians trained to analyze whole genome expression analysis datasets prior to grant application, to ensure that an experimental design enabling suitable statistical approaches is possible, and include extensive target validation experiments in independent samples. Extraction of RNA, target amplification and labeling including quality control, and whole genome array hybridizations should be performed in array facilities or biochip laboratories with proven record of experience, and data analysis and been done by trained statisticians, including statistical approaches to data interpretation. Moreover, state-of-the-art EMF exposure devices should be employed, ensuring exact dosimetry, homogeneous fields and temperature monitoring.

In summary, one must conclude that microarray analyses have not yet added much value to EMF research and that the limited number of microarray studies published cannot fill any of the knowledge gaps defined by Prof. Dr. Meyer and Prof. Dr. Kundi in WP1 of the IMBA.

5. References

- Beissbarth T, Speed TP (2004)
GStat: find statistically overrepresented Gene Ontologies within a group of genes.
Bioinformatics 20(9):1464-5
- Belyaev IY, Koch CB, Terenius O, Roxström-Lindquist K, Malmgren LO, H Sommer W, Salford LG, Persson BR (2006)
Exposure of rat brain to 915 MHz GSM microwaves induces changes in gene expression but not double stranded DNA breaks or effects on chromatin conformation.
Bioelectromagnetics 27(4):295-306
- Chauhan V, Qutob SS, Lui S, Mariampillai A, Bellier PV, Yauk CL, Douglas GR, Williams A, McNamee JP (2007)
Analysis of gene expression in two human-derived cell lines exposed in vitro to a 1.9 GHz pulse-modulated radiofrequency field.
Proteomics 7(21):3896-905
- Guridik E, Warton K, Martin DK, Valenzuela SM (2006)
An in vitro study of the effects of exposure to a GSM signal in two human cell lines: monocytic U937 and neuroblastoma SK-N-SH.
Cell Biol Int 30(10):793-9
- Hirose H, Sakuma N, Kaji N, Nakayama K, Inoue K, Sekijima M, Nojima T, Miyakoshi J (2007)
Mobile phone base station-emitted radiation does not induce phosphorylation of Hsp27.
Bioelectromagnetics 28(2):99-108.
- Hirose H, Sakuma N, Kaji N, Suhara T, Sekijima M, Nojima T, Miyakoshi J (2006)
Phosphorylation and gene expression of p53 are not affected in human cells exposed to 2.1425 GHz band CW or W-CDMA modulated radiation allocated to mobile radio base stations.
Bioelectromagnetics 27(6):494-504.
- Hubbell E, Liu WM, Mei R (2002)
Robust estimators for expression analysis.
Bioinformatics. 18(12):1585-92.
- Keller A, Backes C and Lenhof, HP (2007)
Computation of significance scores of unweighted Gene Set Enrichment Analyses.
BMC Bioinformatics 8:290
- Kerr, Martin and Churchill (2000)
Analysis of variance for gene expression microarray data.
Journal of Computational Biology, 7:819-837.
- Klein-Hitpass L and Möröy T (2005). In: DNA Microarrays (edited by UA Nuber).
Affymetrix GeneChip analyses – the impact of RNA quality, 125-137

