Genome-wide Nucleosome Map and Cytosine Methylation Levels of an Ancient Human Genome

Jakob Skou Pedersen, Eivind Valen, Amhed M. Vargas Velazquez, Brian J. Parker, Morten Rasmussen, Stinus Lindgreen, Berit Lilje, Desmond J Tobin, Theresa K. Kelly, Søren Vang, Robin Andersson, Peter A. Jones, Cindi A. Hoover, Alexei Tikhonov, Egor Prokhortchouk, Edward M. Rubin, Albin Sandelin, M. Thomas P. Gilbert, Anders Krogh, Eske Willerslev, Ludovic Orlando.

Table of Contents

Section SI1. Sequence datasets

Most of the analyses were performed on previously released sequence data sets, consisting of: 1) the first complete ancient human genome [Rasmussen et al. 2010] (section SI1.1); 2) a modern control data set based on a panel of modern samples [Green et al. 2010, Reich et al. 2010] (section SI1.2). 3) the first ancient Aborigine genome [Rasmussen et al. 2011] (section SI1.3); and 4) genomic reads of a 110,000-130,000 year-old polar bear [Miller et al. 2012] (section SI1.5). In addition, we generated genomic reads from an ancient horse dating back to ca. 4.5 ka (section SI1.4) and from modern human hairs (section SI1.6). Those were analyzed to check for the specificity of the patterns observed in hairs and their possible extension to other types of ancient samples, such as calcified bones that represent the vast majority of fossil remains excavated.

1.1. Saqqaq Palaeo-eskimo sequence dataset

The sequence data have been originally generated as part of our characterization of the Saqqaq Palaeo-Eskimo genome [Rasmussen et al. 2010]. Overall, ancient DNA extracts from a 4,000-year-old permafrost-preserved hair were built into Illumina libraries and deep-sequenced at about a 20X-coverage on a GAIIx sequencing platform and Single End sequencing. The vast majority of the sequencing reads (242 lanes out of a total of 245, corresponding to ca. 3.5 billions of reads) were recovered following library amplification with Phusion DNA polymerase (Finnzymes, hereafter referred to as Phusion). A total of 3 GAIIx lanes, however, were generated following library amplification with a *Taq* platinum high-fidelity DNA polymerase (Life technologies, hereafter referred to as Hifi) and resulted in a total number of 50.1 millions of reads [Ginolhac et al. 2011]. Contrarily to Hifi, Phusion has been demonstrated to show poor activity at uracil residues (Figure S3.1) [Fogg et al. 2002]. The latter are generated following *post-mortem* deamination of cytosine residues, mainly at single-stranded overhanging ends. As chemical analogs of thymines, Uracils yield to the mis-incorporation of AT base pairs (instead of GC) during library preparation and amplification, resulting in an excess of GC→AT mismatches when ancient genomes are compared to modern reference genomes [Stiller et al. 2006; Brotherton et al. 2007; Gilbert et al. 2007; Briggs et al. 2007]. Thus, that Phusion, rather than Hifi, has been used for generating the first complete sequence of an ancient genome greatly contributed to the overall quality of the genome by limiting the impact of damage-driven misincorporations in downstream analyses [Rasmussen et al. 2010].

For the nucleosome analysis the original set of 999.3 millions of reads was used, as defined in [Rasmussen et al. 2010]. For the methylation analysis, a mapping procedure allowing for indels was followed as described below (this was performed because the alignments available from the original study were performed with SESAM and were indel-free; [Rasmussen et al. 2010]). Read ends were trimmed for potentially residual adapter sequences, and after the first base showing the poorest quality score (C) using the software AdapterRemoval available for download at http://code.google.com/p/adapterremoval/ [Lindgren 2012]. Trimmed reads were subsequently mapped against nuclear human genome reference (hg18) using bwa 0.5.9 [Li and Durbin 2010] and default values for optional parameters. Unmapped sequences, together with the high-quality reads previously identified, were remapped against the available mitochondrial genome sequence of this individual [Gilbert et al. 2008] using similar parameters; in addition, sequencing reads with a minimal length of 25 nucleotides, a minimal mapping quality of 30, and showing no alternative hits

were further selected using a combination of the samtools suite and awk command lines. Finally, sequence duplicates were collapsed using the sort and rmdup commands available in SAMtools [Li and Durbin 2010], resulting in a final BAM file of unique mitochondrial reads that was used for some of the analyses presented below aiming at detecting cytosine methylation-driven nucleotide mis-incorporation patterns (section SI3). The final dataset consisted of a final number of 696.0 millions and 26.3 millions of non-clonal unique sequences originating from Phusion and *Taq* Platinum high fidelity (Hifi*)* library amplification, respectively.

1.2. Control genome dataset

We constructed a data set with the same properties as the Saqqaq data set in terms of number of reads and read length distribution, except reads were randomly selected (and truncated to match the Saqqaq size distribution) from a panel of sequencing runs from modern genomes. This data set allowed us to (1) evaluate if the Saqqaq data set observations could be ascribed to mapping biases and also to (2) more generally represent a null set to evaluate the significance of the statistical measurements made during the analysis.

To construct this set, we used the combined set of modern genomes sequenced at low read depth as part of the Neanderthal and Desinova projects, which were downloaded from UCSC [Green et al. 2010, Reich et al. 2010]. All samples originate from lymphoblastoid cell lines of the Human Genome Diversity Project (HGDP). A main criterion for the data sets used for the Control was that they were made using the same technology, but based on modern samples that were not degraded or fragmented. All reads, including unmapped, were extracted from the downloaded bam files and converted to fastq format. These were then shuffled to avoid any ordering relative to genomic coordinates. We then constructed a data set with the same length distribution as the Saqqaq, by truncating individual modern reads to the appropriate size, though we initially included about 1.2 times more reads than present in the Saqqaq data set. We then mapped each fragment size separately to hg18 using BWA [Li et al. 2010]. For each fragment-size set of unsorted mapped reads, we then extracted the exact same number of reads of the given size as present in the Saqqaq data set. Finally, all the reads were combined to make our final Control data set.

1.3. Ancient Aborigine sequence dataset

A total number of 311.0 millions of Illumina reads recovered from a *ca*. 100-years old hair sample originating an ancient Aborigine individual [Rasmussen et al. 2011]. These reads were generated following a procedure similar to the one used for characterizing the complete genome of the *Saqqaq* individual, except that DNA libraries were amplified using *Taq* Gold DNA polymerase. The initial unprocessed data set consisted of 2.1 billions of reads (100 nucleotides long), which was reduced by 85.2% following index filtering, adaptor trimming, and selection of uniquely mapping and non-clonal reads. These span 60% of the human genome with an average read depth of 11x (6.4x overall). The average length of endogenous trimmed and unique reads consisted of 64 nucleotides and mapping was done using BWA and standard parameters.

1.4. Ancient horse sequence dataset

The ancient horse sample was provided by one of the co-authors (Dr. Alexei Tikhonov). It has been found in Yakutia at Batagai (Republic of Sakkha) and showed an uncalibrated radiocarbon date of 4,450±35 BP (Gr 50842). A total of 0.225 grams of ulna bone was drilled to powder at low speed in ancient DNA lab facilities available at the Centre for GeoGenetics. DNA was extracted following the silica-based procedure described in [Rohland and Hofreiter 2007], with slight modifications following [Orlando et al. 2009]. DNA was released from silica pellets using a final elution volume of 205ul of EB (QIAgen) and 15 minutes of incubation at 37°C. We constructed one Illumina DNA library following the standard truSeq procedure, except that 16.5ul of DNA extract were used as input and that the End Repair reaction was performed using End-It End Repair kit reagents from Epicentre (catalog reference ER0720). A first incubation at 4°C for 2 minutes was followed by a second incubation at 37°C for 45 minutes. During those incubation steps, 5'-overhangs were filled-in by T4 DNA polymerase and 5'-ends are phosphorylated by PNK. End-repaired DNA templates were purified using QIAgen minelute purification kit and 10 volumes of PN buffer instead of 5 volumes of PB buffer. The whole elution volume was used for klenow exo- polyA-tailing (37°C for 30 min) and adapter ligation using standard indexed truSeq adapter. The DNA library was further purified using Ampure XP beads with 1.8:1 as a volume ratio between beads and DNA, and eluted into 20ul of EB solution. We amplified 10ul of DNA library using 300nM Illumina PCR primers (PCR primer 1.0: 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GA ; PCR primer 2.0: 5'-CAA GCA GAA GAC GGC ATA CGA GAT), 15 cycles of PCR amplification, a concentration of 1mg/ml BSA, 1µM dXTP (Invitrogen) and 5U of Taq Gold (invitrogen). Final PCR volume was 25µl and cycling conditions consisted of a first DNA denaturation at 95°C for 10 minutes, followed by 15 cycles of denaturation (95°C, 30 sec), annealing (60°C, 30sec) and elongation (72°C, 60sec). A final elongation was performed at 72°C for 7 minutes before amplified DNA was purified into 25ul EB using QIAgen minelute purification kit. The DNA library was sequenced on a hiseq2000 platform available at the Danish National DNA Sequencing Centre using 100 cycles Paired-End reactions.

Illumina Paired-end reads were trimmed for adapter sequences at their 3'-end following a pair-wise alignment procedure and using the software AdapterRemoval [Lindgreen 2012]. A minimal sequence identity of 90% over 5-10 nucleotides was required for adapter alignment (this threshold was lowered to 67% for longer alignments). In addition, reads showing stretches of bases with low quality scores and/or Ns at sequence ends were trimmed from the first nucleotide position showing the low quality score. Reads starting with stretches of at least one undetermined base were trimmed, resulting in a final sequence starting at the first position showing a determined base. Following adapter removal, read pairs were aligned and collapsed as long as showing a maximum of one mismatch over a sequence of 11 nucleotides at read ends. A larger fraction of mismatches (33%) were tolerated for longer alignments (>11 nucleotides). Sequence collapse was performed by calling a consensus based on the nucleotide showing the highest quality score; when the two reads showed identical nucleotides, a quality score was re-assigned based on the product of the two previous quality scores but with a maximal limit set-up to a *Phred* score 41 (e.g. when the two nucleotides where sequenced with a 10% chance of error, the base quality score of that position in the consensus was re-assigned to 1%). Read-pairs showing no overlap were discarded. Collapsed-pairs were further

mapped as single-end reads against the horse reference genome EquCab2.0 available for download at UCSC (http://hgdownload.cse.ucsc.edu/downloads.html#horse) using BWA and standard parameters, except that the seed was disabled following the recommendations of [Schubert et al. 2012].

1.5. Ancient polar bear sequence dataset

We followed the full procedure described in section SI1.4 to characterize the size distribution of endogenous DNA inserts using 812,006,446 Illumina Paired-end reads (SRS347019) generated from the DNA extract of a 110,000-130,000 years old polar bear from Svalbard (Norway) [Miller et al. 2012]. The *de novo* assembled scaffolds of the polar bear genome (available for download at ftp://climb.genomics.cn/pub/10.5524/100001_101000/100008/Ursus_maritimus.scaf.fa.gz) were used as a reference for read mapping.

1.6. Modern human hair sequence dataset

We followed the same experimental procedure as for characterizing the Saqqaq genome. We generated a total number of 10.0 millions correctly indexed, adaptor trimmed, uniquely mapping and non-clonal Illumina reads from a single DNA library prepared on modern hairs from a British born Caucasian living in Denmark at present. Average read length was of 61 nucleotides. These sequences span 15.7% of the human genome with an average read depth of 1.4X (0.22X overall). This dataset was generated in order to evaluate if the nucleosome-driven signal detected on previous datasets was specific to ancient DNA. Single-End sequencing was performed to a lower read depth, sufficient to statistically confirm that the same signal is present. DNA extraction and downstream bioinformatics analysis followed the same protocol as for the ancient Saqqaq Palaeo-Eskimo genome [Rasmussen et al. 2010].

1.7. Size distributions

Saqqaq: A subset of 116.1 millions of Paired-End reads generated from 4 indexed libraries was considered for analyzing the size distribution of ancient DNA inserts. Pairs of reads showing a minimum overlap of 11 bases (with 100% identity) were collapsed into single reads using AdapterRemoval [Lindgren 2012] following the procedure described in section SI1.4. This resulted in a total number of 83.6 millions of collapsed reads that were mapped against the human reference genome hg18 using bwa, following the same parameters as described in section SI1.4. We then randomly selected 1.5 million of unique hits per library (for a total of 6.0 millions of reads) and further filtered for collapsed reads 1) ending with base quality scores of 40 and 41 when shorter than the size of sequencing reads (75 nucleotides) and for longer sizes, 2) those that end with base quality scores of 33-40. We found a total number of 4.8 millions selected collapsed reads, corresponding to endogenous library inserts sequenced over their full length, showing unique high-quality alignments against polar bear scaffolds. The resulting size distribution is shown as Figure 2.d, bottom.

Ancient Horse: We focused on collapsed reads showing a sequence length longer than 25 nucleotides and a minimal mapping quality (MQ) of 25. We filtered PCR duplicates using 5'- and 3'-coordinates and keeping the read showing the best

average base quality. This procedure was implemented using a python script kindly provided by Dr. Martin Kircher (*FilterUniqueBAM.py*; for a SAM compatible version, see https://bioinf.eva.mpg.de/fastqProcessing/). The size distribution of those collapsed reads mapping uniquely the horse reference genome was plotted using R [R Development Core Team 2012] and a subset of 100,000 reads randomly selected. Corresponding reads are available for download at SRA (SRA105533). Horse inserts showed a remarkable 10-bp size-periodicity spanning 38bp to 142bp (Figure 2.d, top). This periodicity corresponds to the average size of a DNA helix turn. No 10-bp periodicity could be observed for inserts larger than the size of nucleosome (146bp). This pattern suggests that DNA fragmentation in spacer regions occurs randomly in contrast to nucleosome cores where DNA strands facing away from nucleosome protection are exposed to hydrolytic reactions and fragmentation.

Polar bear: We noticed that within pairs, the quality of the second read was generally of a lower quality than the first read, with a significant fraction showing lowest quality scores over the full read length. This could result in artifact while collapsing reads, with increased false positive and false negative identification of overlapping read ends. In order to avoid such bias and select collapsed reads with starting and ending coordinates reflecting the full length of DNA library inserts, we further selected from the population of collapsed reads 1) those that end with base quality scores of 40 and 41 when shorter than the size of sequencing reads (101 nucleotides) and for longer sizes, 2) those that end with base quality scores of 38-40. We found a total number of 3.8 millions of selected collapsed reads showing unique high-quality alignments against polar bear scaffolds. Those were considered for plotting insert size distribution (Figure 2.d, middle). A periodicity, similar to the one observed using the ancient horse sequence dataset, although admittedly of a lower magnitude, was observed on 110-130 ka polar bear DNA inserts (38bp-153bp), suggesting long term preservation of nucleosomal positioning signal in ancient DNA extracts. We found the size class of 101 nucleotides slightly over-represented. This likely reflects a sequencing or data-handling artifact, as this class of length corresponds to the number of sequencing cycles performed on the hiseq2000 platform.

Of note, size distributions showing similar ca. 10-bp periodicity were observed for a series of ancient DNA extracts from Holocene and Pleistocene bone material preserved in the permafrost (data not shown).

We constructed similar size distributions for the modern hair data generated in this study as well as for a range of other modern genomes generated from fresh blood using a similar methodology (from different horse breeds; Orlando et al. 2013). Those size distributions are presented for 30-60bp in Figure S1.1 in order to facilitate comparison across samples. The top graphs provide size distributions for the Saqqaq, the ancient polar bear, the ancient horse and the modern hair. The bottom graphs provide size distributions for modern horses. Frequencies have been normalized to peak at 1 for the size showing maximum frequency. The ancient horse show a short-range periodicity at 38, 49 and 59 bp whereas the ancient polar bear show a short-range periodicity at 39, 50 and 60 bp. Similarly, the modern hair also shows a short-range periodicity with peaks at 49 and 60 bp (no reads shorter than 39bp and longer than 60bp were experimentally generated, precluding direct observation of a larger size range). As the horse and polar bear datasets have been generated in absence of size-selection, this clearly rules out size selection as a driver for the short-range periodicity observed in fragment size distributions. Given that no short range periodicity is observed in sequence datasets generated from fresh blood, we conclude that the short range periodicity observed in the modern hair sample originates from active degradation of hair DNA following apoptotic processes

occurring during hair differentiation (Botchkavera et al. 2006). The periodicity observed in Saqqaq, the horse and the polar bear, exactly matches the average turn of a B-DNA helix, with nucleotides facing away from nucleosomes becoming more susceptible to *post-mortem* degradation than nucleotides in contact with nucleosomes. In a similar way, apoptotic enzymes are well known to cause DNA laddering by cleaving chromatin DNA into inter-nucleosomal fragments (Nagata et al. 1998). During this process, nucleosome protection also creates ~ 10-bp periodicity patterns (Aruscavage et al. 2010). This is responsible for the short-range periodicity observed in the modern hair data.

That 1-bp offsets are sometimes observed between the Saqqaq, the ancient horse, the ancient polar bear and the modern hair (with peaks either at 38bp or 39bp, 48-50bp), most likely comes from two causes. First, one helix turn is around 10.4 bases, which is expected to create a mixture of both 10-bp and 11-bp periodicities. Second, our trimming procedure removes any stretch of Ns at read starts and depending on the dataset considered, we observed variable fractions of sequencing starting with one N. Read trimming results in the shortening of the sequence by one nucleotide, creating a proportion of fragments offset by 1 nucleotide in their size distribution.
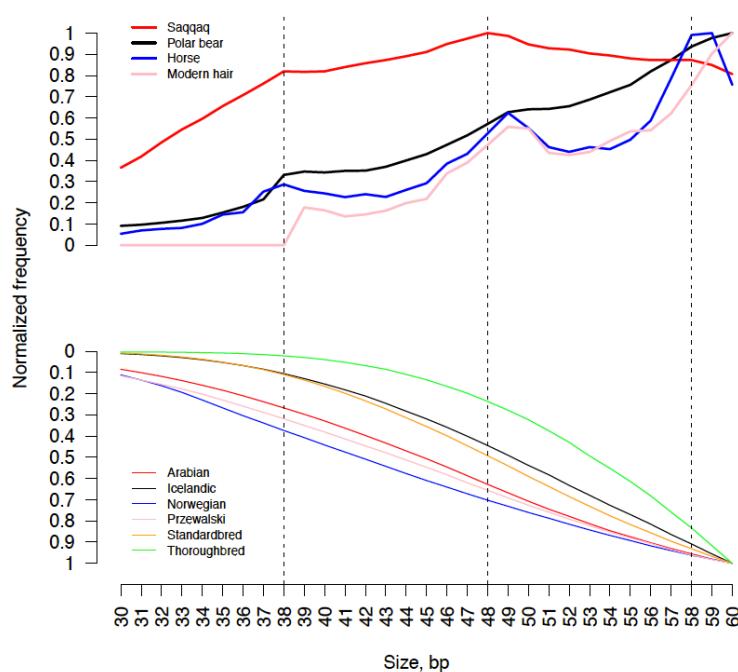
Figure S1.1. Size distribution of ancient and modern sequence datasets. Top: Ancient and modern hair sequence datasets. Bottom: Modern blood sequence datasets.

