Supplementary Information

Supplementary Methods

To gain insight into the degree to which our data may be influenced by variation in the cellular composition of buccal swab samples, we first looked at the beta-values of our twin samples at cg18384097; a CpG site located in the \textit{PTPN7} (protein tyrosine phosphatase non-receptor type 7) gene. Based on a comparison of publicly available DNA methylation data from blood samples and buccal swab samples \cite{1-3}, cg18384097 was found to be one of the most differentially methylated CpGs between buccal swabs samples and blood (beta-value buccal = 0.82 and beta-value blood = 0.05), suggesting that this CpG could be a suitable marker to obtain insight into the relative proportions of buccal versus blood cells within a sample \cite{4}. In our buccal data from twins, the average beta-value of this CpG was 0.89. Beta-values for this CpG for all twin samples are presented in Table S3. Table S3 indicates that the majority of twins had similar methylation levels at this CpG, although some variation is evident. MZ twin 3.2 had the most deviant methylation value, which suggests that the sample of this individual may have contained relatively more blood cells compared to the other samples.

To examine this further, we next turned to another reference dataset consisting of 450k methylation data on multiple tissue types \cite{5} and selected from this dataset all CpGs with an average beta-value difference >0.6 between buccal swab samples (N = 5) and blood samples (N = 5), in order to obtain a set of CpGs of which the methylation value measured in buccal swab samples is presumably reflective of the amount of buccal epithelial cells versus blood cells present in buccal swab samples. This selection yielded 881 CpGs (\textit{p}-value range: 2.58 × 10^{-5}–7.15 × 10^{-4}), after excluding probes containing SNPs in the CpG site. For this set of 881 CpGs, we plotted the methylation beta-values of our buccal samples from twins together with beta-values from the reference set \cite{5} for buccal swab samples (N = 5), blood samples (N = 5) and saliva samples (N = 5) in a heatmap (Figure S2). Figure S2 illustrates that some variation in methylation level at these CpGs is present between the buccal swab samples: some buccal samples show relatively more intermediate methylation, a pattern that is more similar to the saliva reference samples, and is suggestive of a higher proportion of blood cells in the sample. Based on methylation levels at this set of CpGs that differentiates strongly between buccal and blood samples, two smaller clusters were identified among buccal swab samples, containing twin Samples 3.2, 6.2, 7.1, 10.1 and 10.2 and reference buccal Samples 3 and 5 (Figure S2), which demonstrated more intermediate methylation values compared to the other buccal samples, suggesting that the samples in the two small clusters (twin Samples, 3.2, 6.2, 7.1, 10.1 and 10.2) contained higher proportions of leukocytes compared to the other buccal samples. Twin Samples, 3.2, 6.2, 7.1, 10.1 and 10.2 also showed the lowest methylation beta-values at cg18384097 (Table S3). To examine the extent to which our analyses may be affected by heterogeneity across samples related to cell type proportions, we repeated our analyses with twin Pairs 3, 6, 7 and 10 excluded, thus keeping only the most homogenous samples that seemed to have the highest buccal epithelial cell content (based on the approach illustrated in Figure S2) in the analyses, which yielded highly similar results (see Table S1 and Figure S1).
Table S1. Correlations between methylation values of twins, with twin Pairs 3, 6, 7 and 10 excluded.

