

RESEARCH

Open Access

# Ontogeny, distribution and potential roles of 5-hydroxymethylcytosine in human liver function

Maxim Ivanov<sup>1\*</sup>, Mart Kals<sup>2†</sup>, Marina Kacevska<sup>1†</sup>, Isabel Barragan<sup>1</sup>, Kie Kasuga<sup>3</sup>, Anders Rane<sup>4</sup>, Andres Metspalu<sup>2,5,6</sup>, Lili Milani<sup>2,6†</sup> and Magnus Ingelman-Sundberg<sup>1†</sup>

## Abstract

**Background:** Interindividual differences in liver functions such as protein synthesis, lipid and carbohydrate metabolism and drug metabolism are influenced by epigenetic factors. The role of the epigenetic machinery in such processes has, however, been barely investigated. 5-hydroxymethylcytosine (5hmC) is a recently re-discovered epigenetic DNA modification that plays an important role in the control of gene expression.

**Results:** In this study, we investigate 5hmC occurrence and genomic distribution in 8 fetal and 7 adult human liver samples in relation to ontogeny and function. LC-MS analysis shows that in the adult liver samples 5hmC comprises up to 1% of the total cytosine content, whereas in all fetal livers it is below 0.125%. Immunohistostaining of liver sections with a polyclonal anti-5hmC antibody shows that 5hmC is detected in most of the hepatocytes. Genome-wide mapping of the distribution of 5hmC in human liver samples by next-generation sequencing shows significant differences between fetal and adult livers. In adult livers, 5hmC occupancy is overrepresented in genes involved in active catabolic and metabolic processes, whereas 5hmC elements which are found in genes exclusively in fetal livers and disappear in the adult state, are more specific to pathways for differentiation and development.

**Conclusions:** Our findings suggest that 5-hydroxymethylcytosine plays an important role in the development and function of the human liver and might be an important determinant for development of liver diseases as well as of the interindividual differences in drug metabolism and toxicity.

## Background

There are substantial interindividual differences in many liver processes, including intermediary metabolism, protein synthesis, carbohydrate metabolism and the detoxification of drugs and other xenobiotics. With respect to variation in drug metabolism, transport and toxicity, much knowledge has been gained from studies investigating genetic factors responsible for interindividual differences. However, little is known regarding the role of epigenetic factors, such as DNA modifications and their influence on hepatic gene expression, function and interindividual variation [1].

The epigenetic DNA modifications include methylation and hydroxymethylation of cytosine. For many

years, 5-methylcytosine (5mC) was believed to be the only epigenetic modification of genomic DNA, often referred to as 'the fifth base' of the genome. Recently, 5-hydroxymethylcytosine (5hmC) was discovered as a novel epigenetic factor in mammalian DNA that can have a stable effect on gene transcription and has hence been regarded as 'the sixth base' of the genome [2,3]. 5hmC originates from an enzymatic oxidation of 5mC by TET1, TET2 and TET3 proteins [2,4], which recognize their substrate 5mC by the amino-terminal CXXC zinc finger domain [5]. This reaction is dependent on Fe<sup>2+</sup> as well as  $\alpha$ -ketoglutarate; the latter, in turn, is dependent on the activity of either isocitrate dehydrogenase IDH1 in the cytosol and peroxisomes or IDH2 and IDH3 in mitochondria [6,7]. Although other proteins may also have some significance in establishing the presence of genomic 5hmC, the TET and IDH proteins seem to play a very important role as determinants of the global 5hmC content.

\* Correspondence: maxim.ivanov@ki.se

† Contributed equally

<sup>1</sup>Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Nanna Svartz väg 2, 17177 Stockholm, Sweden  
Full list of author information is available at the end of the article

Hydroxymethylcytosine can serve as an intermediate of active DNA demethylation if further oxidized by TET enzymes to 5-formylcytosine (5fC) and then to 5-carboxylcytosine (5caC), which in turn is removed by thymine-DNA glycosylase. The resulting abasic site is repaired by the base excision repair mechanism, producing unmethylated cytosine [8,9]. The level of 5fC was found to be extremely low in mouse embryonic stem cells and brain cortex (0.02 to 0.002% from total cytosine) [9,10]. Despite the recent progress in genome-wide mapping of 5fC in mouse embryonic stem cells, there is still no convincing evidence for its role as an epigenetic modification [11,12]. In contrast, 5hmC is 10 to 100 times more abundant than 5fC and can stably persist *in vivo*, thus serving as an epigenetic mark with unique regulatory functions.

Several functional differences have been demonstrated between the two DNA modifications 5mC and 5hmC. Firstly, the vast majority of CpG sites throughout the genome (except for those located in CpG islands and shores) are constantly methylated. In contrast, 5hmC can be found at only a relatively small subset of CpG sites. Furthermore, variable methylation correlating with gene expression has been described mostly in the promoter regions, whereas the genomic distribution of 5hmC is biased towards exonic regions. In addition, 5mC is, in most cases, a repressive epigenetic mark, whereas the presence of 5hmC correlates with active gene transcription [13-21]. The latter observation can be explained by the different affinity of 5mC and 5hmC for methylated DNA binding proteins. It was demonstrated that MeCP1, MBD1, MBD2 and MBD4 bind to methylated but not to hydroxymethylated loci [22,23], whereas MBD3 binds to 5hmC with even higher affinity than it does to 5mC [24]. Moreover, the composition of DNA binding and histone modifying proteins in 5hmC-containing loci can also be very different from that of methylated DNA, thus resulting in different gene expression outcomes [16,18-20,25].

Highly variable levels of 5hmC have been detected in many human and mammalian tissues and cell types with characteristic tissue-specific patterns [14]. 5hmC is especially enriched in brain tissue and in embryonic stem cells, where it constitutes up to 1 to 1.3% of the total cytosine content [14,17,26-29]. Furthermore, 5hmC has been shown to participate in neurodevelopment and has been associated with pluripotent cell states [30-32]. In cancers, such as prostate, breast, melanoma, glioma and colon carcinoma, the content of 5hmC is significantly decreased compared to normal tissue, is usually associated with decreased *TET* or *IDH* expression and, more importantly, can affect the tumor phenotype [7,14,33,34]. With respect to human liver, very little is known about 5hmC distribution and function.

In this study we demonstrate that adult human livers contain high levels of 5hmC, which are comparable to those found in the brain tissue. The analysis of genome-wide 5hmC hepatic distribution revealed that 5hmC is particularly enriched in coding regions of actively transcribed genes. In addition, the results show profound differences in both genomic distribution and global 5hmC content between fetal and adult liver samples. These differences were further pronounced by pathway analysis of genes enriched for 5hmC elements, which clearly showed consistency with liver development, thus suggesting that 5hmC may play an important role in hepatic development and function.

## Results

### Global 5mC and 5hmC content in fetal and adult human livers

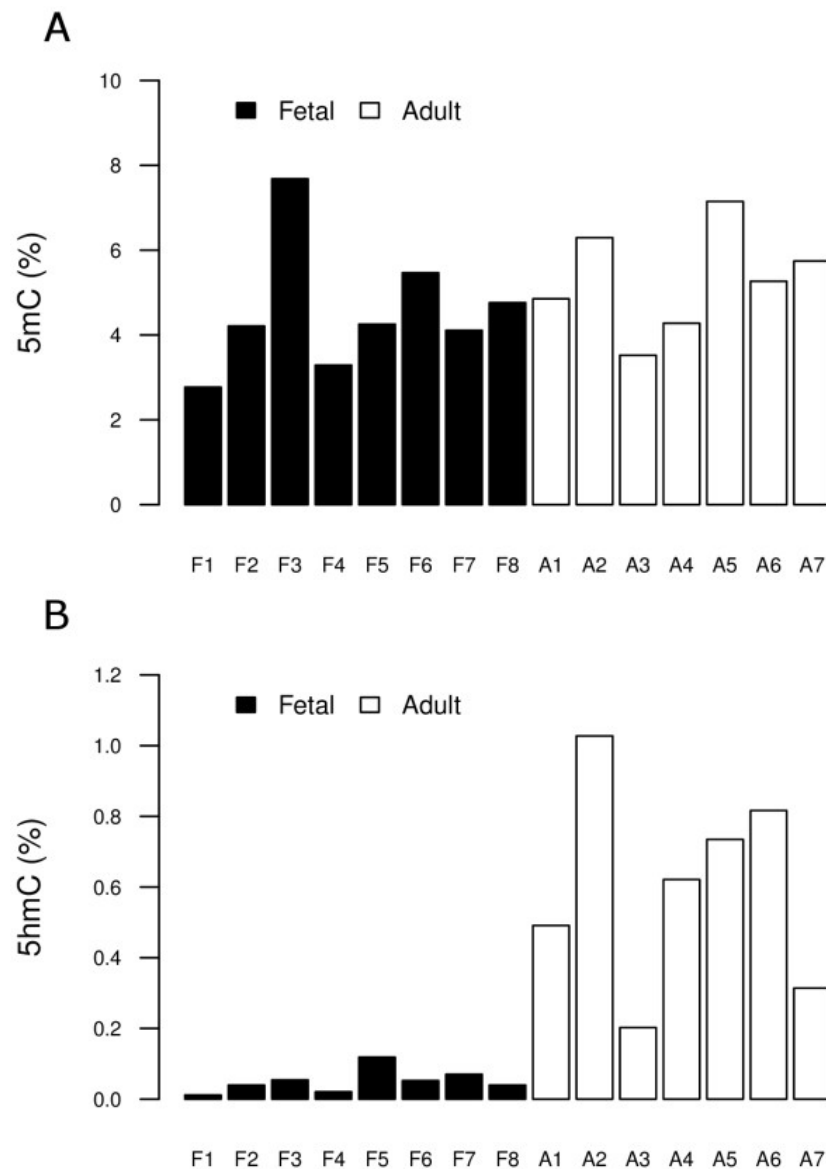
Global levels of 5mC and 5hmC were determined in eight fetal and seven adult human livers using liquid chromatography-mass spectrometry (LC-MS) analysis, based on calibration curves with known 5mC and 5hmC content. The quantification range of 5hmC was from 0.0625 to 1.0% of total cytosine, suggesting our method is highly sensitive. The linear standard curve has excellent fit ( $R^2 > 0.999$ ; Figure S1 in Additional file 1). Therefore, we used this calibration curve to determine global levels of 5mC and 5hmC in eight fetal and seven adult human livers. The 5mC and 5hmC contents of fetal and adult samples are shown in Figure 1 (for details, see Table S2 in Additional file 1). The median 5mC content was 4.2% of total cytosine in fetal livers and 5.3% in adult livers. By contrast, much less 5hmC was found in fetal livers compared to adult livers (Figure 1). All eight fetal samples had a 5hmC content lower than 0.125%, whereas the signal from six out of seven adult samples was above 0.2%, and the median 5hmC content for adults was 0.62% of total cytosine.