- Larsson O, Wahlestedt C, Timmons JA (2005)
Considerations when using the significance analysis of microarrays (SAM) algorithm.
BMC Bioinformatics 6, 129-135
- Lee S, Johnson D, Dunbar K, Dong H, Ge X, Kim YC, Wing C, Jayathilaka N, Emmanuel N, Zhou CQ, Gerber HL, Tseng CC, Wang SM (2005)
2.45 GHz radiofrequency fields alter gene expression in cultured human cells.
FEBS Lett 579(21):4829-36
- Lu J, Tomfohr JK, Kepler TB (2005)
Identifying differential expression in multiple SAGE libraries: an overdispersed log-linear model approach.
BMC Bioinformatics 29; 165-179
- Mariani TJ, Budhreja V, Mecham BH, Gu CC, Watson MA, Sadovsky Y (2003)
A variable fold threshold determines significance for microarray<s.
FASEB J. 17, 321-323
- Mayo MS, Gajewski BJ, Morris JS (2006)
Some statistical issues in microarray gene expression data.
Radiat Res 165(6):745-8
- Nylund R, Leszczynski D (2006).
Mobile phone radiation causes changes in gene and protein expression in human endothelial cell lines and the response seems to be genome- and proteome-dependent.
Proteomics 6(17):4769-80.
- Paparini A, Rossi P, Gianfranceschi G, Brugaletta V, Falsaperla R, De Luca P, Romano Spica V. (2008)
No evidence of major transcriptional changes in the brain of mice exposed to 1800 MHz GSM signal.
Bioelectromagnetics 29(4):312-23
- Port M, Abend M, Römer B, Van Beuningen D (2003)
Influence of high-frequency electromagnetic fields on different modes of cell death and gene expression.
Int J Radiat Biol 79(9):701-8.
- Qutob SS, Chauhan V, Bellier PV, Yauk CL, Douglas GR, Berndt L, Williams A, Gajda GB, Lemay E, Thansandote A, McNamee JP (2006)
Microarray gene expression profiling of a human glioblastoma cell line exposed in vitro to a 1.9 GHz pulse-modulated radiofrequency field.
Radiat Res 165(6):636-44
- Remondini D, Nylund R, Reivinen J, Poullietier de Gannes F, Veyret B, Lagroye I, Haro E, Trillo MA, Capri M, Franceschi C, Schlatterer K, Gminski R, Fitzner R, Tauber R, Schuderer J, Kuster N, Leszczynski D, Bersani F, Maercker C (2006)
Gene expression changes in human cells after exposure to mobile phone microwaves.
Proteomics 6(17):4745-54

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005)

Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.

PNAS 102(43):15545-50

Tusher VG, Tibshirani R and Chu G (2001)

Significance analysis of microarrays applied to ionizing radiation response.

PNAS 98, 5116-5121

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995)

Serial analysis of gene expression.

Science 270(5235):484-7

Whitehead TD, Moros EG, Brownstein BH, Roti Roti JL (2006)

Gene expression does not change significantly in C3H 10T(1/2) cells after exposure to 847.74 CDMA or 835.62 FDMA radiofrequency radiation.

Radiat Res 165(6):626-35

Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN.

GoMiner: a resource for biological interpretation of genomic and proteomic data.

Genome Biol. 2003;4(4):R28.

Zeng Q, Chen G, Weng Y, Wang L, Chiang H, Lu D, Xu Z (2006)

Effects of global system for mobile communications 1800 MHz radiofrequency electromagnetic fields on gene and protein expression in MCF-7 cells.

Proteomics 6(17):4732-8

Zhao R, Zhang S, Xu Z, Ju L, Lu D, Yao G (2007)

Studying gene expression profile of rat neuron exposed to 1800MHz radiofrequency electromagnetic fields with cDNA microassay.

Toxicology 235(3):167-75

Zhao TY, Zou SP, Knapp PE (2006)

Exposure to cell phone radiation up-regulates apoptosis genes in primary cultures of neurons and astrocytes.