<table>
<thead>
<tr>
<th>Category</th>
<th>N CpGs (%)</th>
<th>Mean rho</th>
<th>Median rho</th>
<th>Min rho</th>
<th>Max rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>All CpGs</td>
<td>59,041</td>
<td>0.57</td>
<td>0.60</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Gene-centric annotations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intergenic (&gt;10 kb from TSS)</td>
<td>11,430 (19.4%)</td>
<td>0.56</td>
<td>0.60</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Distal Promoter (~10 kb to −1.5 kb from TSS)</td>
<td>3193 (5.4%)</td>
<td>0.57</td>
<td>0.60</td>
<td>-0.89</td>
<td>1</td>
</tr>
<tr>
<td>Proximal Promoter (~1.5 kb to +500 bp from TSS)</td>
<td>17,880 (30.3%)</td>
<td>0.60</td>
<td>0.66</td>
<td>-0.94</td>
<td>1</td>
</tr>
<tr>
<td>Gene Body (+500 bp to 3’ end)</td>
<td>25,163 (42.6%)</td>
<td>0.56</td>
<td>0.60</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Downstream region (3’ end to +5 kb from 3’ end)</td>
<td>1375 (2.3%)</td>
<td>0.58</td>
<td>0.66</td>
<td>-0.94</td>
<td>1</td>
</tr>
<tr>
<td>CGI annotations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGI</td>
<td>10,576 (17.9%)</td>
<td>0.66</td>
<td>0.77</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>CGI shore</td>
<td>14,803 (25.1%)</td>
<td>0.57</td>
<td>0.60</td>
<td>-0.89</td>
<td>1</td>
</tr>
<tr>
<td>CGI shelf</td>
<td>6001 (10.2%)</td>
<td>0.55</td>
<td>0.60</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Non-CGI</td>
<td>27,661 (46.9%)</td>
<td>0.55</td>
<td>0.60</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Methylation level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypomethylated (average beta &lt;0.3)</td>
<td>17,581 (29.8%)</td>
<td>0.58</td>
<td>0.60</td>
<td>-0.94</td>
<td>1</td>
</tr>
<tr>
<td>Intermediate methylated (average beta ≥0.3–0.7)</td>
<td>29,519 (50.0)</td>
<td>0.58</td>
<td>0.60</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Hypermethylated (average beta ≥0.7)</td>
<td>11,941 (20.2)</td>
<td>0.54</td>
<td>0.60</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure S1. MZ twin correlations for individual CpGs grouped by genomic regions and average methylation level, with twin Pairs 3, 6, 7 and 10 excluded. Hypo = hypomethylated. Inter = intermediate methylation. Hyper = hypermethylated.
Table S2. Names and number of analyzed CpGs of imprinted genes from Yuen et al. [6].

<table>
<thead>
<tr>
<th>Gene No.</th>
<th>Gene Name</th>
<th>N CpGs analyzed</th>
<th>Gene No.</th>
<th>Gene Name</th>
<th>N CpGs analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABCA1</td>
<td>1</td>
<td>24</td>
<td>NNAT</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>ANKRD11</td>
<td>1</td>
<td>25</td>
<td>OSBPL5</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>ATP10A</td>
<td>4</td>
<td>26</td>
<td>PEG10</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>CALCRL</td>
<td>5</td>
<td>27</td>
<td>PEG3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>CDKN1C</td>
<td>1</td>
<td>28</td>
<td>PHLDA2</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>COPG2</td>
<td>1</td>
<td>29</td>
<td>PLAGL1</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>DDC</td>
<td>2</td>
<td>30</td>
<td>PPP1R9A</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>DLX5</td>
<td>4</td>
<td>31</td>
<td>PRIM2A</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>GNAT</td>
<td>17</td>
<td>32</td>
<td>RBP5</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>GRB10</td>
<td>7</td>
<td>33</td>
<td>SGCE</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>H19</td>
<td>34</td>
<td>35</td>
<td>SLC22A18</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>IGF2</td>
<td>1</td>
<td>36</td>
<td>SLC22A18AS</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>IGF2AS</td>
<td>1</td>
<td>37</td>
<td>SLC22A2</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>INPP5F</td>
<td>1</td>
<td>38</td>
<td>SNRPN</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>KCNQ1</td>
<td>1</td>
<td>39</td>
<td>SNURF</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>KCNQ1DN</td>
<td>1</td>
<td>40</td>
<td>TCEB3C</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>KLF14</td>
<td>9</td>
<td>41</td>
<td>TFP12</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>L3MBTL</td>
<td>2</td>
<td>42</td>
<td>TP73</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>MAGEL2</td>
<td>1</td>
<td>43</td>
<td>UBE3A</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>MEG3</td>
<td>3</td>
<td>44</td>
<td>WT1</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>MEST</td>
<td>2</td>
<td>45</td>
<td>ZIM2</td>
<td>6</td>
</tr>
<tr>
<td>22</td>
<td>NDN</td>
<td>2</td>
<td>46</td>
<td>ZNF264</td>
<td>4</td>
</tr>
</tbody>
</table>

