RNA sequencing provides insights into the toxicogenomic response of ZF4 cells to methyl methanesulfonate

Zhouquan Li\textsuperscript{a,b}, Yong Long\textsuperscript{a}, Liqiao Zhong\textsuperscript{a,b}, Guili Song\textsuperscript{a}, Xiaohua Zhang\textsuperscript{a}, Li Yuan\textsuperscript{a}, Zongbin Cui\textsuperscript{a*} and Heping Dai\textsuperscript{a*}

ABSTRACT: Whole genome transcriptomic studies are powerful for characterizing the molecular mechanisms underlying the physiological effects of chemicals, and are informative for environmental health risk assessment. Alkylating agents are an abundant class of chemicals that can damage DNA in the environment, and are used for anticancer treatments. Currently, little is known regarding the molecular mechanisms of toxic alkylating agents in zebrafish cell lines. In this study, RNA-sequencing was used to investigate the transcriptomic responses of zebrafish ZF4 cells following exposure to the model genotoxicant methyl methanesulfonate (MMS). The half-maximal inhibitory concentration (IC\textsubscript{50}) of MMS was $639.16 \pm 61.8 \, \mu\text{m}$, and apoptosis was induced within 24 h of exposure. RNA sequencing identified 3601 differentially expressed genes (DEGs) that were upregulated and 3037 that were downregulated. Gene ontology enrichment analysis revealed that most DEGs belonged to synthesis and metabolism categories. RNA-associated processes were the most upregulated, while cell cycle and adhesion were the most repressed processes, and neuron-related processes were the most downregulated developmental process. Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis identified DNA damage repair, cell cycle, apoptosis and spliceosome as overrepresented terms. Six types of alternative splicing were detected. In total, 1156 alternative splicing DEGs were specifically expressed following MMS treatment, many of which belonged to metabolism and catabolic process categories. Cluster analysis of orthologs was able to extrapolate toxicotranscriptomic data between zebrafish and yeast. These results provide insight into the genome-wide response of ZF4 cells following exposure to MMS, and this knowledge will inform future toxicogenomic data analysis and environmental health risk assessment. Copyright © 2015 John Wiley & Sons, Ltd.

Introduction

The term toxicogenomics was originally coined in 1999 (Nuwaysir \textit{et al.}, 1999), and this discipline aims to study whole genome responses to toxicants or environmental stressors (Waters & Fostel, 2004). Gene expression can be altered either directly or indirectly upon exposure to toxicants (Corton \textit{et al.}, 1999). Toxicogenomics aims to combine proteomic, metabolomic and transcriptomic data to elucidate the molecular mechanisms of toxicity and to determine expression patterns that facilitate prediction of toxicity or genetic susceptibility. For ethical reasons it is desirable to move toxicity studies away from traditional high-dose animal research towards an approach based on perturbation of intracellular toxicity pathways using well-characterized \textit{in vitro} assays (NCR, 2007). It has been suggested that toxicogenomic data could be used to understand the complex effects of chemicals and to inform health risk assessment (Wilson \textit{et al.}, 2013). However, there are significant challenges to overcome before toxicogenomic data can be routinely utilized for this purpose, not least the acquisition, storage and analysis of extensive amounts of data, and extrapolation from model organisms to humans and other species (Burgess-Herbert & Euling, 2013).

Alkylating agents represent an abundant class of chemicals that cause DNA damage, and may be produced endogenously (Rydberg & Lindahl, 1982; Taverna & Sedgwick, 1996), encountered in the environment (Ballschmiter, 2003; Hamilton \textit{et al.}, 2003; Hecht, 1999), or be introduced during medical procedures such as cancer therapies (Christmann \textit{et al.}, 2007). The well-characterized simple monofunctional alkylating agent methyl methanesulfonate (MMS) is used as a model genotoxicant (Bony \textit{et al.}, 2010; Devaux \textit{et al.}, 2011; Deventer, 1996; Solomon & Faustman, 1987). MMS modifies DNA via the N3 or N7 positions of the adenosine or guanosine moiety (Kim & LeBreton, 1994) and causes random lesions that result in single strand (Fortini \textit{et al.}, 2005) at random positions in the genome. MMS also induces damage to proteins, RNA and other macromolecules, causes multiple harmful effects on cells and is mutagenic, carcinogenic, recombinogenic and clastogenic (Vishwakarakash & Vasudev, 2013). Although MMS itself is not relevant in the aquatic environment, it can be
considered representative of several groups of environmental pollutants such as mucohalic acids (Gomez-Bombarelli et al., 2011) and allicy reagents (Kuehl et al., 1994) due to its alkylating mechanism of action. The genome-wide response to MMS has been well characterized in the budding yeast Saccharomyces cerevisiae (Gasch et al., 2000, 2001; Hughes et al., 2000; Jelinsky & Samson, 1999; Lee et al., 2007; Natarajan et al., 2001). One pioneering toxicogenomic study examined the transcription profile of yeast upon exposure to MMS using a microarray approach (Jelinsky & Samson, 1999). Over 30% of all yeast genes responded to MMS treatment, and they represented multiple functional categories (Jelinsky & Samson, 1999; Jelinsky et al., 2000). Specific defense mechanisms protect cells against MMS, and genes associated with stress response/detoxification, drug metabolism, DNA repair/recombination, cell cycle, signal transduction, cell wall biogenesis and membrane transport were all identified. In addition to direct damage to DNA, a large number of RNA processing and ribosomal genes and proteins are also damaged, and splicing proteins were found to govern the sensitivity to the toxic effects of MMS (Svensson et al., 2011). The relatively simple structure and mechanism of action of MMS make it an ideal model toxicant for toxicogenomic research.