### Immunohistostaining of human liver samples

In order to explore the distribution of 5hmC in the liver, we immunostained human liver sections for both this cytosine modification as well as for 5mC. As depicted in Figure 2, both 5mC and 5hmC were detected in the nuclei of most hepatocytes within the liver tissue sections. No significant difference regarding intranuclear localization between 5mC and 5hmC was observed. Three-dimensional analysis of the sections further revealed the abundant distribution of 5hmC to the majority of hepatocytes (data not shown).

### Quantification of *TET1-3* and *IDH1-2* transcripts

The observed differences in global 5hmC content between fetal and adult cohorts may be explained by

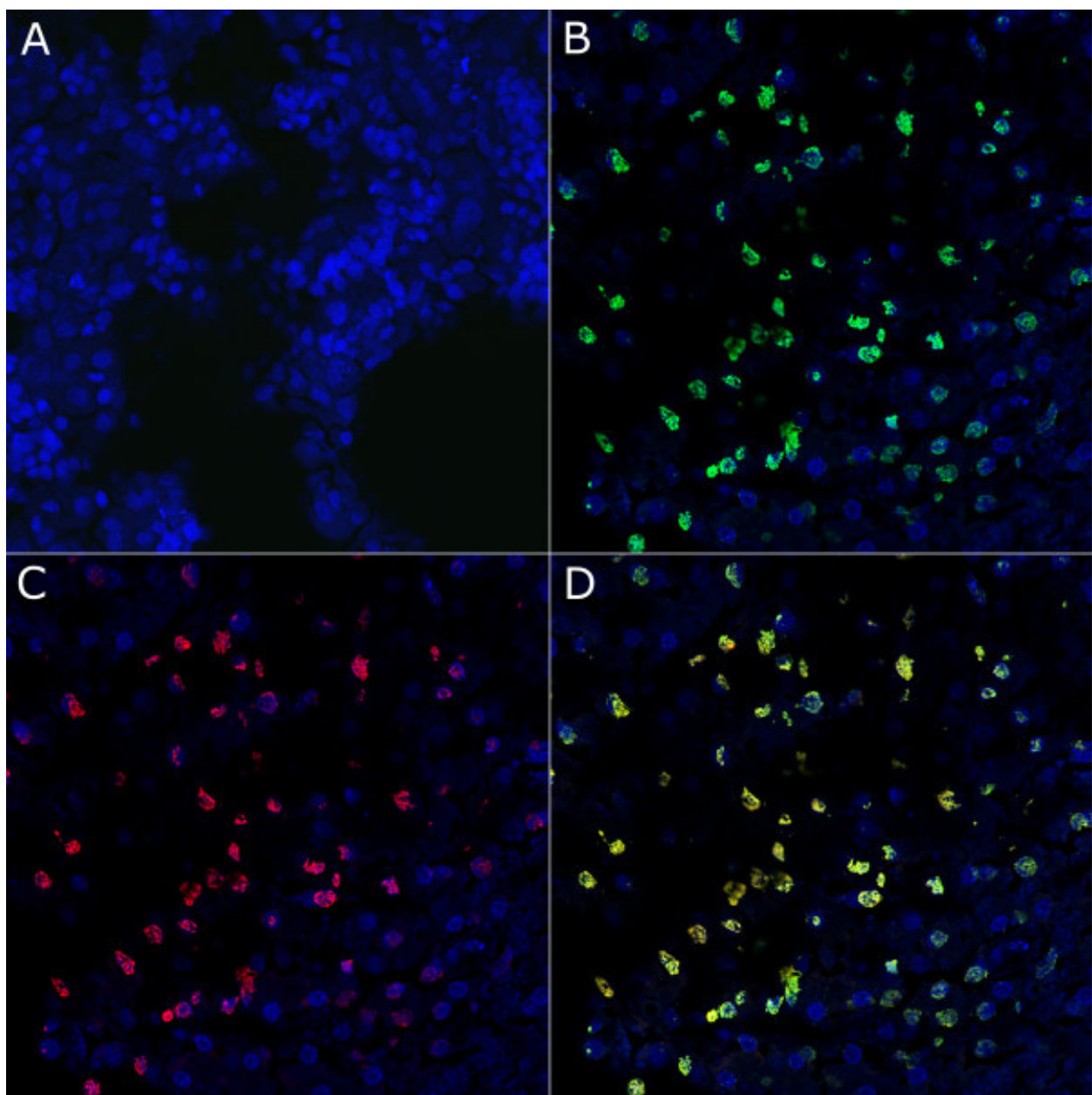


**Figure 1** Total 5mC and 5hmC in fetal and adult livers as determined by LC-MS. The percentage of (a) 5mC and (b) 5hmC in relation to the total cytosine content in eight fetal and seven adult livers is shown.

differences in expression of the TET and/or IDH enzymes, which are involved in the conversion of 5mC to 5hmC. The mRNA levels of *TET1-3* and *IDH1-2* were assessed by quantitative RT-PCR in 14 fetal and 33 adult livers, including the samples investigated by LC-MS. Interestingly, the results showed a significant decrease in all *TET* genes ( $P < 0.0001$ ) in adult livers when compared to fetal (Figure 3). However, the slight increase in *IDH1* ( $P = 0.0024$ ) and *IDH2* ( $P < 0.0001$ ) in adult samples may potentially explain the higher total 5hmC content in these livers compared to fetal 5hmC levels.

#### Genome-wide profiling of 5hmC

To investigate the genome-wide distribution of 5hmC, we enriched the 5hmC-containing fraction of the DNA samples from the eight fetal and seven adult human livers. Both 5hmC-enriched DNA samples and their non-enriched counterparts were ligated with adapters, amplified and then subjected to next-generation sequencing (NGS), as described in Materials and methods. The main quality metrics of the NGS experiment as well as the number of called 5hmC peaks for each liver sample are listed in Table S3 in Additional file 1.