Neurosci Lett 412(1):34-8

Table1 Experimental details of published microarray studies in EMF research

Study	Sample type; species	SAR; duration	Field frequency; modulation	Array system	No of genes	Labelling method	Detection method
Port et al. (2003), Int J Radiat Biol 79, 701	HL-60; human	2.0 W/kg (?); 20 min, 24, 48, 72 h	0.4 GHz; 1 Hz	Clontech Atlas Array; nylon	1176	Incorporation of P-32 labelled dATP into cDNA	Densitometric analysis of autoradiographies
Lee et al. (2005), FEBS Letters 579, 4829	HL-60; human	10 W/kg for 2 or 6 hrs	2.45 GHz	SAGE	n.a.	n.a.	Sequencing
Nylund and Leszczynski (2006), Proteomics 6, 4769	EA.hy929, EA.hy929v1; human	2.8 W/kg; 1h	GSM900 MHz	Clontech Atlas Human 1.2 cDNA Array; nylon	1167	Incorporation of P-32 labelled dATP into cDNA	Densitometric analysis of autoradiographies
Whitehead et al. (2006), Radiat. Res 165, 626	C3H10T1/2; mouse	5 W/kg for 24h	~800 MHz; CDMA, FDMA	Affymetrix U74Av2	~9200	Biotin incorp. into cRNA, post hyb staining	Fluorescence laser scanning
Remondini et al. (2006) Proteomics 6, 4745	NB69, EA.hy926, U937, CHMES, HL-60, primary T-lymphocytes; human	1.0-2.5 W/kg for 1, 24 or 44 h	GSM900 or GSM1800 MHz; Basic or DTX mode	RZPD Unigene RZPD-2 cDNA array, nylon, 3 parts	75000 cDNA clones	Incorporation of P-32 labelled dCTP into cDNA	PhosphorImaging
Qutob et al. (2006), Radiat Res 165 (2006), 626	U87MG; human	0.1, 1 and 10 W/kg ; 4h	1.9 Ghz; 50 Hz, 1/3 duty cycle	Agilent 1A(v1) oligonucleotied 22K	~18000	Dual colour labelling (Cy3, Cy5)	Fluorescence laser scanning
Belayaev et al. (2006), Bioelectromagnetics 27, 295	Cerebellum ; rat	0.4 W/kg for 2 h In vivo exposure	GSM914 MHz;	Affymetrix U34A	~8800	Biotin incorp. into cRNA, post hyb staining	Fluorescence laser scanning
Zeng et al. (2006), Proteomics 6, 4732	MCF-7; human	2 W/kg and 3.5 W/kg for 24 h; 5 min on/10 min off	GSM1800 MHz; 217 Hz	Affymetrix U133A	~14500	Biotin incorp. into cRNA, post hyb staining	Fluorescence laser scanning
Gurisik et al. (2006), Cell Biol Int 30, 793	SK-N-SH; human	0.2 W/kg, 2 hrs, 2 h recovery	GSM900 MHz, 217 Hz	Affymetrix Human Focus	~8400	Biotin incorp. into cRNA, post hyb staining	Fluorescence laser scanning
Hirose et al. (2006), Bioelectromagnetics 27, 494	A172; human	80, 800 mW/kg for 24 h	2.1425 GHz, W-CDMA	Affymetrix U133Plus 2.0	~39000	Biotin incorp. into cRNA, post hyb staining	Fluorescence laser scanning
Hirose et al. (2007), Bioelectromagnetics 28, 99	A172, IMR-90; human	80, 250, 800 mW/kg for 24 or 48 h	2.1425 GHz, W-CDMA, CW	Affymetrix U133Plus 2.0	~39000	Biotin incorp. into cRNA, post hyb staining	Fluorescence laser scanning
Zhao R et al. (2007), Toxicology 235, 167	Primary neuronal cells; rat	2 W/kg for 24 h; 5 min on/10 min off	1800 MHz; 217 Hz	Affymetrix RN U34	~1200	Biotin incorp. into cRNA, post hyb staining	Fluorescence laser scanning
Zhao TY et al (2007), Neurosci Lett 412, 34	Primary neurons; mouse	SAR unclear, 2h; sham-control (standby mode)	GSM 1900 MHz; Samsung SGH-E105	SuperArray Biosc. GEArray Q Mouse apoptosis	96	Incorporation of Biotin-dUTP into cDNA	Chemiluminescent detection with CCD camera
Chauhan et al. (2007), Proteomics 7, 3896	U87MG, Mono-Mac6; human	0.1, 1.0, 10 W/kg for 24 h; 1.0, 10 W/kg for 6 h; 6 h recovery periods	1.9 GHz	Agilent Human 1Av2 22K oligonucleotide array	~18000	Cy3-incorporation into cRNA	Confocal laser scanner
Paparini et al. (2008) Bioelectromagnetics 29, 312	Whole brain, mouse	0.2 W/kg (Average SAR) 1h	GSM 1800 MHz	Affymetrix MG-430A	~15000	Biotin incorp. into cRNA, post hyb staining	Fluorescence laser scanning