The zebrafish (Danio rerio) is a widely used model organism for research on developmental biology, physiology, genetics, toxicology and environmental genomics. At the genetic level, this species shares more than 80% homology with humans. Furthermore, zebrafish and humans are similar in their development, anatomy, physiological responses, metabolism and chemical-induced organ/tissue responses, and many molecular pathways are shared. Over time, abundant biological information and genetic resources have been accumulated for this model organism, including the complete genome sequence, which has been comprehensively annotated. This has greatly assisted high-throughput approaches such as RNA sequencing (RNA-seq). The fibroblast-like ZF4 cells of zebrafish, which are derived from 1-day-old zebrafish embryos therefore could provide early development related toxicity damage (Driever & Rangini, 1993) and can be easily cultivated, have been used in numerous research areas and were used in this study. Next-generation RNA-seq has been applied to transcriptomics (Mortazavi et al., 2008; Nagalakshmi et al., 2008), and has proven to be a powerful tool for genome-wide gene expression profiling. Compared with the microarray platform, RNA-seq can offer a better dynamic range, can detect more subtle changes in gene expression, and can characterize alternative splicing (AS) of mRNAs and detect novel transcripts (Oszolak & Milos, 2011; Wang et al., 2009; Wilhelm & Landry, 2009). This makes RNA-seq one of the best tools for toxicoc transcriptome research.

To develop an appropriate method for toxicogenomic studies and environmental health risk assessment, zebrafish ZF4 cells were exposed to MMS and genome-wide transcriptional responses were investigated using RNA-seq to provide in-depth insight into the molecular mechanisms of MMS toxicity. Zebrafish and yeast toxicoc transcriptomic data were compared, and the results will inform future research into environmental health risk assessment.

Materials and Methods

Cell Culture and Methyl Methanesulfonate Treatment

MMS (Sigma, St. Louis, MO, USA) was dissolved in phosphate-buffered saline (PBS) just before use. Zebrafish ZF4 cells (ATCC CRL-2050) derived from 1-day-old zebrafish embryos were routinely maintained in a humidified chamber (28°C, 5% CO₂) in Dulbecco minimal Eagle’s/F-12 medium (GIBCO, USA), with 10% (v/v) fetal bovine serum (Gibco, Carlsbad, CA, USA), 100 U ml⁻¹ penicillin and 100 mg ml⁻¹ streptomycin (GIBCO). When confluen ce reached 70–80%, cells were treated with MMS or an equal volume of PBS (control group) and incubated for 24 h. Cells were washed three times with cold PBS and detached from the culture plate using trypsin (GIBCO) for further experiments.

MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide] Cytotoxic Assay

To determine the IC₅₀ viability of ZF4 cells after MMS exposure, a colorimetric MTT metabolic activity assay was performed as described (Mosmann, 1983). Briefly, ZF4 cells were plated in a 96-well plate at a density of 2 × 10⁴ cells per well in 100 μl medium and cultured for 24 h before treatment. Cells were then exposed to varying concentrations of MMS (0, 150, 225, 337.5, 506.3, 759.4, 1139.1, 1708.6 μM) or with medium only (negative control group). After 24 h of exposure, 10 μl of MTT (MTT in PBS (Beyotime, Shanghai, China)) was added and incubated for another 4 h, the resultant formazan crystals were dissolved in 150 μl dimethyl sulfoxide, and the absorbance at 490 nm was measured using a microplate reader Elx800 (Bio-Tek, Winooski, VT, USA). IC₅₀ values were calculated by the modified Kärber method (Mu, 2007) using the equation

\[ \text{lgIC}_{50} = X_m - I(P - (3 - P_m - P_n)/4) \]

where \( X_m \) is the logarithm of maximum dose, \( I \) is the difference of two adjacent logarithmic doses, \( P \) is the sum of positive reaction rate, \( P_m \) is the maximum inhibition rate and \( P_n \) is the minimum inhibition rate. The MTT tests were carried out in triplicate and each biological replicate had six technique replicates.