Section SI2. Nucleosomes

2.1. Read depth

We evaluated the average read depth in different genomic regions where read depth is defined as the number of mapped reads covering that region normalized by the region size. We first evaluated read depth at different genic regions, by segmenting the entire genome according to the protein-coding genes of the UCSC Gene set [Fujita et al. 2011]. When multiple genes or isoforms overlapped a region, annotation priority was given as: protein-coding exons (CDS) > 5'UTR > 3'UTR > introns > intergenic regions. In addition, read depth was also evaluated at CpG islands (CGI) as defined by the UCSC Genome Browser.

Read depth was found to vary dramatically between genomic regions for the Saqqaq genome, with the highest read depth observed for CGI. In contrast the Control set displayed a nearly constant read depth across genomic regions, with the lowest read depth observed at CGI. We note that for a wider variety of other genomic datasets, CGI also showed lower mapping rates than the other genomic regions investigated (data not shown). Both the Aboriginal sample and the Modern hair sample largely follow the same read depth pattern as Saqqaq. Finally, a modern human MNase-based nucleosome occupancy data set generally also follows the same pattern, though with less variation between regions.

Region	genome	coding	intergenic	intron	3'UTR	5'UTR	CpG island
Saqqaq	19.4(1.0)	53.8(2.8)	17.2(0.9)	21.1(1.1)	26.9(1.4)	76.8(4.0)	126.1(6.5)
Control	19.4(1.0)	20.7(1.1)	18.7(1.0)	20.3(1.0)	20.8(1.1)	18.1(0.9)	11.5(0.6)
Aboriginal	6.9 (1.0)	17.9(2.6)	6.1 (0.9)	7.6 (1.1)	9.9 (1.4)	22.0(3.2)	32.7(4.8)
Modern hair	0.2 (1.0)	0.5 (2.1)	0.2 (0.9)	0.2 (1.0)	0.3 (1.3)	0.7 (3.2)	1.0 (4.5)
Schones	2.7 (1.0)	3.9 (1.5)	2.6 (1.0)	2.7 (1.0)	3.1 (1.2)	3.9 (1.5)	3.8 (1.4)

Table S2.1.

Regional read depth for main data sets. The average read depth across genomic regions is given for each data set. Read depth normalized by genomic average (enrichment) is given in parenthesis. Schones refers to a mapping of human nucleosomes using MNase [Schones et al. 2008].

Region	genome	coding	intergenic	intron	3'UTR	5'UTR	CpG island
Saqqaq	0.2	5.9	0.0	0.2	-0.2	10.1	22.8
Control	0.6	1.0	0.5	0.8	1.3	0.4	-0.6

Table S2.2.

GC-normalized read depth for Saqqaq and Control.

2.2. Control of mapping and sequencing biases

To eliminate the effect of any potential mapping biases due to the repetitive structure of the human genome, we defined a set of genomically unique regions (Unique). For all mapping procedures used, we retained only uniquely mapping reads. Using ExactRepeats, a mapping program based on suffix arrays (developed by one of us: Anders Krogh), we defined all regions that are uniquely mappable by reads of length 26 or greater. These regions were further trimmed by 100 bases at either end, ensuring that the observed read depth within these is unaffected by mapping issues. The resulting set spans 51.3% of the genome.

The PCR amplification step of NGS library generation is known to introduce biases in the frequency of occurrence of different fragments dependent on GC-content and length [Benjamini and Speed 2012]. To rule out that such biases could create or contribute to the observed variation and to improve the specificity of the nucleosome signal in the downstream analysis, we evaluated their strength at each position along the genome and corrected the read depth signal accordingly. We achieved this by estimating what the read depth would be at every genomic position given the read length distribution; the location of uniquely mappable regions for every read length; and the GC-content for all possible fragments overlapping a given position. The estimated read depth composed of contributions from all available read lengths were then subtracted from the observed read depth. The resulting read depth was thus positive when more read depth is observed than expected from GC content and negative when less read depth was observed than expected from GC content.

More specifically, we based this evaluation on the GCcorrect package [Benjamini and Speed 2012] implemented in R [R Development Core Team 2012]. Briefly, GCcorrect learns the correlation between GC-content and read depth across the genome and uses it to predict the expected number of reads covering every position of the genome. It disregards non-unique regions based on a mappability file. Since mappability is dramatically different depending on the read length (in Saqqaq ranging from length 20 to 76 bp), we supplied a separate mappability file for each length. ExactRepeats (see above) was used to create these. Reads shorter than 26 bp were discarded from the analysis as there were too few reads for GCcorrect to learn the correlation between read depth and GC content. For each read length, we supplied GCcorrect with the corresponding mappability file and the overall number of observed reads. GCcorrect then returned the expected number of reads starting at any given position for the given read length. The resulting expected read depth summed across all read lengths was calculated using an extension to the UCSC Genome Browser command line tool bedItemOverlapCount, which allowed to sum up genome-wide real-valued data across multiple data sets.This overall expected read depth was subtracted from the observed read depth (disregarding reads shorter than 26), to produce the final GC-corrected read depth. The same procedure was applied to the Control data set, while taking the original fragment sizes in the panel of underlying data sets into account.

2.3. Correlation with other nucleosome data and predictions

We evaluated the correlation between different data sets across various genomic regions. Briefly, the following data sets were included in the analysis, (1) The original uncorrected Saqqaq read depth; (2) GC-corrected Saqqaq read depth; (3) The original uncorrected Control read depth; (4) The GC-corrected Control read depth; (4) The original Aboriginal read depth; (5) The read depth of the modern hair sample; (6) GC content in 5 bp windows as available from the UCSC Genome Browser [Meyer et al. 2012]; (7) Computational predictions of nucleosome

occupancy as based on SVM models trained on MNase digestion data from A375 (A375 set) and MDA-kb2 (Dennis set; [Dennis et al. 2007) cell lines; 8) Experimental genome-wide MNase-based occupancy data sets obtained from CD4+ cells [Schones et al. 2008] and lympoblastoid cell lines [Gaffney et al. 2012]. The Schones data is based on read depth of 25 bp long single-end reads, while the Gaffney data is based on midpoints of paired-end reads (126bp < fragment length < 184bp) smoothed by 30bp, following the authors instructions.

Both Pearson's and Spearman's (data not shown) correlation coefficients were evaluated across large subsets of the genome. As the results were highly similar, we only report Pearson's correlation coefficients (PCC).

We first focused on the conservatively defined Unique regions (see Section SI2.2), which span 51% of the genome. These were represented by a random sample of 10,000 regions, in total spanning 7.7 Mb. The results are reported in the main article text.

We next evaluated the correlation across a 20 Kb subsection of a known nucleosome array region (chr12:34,331,000-34,451,000; hg18), where nucleosomes are thought to be consistently and specifically positioned independently of tissue [Gaffney et al. 2012]. To rule out the influence of common mapping biases, we included only the subset of Unique regions (~13 Kb). The resulting correlation matrix (Figure S2.1 a) allows several observations:

1. For both Saqqaq and Control the GC corrected and original read depths are highly correlated.
2. The Saqqaq and Aboriginal data sets are also highly correlated (0.74).
3. Similar levels of correlation are seen for Saqqaq versus the computational nucleosome occupancy predictions (0.77 and 0.74 for uncorrected read depth).
4. The experimental MNase data sets, which are based on related cell types (CD4+ and lymphoblastoid), show a correlation of 0.38 between them and a lower correlation with the computational predictions (0.16-0.28) than Saqqaq. This could suggest some level of noise in state-of-the art MNase-based occupancy maps, potentially from the cutting biases of the MNase enzyme [Gaffney et al. 2012].
5. The correlation of the Saqqaq and Aboriginal data sets against the experimental data sets (pcc of 0.23 to 0.43) is of the same magnitude as the correlation between the two experimental data sets (pcc=0.38).
6. Generally the original, uncorrected Saqqaq data set shows higher levels of correlation with the computational and experimental occupancy maps than the GC corrected Saqqaq data set. This may be attributed to "over-correction" at inherently GC rich nucleosome sites, as GC content is known to be part of the nucleosome-positioning signal [Collings et al. 2010; Valouev et al. 2011]. However, it is also likely that the MNase based data sets, which also underlie the computational predictions, are subject to some level of GC-dependent PCR bias, which is not corrected for.

Finally, we evaluated the correlation in the +/-1Kbp regions around the TSSs of all UCSC transcripts. To illustrate the variance in the correlation coefficients between TSS regions, we plotted the full distribution of PCCs for these. Separate plots were made for the correlation of select data sets against the GC-corrected Saqqaq read depth (Figure 1.f); the original Saqqaq read depth (Figure S2.1.b); and the Schones data set (Figure S2.1.c). Again, the computational predictions stand out with consistently strong positive correlation against the uncorrected

Saqqaq data set across the TSS regions (Figure S2.1.b). The Schones data set shows stronger correlation with the Gaffney data set than with either Saqqaq or the computational predictions (Figure S2.1.c). The nucleosome positioning patterns in TSS regions have been shown to correlate with gene expression levels and hence be cell type specific [Valouev et al. 2011]. The relatively high correlation of Schones and Gaffney in TSS regions may thus be explained by the underlying cell types being related.
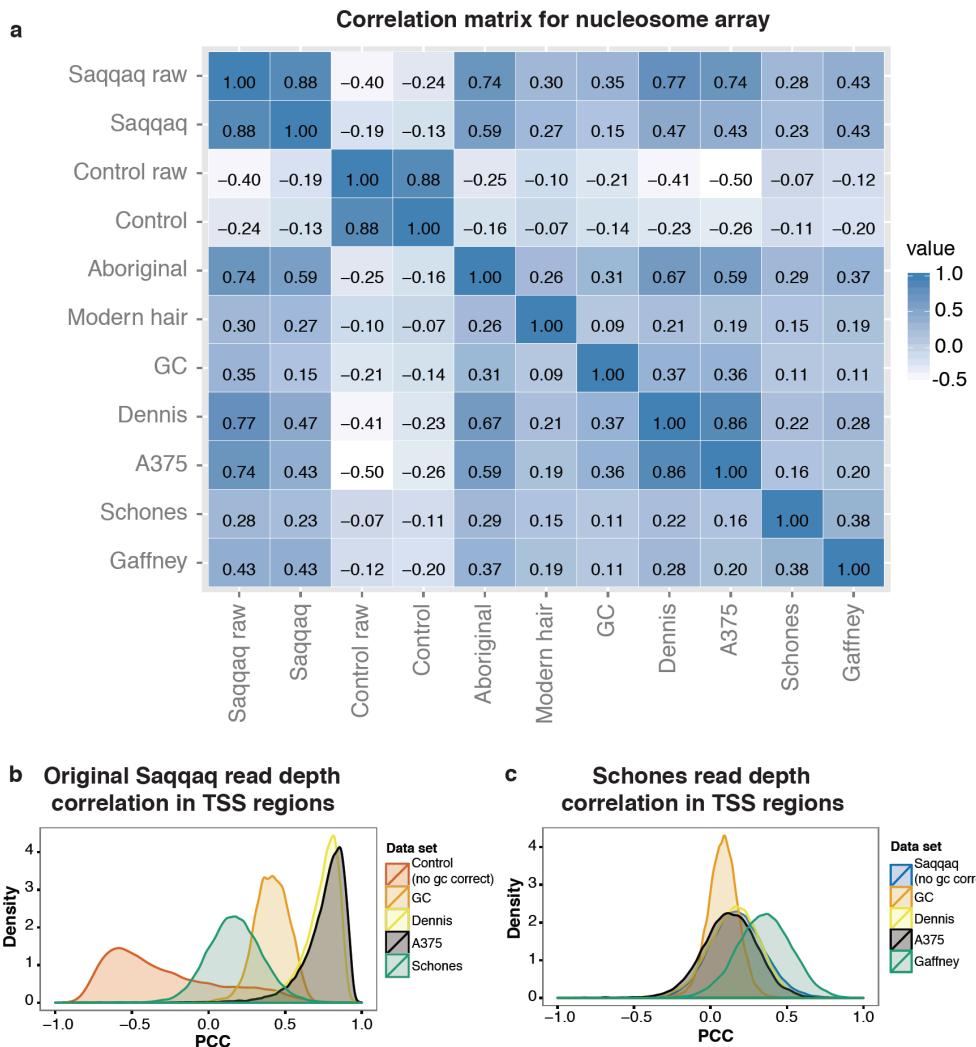
Figure S2.1
Correlation between data sets.
a, Correlation matrix for known nucleosome array region (20 Kb). Pearson's correlation coefficients were calculated for each pairwise combination of data sets and color-coded according to magnitude. All correlations coefficients are significantly different from zero (all p-values < 1.2e-15). **b**, Distributions of Pearson correlation coefficients across TSS regions for original, uncorrected Saqqaq read depth *versus* uncorrected Control set, GC content (GC), two computational occupancy maps (Dennis and A375), and an experimental occupancy map (Schones). **c**, Distribution of Pearson correlation coefficients across TSS regions for Schones versus a subset of data sets from (b) and another experimental occupancy map (Gaffney).

2.4. Nucleosome occupancy at anchor sites

Nucleosome occupancy, represented by the read depth, was plotted around multiple anchor points in the genome: TSSs, CTCF sites and splice sites. All instances of each group of anchor points were aligned and a meta plot of the accumulated read depth in the surrounding regions divided by the number of anchors was plotted. For TSSs and splice sites we used the UCSC genes set.
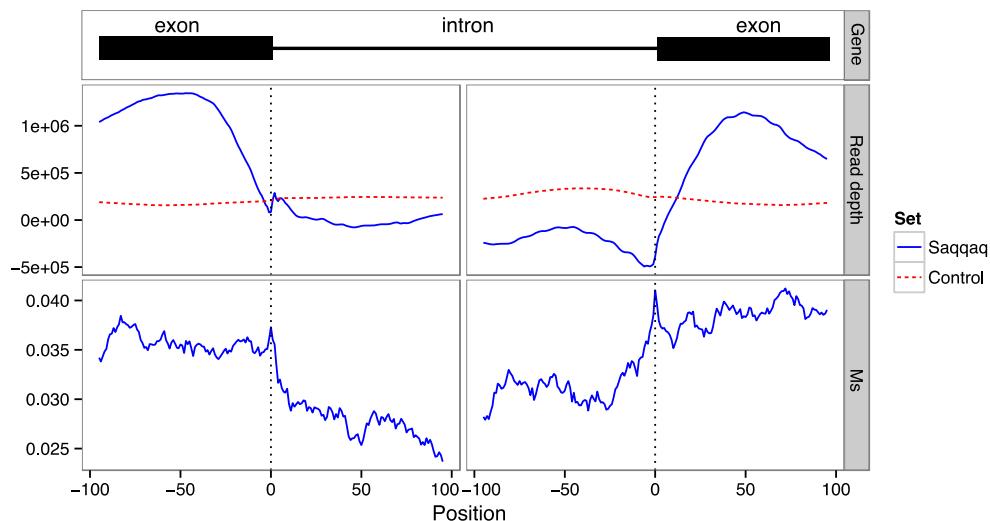
Figure S2.2
Nucleotide and methylation around donor and acceptor splice sites. Methylation signal is described in sections SI3.2 and SI3.3.

2.5. Fourier Transform

We predicted that nucleosome protection would cause a periodicity in read depth variation corresponding to the size of the nucleosome and the spacer region (~200 bp). To unbiasedly evaluate if such periodicity could be observed, we performed a Fourier transform (FT) analysis using Welch's FT method (a windowed Fourier transform). More specifically, we performed a spectral density plot (periodogram) genome-wide across CpG islands, TSS±1000 bp, gene bodies, and CTCF sites ±1000 bp. The raw spectral density was corrected for the background frequency distribution, to remove low frequency variations and constant offsets, by subtracting a background estimated by exponential curve modeling. To visualize the spatial distribution of frequency information, short time Fourier transforms (spectrograms) over known anchor sites, such as Transcription Start Sites (TSS) (source: UCSC genes) and CTCF binding sites (source: [Fu et al. 2008], see section SI3.7), were generated. For Saqqaq and Control the GC-corrected read depth was used.
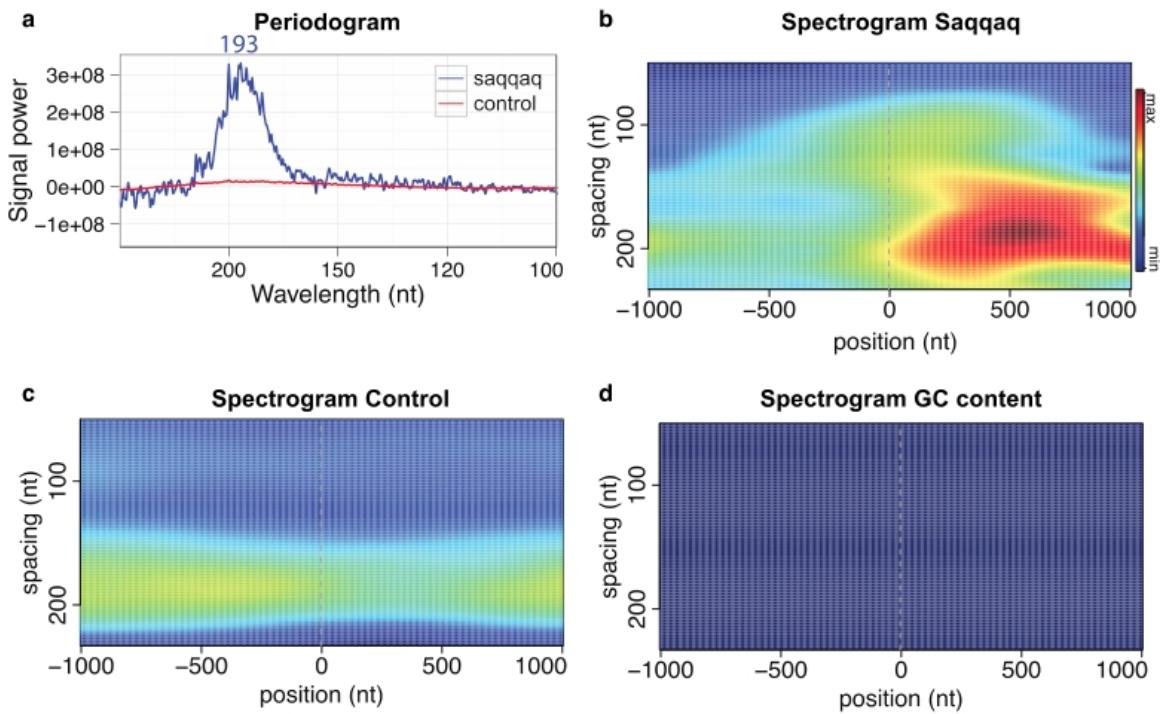
Figure S2.3

Fourier Transform periodicity analyses of transcription start site regions.

a, Read depth periodogram for Saqqaq and Control. Saqqaq peak periodicity is 193 bp. **b**, Spectrogram for GC-corrected Saqqaq read depth. The strongest periodicity signal exists downstream of the TSS. **c,** Spectrogram for Control. Note that some periodicity is found around 200bp in the region, though much weaker than for Saqqaq. **d**, GC content does show any datable periodicity in the range.