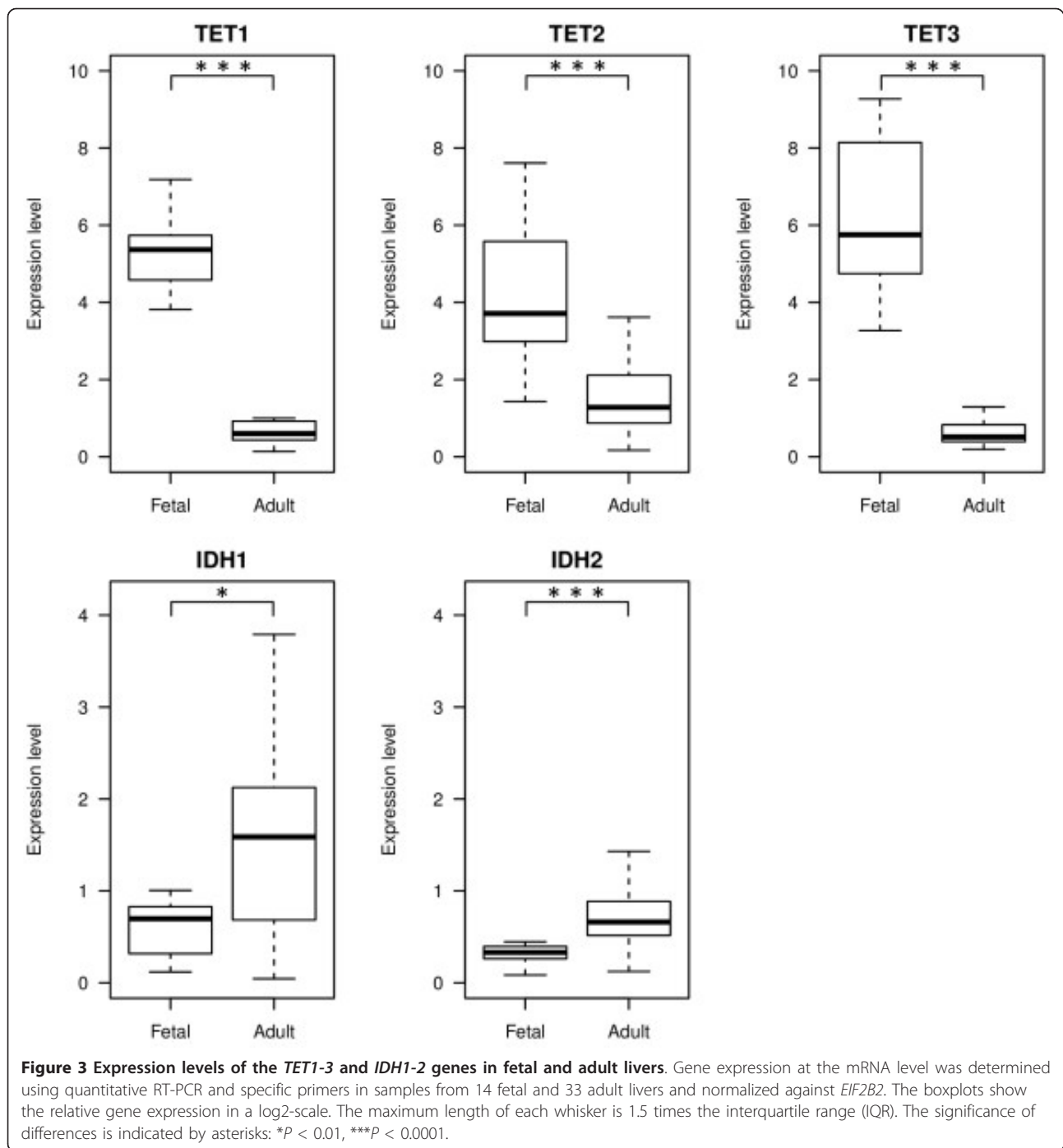


**Figure 2 Immunostaining of human adult liver for 5mC and 5hmC.** Representative stainings of 40  $\mu$ M adult liver sections are shown. **(a)** Negative control without the primary antibody. **(b)** Staining with the monoclonal antibody to 5mC (green) and DAPI (blue). **(c)** Staining with a polyclonal antibody against 5hmC (red) and DAPI (blue). **(d)** Merge of 5mC and 5hmC staining and DAPI. As evident from (d), 5mC and 5hmC show similar patterns of intranuclear distribution, indicating that 5hmC, when present, tends to co-localize with 5mC in the hepatocyte nucleus.

Overall, the fetal liver samples showed significantly fewer 5hmC peaks (11,366 to 32,522) covering a smaller portion of the genome (10 to 40 Mb) compared to adult samples (68,779 to 134,956 peaks covering 80 to 203 Mb). These results are in agreement with the finding by LC-MS that global 5hmC content is lower in fetal livers than in adults. Furthermore, the 5hmC peaks appear to be unevenly distributed among chromosomes, being

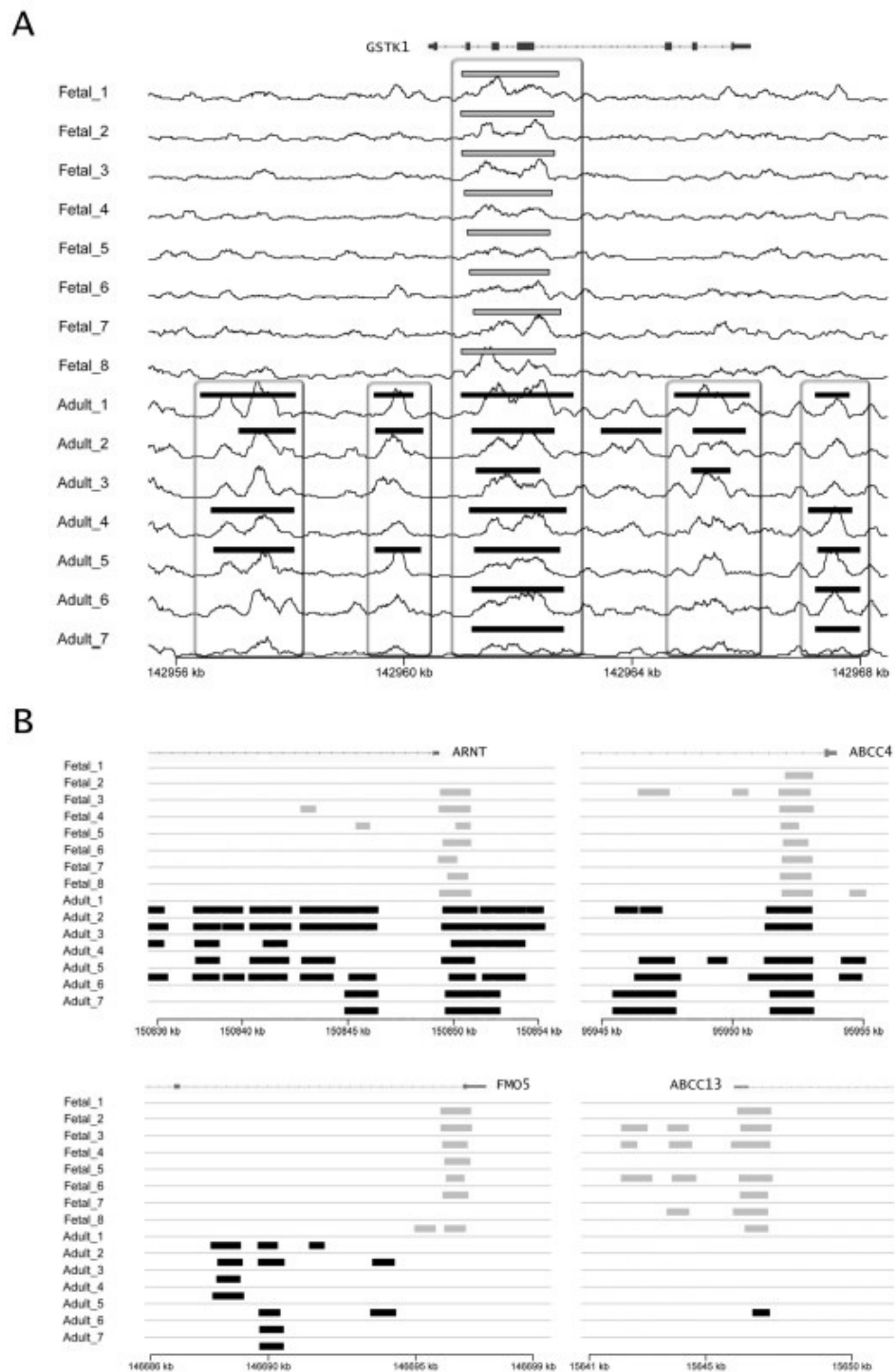
over-represented on chromosomes 16, 17, 19 and 22 and under-represented on the sex chromosomes (Figure S2 in Additional file 1).

When investigating the proportion of peaks that were shared by samples belonging to each cohort, a small fraction of peaks was shared by all fetal samples (2.5%) and all adult samples (8.3%; Figure S3 in Additional file 1). A large majority of the peaks were identified as unique



peaks (54.4% in fetal and 36.1% in adult livers), suggesting high interindividual variability in the genomic distribution of 5hmC. On the other hand, a significant proportion of hydroxymethylation reproducibly appears at the same genomic locations, as seen when comparing 5hmC peaks shared by more than one sample. Examples of the distribution of 5hmC peaks in five different genes are shown in Figure 4. We focused on the continuous

genomic intervals, or '5hmC blocks', where 5hmC occupancy was detected in at least two samples in a given cohort (Figure 4a). In other words, '5hmC blocks' represent genomic intervals that are prone to be hydroxymethylated in a given cohort (although not necessarily hydroxymethylated in every sample in this cohort). In contrast, the probability to detect 5hmC outside of 5hmC blocks is relatively low. Of the five genes illustrated in



**Figure 4** The distribution of 5hmC peaks in five liver genes. **(a)** A detailed view of the *GSTK1* gene. For each 5hmC-enriched sample (y-axis), the read depth is plotted against the genomic coordinates in the given interval (x-axis). The 5hmC peaks called by the MACS software are shown by the short horizontal lines. The 5hmC blocks identified in this study are denoted by vertical rectangles. **(b)** Representative illustration of the genomic positions of 5hmC peaks identified in the genes *ARNT*, *ABCC4*, *FMO5* and *ABCC13*. The fetal peaks are shown in gray and the adult peaks are shown in black.