Apoptosis Analysis by Flow Cytometry

ZF4 cells were seeded in six-well culture dishes in 3 ml medium per well and cultured at 28°C with 5% CO₂. After reaching a confluence of 70–80%, cells were treated with MMS and incubated for 24 h. Cells were then quantified by FACScan flow cytometer (BD, Franklin Lakes, NJ, USA) using Annexin-V-fluorescein isothiocyanate (FITC)/propidium iodide (PI) kit (Beyotime) according to the manufacturer’s instructions. Control tubes contained cells stained with Annexin-V-FITC only, PI only without Annexin-V-FITC or PI only. All experiments were carried out in triplicate. Data were analyzed using FlowJo (v5.7.2) software.

cDNA Library Construction and High-Throughput Sequencing

Total RNA was extracted with TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer’s instructions. Each RNA sample for next-generation sequencing was a mixture of RNA extracted from three independent biological experiments. The cDNA library was constructed and sequenced in a high-throughput manner using the BGI platform (Shenzhen, China). Briefly, RNA samples were analyzed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), then magnetic beads with oligo(dT) were used to isolate poly(A) mRNA. Purified mRNA was fragmented in fragmentation buffer, and cDNAs were synthesized using the mRNA fragments as templates with random hexamer primers. Second-strand cDNA was
synthesized using buffer, dNTPs, RNase H and DNA polymerase I. Short double-stranded cDNA fragments were purified after end repair and addition of an adenine base. Next, the short fragments were ligated to Illumina sequencing adapters. DNA fragments of a selected size were gel-purified and amplified by polymerase chain reaction (PCR). Library quality was measured using an Agilent 2100 Bioanalyzer and an ABI StepOnePlus Real-Time PCR System (Applied Biosystems, Foster, CA, USA). The amplified library was sequenced on one lane of Illumina HiSeq 2000 sequencing platform (Illumina, San Diego, CA, USA) and 90 nt paired-end reads were generated.

Raw Read Filtering

Original image data were converted into sequence data using the base calling pipeline and saved in the fastq format. The sequencing data have been deposited in NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra) and the accession number is SRP053202. During filtering, reads containing adapters were removed, as were reads in which unknown bases (N) accounted for more than 5% or where read quality was low (the percentage of low quality bases is over 30% in a read, we define the low quality base to be the base whose sequencing quality is no more than 10). After filtering, the remaining reads were designated as clean and used for all subsequent analyses.

Mapping of Reads and Calculation of Gene Expression Levels

Reference zebrafish genome sequences were downloaded from the Ensembl database (Zv9.68). All clean reads were aligned to the reference genome using the short oligonucleotide analysis package aligner (SOAPaligner/SOAP2) (Li et al., 2009). No more than five mismatches were allowed in the alignment for each read, and reads that mapped to a unique gene were used to calculate expression levels. Gene expression was calculated with the default value of R/Bioconductor (http://bioconductor.org/) packages (GenomicFeatures, rtracklayer, GenomicRanges and Rsamtools). Normalized expression levels were calculated from the reads per kilobase transcriptome per million mapped reads (RPKM) (Mortazavi et al., 2008) for each library.

RNA Sequencing Background Estimation

To confirm that poorly expressed genes were measurable by RNA-seq, background estimation was performed as previously described (Yang et al., 2013). Briefly, 500 bp were subtracted from both ends of all intergenic regions, and these were then treated as exons for RPKM calculations. Genes for which the RPKM was equal or greater than the median RPKM of the intergenic regions were defined as expressed.

Analysis of Differentially Expressed Genes

Statistical analysis of differential expression was performed as described in the R/Bioconductor (http://bioconductor.org/) package DEGseq (Wang et al., 2010). The MA plot-based method with random sampling model was used to identify differentially expressed genes (DEGs) (Dudoit et al., 2002). Benjamini and Hochberg multiple testing was performed to adjust P values and to generate profiles for DEGs. Genes that were differentially expressed following MMS exposure (compared with PBS controls) were defined as those with a $P < 0.001$, a $q < 0.05$ and a $\log_2$ (MMS_RPKM/PBS_RPKM) of $\geq 1$.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analysis of Differentially Expressed Genes

Cytoscape (version 2.8.3) (Shannon et al., 2003) plugin of BINGO (v3.0.2) (Maere et al., 2005) and ClueGO (v2.0.7) (Bindea et al., 2009) were used for gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis respectively. GO and KEGG pathway data were updated (http://www.geneontology.org/) on March 24, 2014. Enrichment analysis provided all terms that were significantly enriched in the list of DEGs compared with the reference background. A hypergeometric test was used to identify overrepresented GO terms with a significance level at 0.05 and the Benjamini and Hochberg false discovery rate method was used for correction of the P values. A two-sided hypergeometric test was used to identify overrepresented pathway terms with a significance level at 0.05 and the Bonferroni method was used for correction of the P values. Enrichment pathway maps were generated using KegArray (v1.2.3).