.
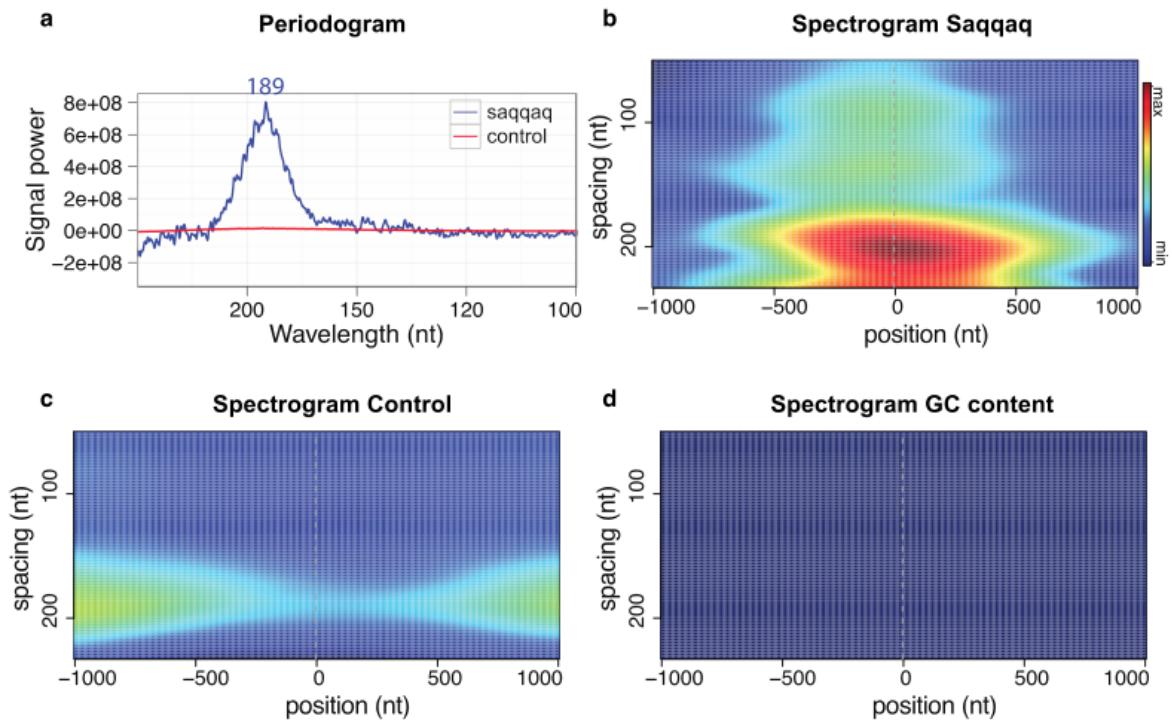
Figure S2.4
Fourier Transform periodicity analyses of CpG islands.
a, Read depth periodogram for Saqqaq and Control. Saqqaq peak periodicity is 189 bp. **b**, Spectrogram for GC-corrected Saqqaq read depth. **c,** Spectrogram for Control. Note that some periodicity is found around 200bp in the region, though much weaker than for Saqqaq. **d**, GC content does show any datable periodicity in the range.
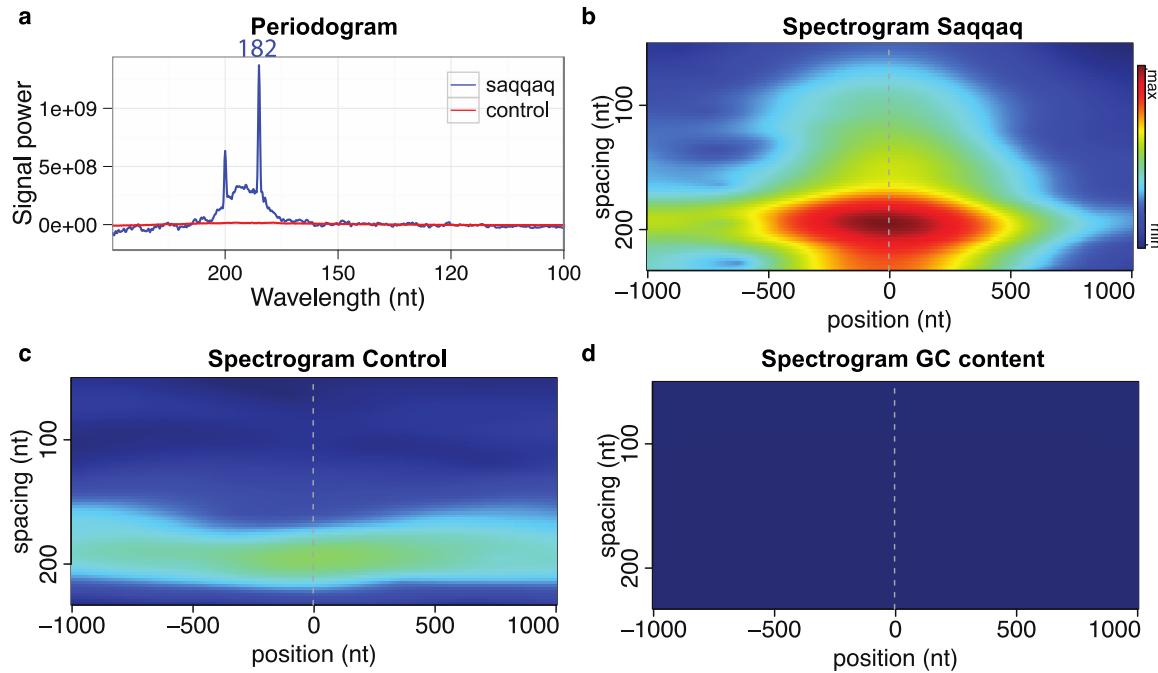
Figure S2.5

Fourier Transform periodicity analyses of CTCF regions.

a, Read depth periodogram for Saqqaq and Control. Saqqaq peak periodicity is 182 bp. **b**, Spectrogram for GC-corrected Saqqaq read depth. **c,** Spectrogram for Control. Note that some periodicity is found around 200bp in the region, though much weaker than for Saqqaq. **d**, GC content does show any datable periodicity in the range.
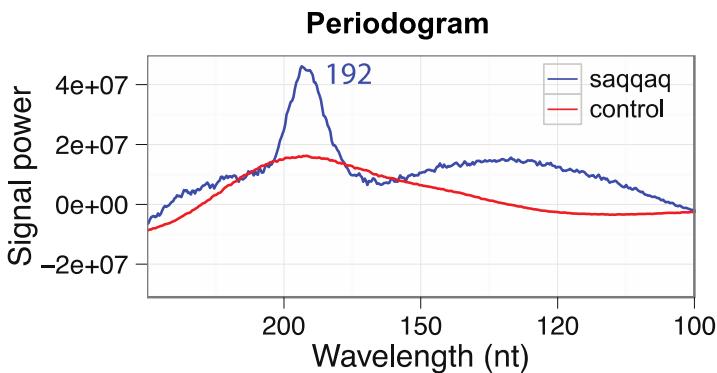
Figure S2.6

Periodogram of gene bodies.
Read depth periodogram for Saqqaq and Control. Saqqaq peak periodicity is 192 bp. Note that Control that exhibits some signal, albeit much weaker than Saqqaq.
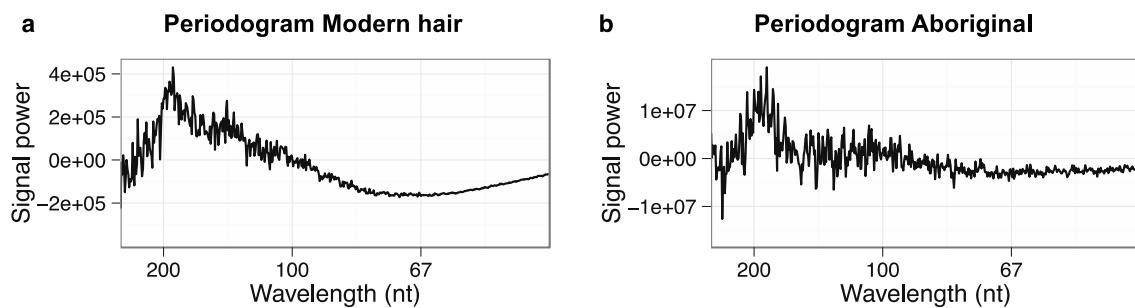
Figure S2.7

Periodograms of CpG islands for Modern hair and Aboriginal data sets.
a, Periodogram for Modern hair data set. Peak periodicity is at 186 bp. Note that modern hair was low-coverage and hence less signal was detected. **b**, Periodogram for Aboriginal data set. Peak periodicity is at 182 bp.

2.6. Phasograms

We predicted that in regions where nucleosomes are consistently positioned, the distance between two successive nucleosomes should equal the size of the nucleosome plus the size of the spacer region (~200 bp). Phasograms were produced in a similar manner as described in [Valouev et al. 2011] in order to test for the presence of such signal.

Using the raw uncorrected data, we counted the distance of pairs of 5' ends on the same strand including only positions with at least at least 5 reads. For both the short and long range plots we used gene bodies (+1Kbp to the termination site) of UCSC protein coding genes

The signal overall follows an exponential distribution due to the local variation in read depth. To estimate the period, we therefore fit and exponential background distribution (Figure S2.8a), which we subtract from the signal to better reveal the periodicity (Figure S2.8b). Finally we analyzed the autocorrelation of these regions identifying the dominant short range phasing to be 10 bp and the long range phasing to be 185 bp (Figure S2.8c).

a

b

c
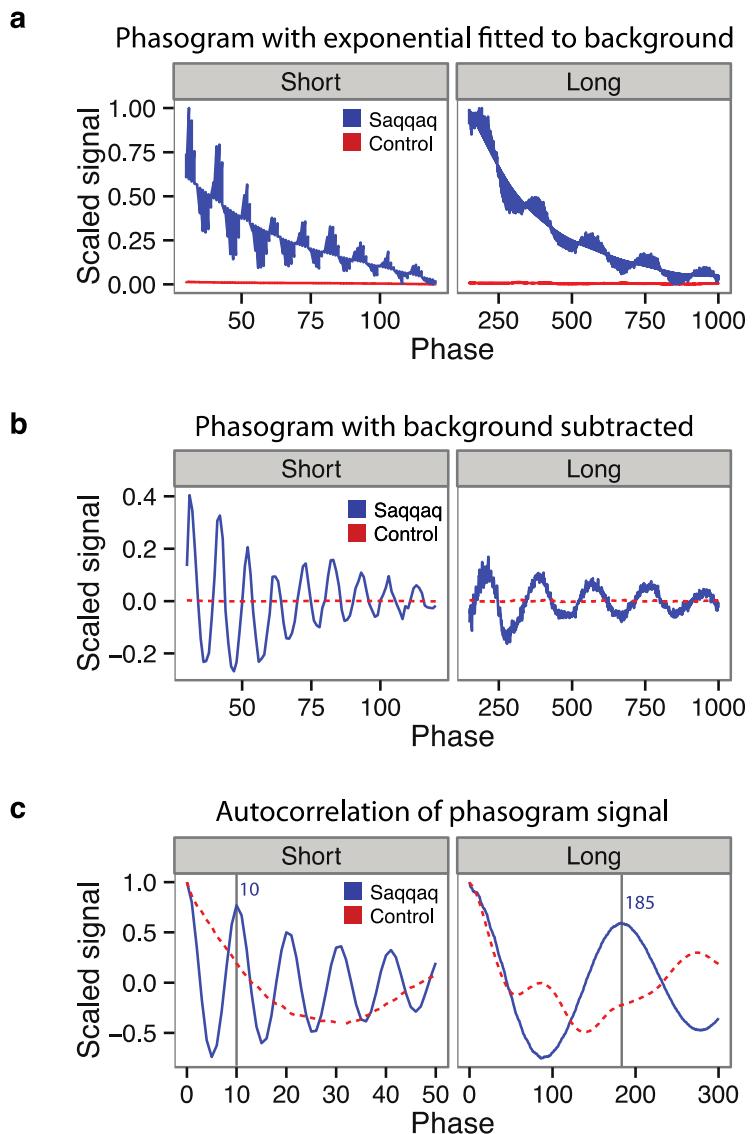
Figure S2.8

Phasogram of Gene bodies.
a, Raw signal of Saqqaq and Control. **b**, Background subtracted. **c**, Autocorrelation plot of phasogram signal with background removed. Peak periodicity is 10 bp for short range and 185 bp for long range (gray lines and blue numbers). Note that the autocorrelation for Control is based on a very weak signal.

2.7. Genome-wide map of nucleosomes

To generate a genome-wide map of nucleosome positions we applied a sliding window to the GC-corrected read depth and called the center position if it has the maximal read depth in that window (Figures 3.a, S2.9). We refer to these as peaks. The window size was chosen to be 147-bp based on the size of a single nucleosome. If the peak score is negative, which is possible when the expected read depth is greater than the observed read depth, the peak is discarded and no call is made. Though some (overcorrected) true calls are likely missed based on this, it avoids calls based simply on variation in the expected read depth.

Each call was assigned a score, defined as the read depth of the peak (p) minus the average read depth of the left flanking region (lf) and right flanking region (rf): score= p – (lf + rf / 2) (Figures 3.a, S2.9). The score is constructed to capture aspects of both occupancy (peak height) and positioning (linker region depletion).

Importantly the expected read depth from our GC-correct procedure takes the place of a control experiment in the peak calling procedure. It is designed to capture and reflect any bias resulting from size and/or GC-content. Furthermore, the mappability of each read length is carefully taken into account and thus modeled in the final expected read depth (Section S2.2). The subtraction of the expected read depth thus corrects for peaks that would be wrongly called due to these effects, similar to the role of a control experiment in say ChIP-seq analysis. To further eliminate any risk mapping biases, the downstream analysis of the nucleosome calls are all restricted to the Unique regions (Section S2.2).

Nucleosome calls based on the original, uncorrected read depth were made using the same approach. However, instead of the difference score used for the GC corrected read depth, we used a log-odds score, which better captures the positioning signal of linker region depletion: score = log(p / ((lf + rf) / 2 + eps)). Eps was chosen to be 1/25. The analyses of these calls gave similar results, as exemplified by their nucleotide and di-nulceotide distrbutions (Figure S2.13).
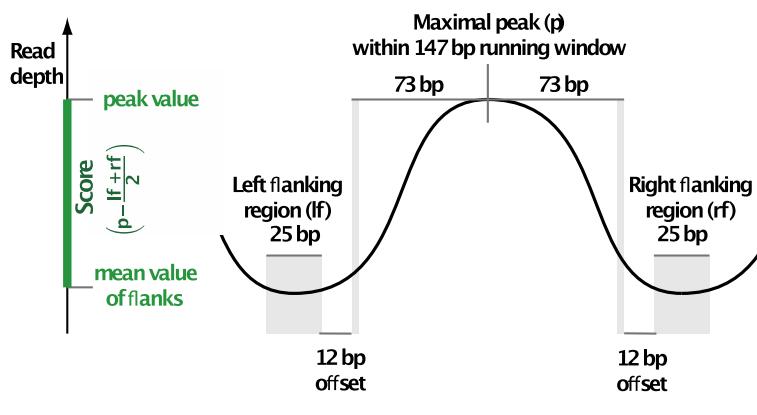
Figure S2.9

Nucleosome calls and scores.
(Similar to Figure 3.a and included for convenience of reference.) Nucleosome center positions (dyads) are called as read depth peaks if maximal at center of running window of nucleosome length (147 bp). Calls are scored by the difference in read depth between the peak and the average read depth of the left and right flanking regions.

2.8. Estimating false positives

The Control data set has the same overall genomic read depth and read length distribution as the Saqqaq by construction. However, it shows only a weak nucleosome signal in the periodicity analysis (section SI 2.5). We therefore made nucleosome predictions (section SI 2.7) based on the Control data set, in the assumption they would all effectively be false positives (FP). We subsequently used those to define the expected number of false positive predictions at a given quality score cutoff.

For a given score cutoff, we thus estimate the expected number of true positive calls (TP) as the difference between the number of Saqqaq calls (S) and the number of Control calls (B): TP = S – FP = S – B (Figure 3.b, Figure S2.10).

The false discovery rate (FDR) for a given score cutoff was estimated using a similar procedure as the ratio between the number of Control calls and the number of Saqqaq calls: FDR = B / S.