Total: 144

A The number of CpGs in each of the imprinted genes described by Yuen et al., which showed an intermediate methylation level in our buccal data from twins (mean β ≥ 0.3–0.7 across subjects).
Figure S2. Beta-values of twin samples (buccal swabs) and reference data from five blood samples, five saliva samples and five buccal samples for CpGs with an average (absolute) beta-value difference >0.6 between blood and buccal in the reference data. Brightest yellow: beta-value = 1. Brightest blue: beta-value = 0. The clustering of CpGs (N = 881, y-axis) and samples (x-axis) was performed using complete linkage based on the Euclidian distance between beta-values.
Table S3. Beta-values of twin samples for cg18384097 in the in PTPN7 gene; a CpG that was previously reported to be highly discriminative between blood samples and buccal swab samples [1–4].

<table>
<thead>
<tr>
<th>Sample</th>
<th>Beta-value cg18384097</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZpair3.2</td>
<td>0.7434</td>
</tr>
<tr>
<td>MZpair10.2</td>
<td>0.8360</td>
</tr>
<tr>
<td>MZpair6.2</td>
<td>0.8513</td>
</tr>
<tr>
<td>MZpair7.1</td>
<td>0.8527</td>
</tr>
<tr>
<td>MZpair10.1</td>
<td>0.8583</td>
</tr>
<tr>
<td>MZpair9.1</td>
<td>0.8723</td>
</tr>
<tr>
<td>MZpair7.2</td>
<td>0.8751</td>
</tr>
<tr>
<td>MZpair4.2</td>
<td>0.8929</td>
</tr>
<tr>
<td>MZpair5.1</td>
<td>0.8949</td>
</tr>
<tr>
<td>MZpair9.2</td>
<td>0.9002</td>
</tr>
<tr>
<td>MZpair8.2</td>
<td>0.9037</td>
</tr>
<tr>
<td>MZpair2.1</td>
<td>0.9055</td>
</tr>
<tr>
<td>MZpair1.2</td>
<td>0.9057</td>
</tr>
<tr>
<td>MZpair6.1</td>
<td>0.9076</td>
</tr>
<tr>
<td>MZpair5.2</td>
<td>0.9129</td>
</tr>
<tr>
<td>MZpair3.1</td>
<td>0.9130</td>
</tr>
<tr>
<td>MZpair4.1</td>
<td>0.9165</td>
</tr>
<tr>
<td>MZpair2.2</td>
<td>0.9187</td>
</tr>
<tr>
<td>MZpair8.1</td>
<td>0.9230</td>
</tr>
<tr>
<td>MZpair1.1</td>
<td>0.9312</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>0.8857</strong></td>
</tr>
</tbody>
</table>

Highlighted in blue are the twin samples that clustered separately from the other twin samples in Figure S2. The samples in the table are sorted by beta-value (lowest to highest).

Table S4. Spearman correlation between the methylation level of MZ twins at individual CpGs based on M-values.

<table>
<thead>
<tr>
<th>Genome-wide</th>
<th>N CpGs</th>
<th>Mean r</th>
<th>Median r</th>
<th>Min r</th>
<th>Max r</th>
</tr>
</thead>
<tbody>
<tr>
<td>All variable CpGs</td>
<td>59,041</td>
<td>0.54</td>
<td>0.54</td>
<td>−0.66</td>
<td>1</td>
</tr>
</tbody>
</table>
**Figure S3.** MZ twin correlations for individual CpGs grouped by genomic region and average methylation level, based on M-values. Hypo = hypomethylated. Inter = intermediate methylation. Hyper = hypermethylated.

**References**


© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).