Alternative Splicing and Functional Clustering of Differentially Spliced Genes

AS analysis was performed with SOAPspice (Huang et al., 2011). AS events were empirically classified into seven different types (A, exon skipping; B, intron retention; C, alternative 5′ splice site; D, alternative 3′ splice site; E, alternative first exon; F, alternative last exon; G, mutually exclusive exon) based on exon structure and as described previously (Wang et al., 2008). GO and KEGG pathway enrichment analysis was performed (as described above) using the identified alternatively spliced DEGs.

Validation of Gene Expression by Quantitative Reverse Transcription–Polymerase Chain Reaction

Total RNA was extracted as described above, and genomic DNA was removed with RNase-free DNase 1 (Fermentas, MD, USA) according to the manufacturer’s protocol. First-strand cDNA was synthesized from 2 μg of total RNA using random hexamer primers with the RevertAidTM First Strand cDNA Synthesis Kit (Fermentas). PCR primers were designed using Primer Premier 6.0 software. The specificity of quantitative reverse transcription–PCR (qRT-PCR) primers was checked by searching against the zebrafish reference transcriptome using the Primer-BlAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome). The qRT-PCR amplification protocol was 95 °C for 1 min, followed by 40 cycles of 95 °C for 10 s, 58–60 °C for 30 s (including plate read) and 72 °C for 10 s. The qRT-PCR products were measured on a CFX Connect Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Primers with 90–110% amplification efficiency were used for qRT-PCR analysis. Primer sequences, amplicon sizes and amplification efficiencies are listed in Supplementary Table S1. All qRT-PCR were carried out in triplicate. GAPDH and smarce1 were stably expressed in all samples and were used as internal references for normalization of gene expression. The mean expression of target genes was normalized as described previously (Vandesompele et al., 2002). The correlation between RNA-seq and qRT-PCR data was analyzed by Spearman’s rho test.
Comparison of Zebrafish and Yeast Transcriptomic Data

Interspecies transcriptomics data were obtained from male and female zebrafish exposed to the estrogen compound, 17α-ethinylestradiol (EE2), and from yeast exposed to MMS. All data were from our laboratory (Zhang, 2014; Zhong, 2014). Those sequencing data have been deposited in NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra) under accession number SRP053279 and SRP053216. Zebrafish and yeast orthologs were downloaded from the OrthoMCL database (http://orthomcl.org/orthomcl/), paired and associated with the transcriptome data based on their sequence similarity rank (Supplementary Table S2). The fold change in expression was used for further bioinformatics analysis. Yeast data were more comprehensively annotated, and GO and KEGG pathway enrichment analysis of the orthologs was performed based on the yeast annotation. The ggplots tool from the R software package was used to cluster the enrichment categories. The heatmaps were generated using Euclidean distance and “complete” agglomeration method.

Results

MTT Cytotoxicity

The IC50 of ZF4 cells exposed to MMS was determined by the MTT method. Cell viability declined rapidly along with increasing drug concentration (Fig. 1a). The IC50 value calculated by a modified Kärber method was 639.16 ± 61.8 μM, and this was the concentration of MMS to which ZF4 cells were exposed in subsequent experiments.

Analysis of Apoptosis by Flow Cytometry

To analyze whether MMS had an effect on apoptosis at the determined IC50 concentration, Annexin-V-FITC/PI flow cytometry was used. A marked increase in apoptotic ZF4 cells was apparent at 600 and 700 μM MMS (Fig. 1b). Quantitative analysis indicated that the percentage of early and later apoptotic cells was 14.6% and 17.1% at 600 and 700 μM MMS, respectively.
Illumina Sequencing and Mapping of RNA Sequencing Reads to the Zebrafish Genome

To obtain a genome-wide view of the ZF4 cell transcriptome and gene expression profile following MMS treatment, high-throughput RNA-seq was performed using Illumina sequencing technology. After filtering adaptor sequences, duplicated sequences, ambiguous reads and low-quality reads, over 51.48 million (M) and 53.73 M pairs of 90 bp length clean reads were obtained from the PBS control and MMS experimental group, respectively. High-quality reads were subsequently mapped to the zebrafish genome (Zv9.68) using SOAP2. More than 77% reads were successfully mapped to the zebrafish genome, and the number of uniquely mapped reads was 38.13 M (70.97%) for the MMS group and 37.58 M (73.01%) for the PBS control group (Table 1).