To investigate if some extreme peaks in the Control could point to problematic regions and lead to spurious calls in the Saqqaq, we evaluated the correlation in ranks between the top-scoring Control peaks and the Saqqaq peaks (figure S2.10c). As can be seen, there is no tendency for the top-ranked Control peaks to be correlated with top-ranked Saqqaq peaks in the Unique regions used for all analysis.

a   All regions

Table (a):

Set	top–25%	top–10%	top–1%
Score cutoff	23	41	99
Count	2.656M	1.059M	0.107M
FDR	4.7%	0.5%	1.3%

b   Unique regions

Table (b):

Set	top–25%	top–10%	top–1%
Score cutoff	20	39	100
Count	1.371M	0.551M	0.054
FDR	8.5%	0.04%	0.03%

c

Top-0.1% Control calls versus Saqqaq calls   Top-1% Control calls versus Saqqaq calls
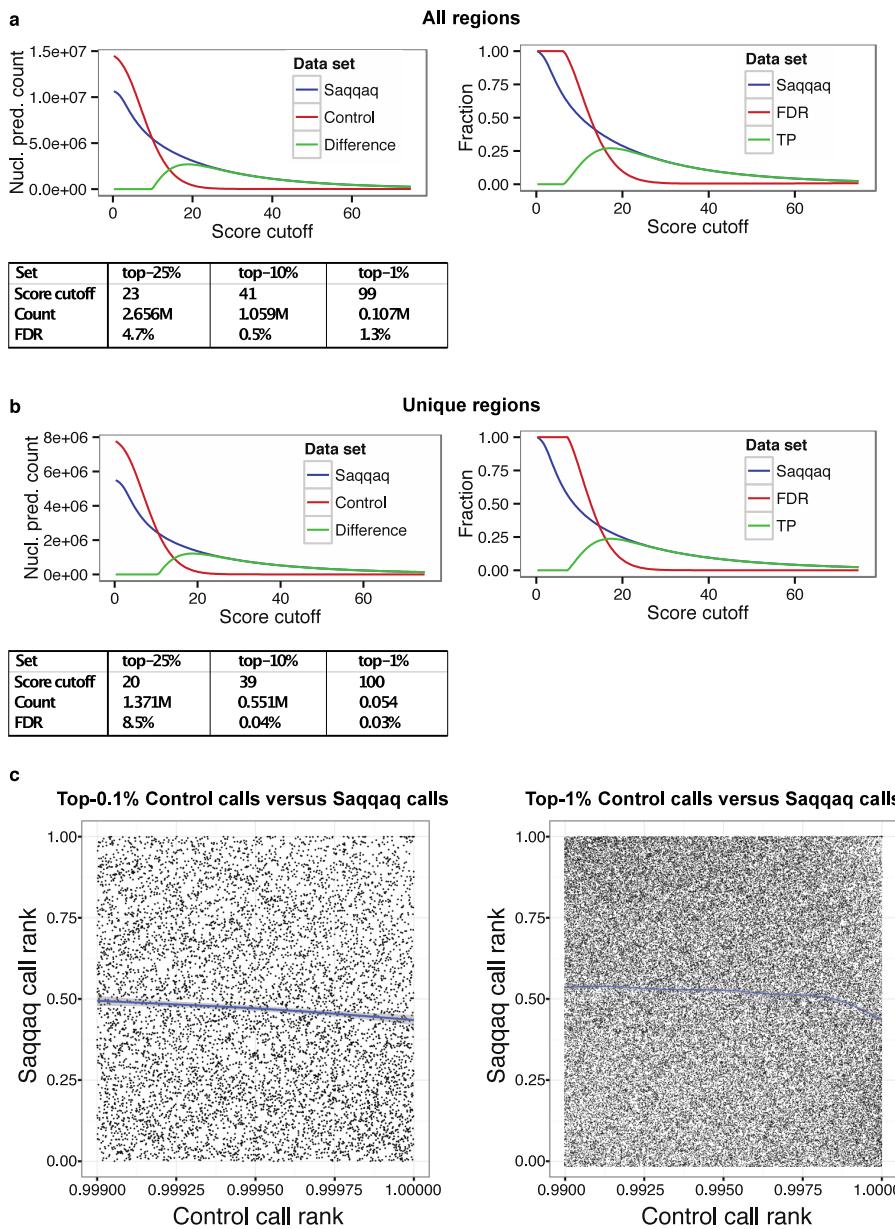
Figure S2.10

Call score distributions and summary statistics. a, Analyses of all calls. Top-left, difference between Saqqaq and Control calls is taken as an estimate of number true positive calls at given cutoff. Top-right, Saqqaq call distribution, false discovery rate (FDR), and fraction of true positive calls (TP) of the total set are given as a function of cutoff. Bottom, summary statistics for different top-scoring subsets (fractiles). **b**, Analyses of calls in Unique regions as in (a). **c**, The rank of the top 0.1% (left) and top 1% (right) peak calls in the null set plotted against their corresponding rank in the Saqqaq set. A smoothed line has been added for visualization of the general trend (stat_smooth from ggplot2 in R). There is no bias for highly ranked null peaks to receive a high rank in the Saqqaq set.

2.9. Nucleotide and di-nucleotide distributions across nucleosomes

Based on the nucleosome calls (section SI 2.8), we evaluate the nucleotide and di-nucleotide distributions across the nucleosomes. For this, we ranked the calls based on the positioning scores and explored the sequence preference at different thresholds by summing the number of mono-nucleotides or di-nucleotides and dividing by the number of nucleosomes. This gave us an average profile for a given threshold (Figures 3.c, S2.10, S2.11).

The resulting profiles exhibit distinct nucleotide level shifts in nucleotide and di-nucleotide frequencies which indicates that the nucleosome sequences are accurately aligned and hence a large fraction of the nucleosome calls are at nucleotide precision. Furthermore, we note distinct nucleotide preferences at positions close to the center (dyad) and the presence of strand-specific patterns nucleotide (Figure S2.11) and di-nucleotide patterns (Figures 3.c). We hypothesize that these profiles reflect sequence-level positioning determinants and note that the strand-specific patterns that shifts across the dyad suggest a method for precise positioning despite overall weak sequence signal.

For comparison, we also include the nucleotide and pyrimidine/purine di-nucleotide distributions around nucleosome calls based on the uncorrected, original Saqqaq read depth data (Figure S2.13). These mirror the characteristics of the corresponding plots for the GC-corrected read depth (Figure S2.11 and Figure 3.c), showing that the nucleotide distribution results are robust. As expected, the top-scoring uncorrected calls show a stronger degree of GC enrichment among top-scoring calls (Figure S2.13), as these will be penalized by the GC correction scheme.
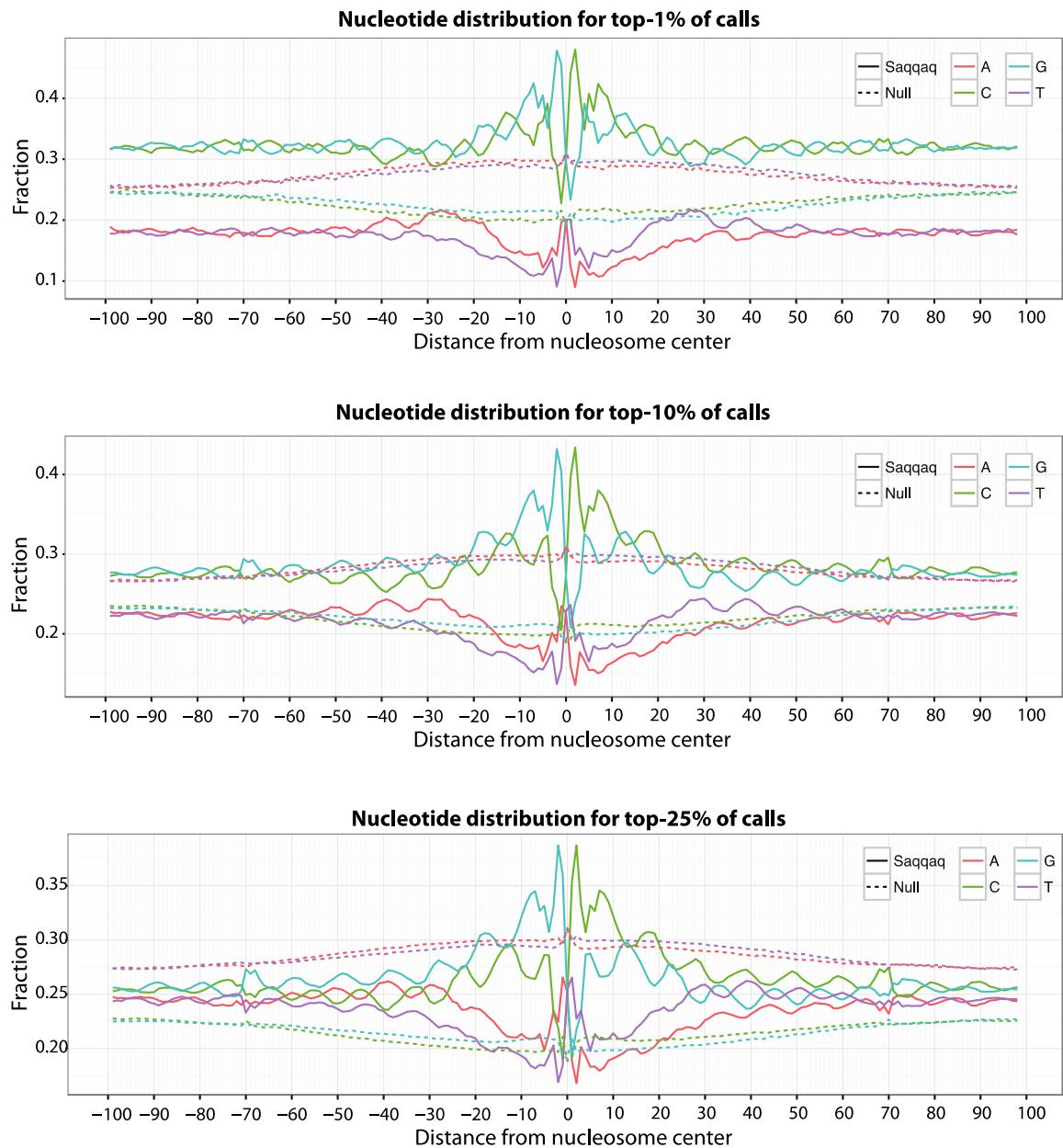
Figure S2.11

Mono-nucleotide distributions across sets of top-scoring nucleosome calls for Saqqaq and Control. The average usage of a mono-nucleotide for positions relative to the nucleosome peak call (dyad) for different top-scoring subsets of nucleosome calls from Unique regions (top: top-1%;middle: top-10%; bottom: top-25%).
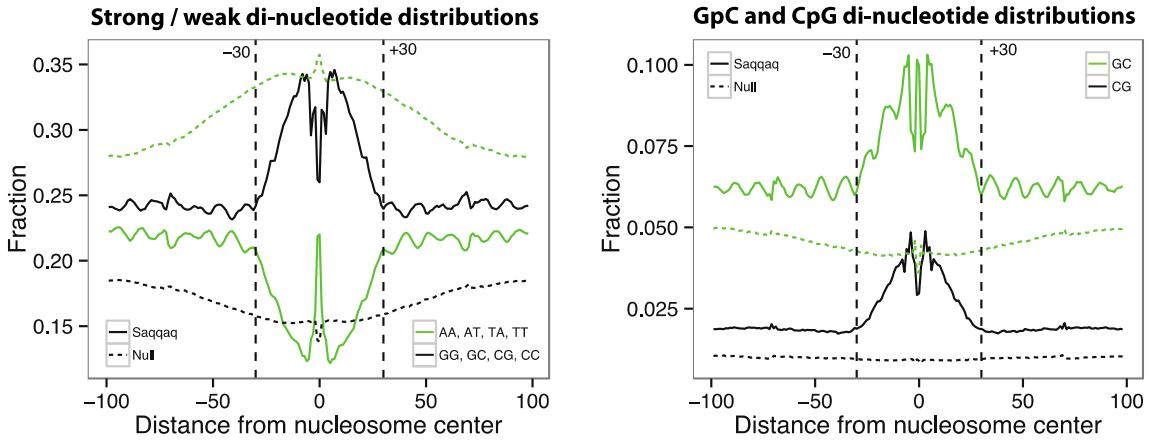
Figure S2.12

Strong / weak and CpG / GpC di-nucleotide distributions across sets of top-25% nucleosome calls for Saqqaq and Cotrol. Left, strong (GG, CG, GC, CC) / weak (AA, AT, TA, TT) di-nucleotide profiles. Boundary of peak is denoted with vertical lines (at -30 and +30). Right, CpG and GpC di-nucleotide profiles. Boundary of peak denoted with vertical lines (at -30 and +30).

a

Nucleotide distribution for top-25% of calls on uncorrected Saqqaq and Control read depth

b

Dinucleotide distribution for top-25% of calls on uncorrected Saqqaq and Control read depth
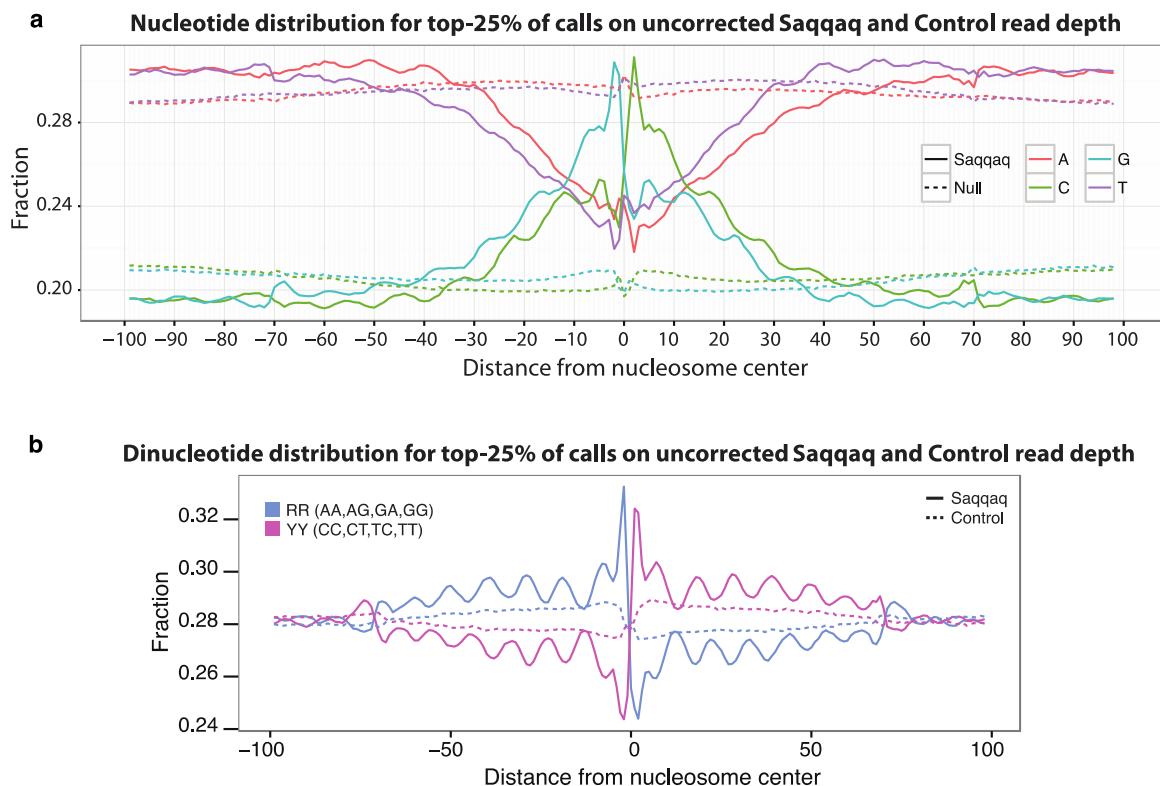
Figure S2.13

Mono-nucleotide and pyrimidine/purine di-nucleotide distributions across top-25% scoring nucleosome calls for the original, uncorrected Saqqaq and original, uncorrected Control.

a, The average usage of mono-nucleotides at positions relative to the nucleosome peak call (dyad) for top-25% scoring subset of nucleosome calls from Unique regions. **b**,The distribution of purine/pyrimidine di-nucleotide usage across top 25% called nucleosomes from Unique regions.

Section SI3. Methylation footprint of the Saqqaq genome

3.1. Background

While *post-mortem* deamination of cytosine residues results in the formation of uracils, the deamination of methylated cytosines (5mC) results in thymine residues that are adequate templates for both Phusion and *Taq* platinum high-fidelity (Hifi) DNA polymerases (Figure S3.1). Consequently, DNA strands with methylated CpG dinucleotides will partake to the pool of library molecules that will generate sequence data regardless of which of the two DNA polymerases was used, should cytosine residues be deaminated or not (Figure S3.1). Conversely, DNA strands with unmethylated but deaminated CpG dinucleotides (ie. UpG) will provide sequence data following amplification with Hifi but not Phusion (Figure S3.1). In the next sections, we exploited this versatile behavior of 5mCpG and CpG dinucleotides as a direct experimental signature of ancient DNA methylation tracts.
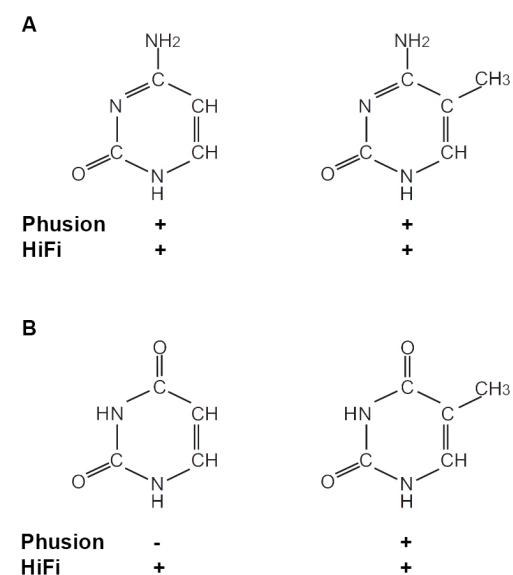
Figure S3.1.

Unmethylated and methylated CpG dinucleotides show different sequencing features following *post-mortem* deamination. A, Biochemical structure of native cytosine (left) and 5mC (right) residues. **B,** Biochemical structure of deaminated cytosine (ie. uracil, left) and deaminated 5mC (thymine, right) residues. The ability of Phusion and Hifi DNA polymerases to by-pass the different types of nucleotides during PCR amplification is indicated with plus/minus signs.