Figure 4, most had many more 5hmC blocks in adult livers, whereas in the *ABCC13* gene, 5hmC blocks were mainly identified in fetal livers. The relation between this variability and the differences in gene expression between the livers has to await analyses in larger cohorts. The full lists of fetal ( $n = 29,917$ ) and adult 5hmC blocks ( $n = 116,911$ ) identified in this study are available in Additional files 2 and 3, respectively.

#### Locus-specific validation of next-generation sequencing data

To validate our NGS data, we profiled 5mC and 5hmC content at single-base resolution for two randomly selected genomic intervals in four fetal and four adult liver DNA samples. The first interval (containing four CpG sites overlapping with a fetal-only 5hmC block) was located in intron 15 of the *DROSHA* gene, whereas the second one (containing seven CpG sites) was located in intron 2 of the *CDH2* gene and manifested adult-only hydroxymethylation. The results of the single-base analysis are shown in Figure S5 in Additional file 1. The fetal hydroxymethylation in *DROSHA* (averaged over four CpG sites in each sample) could reach up to 35%, whereas in adult samples it did not exceed 4%. In contrast, fetal hydroxymethylation in *CDH2* (averaged over seven CpG sites) was never higher than 5%, whereas in adult samples it varied between 49 and 56%. Thus, despite the single false positive and false negative 5hmC peaks identified for both genes, the 5hmC values obtained from the validation experiment are in good agreement with the NGS data.

#### Genomic and functional localization of 5hmC

Next, we investigated the possible enrichment of 5hmC blocks in RefSeq genes, microRNA (miRNA) and long non-coding RNA (lncRNA) genes, hepatic enhancers, CpG islands (CGIs), CGI shores and repetitive sequences. We found that both fetal and adult 5hmC blocks are slightly enriched in RefSeq genes and strongly enriched in enhancers, miRNA genes, CGIs and CGI shores, but under-represented in lncRNA genes and repetitive regions (Figure 5a). Furthermore, the relative enrichment of 5hmC blocks in enhancers, miRNA genes, CGIs and CGI shores is approximately double in fetal samples compared to adults, whereas in RefSeq genes and lncRNA genes as well as in repetitive regions the rate of 5hmC enrichment remains nearly constant during liver development.

To investigate if cytosine hydroxymethylation in human liver correlates with active gene transcription, we analyzed how 5hmC blocks are distributed among genes with different expression levels. To this end, we quantified the genome-wide mRNA expression levels of 14 fetal and 86 adult liver samples (including all samples

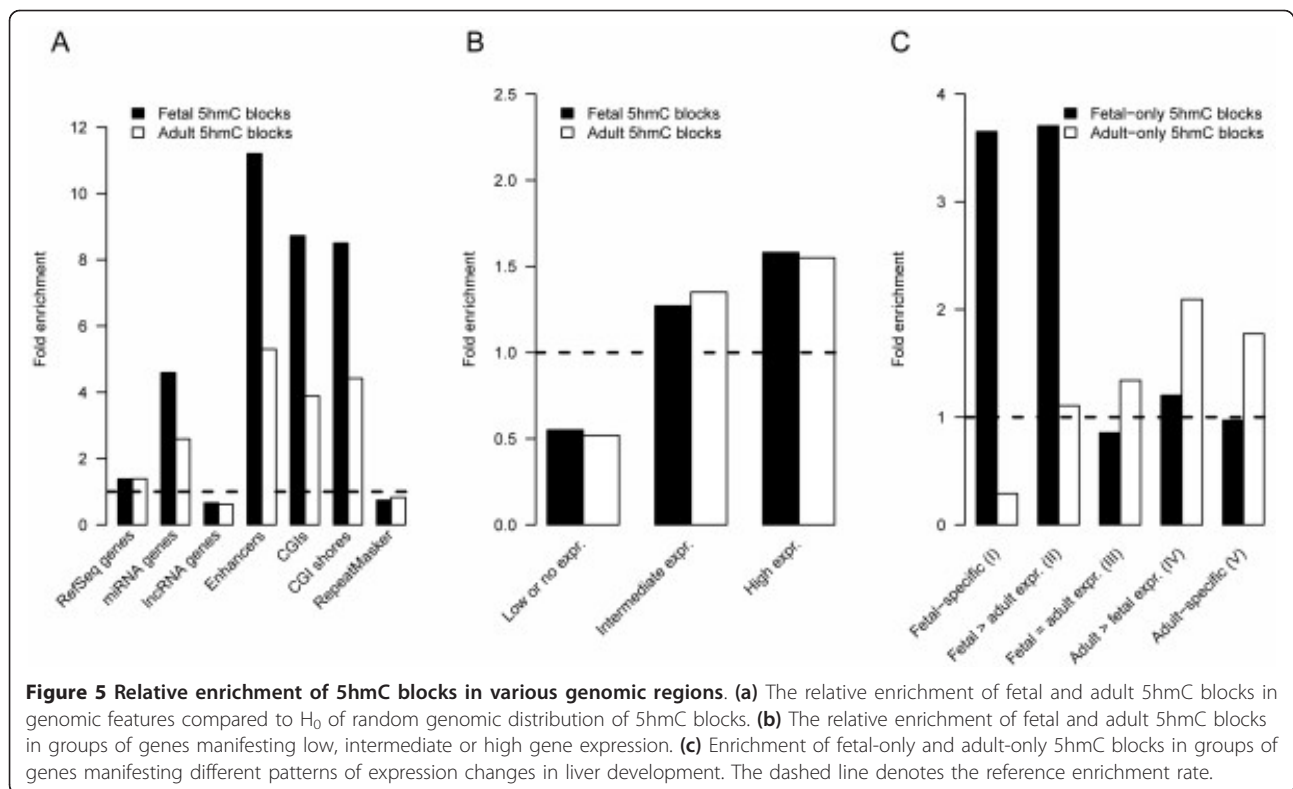
employed in the NGS study). The analyzed 17,771 RefSeq genes were divided into low, intermediate and highly expressed genes, separately for fetal and adult cohorts. The relative enrichment of 5hmC blocks in each of these three categories of genes is shown in Figure 5b. Both fetal and adult 5hmC blocks are under-represented in the genes with low expression or silent genes, and over-represented in genes with intermediate or high expression, suggesting a likely role for hydroxymethylation in active gene transcription.

#### 5hmC localization during development

The observed profound differences in both global 5hmC content and the number of called 5hmC peaks between the fetal and adult liver samples suggest that 5hmC distribution is significantly modified during liver development throughout the genome. When comparing the fetal and adult hydroxymethylomes, we found 5,038 5hmC blocks present in fetal livers only and 96,097 blocks detectable only in adult samples. These fetal- and adult-specific 5hmC blocks were then investigated for their distribution among the following five gene groups: I) fetal-specific genes (expressed in fetal samples, but silent in adults;  $n = 464$ ); II) genes expressed in fetal samples at least two-fold higher than in adults ( $n = 805$ ); III) non-developmentally expressed genes ( $n = 8,731$ ); IV) genes expressed in adult samples at least two-fold higher than in fetal samples ( $n = 1,079$ ); and V) adult-specific genes (expressed in adult cohort, but silent in fetal cohort;  $n = 306$ ). Relative enrichment values of genic fetal-specific and adult-specific 5hmC blocks in these groups of genes are shown in Figure 5c. As evident from these data, fetal-specific 5hmC blocks were seen strongly enriched in fetal expressed genes (groups I and II) but not in non-developmental (group III) or adult-specific genes (groups IV and V). Accordingly, adult-only 5hmC blocks are strongly under-represented in fetal-only genes (group I), but skewed towards genes that are exclusively expressed in the adult livers (groups IV and V). Hence, developmental patterns of hydroxymethylation correlate with developmental patterns of gene transcription in human livers.

#### 5-Hydroxymethylation in relation to gene function

Finally, genes intersecting with fetal-only or adult-only 5hmC blocks were annotated to specific biological pathways using the GREAT software. Those genes that acquired hydroxymethylation in the adult state were found to be involved in active catabolic and metabolic processes, such as mRNA catabolic process, carbohydrate transport, fatty acid oxidation and lipid oxidation, which is well in line with the function of the adult liver. Meanwhile, the genes containing fetal-only 5hmC blocks were more specific to pathways for differentiation and



development, such as regulation of cell differentiation, stem cell development, and establishment or maintenance of cell polarity. The top 10 biological processes for these two groups of genes are given in Table 1.

#### The comparison of human hepatic and cerebellum hydroxymethylomes

We compared our data on fetal and adult hepatic hydroxymethylomes with the previously published 5hmC map of the human adult cerebellum [30]. The study of Szulwach and colleagues [30] revealed 43,272 5hmC-containing genomic intervals in the cerebellum, covering 15.1 Mb. Among them, 8.4 Mb (56%) do intersect with the 5hmC blocks in adult liver. In addition, 3.3 Mb of 5hmC-containing sequences are shared between cerebellum and fetal liver, and 2.7 Mb are common for all these three tissues. Thus, these data indicate that, even though 5hmC distribution is tissue-specific, some 5hmC elements are conserved between different tissues and developmental stages. Pathway analysis of these conserved 5hmC elements reproducibly identified sterol metabolism and insulin response genes among the top 10 pathways (Table S4 in Additional file 1).