Gene Expression Profile and Differentially Expressed Gene Analysis

A total of 25 059 genes detected with more than one read in the MMS-exposed group, while 24 317 genes were identified in the PBS control group. A total of 32 677 genes were detected with more than one read in both samples (Supplementary Table S3). The abundance of gene transcripts was measured by RPKM, and both experimental and control groups exhibited a similar RPKM distribution (Fig. 2a); the mean RPKM value was 11.92 in the MMS group, and 9.53 in the control group. To confirm the low RPKM gene truly expressed, the background coverage of intergenic regions was calculated. The median background coverage was 0.2273 RPKM in the MMS group and 0.1864 RPKM in the control group. Therefore, genes with RPKM ≥0.2273 in the MMS group were regarded as expressed and 20 322 genes were detected (81.1% detection rate). Genes with RPKM ≥0.162 in the PBS control group totaled 20 329 (83.6% detection rate). DEGs were calculated using the R/Bioconductor (http://bioconductor.org/) package DEGseq. Genes with a fold change ≥2, a P < 0.001 and a q < 0.05 were defined as DEGs (Fig. 2b). The numbers of up- and downregulated genes were 3601 and 3037, respectively (Supplementary Table S4). Genes regulated by DNA damage such as p53R2, Gadd45, P48 and Sestrins were specifically expressed following MMS treatment. Many of the DEGs that were upregulated were uncharacterized (Supplementary Table S4).

Gene Ontology Enrichment Analysis of Differentially Expressed Genes Regulated by Methyl Methanesulfonate Treatment

A total of 6638 DEGs were annotated across the GO database, and genes associated with enriched GO terms were grouped into 71

**Table 1.** Statistics for read mapping

<table>
<thead>
<tr>
<th>Sample name</th>
<th>PBS</th>
<th>PBS%</th>
<th>MMS</th>
<th>MMS%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total reads</td>
<td>51481868</td>
<td>100.00%</td>
<td>53737650</td>
<td>100.00%</td>
</tr>
<tr>
<td>Total base pairs</td>
<td>4633368120</td>
<td>100.00%</td>
<td>4836388500</td>
<td>100.00%</td>
</tr>
<tr>
<td>Total mapped reads</td>
<td>39948691</td>
<td>77.60%</td>
<td>41586308</td>
<td>77.39%</td>
</tr>
<tr>
<td>Perfect match</td>
<td>24184196</td>
<td>46.98%</td>
<td>25260436</td>
<td>47.01%</td>
</tr>
<tr>
<td>&lt; = 5 bp mismatch</td>
<td>15764495</td>
<td>30.62%</td>
<td>16325872</td>
<td>30.38%</td>
</tr>
<tr>
<td>Unique match</td>
<td>37586280</td>
<td>73.01%</td>
<td>38137602</td>
<td>70.97%</td>
</tr>
<tr>
<td>Multiposition match</td>
<td>236241117</td>
<td>4.59%</td>
<td>3448706</td>
<td>6.42%</td>
</tr>
<tr>
<td>Total unmapped reads</td>
<td>11533177</td>
<td>22.40%</td>
<td>12151342</td>
<td>22.61%</td>
</tr>
</tbody>
</table>

MMS, methyl methanesulfonate; PBS, phosphate-buffered saline.

**Figure 2.** Statistics of RNA-seq data. (a) Boxplot showing the distribution of expression RPKM values between PBS control and MMS treatment groups. The range of the box represented the first quartiles (Q1) and the third quartiles (Q3). A horizontal line in the box presented the median RPKM value. The ends of vertical bars represented 5th and 95th percentiles, and the dots represented outliers defined as the RPKM values ≥ (Q3 + 1.5(Q3 – Q1)); (b) MA plot showed gene expression in PBS control and MMS-treated groups. Up- and down-regulated differentially expressed genes are in red and blue, respectively. Genes not regulated by MMS treatment are in black.
biological processes, five cellular component and 20 molecular function categories (Supplementary Table S5). Among the biological processes category (Fig. 3), DEGs mainly populated the synthesis and metabolism function. There were 13 categories associated with metabolic processes and 14 categories related to biosynthesis, with a similar proportion of DEGs up- and downregulated (Supplementary Fig. S1). It was found that 17 RNA-related processes were the most enriched following MMS exposure, many of which were upregulated (Supplementary Fig. S2). About 18 developmental and regulatory categories were also enriched, the majority of which were downregulated (Supplementary Fig. S3). Four categories associated with the regulation of neuronal development were among the most downregulated DEGs (Supplementary Fig. S3). In addition, it was noticed that cell cycle and adhesion categories such as homophilic cell adhesion, cell–cell adhesion, cell cycle and cell cycle processes were dominated by downregulated DEGs (Supplementary Fig. S3). The results of GO enrichment indicated that multiple biological processes were involved in the response to MMS exposure.