3.2. Nucleotide mis-incorporation patterns

Nucleotide mis-incorporation patterns were recovered from the mapDamage package [Ginolhac et al. 2011]. This package reports the frequencies of all mismatches observed between sequence reads and the genome used as a reference (here, hg18) for all nucleotidic positions along sequencing reads. Even though Phusion cannot by-pass deaminated cytosine residues, nucleotide mis-incorporation patterns clearly mimicked cytosine deamination, with C→T mis-incorporation rates increasing towards the 5'-end of the sequencing reads (Figure 4.a, bottom left), in agreement with previous reports [Briggs et al. 2007; Krause et al. 2010; Ginolhac et al. 2011; Orlando et al. 2011] and the presence of 5'-overhangs that favor cytosine deamination after death. We hypothesized that the signal originated from 5mC residues that were deaminated after death and were replicated

as thymine analogs, hence generating C→T mismatches post-sequencing when compared to the hg18 reference.

Encouragingly, this signal was absent from the promoter of the constitutively-expressed house-keeping gene GAPDH (Glyceraldehyde 3-phosphate dehydrogenase; uc001qor.1, chr12:6,513,918-6,517,797), in agreement with the documented absence of cytosine methylation on that region [Noer et al. 2006]. Conversely, the signal was present on Phusion reads respectively mapping the genomic regions corresponding to TSH2A (chr6:25,834,270-25,834,769; for a total of 484 reads) and TSH2B (chr6:25,835,116-25,835,552; for a total of 244 reads) (data not shown), two testis-specific genes that are known to be silent in somatic tissues and hypermethylated [Zalensky et al. 2002; Weber et al. 2007].

We further recorded mis-incorporation rates in different genomic regions (mitochondrial DNA; nuclear DNA; exons located outside CGIs; introns located outside CGIs; CGIs). These genomic regions were identified using the IRanges R library [Pages et al. 2010] and UCSC Genome Browser tracks. Since C→T mismatch rates decrease from 5' ends of ancient DNA templates due to the presence of 5'-overhangs [Briggs et al. 2007], CpN dinucleotides located at sequence starts are expected to show maximal level of cytosine deamination. Following our hypothesis, none of such Phusion reads would show C→T misincorporations at the first position unless it was methylated (UpG are not by-passed in contrast to TpG, the *post-mortem* by-product of 5mCpG deamination; Figure S3.1). Therefore at read starts, CpG→TpG conversion rates are expected to be significantly increased at CpG due to the methylation background, but not at CpA, CpT and CpC (70-80% of cytosine methylation occurs at CpG sites in the human genome; [Lister et al. 2009]). We tested this prediction by binning sequence reads generated with Hifi and Phusion according to which of those four types of dinucleotides was present in the reference genome at sequencing starts (Figures 4.a and 4.b).

For genomic regions starting with CpG dinucleotides, a higher fraction of Phusion reads were found to start with TpG dinucleotides, especially in exonic and intronic regions located outside CGI (Figure 4.b). Conversely, CGI exhibit reduced proportions of Phusion reads starting with TpG dinucleotides, in agreement with the fact that most, but not all, CGI represent hypomethylated genomic regions [Deaton and Bird 2011]. From all genomic regions considered, the mitochondrial genome appeared as an outlier showing the lowest levels in agreement with the global unmethylated level of this genome [Rebello et al. 2009]. Interestingly, the same trend in CpG→TpG conversion rates was observed but with reduced frequencies when considering sequencing reads that covered genomic CpG dinucleotides at positions 20 and 21 (Figure S3.2). Compared to sequencing starts, these positions are affected by lower cytosine deamination rates [Briggs et al. 2007], which results in lower *post-mortem* 5mCpG to TpG conversion rates, and consequently, lower CpG→TpG conversion rates. Importantly, CpA, CpC and CpT dinucleotides exhibit similar levels of TpA, TpC, and TpT conversions, respectively, regardless of the genomic region considered, suggesting that the observed patterns were indeed CpG-specific and therefore driven by the combination of methylation and deamination of cytosines.

All in all, we found that C→T mis-incorporations were preferentially located in genomic regions known to undergo methylation (CpG, resulting when methylated in tractable 5mCpG→TpG Phusion conversions); as expected (Figure S3.1), the strength of the signal detected appeared directly dependent of *post-mortem* deamination levels (see conversion rates at positions 1-2 and 20-21), suggesting that methylation could only be detected following *post-mortem* deamination of 5mC

residues. This confirmed our hypothesis and suggested that Phusion reads could provide a genuine a signature of methylation patterns in ancient genomes.
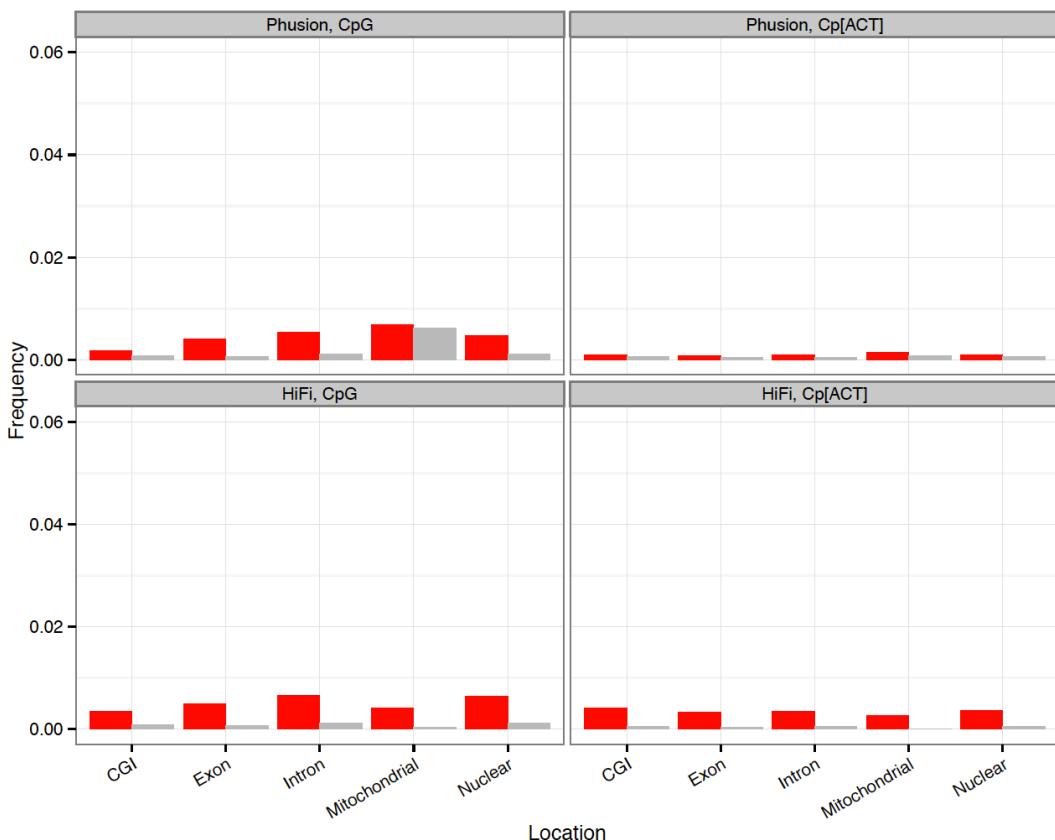
Figure S3.2.
Substitution patterns at CpN dinucleotides. We calculated the proportion of alignments showing a CpN dinucleotide in the reference genome and TpN at positions 20-21 within sequencing reads. Such CpG→TpG conversions at positions 20-21 are reported in red and have been calculated for different genomic regions, including: mitochondrial DNA (Mitochondrial); the whole nuclear genome (Nuclear); exons outside of CGI (Exon); introns outside of CGI (Intron); and CGI. Additionally, the rates of other types of dinucleotide conversions were investigated (Other: CpA→TpA, CpT→TpT, CpC→TpC) and are reported in grey. Top: Phusion reads. Bottom: Hifi reads. Figure 4.b provides similar calculations at read starts (ie. positions 1-2).

Overall, relatively high CpG→TpG conversion rates (over ca. 3%) were observed with Hifi in all genomic regions, including the mitochondrial genome (Figure 4.b). This appears in striking contrast with what observed on Phusion reads and confirms the ability of Hifi polymerase to by-pass uracil residues, regardless of their methylation status (Figure S3.1). Of note, that higher CpG→TpG conversion rates were observed with Hifi in Introns and Exons than in CGIs and in the mitochondrial genome suggests preferential elongation for that enzyme at deaminated 5mC (ie. analog of native thymines) than at deaminated cytosines (ie. uracils).

In this study, we identified a simple means of recovering information about the

28

methylation status of CpG dinucleotides in ancient genomes by looking at CpG→TpG conversions events at starting positions of Phusion reads. At these positions, *post-mortem* cytosine deamination rates are maximal, and methylation preserves our ability to PCR amplify and sequence damaged endogenous ancient templates. Note that the same holds true at sequence ends where complementary G→A misincorporations will be observed [Briggs et al. 2007], resulting in CpG→CpA conversion events. However, as the Saqqaq sequence data mostly corresponded to Single End reads, the large fraction of DNA inserts were not sequenced over their full length. In addition, a significant drop in base qualities was observed towards read ends. We therefore decided to disregard CpG→CpA conversion events at read ends (those result from the deamination of 5mCpG on the complementary strand), despite the information those carry about methylation. We anticipate that this information could be fully used in case where extensive Paired-End sequence data is available.

3.3. Quantifying ancient regional methylation levels

We used CpG→TpG at read starts in order to define M_s, a proxy for regional methylation levels (Figure S3.3). Specifically, M_s is defined as the total counts of CpG→TpG conversions at read starts normalized by the number of CpG dinucleotides present within a given genomic region of the reference human genome hg18. We disregarded conversions that possibly colocated with positions described as polymorphic in dbSNP130 in order to filter for SNP hotspots, which provide a possible source of methylation-unrelated CpG→TpG conversions. Of note, even if CpG→TpG conversions represent a footprint of CpG methylation at a given location of the human genome, such conversions could only be detected in situations where the 5mC residue is affected by *post-mortem* deamination (see sections SI3.1 and SI3.2). For the Saqqaq data, this occurs only in ca. 4% of the cases (Figure 4.a), representing on average one chance event every 25 observations. This means that in 96% of the cases a given 5mCpG will not be deaminated and will therefore be read as a regular CpG, leaving no CpG→TpG conversion mark. Therefore, the absence of CpG→TpG conversions at a given CpG cannot be used to demonstrate the absence of methylation, unless this position shows an extremely high depth-of-coverage. Our methodology will consequently be most generally innapropriate for reconstructing genome-wide methylation maps at single nucleotide levels. However, by summing over particular genomic regions instead of unique CpGs (eg. promoters, gene bodies, exons and introns; see below), we could increase the chances to observe CpG→TpG conversions and predict regional methylation levels. Since CpG methylation levels have been shown to be significantly correlated within 1,000-2,000 bp [Eckart et al. 2006; Down et al. 2008], our method provides a means to scan genomes for differences in regional methylation levels, despite its relative lack of sensitivity at the single nucleotide level.

We first validated M_s as a methylation proxy by looking at known methylation patterns in specific genomic regions: gene promoters and exon-intron boundaries. As promoter methylation levels have been reported to be negatively correlated with their CpG density, we first binned promoters into three groups according to CpG density following [Ball et al. 2009]. A promoter region (PM) was defined following [Ball et al. 2009] as spanning 500 nucleotides upstream and 2,000 nucleotides downstream of the Transcription Start Site (TSS). High CpG promoters were defined to contain a 500-bp interval with a GC content of at least 0.55 and a CpG observed/expected ratio of at least 0.75. Low CpG promoters were defined as containing no 500-bp interval with a CpG observed/expected ratio of at least 0.48; all remaining promoters were defined as intermediate CpG promoters. Of the 51,939 genes analyzed, 16,186 (31.2%) were found to have high CpG promoters, 11,830 (22.7%) had low CpG

density promoters, and the remaining 23,923 (46.1%) had intermediate CpG promoters (Figure 4.c). For each category we calculated the distribution of M_s values. Given the relatively low *post-mortem* deamination rates observed in the Saqqaq genome (leading to ~4% C$\rightarrow$T mismatch rates at sequencing starts; Figure 4.a), we required each promoter to have a minimum of 50 CpG dinucleotides in the reference in order to increase the likelihood of detecting deamination-driven 5mCpG$\rightarrow$TpG conversions. In addition, we required a minimal number of 2 C$\rightarrow$T mismatches per promoter in order to limit the impact of sequencing error in our estimates. In total, we considered a final number of 2,444, 5,973 and 10,026 promoters in the three classes of promoters (low, intermediate and high CpG, respectively).

We found that on average, genes having promoters with high CpG densites exhibited lower M_s values than genes having promoters with intermediate CpG densities (median R_s values for the global gene dataset: 1.35% *versus* 2.06%; Figure 4.c). Similarly, genes having promoters with intermediate CpG densities showed lower M_s values than promoters from the low CpG class (median M_s values of 2.06% *versus* 5.46%). Overall, this supported the expected inverse relationship between CpG density and promoter methylation levels as reported in Ball et al. [2009] and therefore confirmed M_s as a good proxy for gene methylation at the regional level.

$$M_s(\text{region}) = \frac{N2+M1}{N1+N2+M1+M2}$$
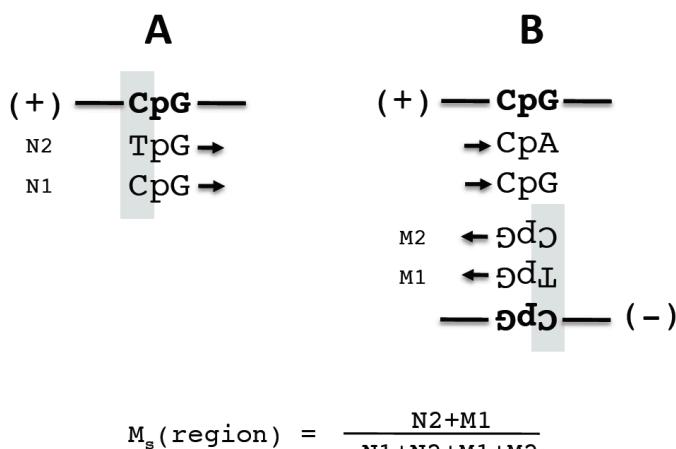
Figure S3.3

Defining M_s, a proxy for methylation levels of a given genomic region.
As cytosines at sequencing starts show maximal but still moderate deamination levels in the Saqqaq sequence dataset (Figure 4.a, bottom left), we decided to focus only on the first position of sequencing reads, outlined in grey, and ignore possible CpG$\rightarrow$TpG conversions located further down in the sequence. The direction of sequencing reads is indicated with arrows. For Phusion reads mapping the human reference hg18 on the (+) strand and at a location starting with a CpG dinucleotide, we first counted N1 and N2 as the number of reads starting with CpG and TpG, respectively (panel A). The latter results from *post-mortem* deamination of 5mC. We next focused on reads mapping the (-) strand and counted the number of Phusion reads starting with CpG and TpG. These were flagged as ending with CpG and CpA when reverse completed and aligned to the (+) strand (panel B). The M_s statistics of a given region consists of the ratio of methylation-derived mis-incorporation signatures by the total number of reads starting/ending at a CpG location in hg18. Note that this measure could be extended to other sequence positions in cases ancient DNA templates show higher deamination rates than what observed on the Saqqaq individual (and to sequencing ends if full length sequence information is available).

We next performed similar analyses at splice sites boundaries, in an attempt to recapitulate the known enrichment in methylated Cytosines at the splice site boundary and the overall decrease in methylation rates from exons to introns [Laurent et al. 2010]. Exon coordinates of all genes from the hg18 reference genome were retrieved from UCSC Genome browser database [Meyer et al 2012] resulting in 95,746 splice sites. We calculated M_s for each position within 100 bp of each splice site. Final M_s profiles were plotted as a function of their position relative to the distance to the splice site boundary using a smoothing procedure based on 10bp long sliding windows (Figure S2.2). M_s successfully recovered the expected decline from relatively higher methylation levels in exons *versus* introns, as well as a sharp local increase in methylation at the splice site. This further confirmed M_s as a good proxy for monitoring methylation levels.

We further focused on CGIs, which function as promoters for about 60% of all human genes. Recently, Ginno et al. [2012] reported that CGI promoters showing significant strand asymmetry in the distribution of guanines and cytosines (GC skew) immediately downstream of their transcription start (TSS) are methylation-resistant. This feature capacitates the formation and the thermodynamic stability of R loops during transcription (with G-rich RNA strands reannealing to template C-rich DNA strands, forcing G-rich DNA strands into a single strand conformation) and is highly predictive of the unmethylated state of CGIs. Consequently, at CGI promoters showing high GC skew, methylation is expected to show minimal levels at TSS and to gradually increase upstream and downstream the TSS before reaching average levels ca. 1,000-2,000 bp away [Ginno et al. 2012]. This pattern is virtually absent at CGI promoters showing no strand asymmetry in the distribution of guanines and cytosines immediately downstream from their TSS. We therefore retrieved all 7,820 genomic coordinates covered by GC skew CGI promoters as defined from Ginno et al. [2012] and binned each position relative to the distance to the TSS. We then calculated M_s levels at each of those positions summing across all GC skew promoters. Final M_s profiles were plotted using a smoothing procedure based on 10bp long sliding windows (Figure S3.4). GC skew CGI promoters were found to show the expected methylation profile, with minimal methylation at TSS followed by a gradual increase upstream and downstream the TSS, suggesting that M_s provides a good measure of ancient cytosine methylation levels.

Ms(CG skew CGI promoters)
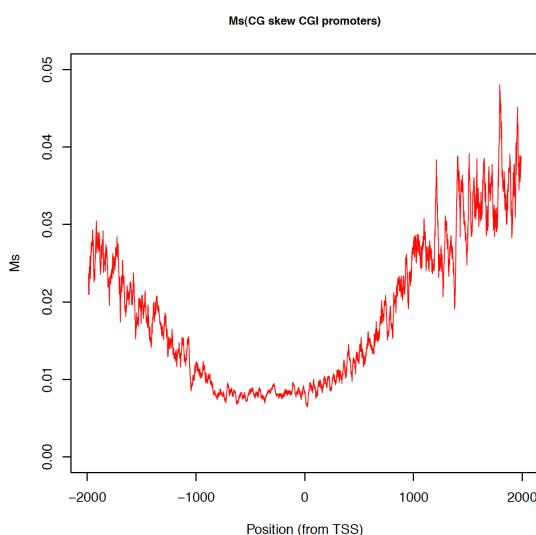
Figure S3.4

M_s methylation profiles at GC skew CGI promoters.
GC skew CGI promoters show significant strand asymmetry in the distribution of guanines and cytosines immediately downstream from their TSS [Ginno et al. 2012]. We calculated M_s values across GC skew CGIs at each position located within 2,000 bp upstream and downstream of the TSS. The final plot was generated following a smoothing procedure. A characterization of the promoters showing a pattern similar to GC skew CGI promoters is provided in Supplemental Section SI3.5 and Table S3.5.