#### Discussion

Our study reports, for the first time, the characteristics of the 5hmC epigenetic modification in fetal and adult

human livers. We have quantified the global 5hmC content in fetal and adult human livers by LC-MS and showed significantly higher levels of 5hmC in adult compared to fetal samples, accompanied by a correspondingly higher expression of *IDH* genes, known 5hmC regulators. Our investigations of the global genomic distribution patterns of 5hmC using NGS categorize this DNA modification as an important mark in hepatic gene expression and development.

Until recently, quantification of 5hmC relied on less precise methods, mainly antibody detection techniques, and utilized a very small sample size. For example, Li and Liu [35] reported the 5hmC human liver content to be 0.45% of total DNA (0.09% of total cytosine), whereas Nestor *et al.* [14] reported values of approximately 0.22% of total cytosines. LC-MS enables sensitive and reliable quantification of 5hmC. Previous studies in murine livers found that 5hmC content quantified by LC-MS was very low (0.05 to 0.07% from the total cytosine) compared to other murine tissues (for example, 0.3 to 0.7% in the central nervous system [26]). Importantly, LC-MS quantification of 5hmC in human liver, to our knowledge, has not been carried out before. Utilizing LC-MS together with a protocol developed to detect both 5mC and 5hmC within the same tissue sample, we were able to quantify the 5mC and 5hmC content in seven adult and eight fetal livers. In contrast to the data reported in mice, our results demonstrate that



**Table 1 Functional annotation of developmentally hydroxymethylated regions**

Biological process	Binomial test			Hypergeometric test	
	Raw <i>P</i> -value <sup>a</sup>	FDR <i>Q</i> -value <sup>b</sup>	Fold enrichment <sup>c</sup>	FDR <i>Q</i> -value <sup>d</sup>	Fold enrichment <sup>e</sup>
<b>Genes enriched with adult-only 5hmC blocks</b>					
Nuclear-transcribed mRNA catabolic process	4.30E-182	3.22E-180	2.49	8.82E-03	1.16
mRNA catabolic process	8.88E-165	5.52E-163	2.25	1.77E-02	1.14
Carbohydrate transport	3.29E-125	1.51E-123	2.06	5.91E-03	1.15
Vesicle localization	9.48E-113	4.05E-111	2.20	3.73E-02	1.16
Fatty acid oxidation	5.52E-102	2.12E-100	2.22	4.59E-03	1.20
Lipid oxidation	1.44E-100	5.39E-99	2.20	3.97E-03	1.20
mRNA 3'-end processing	1.73E-98	6.27E-97	2.24	1.11E-02	1.17
Monocarboxylic acid catabolic process	1.35E-95	4.80E-94	2.21	5.31E-03	1.20
RNA 3'-end processing	2.92E-94	1.02E-92	2.07	2.13E-02	1.15
Fatty acid catabolic process	2.99E-90	9.99E-89	2.22	1.00E-02	1.20
<b>Genes enriched with fetal-only 5hmC blocks</b>					
Regulation of granulocyte differentiation	5.88E-23	2.45E-20	6.20	3.56E-02	2.58
Regulation of myeloid cell differentiation	9.74E-17	1.67E-14	2.19	1.84E-03	1.64
Induction of apoptosis by extracellular signals	5.49E-15	7.18E-13	2.18	9.28E-03	1.54
Positive regulation of steroid metabolic process	1.76E-12	1.61E-10	3.87	9.82E-03	2.37
Stem cell development	6.00E-12	5.05E-10	2.09	4.16E-04	2.06
Establishment or maintenance of cell polarity	9.26E-11	6.14E-09	2.02	1.97E-02	1.63
Regulation of protein binding	6.32E-10	3.38E-08	2.03	2.36E-02	1.64
Stem cell maintenance	8.22E-10	4.31E-08	2.00	9.44E-04	2.04
Somatic stem cell maintenance	6.28E-09	2.59E-07	2.15	2.15E-02	2.04
Negative regulation of lipoprotein particle clearance	1.99E-07	6.11E-06	6.52	3.08E-02	3.73

Shown are the top 10 significantly enriched biological processes sorted according to the binomial *P*-values. <sup>a</sup>Uncorrected *P*-value from the binomial test over genomic regions; <sup>b</sup>false discovery rate *q*-value; <sup>c</sup>fold enrichment of number of genomic regions in the test set; <sup>d</sup>false discovery rate *q*-value from the hypergeometric test over genes; <sup>e</sup>fold enrichment of number of genes in the test set.

5hmC levels in the majority of human adult livers (0.5 to 1% 5hmC) remain comparable to those found in the cerebral cortex (1 to 1.3% 5hmC), the most DNA-hydroxymethylated human tissue identified to date [14,29]. As adult human livers have notably higher global 5hmC content than murine ones, mice are unlikely to serve as relevant models for the human hepatic hydroxymethylome. Moreover, LC-MS quantification of multiple liver samples also demonstrated that the overall content of both 5mC and 5hmC is variable between the samples.

Interestingly, fetal samples showed a significantly lower 5hmC content and very similar levels of total 5mC when compared to adult livers. Drastic differences in 5hmC levels between adult and fetal livers can suggest that DNA hydroxymethylation is an important epigenetic phenomenon for liver development. The observed increase of global 5hmC content in adult compared to fetal livers is interesting because, in the majority of studies, a high 5hmC content has been associated with an embryonic or undifferentiated state of tissues or cells. For example, Ruzov and co-authors [32] report changes of 5hmC abundance in the development of

mouse and postulate that 5hmC is enriched in embryonic context compared to adult tissues. Szwagierczak and colleagues [28] showed significantly more 5hmC in undifferentiated mouse embryonic stem cells than in corresponding embryoid bodies, where the decrease in 5hmC content during differentiation was attributed to decreased expression of *TET1*. In addition, reprogramming of differentiated cells into induced pluripotent stem cells activates TET enzymes and leads to accumulation of 5hmC, where the link between 5hmC levels and development was explained by the presence of binding sites for pluripotency-associated transcription factors (OCT4, SOX2) in the promoter of *TET1* [13,31]. In contrast to these studies, our data indicate a decrease in the extent of 5hmC modifications from the early embryonic stages to adult liver. Similarly, Hahn and colleagues [36] showed that 5hmC is more abundant in terminally differentiated neurons than in neural progenitor cells, thus suggesting that the global 5hmC content increases with terminal maturation in specific tissues.

Our studies revealed that the lower 5hmC content in the fetal livers was accompanied by lower expression of

the *IDH1* and *IDH2* genes, known regulators of 5hmC level. Hence, it is tempting to speculate that the increase of 5hmC in adult samples is due to the developmental increase of IDH enzymes. However, we did not observe any correlation between the individual 5hmC content and the levels of either *TET1-3* or *IDH1-2* transcripts in individual livers (data not shown), indicating that other factors contribute to the global 5hmC content of human liver. Thus, we cannot completely rule out the possibility that the level of TET and/or IDH enzymes can be regulated in a post-transcriptional or post-translational manner, and the developmental burst of 5hmC is controlled by mechanisms other than the increased expression of *IDH1* and *IDH2* mRNA.

The observed abundance of 5hmC in adult livers and its preferential localization in actively transcribed genes suggest that bisulfite sequencing data from adult livers should be considered with caution, as bisulfite conversion does not distinguish between 5mC and 5hmC [37]. Hence, specific methods allowing discrimination between these two cytosine modifications become especially important in human liver research.

NGS of 5hmC-enriched genomic DNA obtained from fetal and adult liver samples revealed profound differences in the genomic distribution of 5hmC peaks between the cohorts and significant interindividual variation among samples within the groups. Interestingly, the genomic distribution of 5hmC peaks did not occur at random but appeared to be dependent on the genomic context and on the stage of liver ontogeny. In general, 5hmC is relatively under-represented in repetitive sequences and enriched in coding regions of actively transcribed genes (Figure 5a). In addition, 5hmC peaks are even more enriched in CGI shores (and hence frequently overlap with CGIs), which is in agreement with the tendency of 5hmC to occur in regions with increased CpG density (Figure S4 in Additional file 1). These 5hmC distribution patterns are quite consistent with previous reports for mammalian nervous tissue and embryonic stem cells [13-17,19-21], suggesting a functional role for 5hmC in the regulation of gene expression. In addition, the observed tendency of 5hmC peaks to occur in coding genes and regions with increased CpG density may explain the over-representation of 5hmC blocks on chromosomes 16, 17, 19 and 22 (Figure S2 in Additional file 1). Moreover, the depletion of 5hmC from sex chromosomes as seen here has been observed before in human embryonic stem cells and mouse brain [16,30].