**KEGG Pathway Enrichment Analysis of Differentially Expressed Genes Regulated by Methyl Methanesulfonate Treatment**

To investigate the biological function of DEGs, KEGG pathways were analyzed by Cytoscape (v2.8.3) with the plugin ClueGO (v1.71). A total of 114 pathways were overrepresented following MMS treatment (Supplementary Table S6) and 94 KEGG pathway categories had \( P < 0.005 \). Up- and downregulated DEGs were divided into three groups (upregulated pathway: upregulated DEG > 66%, Supplementary Fig. S4; up- and downregulated DEG commonly regulated pathway, Supplementary Fig. S5; downregulated pathway: downregulated DEG > 66%, Supplementary Fig. S6) based on their distribution percentage in the KEGG categories. RNA polymerase, ribosome, proteasome, drug metabolism–associated enzymes, spliceosome, citrate (TCA) cycle and nucleotide excision repair (NER) pathway terms were upregulated. In contrast, homologous recombination, cell cycle, base excision repair (BER), DNA replication, mismatch repair, cell adhesion molecules and adherens junction were downregulated. Finally, up- and downregulated DEG commonly regulated pathway included metabolism of xenobiotics by cytochrome P450, p53 signaling pathway, lysosome, drug metabolism–cytochrome P450, ubiquitin-mediated proteolysis and apoptosis pathway. A total of 40 DEGs were associated with the spliceosome, of which 38 were upregulated by MMS (Fig. 4). U2, U4/U6 and US were the apoptosis-associated genes most upregulated by MMS. The major extrinsic apoptosis pathway genes IL-1, TRAF2, IRAK and Bax, and the intrinsic (mitochondria-associated) pathway genes CytC, AIF and ENDO-G were also all upregulated. In contrast, the apoptosis pathway survival factor genes NGF, PI3K, Akt/PKB and Bcl-2 were downregulated. In all, 33 of the 45 zebrafish cell cycle pathway DEGs were downregulated.

*Figure 3. GO enrichment analysis of DEGs. The size of circles is proportional to the number of genes associated with the GO term. Arrows represent the relationship between parent-child terms. The color scale indicates corrected p-values from the enrichment analysis.*
Cyclin-dependent kinases (CDKs) are key regulators in the cell cycle, but only CDK7 was upregulated, while CDK1, CDK2, CDK4 and CDK6 were all downregulated. Many of the affected pathways identified were already known to respond to MMS exposure; however, a number of previously unknown pathways were also identified, namely oocyte meiosis, endocytosis and tryptophan metabolism pathways.

**Figure 4.** DEGs associated with the spliceosome pathway. The gene expression value was mapped to the reference pathway using the KegArray. Downregulated genes are in blue; Up-regulated genes are in red.

**Table 2.** Statistics for alternatively spliced genes

<table>
<thead>
<tr>
<th></th>
<th>A5S S</th>
<th>A5S S</th>
<th>SE</th>
<th>AFE</th>
<th>ALE</th>
<th>MXE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagram</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific AS in PBS</td>
<td>1872</td>
<td>1248</td>
<td>1024</td>
<td>75</td>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>Common AS in PBS and MMS</td>
<td>1409</td>
<td>1124</td>
<td>779</td>
<td>18</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>Specific AS in MMS</td>
<td>2235</td>
<td>1689</td>
<td>1281</td>
<td>80</td>
<td>43</td>
<td>4</td>
</tr>
<tr>
<td>Specific AS DEGs in MMS</td>
<td>781</td>
<td>540</td>
<td>443</td>
<td>34</td>
<td>18</td>
<td>0</td>
</tr>
</tbody>
</table>

AS, alternative splicing; DEGs, differentially expressed genes; MMS, methyl methanesulfonate; PBS, phosphate-buffered saline.

**Alternative Splicing and Functional Clustering of Differentially Spliced Genes**

AS plays a major role in the generation of proteomic and functional complexity in higher organisms (Black, 2003; Blencowe, 2006; Matlin et al., 2005; Reddy, 2007; Zhang et al., 2010). We carried out computational analysis to determine the theoretical
splicing junctions. There were 4788 alternatively spliced genes and 8685 AS events following MMS treatment, compared to 4438 AS genes and 7607 AS events in the control group (Table 2). In the MMS group, 40.4% of genes shared more than one AS event, and this number was 37.4% in the control group. We identified six of the seven AS models reported previously (Wang et al., 2008), the A3SS, A5SS and SkippedExon were the dominant AS models in both groups. A total of 1869 AS genes were specific to the MMS treatment group, and 1156 of these were DEGs (Supplementary Table S7). Next, GO enrichment was analyzed using all expressed genes as background, and 121 biological process terms were enriched (Supplementary Table S8). The most enriched categories were gene expression, ncRNA metabolic process, ncRNA processing, tRNA metabolic process and cellular nitrogen compound metabolic processes. Two pathways related to changes in subcellular localization (protein targeting to the endoplasmic reticulum and translational protein targeting to the membrane) were identified, and many other biological processes were also enriched, reflecting the strong influence of AS on gene expression in ZF4 cells.