Table2 (part1) Evaluation of published microarray studies in EMF research

Study	Experimental design	Sample quality control (QC)	Target gene identification method	No. of reported potential target genes	Data mining process documentation	Experiment al. validation	Statistical approach to data interpretation	Raw data accessibility	Remarks
Port et al. (2003), Int J Radiat Biol 79, 701	n=2; for 4 different time points	RNA and cDNA checked (gel)	Fold change (Ratio) method	1 to 6 at different time points; none confirmed in 2 nd array experiment	Comprehensive	no	n.a.	no	No sample QC reported; few replicates > lack of statistical approach; raw data not published
Lee et al. (2005), FEBS Letters 579, 4829	n=1; two different time points	Not indicated	Special statistic for SAGE used, p<0.05 and FC>4 used as cut-offs	221 and 759 with FC>4 at 2 and 6 hrs; 84 candidates are commonly changed at 2 and 6 h by EMF	Comprehensive	no	no	NCBI GEO	SAR clearly beyond ICNIRP threshold, no sample QC reported; single replicates; lack of statistical approach to data interpretation
Nylund and Leszczynski (2006), Proteomics 6, 4769	n=3 for both cell lines	Not indicated	Fold change (Ratio) method; t-test; signals<3000 excluded; FC _{Avg} threshold >2	In EA.hy929, 4 out of 110 were up- and 89 down-regulated >2-fold; 2 were significant. In EA.hy929v1, 61 out of 80 were up- and 2 down-regulated >2-fold, 12 genes significant	Comprehensive			no	Unsentive method associated with large variations; no independent validation of candidate genes; raw data not published
Whitehead et al. (2006), Radiat. Res 165, 626	n=3 for both modulation forms	RNA checked on Agilent Bioanalyzer	t-test; only genes showing present detection calls in all 3 controls or exposed samples were considered	CDMA- and FDMA-exposed cells show even less differentially expressed genes than the expected no. of false positives determined > no effect of RMF. X-ray clearly induces a greater number of genes.	Comprehensive	n.a.	n.a.	no	FDR estimation by a permutaion based approach ; raw data not published
Remondini et al. (2006) Proteomics 6, 4745	n=1-2, depending on cells and exposure; two independent hybridizations	RNA checked on Agilent Bioanalyzer	Intersection of genes found by SAM and those showing a normalized difference >3 SD of the difference distribution	In 3/11(9) exposure settings between 12, 32 and 34 fifferentially expressed genes were identified, only few genes with FC>2; no consistent changes observed	Inconsistencies between explained data mining strategy and presented results	no	no	no	Method obviously associated with large variations; too few replicates for SAM based FDR approach; documentation of data mining process and results show inconsistencies, no independent validation of candidate genes by different methods; raw data not published
Qutob et al. (2006), Radiat Res 165, 626	n=5; independent experiments	RNA checked on Agilent Bioanalyzer	ANOVA, SAM (FDR)	Both approaches did not identify significant differences between exposed and controls, FDR threshold unclear	Not easily comprehensive	qRT-PCR confirmed absence of HSP regulation	n.a.	no	Sophisticated normalization necessary to eliminate day-to-day variability ; data mining steps not fully documented; raw data not published
Belyaev et al. (2006), Bioelectromagnetics 27, 295	n=3	RNA checked on Agilent Bioanalyzer	Cross comparison analysis (MAS5); 100% consistency threshold	11 up-regulated gens (up to 2.74-fold) and 1 down-regulated (0.48-fold) gene	Comprehensive	no	no	no	Lack of statistical approach; no independent validation of candidate genes by different methods; raw data not published