We found that 5hmC blocks in both fetal and adult livers are strongly enriched in genes encoding miRNAs and depleted within lncRNA genes, thus suggesting a role for 5hmC in the regulation of the expression of miRNA genes (Figure 5a). In addition, the observation

of high enrichment of 5hmC in hepatic enhancers suggests that 5hmC could play a role in the regulation of gene expression through distal regulatory elements. Moreover, our results evidently indicate the involvement of 5hmC in the developmental expression of protein-coding genes in human liver. Fetal-only 5hmC blocks (that is, genomic intervals that lose 5hmC during liver development, despite the increasing genomic 5hmC content) are especially enriched within genes annotated as belonging to developmental and differentiation-related processes, whereas in adult livers, *de novo* acquired 5hmC appears to predominantly reside in genes associated with hepatic metabolism. These data are in line with the observation of Thomson and co-authors [38], who recently described the dynamics of 5hmC and its effect on gene expression in mouse livers. The authors showed that following exposure to phenobarbital, rapid, dynamic and reciprocal changes in the level of 5hmC and 5mC occurred over the promoter regions of the murine genes known to be transcriptionally regulated by phenobarbital. The changes in hydroxymethylation and methylation further coincided with changes in the histone marks H3K4me2, H3K27me3 and H3K36me3, and indicate that cytosine hydroxymethylation may be crucial in the acute transcriptional regulation of specific hepatic genes.

Here we provide the evidence that adult human liver DNA contains a high level of 5hmC. The brain is another differentiated human tissue that is highly hydroxymethylated [14,29]. We compared our hepatic 5hmC data with the available cerebellum dataset, obtained using the same chemical labeling-based method for 5hmC capture as in our study [30]. The results of pathway analysis of 5hmC intervals that are shared between cerebellum and fetal and adult human livers suggest that the presence of specific 5hmC signatures within genes related to sterol metabolism and insulin response could be conserved between various tissues.

## Conclusions

In this study, we for the first time demonstrate that, in contrast to earlier reports, 5hmC is an abundant epigenetic modification in adult human liver. Our findings on the genomic distribution of 5hmC suggest that this epigenetic mark may be an important determinant of the expression of both protein-coding and miRNA genes in hepatocytes. Furthermore, 5hmC appears to play a significant role in hepatic gene expression changes during liver development. Hence, cytosine hydroxymethylation may be a missing link in the full understanding of the functional as well as the developmental processes of the liver, and also can contribute to interindividual differences in liver function and susceptibility to liver disease. Moreover, our findings suggest the importance of using

methods that permit the discrimination between 5mC and 5hmC when investigating the liver epigenome, as the traditional methods based on bisulfite sequencing can cause erroneous conclusions.

## Materials and methods

### Human tissues

Liver samples from fourteen 8- to 12-week-old fetuses were acquired from Karolinska University Hospital (Huddinge, Sweden). Fifty-two liver samples originate from adult organ donors who met accidental death. Of these, 33 were acquired from Karolinska University Hospital, and 19 were commercially purchased from XenoTech (Lenexa, KS, USA) and the International Institute for the Advancement of Medicine (IIAM; Edison, NJ, USA). This work was carried out in compliance with the methods within the Helsinki declaration. The use of fetal and adult liver tissues for the purposes of this study was approved by the Regional Ethics Committee. The reference numbers for the ethical approvals by the Karolinska Institutet Internal Review Ethical Board are 2010/541-23/1, 2010/541-31/1, 2010/678-31/3 and 280/00.

### Nucleic acid purification and quantification

DNA and total RNA were isolated from the liver samples using a AllPrep DNA/RNA/Protein Mini kit (QIAGEN, Germantown, MD, USA; catalogue number 80004). DNA was quantified using a Quant-iT PicoGreen dsDNA Assay kit (Invitrogen, Carlsbad, CA, USA; catalogue number P11496) and a SpectraMax Gemini XPS/EM Fluorescence Microplate Reader (Molecular Devices, Sunnyvale, CA, USA). The RNA samples were quantified and assessed for integrity using the Agilent Bioanalyzer 2100 with RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA, USA; catalogue number 5067-1511). The RIN (RNA integrity number) values of the RNA samples used for cDNA amplification were  $\geq 8$ .

### Liquid chromatography-mass spectrometry

In order to achieve monomers, 100 ng of each DNA sample were treated with DNA Degradase Plus (ZymoResearch, Irvine, CA, USA; catalogue number E2020) according to the manufacturer's instructions. In parallel, 12 calibration samples containing 0.3125% to 10% 5mC and 0.0625% to 1% 5hmC were prepared from 5-Methylcytosine and 5-Hydroxymethylcytosine DNA Standard Sets (ZymoResearch, catalogue number D5405). The standards were degraded to monomers in the same manner as the DNA samples. Prepared samples were placed into the auto sampler and 5  $\mu$ l of samples were injected onto the LC-MS system.

Quantification of C, 5mC and 5hmC was accomplished using an Agilent UPLC 1290 system coupled to an Agilent 6460 triple quadrupole (QQQ) equipped with JetStream

electrospray ionization (ESI). Chromatographic separation was performed on a Waters Atlantis T3 (C18, 3  $\mu$ m particles, 100 mm length  $\times$  2.1 mm inner diameter; Waters, Milford, MA, USA) at a flow rate of 300  $\mu$ l/minute and 17°C. Mobile phase A consisted of 10 mM ammonium acetate in MilliQ water, and mobile phase B consisted of 10 mM ammonium formate in methanol with 0.1% (v/v) formic acid. The initial mobile phase composition was 0% of B for 10.2 minutes. From 10.2 to 13.0 minutes, mobile phase B was increased to 20%, and held for 1 minutes. From 14.1 minutes, mobile phase B was decreased to 0% and held until 19.0 minutes.

Mass spectrometric detection was carried out using 6460 QQQ in positive ion multiple reaction monitoring mode. The first quadrupole (Q1) was set to transmit the  $[M+H]^+$  ions of the analytes and Q3 was set to transmit selected fragment ions of analytes. Multiple reaction monitoring conditions, such as dwell time, Fragment and CE, are given in Table S1 in Additional file 1. The sheath gas heater was set to 350°C, gas flow was 5 (l/minute) and the sheath gas flow was 10. The instrument was controlled by Mass Hunter Workstation software (Agilent Technologies). 5mC and 5hmC contents were calculated based on the calibration curve constructed by serial standard samples (the standard samples were run twice, before and after real sample analysis). The linear fits of determined area over standard amounts ratio gave  $R^2$  values of 0.99981 for 5mC and 0.99941 for 5hmC (Figure S1 in Additional file 1).

### Immunohistochemistry

Sections (40  $\mu$ m) of snap frozen liver were cut with a microtome cryostat (Cryo-Star HM 560M, MICROM International GmbH, Walldorf, Germany) and fixated in 4% formaldehyde. The denaturation of sections was accomplished by treating the tissues with 1 M HCl for 30 minutes at 37°C. Subsequently, the sections were neutralized by incubating them in 100 mM Tris-HCl (pH 8.5) for 10 minutes. As primary antibodies, a rabbit polyclonal antibody to 5hmC (1:5,000; ActiveMotif, Carlsbad, CA, USA; catalogue number 39769) and a mouse monoclonal antibody to 5mC (1:500; ActiveMotif, catalogue number 39649) were used. For secondary antibodies, we used a goat antibody to rabbit Alexa 647 (1:500; Invitrogen, Carlsbad, CA, USA; catalogue number A21244) and a rabbit antibody to mouse Alexa 488 (1:500; Invitrogen, catalogue number A11059). Samples were counter-stained with the fluorescent nuclear dye 4',6-diamidino-2-phenylindole included in the ProLong<sup>®</sup> Gold Antifade Reagent mounting media (Invitrogen, catalogue number P-36931). Three-dimensional images were visualized with a Zeiss confocal microscope and ZEN 2009 software. No processing of the images was performed.

### cDNA synthesis and real-time PCR

cDNA was obtained from 1 µg of each total RNA sample using Superscript III First-Strand Synthesis System (Invitrogen, catalogue number 18080-051). Real-time PCR was done with PerfeCTa SYBR Green SuperMix, Low ROX (Quanta Biosciences, Gaithersburg, MD; catalogue number 95056-500) on an ABI 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Relative expression values were calculated from Ct values using the  $\Delta\Delta$ Ct method. The housekeeping gene *EIF2B2* was used for the normalization of expression values. Primer sequences were published before [34,39]: TET1-F, GCTATACACAGAGCTCACAG; TET1-R, GCCAAAAGAGAATGAAGCTCC; TET2-F, CTTTCTCCCTGGAGAACAGCTC; TET2-R, TGC TGGGACTGCTGCATGACT; TET3-F, GTTCCTGGAG CATGTACTTC; TET3-R, CTTCTCTTTGGGAT TGTCC; IDH1-F, TCCGTCACCTGGTGTGTAGG; IDH1-R, GGCTTGTGAGTGGATGGGTA; IDH2-F, TGAAGTCCAGATAATACGGG; IDH2-R, CTGACAGCCCCACCTC; EIF2B2-F, TCAAGATTATCC GGGAGGAG; EIF2B2-R, ATGGAAGCTGAAAT CCTCGT.