We next followed the procedure described in Straussman et al. [2009] to predict CGIs that are protected from *de novo* methylation and that thus remain constitutively under-methylated throughout development. At such regions, M_s is expected to show lower values that at methylation-rich CGIs. The latter were defined using the data presented in Straussman et al. [2009] as CGIs showing methylation in every cell type tested. Although no specific data were available for hair shafts and hair follicles in this study, the methylation levels were monitored across a wide range of tissues as diverse as brain, liver, bone, colon, sperm and blood and CGIs showing methylation in every cell type tested were recorded. Similar levels of methylation at those CGIs are likely to be found in hairs. Predicted constitutively under-methylated CGIs represented a total number of 13,372 genomic regions while methylation-rich CGIs represented 2,583. As expected, we found a ca. 7.4-fold reduction in methylation levels for predicted under-methylated CGIs as compared to ubiquitously methylated CGIs (Figure S3.5). This again validated M_s as a good proxy for regional methylation levels.
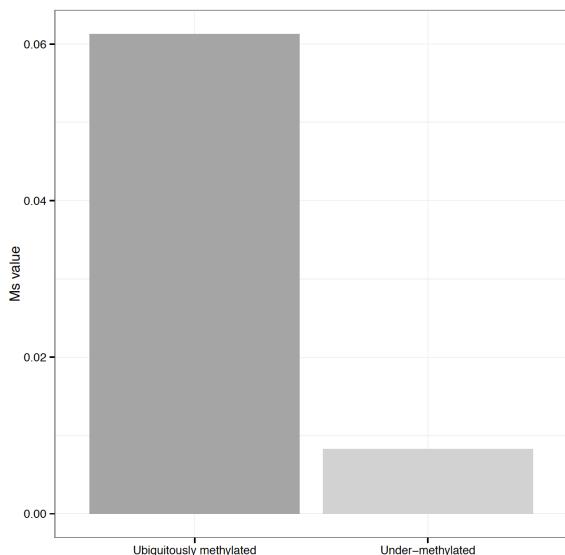
Figure S3.5

M_s methylation levels at predicted under-methylated and ubiquitously methylated CGIs. GCIs were binned in two main classes following Straussman et al. [2009]. The first corresponds to a class of CGIs that remain significantly under-methylated during development while the second corresponds to CGIs showing significant methylation levels across a full range of somatic tissues.

Finally, we focused on two specific genes for which methylation levels have been quantified in a range of hairs sampled from human individuals at different stages of their life (from birth to >80 years). The first gene, SOX10, has been shown to reach high methylation levels within its promoter region (chr22:36,710,485-36,711,485) following two years of age [Kim et al. 2006]. Similarly, the second gene, CSX, follows similar and synchroneous age-dependent methylation changes, however reaching lower methylation levels than SOX10 at its 3'-untranslated region [Kim et al. 2006]. We therefore used M_s to calculate methylation levels within both regions. For CSX, coverage was limited, precluding restricting M_s calculation to the sole 3'-UTR. Ms was therefore calculated for a larger region (chr5:172,591,951-172,592,951) where we found evidence for methylation in both genes. We estimated ca. 18-fold greater methylation levels at SOX10 than CSX. This suggests that besides large-scale trends at different classes of genomic regions, M_s can also capture hair-specific methylation information at specific genes.

3.4. Methylation-based unsupervised hierarchical clustering

Convinced that M_s could capture genuine information regarding ancient regional methylation levels of the human genome, we next compared the methylation levels as estimated from the Saqqaq data to known methylation levels of different human tissues, including hairs, blood, buccal, fat, liver, muscle, omentum, pancreas and saliva [Slieker et al. 2013]. For hair, blood, buccal and saliva, those methylation levels were surveyed in five human donors (PT1, PT2, PT3, PT4 and PT5) ranging in age from 22 to 40 years. For blood, fat, liver, muscle, omentum and pancreas, we restricted our comparisons to the only two individuals from the original study surveyed for all tissues (IT13 and IT15 of age 64 and 65, respectively). This comparative dataset corresponded to normalized methylation levels at 450k CpG sites as available from the Illumina 450k array [Dedeurwaerder et al. 2011; Roessler et al. 2012]. Taking advantage of the known correlation between CpG sites located within 1,000bp-2,000 bp [Eckart et al. 2006; Down et al. 2008], we first calculated M_s for 2,000 bp-wide regions centered at each CpG. Of note, at least 50% (95%) of such regions include 124 (24) CpGs a large proportion of CpGs and are therefore suitable for M_s calculations.

Hereafter (Supplemental Sections SI3.4-SI3.7, SI4), we will refer to coverage as the total number of counts for reads starting at a CpG location within a given genomic region (N1+N2+M1+M2; see Figure S3.3). This represents the total number of reads observed within a given region that enabled us to calculate a regional M_s value for each CpG represented on the Illumina 450k array. We excluded CpG sites located on sexual chromosomes in order to avoid possible bias related to chromosomal X inactivation in females (individuals PT1, PT2 and PT3) and/or coverage differences (the Saqqaq individual was a male). We then filtered CpG regions for a range of minimal coverage (every 50, from 100 to 500) and tested for a possible correlation between M_s and methylation levels observed in each tissue and each individual described above. Linear models were built using the lm() function in R [R Development Core Team 2012] and adjusted-R squared values were calculated for three different models. In the first model, M_s and methylation levels known for a given individual and a given tissue were the two measurements considered (y~x). In the second model, we also introduced the coverage observed as an extra measurement, as the genomic regions considered show different ranges of CpG densities; this measurement was considered as independent from M_s (y~x+z) Finally, in the third model, this measurement was allowed to covary with M_s (y~x*z). Tables S3.1 and S3.2 report the models that received higher support in each of the tests performed together with respective adjusted-R squared values. Of note, we found high adjusted R-squared values across all tissues and individuals investigated (0.440-0.785), supporting significant correlation levels between known biological methylation data and the Saqqaq M_s levels. Importantly, the Saqqaq data showed a higher fit with models built with the methylation data originating from hair tissues (adjusted-R squared = 0.618-0.785 *versus* 0.440-0.778 for non-hair tissues; Tables S3.1-S3.2), in agreement with the nature of the sample used for reconstructing the Saqqaq genome [Rasmussen et al. 2010].

Individual	Coverage	Model	Blood	Model	Buccal	Model	Saliva	Model	**Hair**	Number
PT5	500	y~x*z	0.707	y~x*z	0.770	y~x*z	0.738	y~x*z	**0.778**	937
	450	y~x+z	0.656	y~x+z	0.730	y~x*z	0.706	y~x*z	**0.743**	2139
	400	y~x+z	0.623	y~x+z	0.694	y~x+z	0.664	y~x+z	**0.707**	4827
	350	y~x+z	0.603	y~x+z	0.666	y~x+z	0.637	y~x+z	**0.675**	10421
	300	y~x*z	0.577	y~x*z	0.636	y~x*z	0.606	y~x*z	**0.659**	20830
	250	y~x*z	0.555	y~x*z	0.618	y~x*z	0.588	y~x*z	**0.644**	390007
	200	y~x*z	0.558	y~x*z	0.613	y~x*z	0.588	y~x*z	**0.625**	68873
	150	y~x*z	0.577	y~x*z	0.623	y~x*z	0.601	y~x*z	**0.635**	114561
	100	y~x*z	0.573	y~x*z	0.613	y~x*z	0.593	y~x*z	**0.621**	182064
PT4	500	y~x*z	0.717	y~x*z	0.768	y~x*z	0.734	y~x*z	**0.775**	937
	450	y~x+z	0.677	y~x+z	0.715	y~x+z	0.696	y~x*z	**0.750**	2139
	400	y~x+z	0.640	y~x+z	0.681	y~x*z	0.656	y~x+z	**0.710**	4827
	350	y~x+z	0.621	y~x+z	0.658	y~x*z	0.628	y~x+z	**0.694**	10421
	300	y~x*z	0.592	y~x*z	0.634	y~x*z	0.601	y~x*z	**0.672**	20830
	250	y~x*z	0.577	y~x*z	0.621	y~x*z	0.581	y~x*z	**0.654**	390007
	200	y~x+z	0.573	y~x*z	0.613	y~x+z	0.584	y~x*z	**0.633**	68873
	150	y~x*z	0.590	y~x*z	0.622	y~x*z	0.596	y~x*z	**0.638**	114561
	100	y~x*z	0.582	y~x*z	0.611	y~x*z	0.591	y~x*z	**0.627**	182064
PT3	500	y~x*z	0.698	y~x*z	0.762	y~x*z	0.739	y~x*z	**0.785**	937
	450	y~x+z	0.648	y~x+z	0.715	y~x+z	0.702	y~x*z	**0.752**	2139
	400	y~x*z	0.613	y~x+z	0.681	y~x+z	0.662	y~x+z	**0.709**	4827
	350	y~x+z	0.594	y~x+z	0.658	y~x+z	0.638	y~x+z	**0.692**	10421
	300	y~x*z	0.569	y~x*z	0.630	y~x*z	0.612	y~x*z	**0.669**	20830
	250	y~x*z	0.555	y~x*z	0.615	y~x*z	0.580	y~x*z	**0.644**	390007
	200	y~x+z	0.558	y~x+z	0.612	y~x+z	0.587	y~x*z	**0.635**	68873
	150	y~x*z	0.575	y~x*z	0.619	y~x*z	0.606	y~x*z	**0.641**	114561
	100	y~x*z	0.570	y~x*z	0.610	y~x*z	0.598	y~x*z	**0.624**	182064
PT2	500	y~x*z	0.691	y~x*z	0.778	y~x*z	0.755	y~x*z	**0.776**	937
	450	y~x+z	0.639	y~x+z	0.720	y~x+z	0.715	y~x*z	**0.747**	2139
	400	y~x+z	0.607	y~x+z	0.683	y~x+z	0.675	y~x+z	**0.706**	4827
	350	y~x*z	0.591	y~x*z	0.635	y~x*z	0.618	y~x+z	**0.688**	10421
	300	y~x*z	0.569	y~x*z	0.625	y~x*z	0.606	y~x*z	**0.665**	20830
	250	y~x+z	0.555	y~x*z	0.606	y~x*z	0.601	y~x*z	**0.643**	390007
	200	y~x+z	0.554	y~x*z	0.609	y~x*z	0.601	y~x*z	**0.633**	68873
	150	y~x*z	0.573	y~x*z	0.624	y~x*z	0.613	y~x*z	**0.638**	114561
	100	y~x*z	0.569	y~x*z	0.613	y~x*z	0.603	y~x*z	**0.622**	182064
PT1	500	y~x*z	0.705	y~x*z	0.769	y~x*z	0.718	y~x*z	**0.782**	937
	450	y~x+z	0.650	y~x+z	0.703	y~x+z	0.687	y~x+z	**0.741**	2139
	400	y~x+z	0.617	y~x+z	0.670	y~x+z	0.655	y~x+z	**0.703**	4827
	350	y~x+z	0.600	y~x*z	0.646	y~x	0.542	y~x+z	**0.686**	10421
	300	y~x*z	0.575	y~x*z	0.626	y~x*z	0.568	y~x*z	**0.661**	20830
	250	y~x*z	0.562	y~x*z	0.612	y~x*z	0.537	y~x*z	**0.632**	390007
	200	y~x+z	0.560	y~x*z	0.607	y~x+z	0.560	y~x*z	**0.618**	68873
	150	y~x*z	0.577	y~x*z	0.620	y~x*z	0.553	y~x*z	**0.632**	114561
	100	y~x*z	0.572	y~x*z	0.607	y~x*z	0.548	y~x*z	**0.622**	182064

Table S3.1

Saqqaq M_s methylation levels show stronger correlation with modern hair methylation levels than blood, buccal and saliva tissues.

We considered the normalized methylation data available from [Slieker et al. 2013] for individuals PT1, PT2, PT3, PT4 and PT5. Parameters x, y, and z are described as follows: y = observed methylation levels in a given tissue for a given modern human individual; x = Saqqaq M_s values calculated for a 2,000 bp-wide genomic region centered on a CpG site; z = coverage. The model showing the best fit was selected amongst three different models (y~x; y~x+z, and; y~x*z). Corresponding adjusted-R squared values are provided as well as the total number of CpG loci considered in each analysis.

Individual	Coverage	Model	Blood	Model	Liver	Model	Muscle	Model	Omentum	Model	Pancreas	Model	Fat	Number
IT15	500	y~x*z	0.680	y~x*z	0.661	y~x*z	0.727	y~x*z	0.714	y~x*z	0.685	y~x*z	0.722	937
	450	y~x+z	0.615	y~x+z	0.612	y~x+z	0.686	y~x+z	0.659	y~x+z	0.639	y~x+z	0.697	2139
	400	y~x+z	0.588	y~x+z	0.465	y~x+z	0.626	y~x+z	0.525	y~x*z	0.620	y~x+z	0.470	4827
	350	y~x+z	0.539	y~x	0.455	y~x	0.629	y~x+z	0.534	y~x+z	0.601	y~x+z	0.542	10421
	300	y~x+z	0.482	y~x+z	0.485	y~x*z	0.582	y~x*z	0.548	y~x*z	0.482	y~x*z	0.562	20830
	250	y~x+z	0.498	y~x+z	0.440	y~x*z	0.547	y~x*z	0.491	y~x*z	0.486	y~x*z	0.543	390007
	200	y~x+z	0.496	y~x+z	0.458	y~x*z	0.521	y~x*z	0.519	y~x*z	0.521	y~x+z	0.528	68873
	150	y~x*z	0.506	y~x*z	0.487	y~x*z	0.545	y~x*z	0.542	y~x*z	0.543	y~x*z	0.525	114561
	100	y~x*z	0.499	y~x*z	0.485	y~x*z	0.530	y~x*z	0.534	y~x*z	0.541	y~x*z	0.535	182064
IT13	500	y~x*z	0.683	y~x*z	0.663	y~x*z	0.723	y~x*z	0.703	y~x*z	0.695	y~x*z	0.722	937
	450	y~x+z	0.629	y~x+z	0.628	y~x+z	0.682	y~x+z	0.643	y~x+z	0.652	y~x+z	0.696	2139
	400	y~x+z	0.574	y~x*z	0.609	y~x+z	0.623	y~x*z	0.625	y~x*z	0.632	y~x+z	0.673	4827
	350	y~x*z	0.581	y~x*z	0.589	y~x	0.572	y~x*z	0.593	y~x	0.616	y~x+z	0.538	10421
	300	y~x+z	0.518	y~x+z	0.576	y~x*z	0.560	y~x+z	0.554	y~x*z	0.584	y~x+z	0.540	20830
	250	y~x+z	0.481	y~x+z	0.550	y~x*z	0.546	y~x*z	0.546	y~x*z	0.571	y~x*z	0.524	390007
	200	y~x*z	0.503	y~x*z	0.556	y~x*z	0.509	y~x+z	0.542	y~x*z	0.568	y~x+z	0.515	68873
	150	y~x+z	0.515	y~x*z	0.548	y~x*z	0.537	y~x*z	0.532	y~x*z	0.580	y~x*z	0.549	114561
	100	y~x*z	0.515	y~x*z	0.538	y~x*z	0.521	y~x*z	0.525	y~x*z	0.573	y~x*z	0.534	182064

Table S3.2

Levels of Saqqaq M_s methylation with the methylation levels measured in a variety of somatic tissues from modern human individuals.
We considered the normalized methylation data available from [Slieker et al. 2013] for individuals IT13 and IT15. See legends of Table S3.1 for further details.

We next performed unsupervised hierarchical clustering analyses in order to evaluate whether methylation levels as measured in the Saqqaq sequence data would support clustering with modern hair tissues. Analyses were first restricted to individuals PT1, PT2, PT3, PT4 and PT5 where methylation levels were surveyed across four different tissues, including hairs (this information was not available for any other individual from the original study, including IT13 and IT15) [Slieker et al. 2013]. We selected the most supported linear model described above in order to predict Saqqaq methylation values from M_s. We then selected from the Illumina 450k array data all CpG showing at least a 2-fold average difference in methylation levels between pairs of tissues (note that we also found similar clustering results while investigating a more conservative selection scheme where CpGs showing at least a 2-fold average differences between one tissue and all the other tissues altogether; Figure S3.10 and S3.11). Interestingly, CpG showing at least 2-fold higher methylation levels in modern hairs than other tissues ('High') showed greater M_s values for the Saqqaq than CpG showing at least 2-fold higher methylation levels in other tissues than in modern hairs ('Low'; Figures S3.6-S3-7), regardless of the threshold considered as minimal coverage (even though Figures S3.6-S3-7 refer 100 or 500 coverage thresholds, the full range of possible values from 100 to 500, iterating every 50, was investigated and resulted in similar findings; data not shown). This suggests that for those CpG, the Saqqaq hair M_s values better fit an expected hair methylation profile than any other tissue profile investigated. Finally, we used the subset of CpG showing at least a 2-fold average difference in methylation levels between pairs of tissues to perform unsupervised clustering using the heatmap() function from R [R Development Core Team 2012]. Predicted Saqqaq methylation values were ignored whenever the absolute values of residuals were greater than 0.1 and forced to zero when negative and to one when superior to one (note that similar clustering was also found in absence of filtering on residuals; data not shown). The analyses were repeated over a full range of minimal coverage (every 50, from 100 to 500), and resulted in similar outcomes. Therefore, the analyses performed with the lower coverage investigated are shown as Figure 5 (information about the specific of CpG sites considered is available upon request). For all coverage threshold

investigated, the Saqqaq specimen was found to cluster together with hair tissues from 5 modern human individuals, in agreement with the nature of the biological material used for reconstructing the Saqqaq genome [Rasmussen et al. 2010]. GREAT version 2.0.2 analysis [McLean et al. 2010], where the top-10% CpGs showing a maximal absolute methylation difference between Saqqaq hairs and the closest of the five modern individuals was challenged against the background of remaining CpGs, revealed no significant annotation categories related to hair biology (hypergeometric test; data not shown). Interestingly, the top-10% array CpGs that showed divergent methylation profiles apparently seemed more often located 50-500 bp away from the TSS than other array CpGs passing our filtering criteria (29.6% vs 27.5%; Figure S3.8). However, this difference was not found significant (chi-square p-value = 0.396). This suggests that the top-10% most divergent CpG sites represent a random sample of the other array CpG sites. That lower coverage did not disrupt this clustering, despite fewer chance events of observing deaminated 5mC, indicates that the *post-mortem* levels of cytosine deamination observed here are compatible with the recovery of a tissue-specific methylation proxy as long as coverage values are superior or equal to 100. Lower minimal coverage requirements were not investigated but could likely be achieved in situations where *post-mortem* deamination levels are higher than those observed in the Saqqaq individual.