Table2 (part2) Evaluation of published microarray studies in EMF research

Study	Experimental design	Sample quality control (QC)	Target gene identification method	No. of reported potential target genes	Data mining process documentation	Exp. validation	Statistical approach to data interpretation	Raw data accessibility	Remarks
Zeng et al. (2006), Proteomics 6, 4732	n=2; two independent experiments	RNA checked on gel; hybr. targets on Test chip	Cross comparison analysis (MAS5); 100% consistency threshold	No candidates at 2 W/kg ; 5 upregulated (<2-fold) genes at 3.5 W/kg	Comprehensible	qRT-PCR on all candidates failed to confirm regulation	n.a.	no	Single sample study > lack of statistical approach; raw data not published
Gurisik et al. (2006), Cell Biol Int 30, 793	n=1	Not indicated	not indicated (Affymetrix comparison analysis?)	6 genes down-regulated by EMF	Some steps not comprehensibly described	qRT-PCR (n=3) failed to confirm 2 candidates	n.a.	no	Single sample study > lack of statistical approach; raw data not published
Hirose et al. (2006), Bioelectromagnetics 27, 494	n=6 split into two independent sets	Not indicated	Comparison analysis (MAS5); FC threshold (>2 in both exp.) and \geq one P detection call required; Welch's t-test on FC values	No EMF targets reported, since statistical significance was not confirmed in 2 nd experiment; data analysis and presentation restricted to 21 p53 related genes	Comprehensible	qRT-PCR on some selected candidates confirmed negative results	n.a.	no	No sample QC reported; reduced sensitivity due to FC threshold; no statistical analysis on pooled dataset (n=6) reported; analysis and data presentation restricted to 21 out of 54000 probesets; raw data not published
Hirose et al. (2007), Bioelectromagnetics 28, 99	n=6 split into two independent sets	Not indicated	Welch's t-test, only probesets showing P or M calls were considered	No EMF targets reported, as significance was not confirmed in 2 nd exp.; heat treatment at 43°C changed expression of hundreds of genes in both cell lines	Comprehensible	partial	n.a.	no	No sample QC reported; no statistical analysis on pooled dataset (n=6) reported; analysis and data presentation is restricted to 4 out of 54000 probesets ; raw data not published;
Zhao R et al. (2007), Toxicology 235, 167	n=1 ?	RNA checked on gel; hybr targets on Test chip	Comparison analysis (MAS5)	24 up- and 10 down-regulated genes with FC<1.75	Several steps not comprehensibly described	qRT-PCR (n=3) confirmed significant differences on 23/25 genes	no	no	Single samples; lack of statistical approach; raw data not published
Zhao TY et al. (2007), Neurosci Lett 412, 34	n=2; two separate experiments	Not indicated	Genes with increase or decrease of \Rightarrow 35% in both experiments were selected	8 up- and 1 down-regulated genes with FC up to 2.76	Comprehensible	3/5 candidates confirmed by qRT-PCR in neurons (n=?), in astrocytes 4/6	n.a.	no	Lack of dosimetry data, no sample QC reported; few replicates >lack of statistical approach; theme-specific array causing biased interpretation; raw data not published
Chauhan et al. (2007), Proteomics, Epub ahead of print	n=5 for each of the conditions	RNA checked on Agilent Bioanalyzer	MAANOVA	No differentially expressed genes could be identified. Heat treatment changed several hundreds of genes in both cell lines	Comprehensible	qRT-PCR on six Hsp genes confirmed in EMF and heatshock exp.	n.a.	no	Spearman correlations within treatment groups were somewhat low and variable; raw data not published
Paparini et al. (2008) Bioelectromagnetics 29, 312	n=3; pools of 5 samples each	RNA checked on gel	Welch t-test w/o correction	Without correction 75 probesets were significant (p<0.05)	Comprehensible	qRT-PCR on 30 genes showed little correlation with microarray	n.a.	no	Pre-filtering on detection calls may have introduced a bias. No statistical analysis of qRT-PCR data