Comparison of Zebrafish and Yeast Transcriptomic Data

To use the toxicogenomic data for environmental health risk assessment, transcriptome data from zebrafish exposed to the non-genotoxicant estrogen compound EE2 were compared with data from yeast exposed to MMS. E-ZF and E-ZM represent the EE2 treatment of female and male zebrafish, respectively, and M-Y corresponds to the MMS treatment of yeast, while E-MC relates to the MMS treatment of ZF4 cells. Yeast and zebrafish share 2264 groups of orthologs, and 2513 pairs of genes were identified based on a sequence similarity rank. A total of 2377 pairs remained after combining the transcriptomic data. We used yeast gene annotation for GO along with KEGG pathway enrichment analysis because this gave the most comprehensive annotation. A total of 601 biological processes categories and 81 pathways were enriched (Supplementary Table S9). There were more genes and categories in the GO enrichment than in the KEGG pathway enrichment; therefore, GO enrichment categories were used for subsequent cluster analysis. To identify the more significant transcriptome data results, various normalizing methods, including z-score, mean and median were tested (data not shown), and the log2 (ratio) was found represented the best diversity between different drugs or species. The results showed that the effects of MMS on cellular responses to DNA damage, DNA repair and stress responses were clearly clustered in zebrafish and yeast, whereas the effects of EE2 were associated with lipid biosynthesis and metabolism, and formed a separate cluster. Comparison of interspecies differences following exposure to the same chemical (MMS) showed that both ZF4 cells and yeast shared similar expression patterns that centered around DNA damage and repair, although some clear differences between zebrafish and yeast were apparent (Fig. 5b).

Validation of RNA Sequence Data by Quantitative Reverse Transcription–Polymerase Chain Reaction

To validate the expression profiles obtained from RNA-seq, 18 genes exhibiting different expression patterns were measured by qRT-PCR. Expression data and transcript detection by RNA-seq and qRT-PCR (Fig. 6) was in excellent agreement for both up- and downregulated genes. The correlation between RNA-seq and qRT-PCR data was analyzed by Spearman’s rho test and a highly statistical significance ($r = 0.851$, $P = 2.2e-16$) was observed.

Discussion

In this study, the aim was to investigate the toxicotranscriptome of ZF4 cells in response to MMS to gain insight into the molecular mechanisms of MMS toxicity. RNA-seq was used to generate genome-wide toxicotranscriptome data from ZF4 cells. A previous study found that over a third of yeast genes respond to MMS damage (Jelinsky et al., 2000). In this study, 3601 and 3037 genes were up- and downregulated in ZF4 cells following exposure to MMS.
This reflects the global effects of DNA damage induced by MMS. After GO enrichment analysis, DEGs were mainly grouped into metabolic and biosynthetic process, RNA-related process, cell adhesion, cell cycle and development categories. After KEGG pathway enrichment analysis, the biological functions were distributed across 114 pathways related to metabolism, biosynthetic, DNA damage and repair, cell cycle, apoptosis, development, AS and proteasome categories. Therefore, multiple genes responded to MMS in ZF4 cells, which indicates the usefulness of this approach for further studies.

MMS primarily acts by damaging DNA directly and induces DNA damage repair. All major DNA repair processes such as BER, NER, mismatch repair and homologous recombination repair are affected by alkylating agents (Hoeijmakers, 2001; Kondo et al., 2010). BER and NER pathways and alkytransferases are reported to be the most important processes (Helleday et al., 2008; Kondo et al., 2010; Lundin et al., 2005). In this study, all the aforementioned DNA repair pathways were well represented among DEGs. However, DEGs in the BER, mismatch repair and homologous recombination pathways were downregulated, and only NER pathway DEGs were upregulated. Indeed, 10 of 13 NER pathway-associated DEGs were upregulated, suggesting this may be the most important repair pathway in ZF4 cells following MMS exposure, at least at 24 h after exposure. It is possible that the BER pathway was initially upregulated but became downregulated after 24 h, and further investigations including different exposure time points are necessary to clarify this.

It was expected that the toxicotranscriptomic data would reflect changes in cell cycle and apoptosis following MMS exposure, as cells responding to DNA damage would arrest the cell cycle in the S phase and modulate gene expression to ensure efficient DNA repair (Gasch et al., 2001). MMS induced apoptosis (Jiang et al., 2012; Lackinger et al., 2001; Ryu et al., 2001), which affects cell survival, reproduction and development. The DNA damage response is a complex signal transduction pathway (Ciccia & Elledge, 2010), and DNA damage response, cell cycle arrest and apoptosis are all induced by signal transduction. In this study, it was found that 16 signaling pathways participating in cell repair, cell cycle, apoptosis and development were enriched. Indeed, 12 of 28 p53 signaling pathway-associated DEGs were upregulated, including p53R2, Gadd45, P48 and Sestrins that are all important for DNA repair and consistent with other studies (Datta et al., 2001; Tanaka et al., 2000; Zhan et al., 1998). Also enriched were the important MAPK, FoxO and Wnt signaling pathways that are directly related to the DNA damage response, as were the insulin Jak-STAT, calcium and transforming growth factor-beta signaling pathways that are indirectly related to the DNA damage response. As expected, the drug metabolism–other enzymes and drug metabolism–cytochrome P450 pathway categories were also enriched in response to MMS treatment. Eleven of 14 drug metabolism–other enzymes pathway DEGs were upregulated, and these pathways may play an important role in MMS detoxification. These results demonstrate that DNA damage caused by MMS is a chain reaction that causes a global response.