### Gene expression profiling

The TargetAmp™-Nano Labeling Kit for Illumina® Expression BeadChip (Epicentre Biotechnologies, Madison, WI, USA; catalogue number TAN07924) was used to amplify and biotinylate the RNA samples, according to the manufacturer's instructions. For each sample, 750 ng of biotinylated cRNA were hybridized on Illumina HumanHT-12 v4 BeadChips, according to the standard protocol. The BeadChips were scanned within 24 h using a HiScanSQ scanner. The raw signals were exported using the GenomeStudio software, quantile normalized and further analyzed in R [40].

### Next-generation sequencing

Genomic DNA samples (3 µg in 120 µl) were sonicated using a Covaris S2 with the following settings: duty cycle = 10%, intensity = 5, cycles per burst = 200, time = 6 cycles (60 s each), mode = frequency sweeping, temperature = 4°C. Sonicated DNA was purified with 180 µl of Agencourt AMPure XP beads (Beckman-Coulter, Miami, FL, USA; catalogue number A63880) and assessed for concentration and size distribution with the DNA 1000 kit (Agilent Technologies, catalogue number 5067-1504) on an Agilent Bioanalyzer 2100. All samples manifested a size distribution between 100 and 250 bp, with peak at 150 to 200 bp.

Sheared DNA samples were enriched for the 5hmC-containing fraction with the Hydroxymethyl Collector kit (ActiveMotif, catalogue number 55013) following the

manufacturer's instructions. In parallel, non-enriched aliquots of sheared DNA were diluted 1:100. 5hmC-enriched samples and their non-enriched counterparts were then ligated to TruSeq adapters and PCR amplified according to the protocol detailed in Text S1 in Additional file 1. The resultant NGS libraries were assessed for size distribution with the DNA High Sensitivity kit (Agilent Technologies, catalogue number 5067-4626) on an Agilent Bioanalyzer 2100 and quantified using a Agilent QPCR NGS Library Quantification kit for Illumina Genome Analyzer (Agilent Technologies, catalogue number G4880A). The 5hmC-enriched adult and fetal samples were pooled separately and sequenced on two separate lanes. The control non-enriched samples were also pooled separately for the fetal and adult samples, and each pool was sequenced on seven lanes. The flow-cells (v3) were sequenced on an Illumina HiSeq2000 using paired-end sequencing of 100 bp and 6 bp for the index read.

### Sequencing data analysis

The FASTQ sequences of all samples were aligned to the reference human genome (NCBI37/hg19) using Burrows-Wheeler Aligner (BWA) [41]. Parameter -q 20 was used for read trimming, and all reads with mapping quality <20 were filtered out. Duplicate reads generated during the PCR amplification were removed using SAM-tool [42]. 5hmC peak identification was performed using MACS with the following parameters: tag size = 100; effective genome size = 2.70e+09; band width = 200; model fold = 10, 30; *P*-value cutoff = 1.00e-05; ranges for calculating regional lambda = 1,000 bp and 10,000 bp [43]. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE49291 [44].

All subsequent steps of data analysis were done using custom Python3 scripts, which are available on request. The coordinates of genomic features were downloaded from the UCSC Table Browser (tables refGene, lincRNAsTranscripts, cpgIslandExt, rmsk, wgEncodeBroadHmmHepg2HMM). The coordinates of 5hmC peaks in human cerebellum were downloaded as Supplementary Data Set 10 of the original paper of Szulwach and co-authors [30] and converted to Hg19 using the Galaxy LiftOver tool [45]. The relative enrichment of 5hmC elements in different functional categories of genes was evaluated using the Genomic Regions Enrichment of Annotations Tool (GREAT) [46], setting the basal gene regulatory domains within 5 Kb upstream and 1 Kb downstream of the transcription start site, and up to 1 Mb extended regulatory domains in both directions to the nearest gene's basal domain [47].

### Locus-specific validation of next-generation sequencing data

To validate NGS data at single-base resolution in selected genomic intervals, we employed the TAB-Seq method [48]. Briefly, 1 µg of DNA was oxidized with the WiseGene 5hmC TAB-Seq kit (WiseGene, Chicago, IL, USA) following the manufacturer's instructions [49]. Then, both the oxidized DNA and the corresponding untreated 500 ng aliquots were bisulfite converted with EZ DNA Methylation-Gold kit (ZymoResearch, catalogue number D5005), and the resultant DNA samples were amplified with the following primers: DROSHA-F, TTTAGTTGGGTGGTTTTATTTG; DROSHA-R, CAACTACTTTTATACCAAC; CDH2-F, GTGGTAGTGGTTGTAATTATATA; CDH2-R, CTAAAAAATAATCATTCTCCC. To increase the specificity of amplification, the first-round PCR products were purified, diluted 1:10,000 and amplified again with the nested primers: DROSHA(n)-F, GTTTTATGTTTTGTGGTAGA; DROSHA(n)-R, ACTTTTATACCAACCTAACCA; CDH2(n)-F, ATGAGAAGAGTTATGATATGGGAAT; CDH2(n)-R, AAAAATAAATCATTCTCCC. The second-round PCR products were purified and subjected to direct Sanger sequencing (Eurofins MWG Operon, Ebersberg, Germany) [50]. The C/(C+T) ratios from the oxidized samples reflect the 5hmC values, whereas such ratios from corresponding non-oxidized aliquots are equal to the sum of 5mC and 5hmC at the same CpG sites. Hence, the subtraction of the resultant C/(C + T ratios) permits the calculation the 5mC and 5hmC values for each CpG site.

### Additional material

**Additional file 1: Text S1 (the detailed NGS library preparation protocol), Figures S1 to S5 and Tables S1 to S4.**

**Additional file 2: Tab-delimited text file describing all 5hmC blocks discovered in the epigenomes of fetal livers.** Fields 1 to 3 contain genomic coordinates (Hg19) of each 5hmC block. Field 4 indicates the number of 5hmC-positive samples in a given age cohort (that is, those samples that manifest a 5hmC peak in a given genomic interval). Fields 5 and 6 provide some metrics of 5hmC occupancy of given 5hmC blocks; field 5 indicates the sum of 5hmC peak lengths (bp), whereas field 6 shows the sum of NGS reads in 5hmC peaks among all 5hmC-positive samples.

**Additional file 3: Tab-delimited text file describing all 5hmC blocks discovered in the epigenomes of adult human livers.** Fields 1 to 3 contain genomic coordinates (Hg19) of each 5hmC block. Field 4 indicates the number of 5hmC-positive samples in a given age cohort (that is, those samples that manifest a 5hmC peak in a given genomic interval). Fields 5 and 6 provide some metrics of 5hmC occupancy of given 5hmC blocks; field 5 indicates the sum of 5hmC peak lengths (bp), whereas field 6 shows the sum of NGS reads in 5hmC peaks among all 5hmC-positive samples.

### Abbreviations

5fC: 5-formylcytosine; 5mC: 5-methylcytosine; 5hmC: 5-hydroxymethylcytosine; bp: base pair; CGI: CpG island; IDH: isocitrate

dehydrogenase; LC-MS: liquid chromatography-mass spectrometry; lncRNA: long non-coding RNA; miRNA: microRNA; NGS: next-generation sequencing.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

MI conceived the study, carried out sample preparation and data interpretation and drafted the manuscript. MKal carried out the data analysis. MKac participated in the design of the study and in sample preparation. IB carried out the immunohistochemistry assays and the locus-specific validation of NGS data. KK carried out the LC-MS analysis. AR provided the human fetal liver samples. LM carried out the next-generation sequencing and microarray assays. AM, LM and MIS participated in planning of the study and coordinated and financed the work from their grants. All authors carried out manuscript revisions. All authors read and approved the final manuscript.

### Acknowledgements

The authors would like to acknowledge support from Mr Rickard Wahlström and Mr Gunnar Häggglund from Q&Q Labs AB (Göteborg, Sweden) and Clinical Proteomics Mass Spectrometry core facility at Karolinska University Hospital and Science for Life Laboratory (Stockholm, Sweden) for providing assistance in mass spectrometry and data analysis. We also acknowledge the important help provided by Dr Lena Ekström, Dr Inger Jonasson, Dr Chunxiao Song, Dr Chuan He, Ms Susanne Virding, Ms Heidi Martikainen, Ms Silva Kasela, Dr Paula Ann Kivistik and Mr Viljo Soo. This work is supported by grants from The Swedish Research Council, the IMI-JU project MIP-DILI (grant agreement 115336), the Seurat-1 project NOTOX, the Estonian Science Foundation (ETF9293), the European Union through the European Social Fund (MJD71) and the European Regional Development Fund, in the frame of the Centre of Excellence in Genomics; and targeted financing from the Estonian Government [SF0180142s08].

### Authors' details

<sup>1</sup>Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Nanna Svartz väg 2, 17177 Stockholm, Sweden.

<sup>2</sup>Estonian Genome Center, University of Tartu, Riia 23b, 51010 Tartu, Estonia.