Performing the full set of analyses described above but using 1,500 bp-wide regions centered at each CpG (instead of 2,000 bp), we found similar support for stronger correlations and clustering between the Saqqaq and hair methylation levels (minimal coverage investigated ranged from 100 to 400, iterating every 50; Figure S3.9). Using the same approach, we also found that the clustering of the Saqqaq together with modern hair tissues was maintained when considering all CpGs, regardless of possible methylation differences across tissues (one example corresponding to regions with coverage greater or equal to 100 is provided in Figure S3.12). This suggests that CpG sites showing no strong tissue specificity in their methylation patterns do not preclude genuine clustering. Finally, since M_s provides regional methylation levels as opposed to the single-base methylation values gathered from the Illumina 450K array, we calculated regional methylation values from the Illumina 450K array data. We achieved this by parsing the genomic coordinates of all CpGs from the array and collapsing those that are distant by at best 1,000bp (i.e. the physical distance used on each side of the CpG when calculating M_s values). We then calculated a regional methylation value for those CpGs by averaging normalized methylation levels. Applying a minimal coverage threshold of 100 for M_s calculations, such regions exhibited 4.76 CpGs on average (median = 4, min = 2, max = 28). Only regions showing a minimum number of 4 CpGs were considered in order to insure that the average values could represent regional methylation, thereby limiting the impact of potentially aberrant single CpGs (eg. a hyper-methylated CpG located on the shore of an otherwise unmethylated island). We then used M_s values for Saqqaq at the first array CpG present in a group of collapsed CpGs and performed unsupervised clustering following the same procedure as above. The results are shown in Figure S3.13. This analysis supports the clustering of the Saqqaq data together with modern hairs despite the presence of a fraction of loci where the Saqqaq exhibits estimated methylation values apparently outside the range of the variation observed amongst hair tissues of the five modern individuals analyzed here (Figure S3.13). Such loci were identified as the top-10% CpGs showing a maximal absolute methylation difference for hairs between the Saqqaq and the closest of the five modern individuals. GREAT analysis was performed against the background of CpGs used for clustering in order to test for possible enrichment in specific GO molecular functions (hypergeometric test) [McLean et al. 2010]. No significant GO category was identified following Bonferroni correction for multiple testing (data not shown). All in all, those analyses support the validity of our methylation estimates and demonstrate that genuine methylation

information can be recovered from past individuals using next-generation sequencing information in absence of bisulfite treatment.
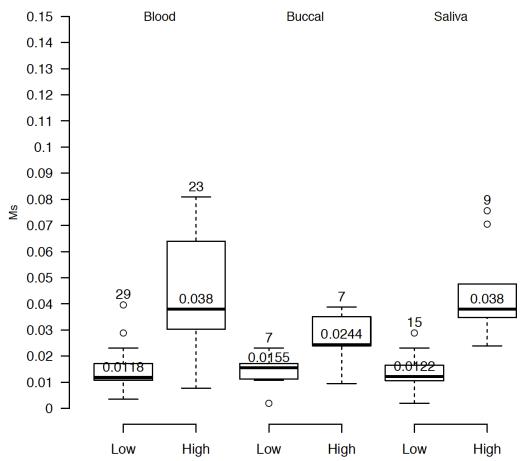
Figure S3.6

Comparing methylation levels in the Saqqaq hairs at CpG showing 2-fold difference methylation in modern tissues.

We identified CpG from the Illumina 450k array showing at least 2-fold higher methylation in modern hairs than other tissues (blood, buccal or saliva; 'High') and CpG from the Illumina 450k array showing at least 2-fold lower methylation in modern hairs than other tissues ('Low'). M_s values were greater for the 'High' CpG set and lower for the 'Low' CpG set, in agreement with known methylation levels in modern hairs. We considered a minimal coverage threshold of 500. The total number of CpG regions from the array is reporting above the boxplots. Median M_s values are also indicated.
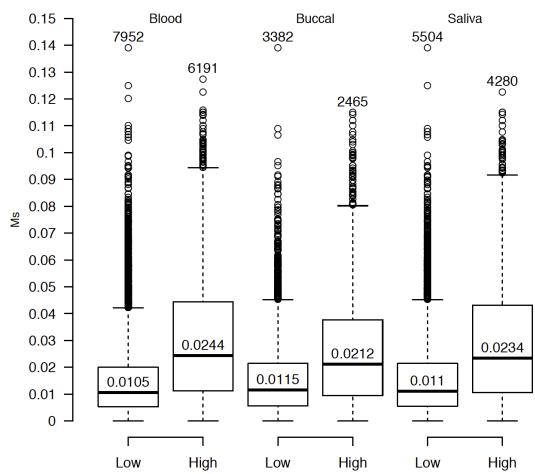
Figure S3.7

Comparing methylation levels in the Saqqaq hairs at CpG showing 2-fold difference methylation in modern tissues.

We identified CpG from the Illumina 450k array showing at least 2-fold higher methylation in modern hairs than other tissues (blood, buccal or saliva; 'High') and CpG from the Illumina 450k array showing at least 2-fold lower methylation in modern hairs than other tissues ('Low'). M_s values were greater for the 'High' CpG set and lower for the 'Low' CpG set, in agreement with known methylation levels in modern hairs. We considered a minimal coverage threshold of 100. The total number of CpG regions from the array is reporting above the boxplots. Median M_s values are also indicated.

Figure S3.8

Distribution of the absolute distance of array CpGs from TSS.

Red: top-10% CpG sites with a methylation profile divergent to modern hair methylation profiles. Black: other CpG sites. CpG sites correspond to the clustering analysis presented in Figure 5.
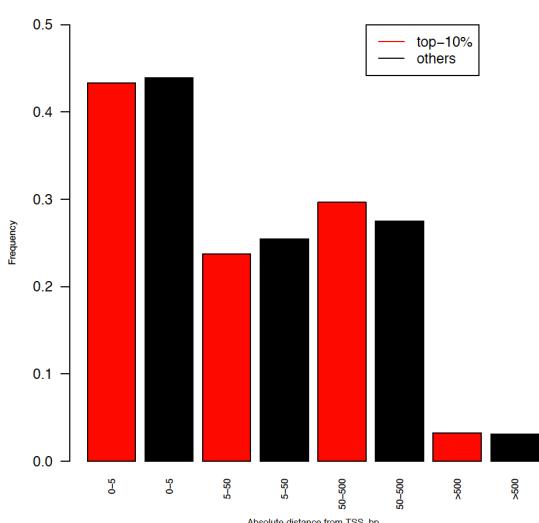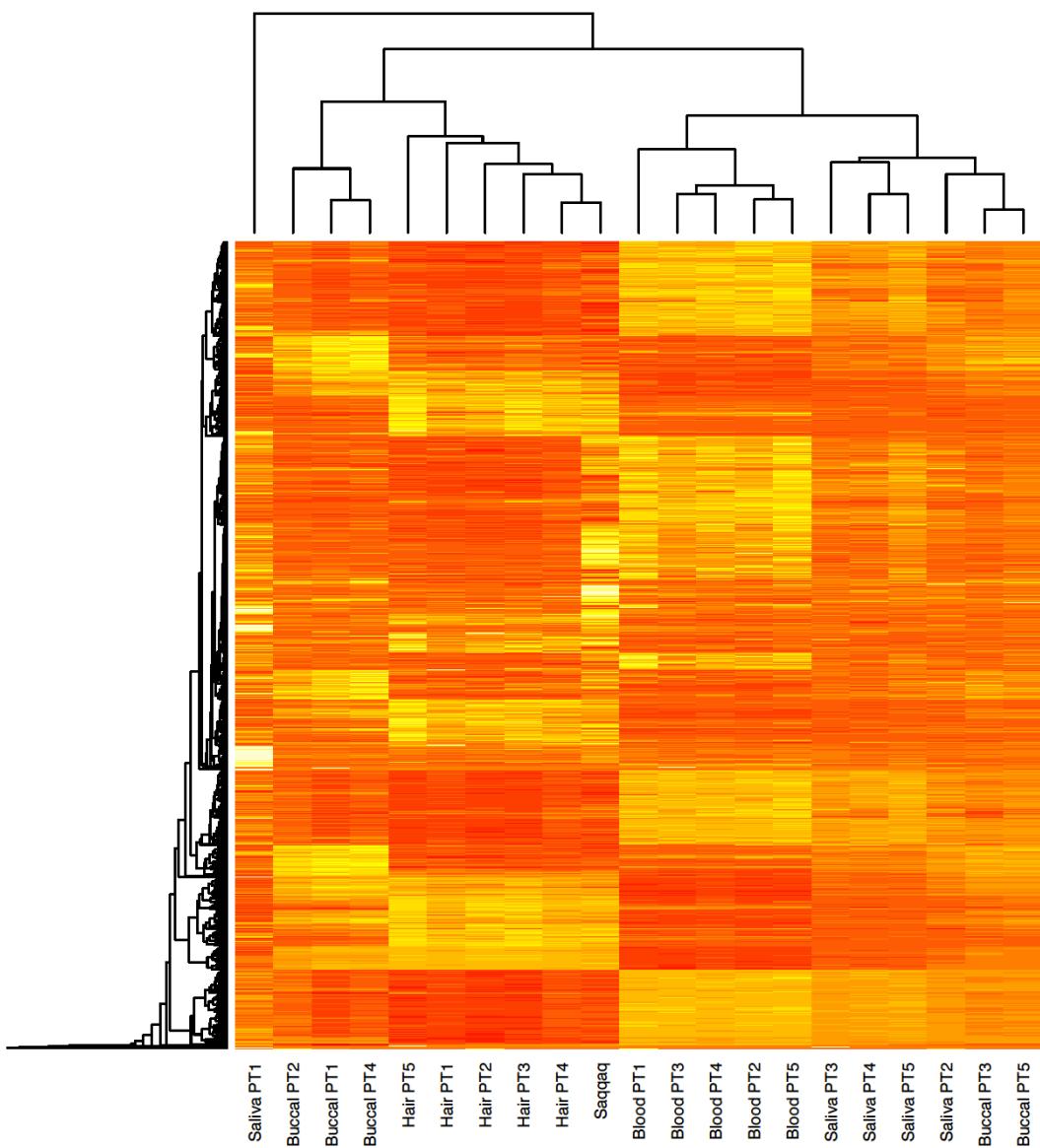
Figure S3.9

Unsupervised hierarchical clustering of the Saqqaq and modern human tissues based on methylation levels.

See Figure S3.10 for details, except that the regions considered were 1,500 bp-wide. A minimal coverage of 100 was required. A total number of 4,939 CpG showing at least a 2-fold difference in methylation levels between pairs of tissues for modern humans PT1, PT2, PT3, PT4 and PT5 were considered for unsupervised hierarchical clustering. The list of CpGs considered is available upon request.
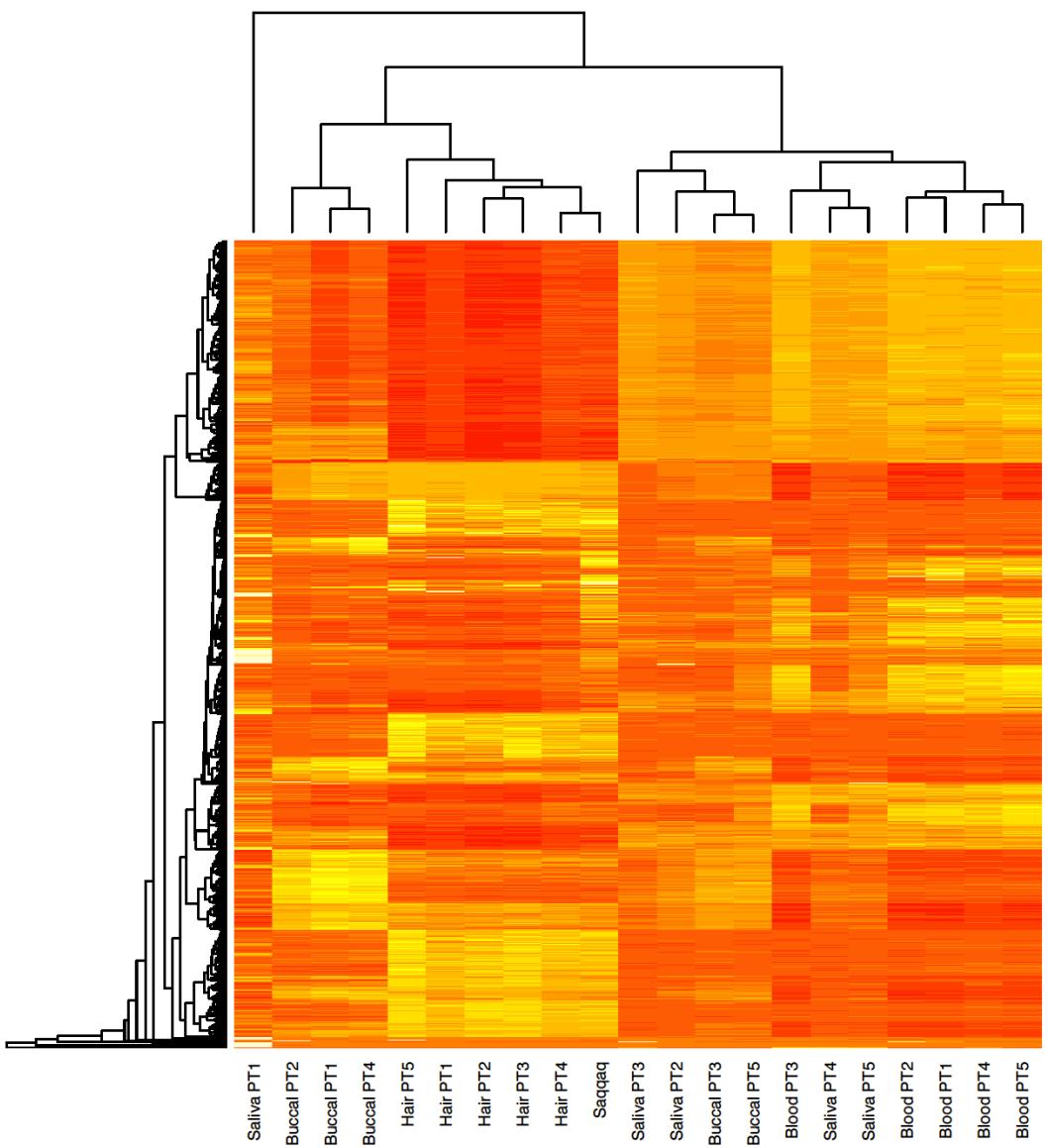
Unsupervised hierarchical clustering of the Saqqaq and modern human tissues based on methylation levels.