Other macromolecules such as RNA and proteins are also damaged by MMS in addition to DNA (Boffa & Bolognesi, 1985), and these must be eliminated and re-synthesized (Boffa & Bolognesi, 1985; Jelinsky & Samson, 1999; Jelinsky et al., 2000). Many of the processes affected by MMS were related with metabolism and biosynthesis, and this may be explained by the need for cells to recover from the macromolecular damage following MMS exposure (Boffa & Bolognesi, 1985; Jelinsky & Samson, 1999; Jelinsky et al., 2000). In addition to the DNA damage response-associated processes pathways related to oocyte meiosis, endocytosis and tryptophan metabolism were also affected, which is more difficult to explain. Further experiments are clearly needed to explain the molecular mechanisms underlying the toxicogenomic data.

AS is a major mechanism that contributes to both transcription and proteome diversity (Keren et al., 2010; Mastrandelo et al., 2012). Accumulating evidence suggests that AS is an important adaptation response to a wide range of stress conditions (Ali & Reddy, 2008; Fujikake et al., 2005; Keren et al., 2010; Long et al., 2013). RNA splicing is required for cellular survival following MMS damage in yeast (Svensson et al., 2011). In this study, 40 DEGs were related to the spliceosome, 38 of which were upregulated (Fig. 4). U2, U4/U6 and U5 play an important role in the spliceosome pathway, and A35S, A55S and SkippedExon are the dominant AS events, suggesting this is among the most important mechanisms at work following MMS exposure. A total of 1869 AS genes were specifically expressed following MMS treatment, and 1156 of these were DEGs. AS in response to stress is mainly associated with genes encoding protein kinases, transcription factors, splicing regulators and pathogen-resistance factors, and usually leads to changes in subcellular localization, binding properties and activity or stability of damaged proteins (Mastrangelo et al., 2012). To our surprise, only the “establishment of protein localization to endoplasmic reticulum” and “translational protein targeting to membrane” categories were enriched in this study. Even so, the AS DEGs identified were associated with complex processes and pathways, and most were related to multiple metabolic and specifically catabolic processes. This may reflect the macromolecular damage occurring, and the AS genes may be important for cellular recovery from MMS toxicity.

Some of the genes that exhibited responses to MMS exposure are commonly associated in different species, and it is important for environmental risk assessment purposes to identify genes that respond specifically to MMS in different species. Those genes might extrapolate toxicogenomic data between species and used for environmental health risk assessment. To this end, zebrafish was compared with yeast to identify more general characteristics related to DNA damage after MMS exposure. Comparative genomics provides a method for comparing different species at a...
Toxicogenomic response of ZF4 cells to methyl methanesulfonate
genome-wide level to include genes/proteins, pathways and networks (Burgess-Herbert & Euling, 2013). Both GO and KEGG pathway enrichment are important bioinformatic tools for “omics” data analysis, while clustering is a rapid and efficient method for toxicogenomic data analysis (Buczynski et al., 2000; Irwin et al., 2004). Clustering algorithms can be used to group genes with similar expression patterns. In the present study, we combined gene/protein level and GO or pathway level orthologs and clustered GO or pathway categories for interspecies analysis of zebrafish and yeast toxicogenomic data. GO categories gave a greater number of genes and higher specificity than KEGG pathway categories. The log2 (ratio) value was found to be appropriate for data standardization. Processes and pathways specific to MMS were clustered together using this parameter in both species. The results demonstrated that in spite of interspecies differences, there were characteristic responses that were common to both organisms following exposure to the same chemical toxicant. Genes with similar expression patterns responded specifically to MMS exposure. The toxicogenomic data from different species suggest that this type of approach can be valuable for environmental health risk assessment, and may be used to extrapolate from model organisms to humans and other species. The approach developed in this study can identify common characteristics from toxicogenomic data using bioinformatics methods, and the results can be used to construct a database that will benefit environmental health risk assessment.

Acknowledgments

This work was supported by grants from the Natural Science Foundation of China (no. 21037004), and the State Key Laboratory of Fresh Water Ecology and Biotecnology (2011FBZT10).

Conflict of Interest

The authors declare that there are no conflicts of interest.

References


Kim HS, LeBreton PR. 1994. UV photoelectron and ab initio quantum mechanical characterization of valence electrons in Na(1)-water-2H2O-deoxyguanosine 5'-phosphate clusters: electronic influences on DNA