<sup>3</sup>Cancer Proteomics Mass Spectrometry, Department of Oncology-Pathology, Science for Life Laboratory, Karolinska Institutet Science Park, Tomtebodavägen 23A, 17165 Solna, Sweden. <sup>4</sup>Division of Clinical Pharmacology, Department of Laboratory Medicine, Karolinska University Hospital at Huddinge, Medicingatan 5, 14186 Stockholm, Sweden. <sup>5</sup>Estonian Biocentre, Riia 23b, 51010 Tartu, Estonia. <sup>6</sup>AQ2 Institute of Molecular and Cell Biology, University of Tartu, Riia 23b, 51010 Tartu, Estonia.

Received: 24 June 2013 Revised: 30 July 2013

Accepted: 19 August 2013 Published: 19 August 2013

### References

- Ivanov M, Kacevska M, Ingelman-Sundberg M: **Epigenomics and interindividual differences in drug response.** *Clin Pharmacol Ther* 2012, **92**:727-736.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.** *Science* 2009, **324**:930-935.
- Kriaucionis S, Heintz N: **The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.** *Science* 2009, **324**:929-930.
- Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y: **Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.** *Nature* 2010, **466**:1129-1133.
- Iyer LM, Tahiliani M, Rao A, Aravind L: **Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids.** *Cell Cycle* 2009, **8**:1698-1710.
- Reitman ZJ, Yan H: **Isocitrate dehydrogenase 1 and 2 mutations in cancer: alterations at a crossroads of cellular metabolism.** *J Natl Cancer Inst* 2010, **102**:932-941.
- Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, Ito S, Yang C, Xiao MT, Liu LX, Jiang WQ, Liu J, Zhang JY, Wang B, Frye S, Zhang Y, Xu YH, Lei QY, Guan KL, Zhao SM, Xiong Y: **Oncometabolite 2-hydroxyglutarate is a**

- competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* 2011, **19**:17-30.
8. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL: **Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.** *Science* 2011, **333**:1303-1307.
  9. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y: **Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.** *Science* 2011, **333**:1300-1303.
  10. Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeyer C, Carell T: **The discovery of 5-formylcytosine in embryonic stem cell DNA.** *Angewandte Chemie* 2011, **50**:7008-7012.
  11. Raiber EA, Beraldi D, Ficiz G, Burgess HE, Branco MR, Murat P, Oxley D, Booth MJ, Reik W, Balasubramanian S: **Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase.** *Genome Biol* 2012, **13**:R69.
  12. Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H, Gao J, Liu P, Li L, Xu GL, Jin P, He C: **Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming.** *Cell* 2013, **153**:678-691.
  13. Ficiz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W: **Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation.** *Nature* 2011, **473**:398-402.
  14. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, Meehan RR: **Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes.** *Genome Res* 2012, **22**:467-477.
  15. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A: **Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells.** *Nature* 2011, **473**:394-397.
  16. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE: **5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells.** *Genome Biol* 2011, **12**:R54.
  17. Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C: **Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.** *Nat Biotechnol* 2011, **29**:68-72.
  18. Szulwach KE, Li X, Li Y, Song CX, Han JW, Kim S, Namburi S, Hermetz K, Kim JJ, Rudd MK, Yoon YS, Ren B, He C, Jin P: **Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells.** *PLoS Genet* 2011, **7**:e1002154.
  19. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K: **TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.** *Nature* 2011, **473**:343-348.
  20. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y: **Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells.** *Genes Dev* 2011, **25**:679-684.
  21. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H, Huang S, Min J, Nicholson T, Chen T, Xu G, Shi Y, Zhang K, Shi YG: **Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells.** *Mol Cell* 2011, **42**:451-464.
  22. Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC: **Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2).** *Nucleic Acids Res* 2004, **32**:4100-4108.
  23. Jin SG, Wu X, Li AX, Pfeifer GP: **Genomic mapping of 5-hydroxymethylcytosine in the human brain.** *Nucleic Acids Res* 2011, **39**:5015-5024.
  24. Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, Fazio TG: **Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells.** *Cell* 2011, **147**:1498-1510.
  25. Serandour AA, Avner S, Oger F, Bizot M, Percevault F, Lucchetti-Miganeh C, Palierno G, Gheeraert C, Barloy-Hubler F, Peron CL, Madigou T, Durand E, Froguel P, Staels B, Lefebvre P, Metivier R, Eeckhoutte J, Salbert G: **Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers.** *Nucleic Acids Res* 2012, **40**:8255-8265.
  26. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T: **Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates.** *PLoS One* 2010, **5**:e15367.
  27. Munzel M, Globisch D, Carell T: **5-Hydroxymethylcytosine, the sixth base of the genome.** *Angewandte Chemie* 2011, **50**:6460-6468.
  28. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H: **Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA.** *Nucleic Acids Res* 2010, **38**:e181.
  29. Kraus TF, Globisch D, Wagner M, Eigenbrod S, Widmann D, Munzel M, Muller M, Pfaffeneder T, Hackner B, Feiden W, Schuller U, Carell T, Kretzschmar HA: **Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors.** *Int J Cancer* 2012, **131**:1577-1590.
  30. Szulwach KE, Li X, Li Y, Song CX, Wu H, Dai Q, Irier H, Upadhyay AK, Gearing M, Levey AI, Vasanthakumar A, Godley LA, Chang Q, Cheng X, He C, Jin P: **5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging.** *Nat Neurosci* 2011, **14**:1607-1616.
  31. Koh KP, Yabuuchi A, Rao S, Huang Y, Cunruff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, Rodig SJ, Daley GQ, Rao A: **Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells.** *Cell Stem Cell* 2011, **8**:200-213.
  32. Ruzov A, Tsenkina Y, Serio A, Dudnukova T, Fletcher J, Bai Y, Chebotareva T, Pells S, Hannoun Z, Sullivan G, Chandran S, Hay DC, Bradley M, Wilmot I, De Sousa P: **Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development.** *Cell Res* 2011, **21**:1332-1342.
  33. Haffner MC, Chau A, Meeker AK, Esopi DM, Gerber J, Pellakuru LG, Toubaji A, Argani P, Iacobuzio-Donahue C, Nelson WG, Netto GJ, De Marzo AM, Yegnasubramanian S: **Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers.** *Oncotarget* 2011, **2**:627-637.
  34. Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu W, Tan L, Hu Y, Zhan Q, Lee CW, Hu D, Lian BQ, Kleffel S, Yang Y, Neiswander J, Khorasani AJ, Fang R, Lezcano C, Duncan LM, Scolyer RA, Thompson JF, Kakavand H, Houvras Y, Zou LI, Mihm MC Jr, Kaiser UB, Schatton T, Woda BA, Murphy GF, et al: **Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma.** *Cell* 2012, **150**:1135-1146.
  35. Li W, Liu M: **Distribution of 5-hydroxymethylcytosine in different human tissues.** *J Nucleic Acids* 2011, **2011**:870726.
  36. Hahn MA, Qiu R, Wu X, Li AX, Zhang H, Wang J, Jui J, Jin SG, Jiang Y, Pfeifer GP, Lu Q: **Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis.** *Cell Rep* 2013, **3**:291-300.
  37. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A: **The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing.** *PLoS One* 2010, **5**:e8888.
  38. Thomson JP, Lempiainen H, Hackett JA, Nestor CE, Muller A, Bolognani F, Oakeley EJ, Schubeler D, Terranova R, Reinhardt D, Moggis JG, Meehan RR: **Non-genotoxic carcinogen exposure induces defined changes in the 5-hydroxymethylome.** *Genome Biol* 2012, **13**:R93.
  39. Fedrigo O, Warner LR, Pfefferle AD, Babbitt CC, Cruz-Gordillo P, Wray GA: **A pipeline to determine RT-QPCR control genes for evolutionary studies: application to primate gene expression across multiple tissues.** *PLoS One* 2010, **5**:e12545.
  40. **The R Project for Statistical Computing.** [<http://www.r-project.org/>].
  41. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**:1754-1760.
  42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**:2078-2079.
  43. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-seq (MACS).** *Genome Biol* 2008, **9**:R137.
  44. **Gene Expression Omnibus publication GSE49291.** [<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49291>].
  45. **Galaxy.** [<https://main.g2.bx.psu.edu/>].
  46. **Genomic Regions Enrichment of Annotations Tool (GREAT).** [<http://bejerano.stanford.edu/great/public/html/>].
  47. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: **GREAT improves functional interpretation of cis-regulatory regions.** *Nat Biotechnol* 2010, **28**:495-501.

48. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min JH, Jin P, Ren B, He C: **Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome.** *Cell* 2012, **149**:1368-1380.
49. **WiseGene..** [<http://www.wisegeneusa.com/>].
50. **Eurofins MWG Operon..** [<http://www.eurofinsgenomics.eu/>].

doi:10.1186/gb-2013-14-8-r83

**Cite this article as:** Ivanov *et al.*: Ontogeny, distribution and potential roles of 5-hydroxymethylcytosine in human liver function. *Genome Biology* 2013 **14**:R83.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