Same as Figure 5, except that a more conservative selection scheme was considered where CpGs showing at least a 2-fold average differences between one tissue and all the other tissues. This represented a total number of 2,607 CpG sites.
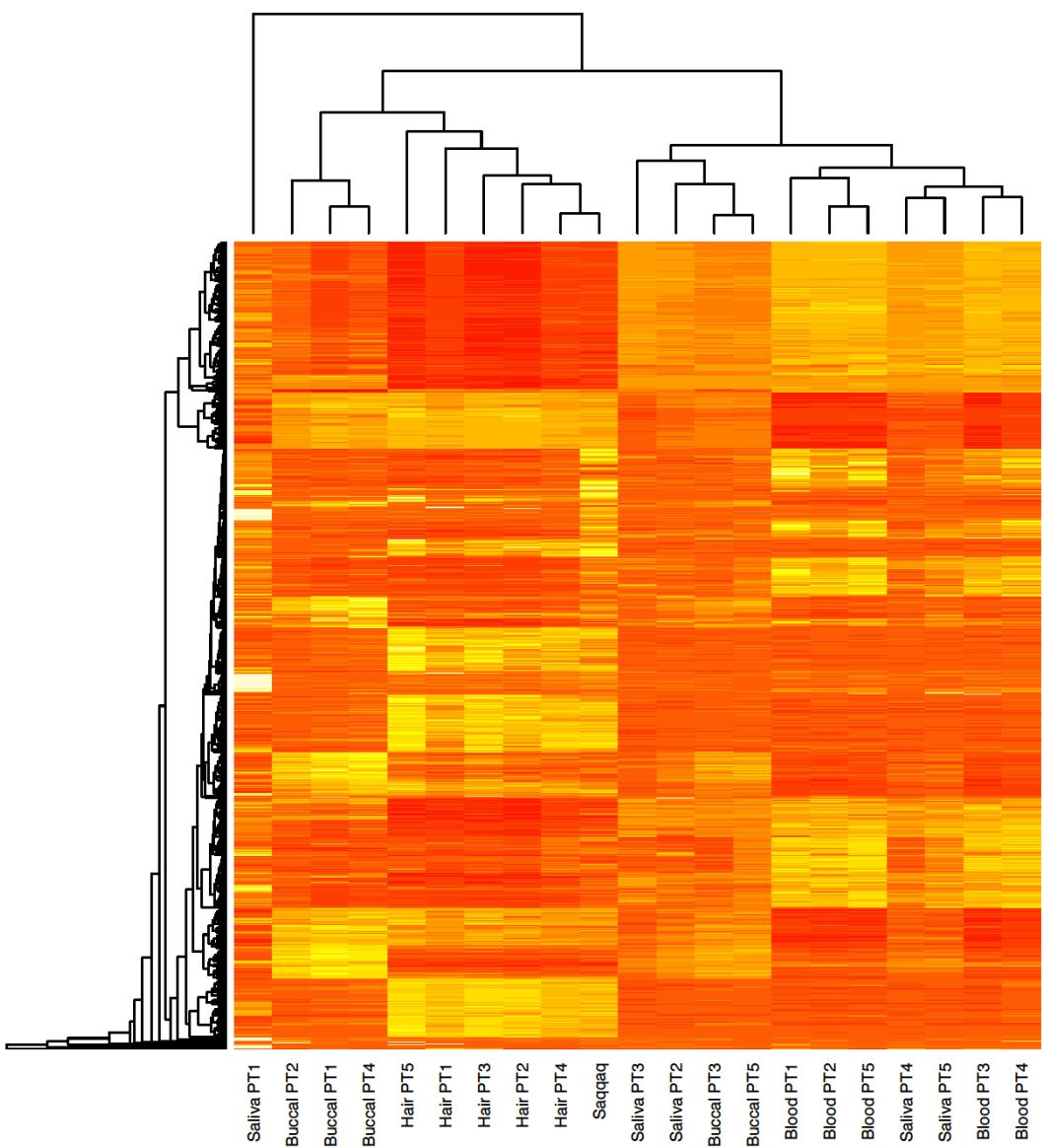
Figure S3.11
Unsupervised hierarchical clustering of the Saqqaq and modern human tissues based on methylation levels.
Same as Figure S3.9, except that a more conservative selection scheme was considered where CpGs showing at least a 2-fold average differences between one tissue and all the other tissues. This represented a total number of 1,722 CpG sites.
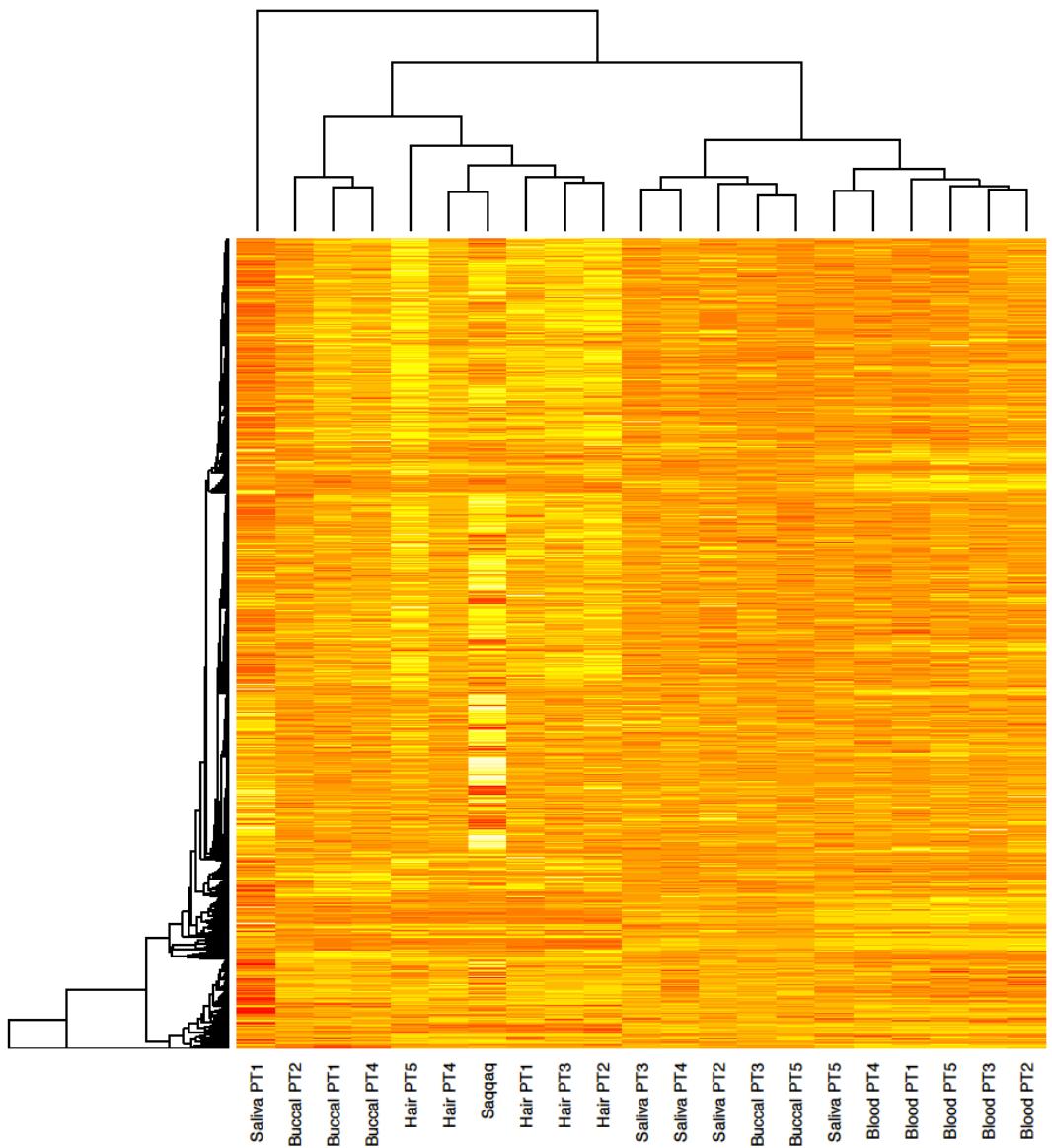
Unsupervised hierarchical clustering of the Saqqaq and modern human tissues based on methylation levels.
Same as Figure S3.10, except that all CpGs passing coverage threshold were considered, regardless of their respective level of methylation difference across tissues. A minimal coverage threshold of 100 was applied and a total number of 45,016 CpG was considered.
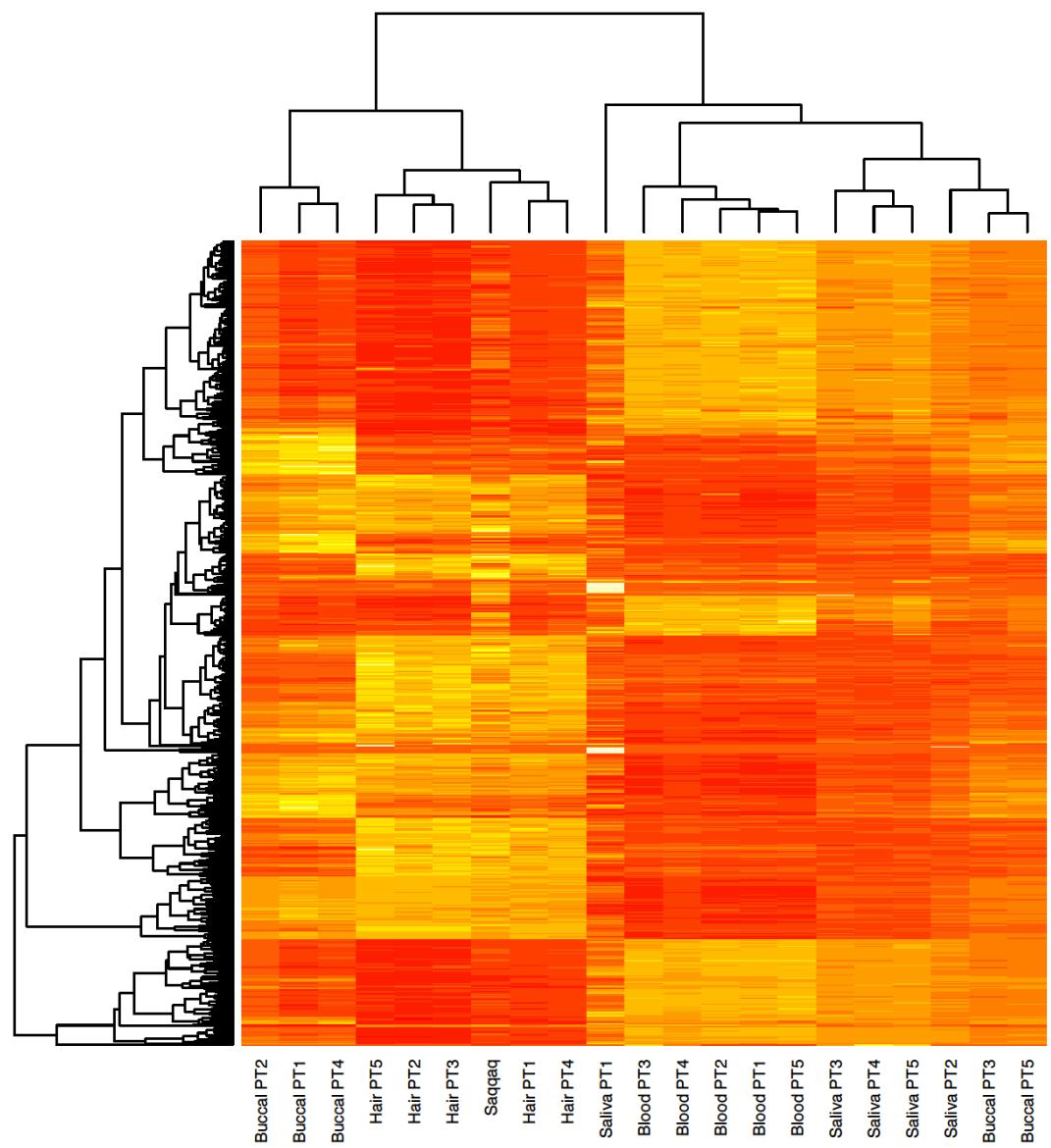
Figure S3.13

Unsupervised hierarchical clustering of the Saqqaq and modern human tissues based on methylation levels.

Same as Figure S3.10, except that array CpGs located within 1,000 bp were collapsed in order to estimate regional methylation levels for modern individuals. A minimal coverage threshold of 100 was applied and a total number of 685 collapsed regions containing 4,102 CpGs was considered.

3.5. Promoter and first exon methylation levels

We downloaded from the UCSC genome browser all gene annotations available for the human reference genome hg18 and calculated the distribution of M_s values for each gene within their promoter region. Promoters were defined as the 1,000 bp located upstream of the transcription starting site (TSS). The top-1% promoters showing highest coverage but M_s values equal to zero are provided in Table S3.3. Those represent a total number of 266 UCSC candidate promoters (of which 168 are not redundant) virtually devoid of cytosine methylation in the Saqqaq genome (this list could be extended to 1,311 genes if the top-5% was considered instead of the top-1%). The lack of detectable methylation levels despite extensive coverage excludes promoter cytosine methylation as one important mechanism controlling the expression of those genes. Likewise, we identified the top-1% promoters showing highest methylation levels (Table S3.4). In order to correct for possible coverage issues, we first selected all genes showing a minimum coverage of 75 over their full gene body. We also required a minimum coverage of 25 for the promoter region (given the observed deamination levels, ca. 4% at read starts, this gives on average one chance to observe one CpG→TpG conversion). We then selected genes where coverage at promoters was at least equivalent to that observed within gene bodies (correcting for length differences between both regions). Our list of 300 UCSC candidates for highest methylation levels is provided in Table S3.4 (214 are not redundant). This list likely includes a substantial fraction of loci (but not uniquely) whose expression was silenced in Saqqaq hairs.

Additionally, and encouraged by the overall methylation profile observed at GC skew CGI promoters (Supplementary Section S3.4; Figure S3.4), we also attempted to identify which promoters were closely matching such a pattern. M_s values were calculated within 4 successive 1kb windows located within 2kb upstream and downstream of TSS. We then identified which promoters showed minimal M_s values in the second region (i.e. the first kb upstream of TSS) and were associated with (i) at least doubled M_s values in the preceding kb, (ii) a 50% M_s increase in the third region (the first kb downstream of TSS) and (iv) with at least a further 2-fold increase in the following region. This pattern was in agreement with average methylation differences observed across those regions at GC skew CGI promoters (Figure S3.4). Requiring a minimum coverage of 25 for each of the four regions, we identified a total number of 402 promoters matching the expected profile, which represented ca. 5.6% (402/7,214) of the promoters considered (Table S3.5).

Finally, we calculated the distribution of M_s values for each gene within their first exonic region, following criteria similar to those used for identifying promoters with highest methylation levels. Interestingly, M_s values at the first exon were found to be significantly correlated with promoter M_s values (Spearman correlation coefficient = 0.318, p-value < 2.2 10^{-16}), in agreement with recent reports [Brenet et al. 2011]. Given the tight association between transcriptional silencing and methylation levels at the first exon [Brenet et al. 2011], the 214 genes showing highest (top-1%) methylation in this region (159 of which are not redundant; Table S3.6) could potentially provide insights about downregulated cellular functions in the Saqqaq hairs. Potential functional insights inferred from methylation levels and nucleosomal positionning are further discussed in an independent section (SI4).

Name	UCSC	Coordinates	Cov(PM)
ZNF362	uc001bxc.1	chr1:33493761-33494761	254
AK3L1	uc001dca.1	chr1:65385474-65386474	261
SYDE2	uc009wcm.1,uc001dku.2	chr1:85439316-85440316	289
WNT3A	uc001hrp.1,uc001hrq.1	chr1:226260375-226261375	234
C1orf124	uc001hur.1,uc001hus.1,uc001hut.1	chr1:229539305-229540305	242
PLD5	uc001hzn.1	chr1:240754328-240755328	228
AHCTF1	uc001ibv.1	chr1:245161346-245162346	252
GEN1	uc002rcu.1	chr2:17797895-17798895	248
SOS1	uc002rrk.2	chr2:39201108-39202108	230
PRKCE	uc002rut.1	chr2:45731547-45732547	275
RTN4	uc002ryd.1	chr2:55129830-55130830	231
BCL2L11	uc002tqt.1,uc002tqu.1,uc002tqv.1	chr2:111593962-111594962	232
DLX1	uc010fqj.1,uc002uhl.1,uc002uhm.1	chr2:172657454-172658454	241
OSBPL6	uc002ulw.2,uc002ulx.1,uc002uly.1	chr2:178766620-178767620	246
GULP1	uc002uqg.1	chr2:188864804-188865804	227
SATB2	uc002uuy.1	chr2:200031064-200032064	253
FAM119A	uc002vce.2,uc010fuk.1	chr2:208198218-208199218	243
DNPEP	uc002vlf.1	chr2:219960777-219961777	235
GBX2	uc002vvw.1	chr2:236741391-236742391	298
CRELD1	uc003buf.1,uc003buq.1,uc003buh.1,uc003bui.1	chr3:9949506-9950506	234
slac2-c	uc010hhx.1	chr3:39825982-39826982	258
TCTA	uc003cwv.2	chr3:49423643-49424643	250
DOCK3	uc003dbb.1	chr3:50686676-50687676	278
THOC7	uc003dlt.2,uc003dlu.2	chr3:63824637-63825637	255
SCA7	uc010hnu.1	chr3:63824273-63825273	332
ATXN7	uc003dlv.2,uc010hnv.1,uc003dlw.2	chr3:63824273-63825273	332
CDV3	uc003epr.1	chr3:134774991-134775991	326
TMEM175	uc010ibl.1,uc003gbq.1,uc003gbr.1,uc003gbs.1,uc003gbt.1,uc003gbu.1,uc003gbv.1	chr4:915262-916262	260
DKFZp547K246	uc003gbp.1	chr4:915262-916262	260
NOP14	uc003ggj.1,uc003ggk.2,uc003ggl.2,uc010icq.1	chr4:2934916-2935916	228
RGS12	uc003ggu.2,uc010ics.1	chr4:3263553-3264553	246
KLF3	uc003gtq.2,uc003gth.2	chr4:38341185-38342185	317
FAM47D	uc003hjv.1	chr4:77353217-77354217	248
PGRMC2	uc003igg.1	chr4:129428398-129429398	232
PDGFC	uc003iph.1,uc003ipi.1	chr4:158111996-158112996	251
ANKRD37	uc003ixm.1	chr4:186553834-186554834	231
PLK2	uc003jrn.1	chr5:57791670-57792670	253
SFRS12	uc003juo.1	chr5:65475419-65476419	229
PRDM6	uc003kti.1,uc003ktj.1	chr5:122452916-122453916	250
VDAC1	uc003kyp.1,uc003kyq.1	chr5:133368332-133369332	263
TCF7	uc003kyz.1,uc003kza.1	chr5:133478499-133479499	260
EIF4E1B	uc010jkf.1	chr5:175989289-175990289	253
BMP6	uc003mxu.2	chr6:7671010-7672010	238
HS3ST5	uc003pwh.2	chr6:114770233-114771233	282
TCBA1	uc010kep.1,uc003pzp.1,uc010keq.1	chr6:124165768-124166768	323
NKAIN2	uc003pzn.1,uc003pzo.1	chr6:124165768-124166768	323
MAP3K7IP2	uc003qmj.1	chr6:149679756-149680756	228
FGFR1OP	uc003qvj.1,uc003qvk.1	chr6:167331806-167332806	227
CCR6	uc003qvl.1	chr6:167331806-167332806	227
C7orf23	uc003uio.1	chr7:86686967-86687967	255
RELN	uc003vca.1,uc010liz.1	chr7:103417198-103418198	263
KIAA1862	uc010lpj.1	chr7:149042035-149043035	270
KRBA1	uc003wfz.1	chr7:149042081-149043081	282
WHSC1L1	uc003xli.1,uc010lwe.1,uc003xlj.1	chr8:38358947-38359947	260
XKR4	uc003xsf.1	chr8:56176571-56177571	236
PLAG1	uc003xsr.2,uc010lyi.1,uc010lyj.1	chr8:57286413-57287413	271
MAFA	uc003yyc.1	chr8:144583719-144584719	343
EEF1D	uc003yyr.1,uc003yys.1,uc003yyt.1,uc003yyu.1,uc003yyv.1	chr8:144750726-144751726	329
TIGD5	uc003yyx.1	chr8:144750364-144751364	328
SCRIB	uc003yzo.1,uc003yzp.1	chr8:144969537-144970537	355
PUF60	uc003yzq.1,uc003yzr.1	chr8:144983197-144984197	302
pp9320	uc010mfm.1	chr8:144994389-144995389	263
KIFC2	uc003zcz.1	chr8:145661546-145662546	235
ZNF7	uc003zeg.2,uc010mge.1,uc003zeh.2	chr8:146022707-146023707	284
RNF38	uc003zzh.1,uc003zzi.1	chr9:36390296-36391296	388
PIP5K1B	uc004agu.1	chr9:70509436-70510436	269
C9orf85	uc004ain.1,uc010mot.1,uc010mou.1,uc010mov.1	chr9:73715243-73716243	268
PTCH1	uc010mrq.1,uc004avl.2	chr9:97310143-97311143	285
PPP2R4	uc004bxm.1,uc004bxn.1,uc004bxo.1	chr9:130912434-130913434	240
DKFZp781M17165	uc010myr.1	chr9:130912434-130913434	240
CAMSAP1	uc004cgr.2	chr9:137938826-137939826	290
SEC16A	uc004chx.1,uc010nbo.1	chr9:138497328-138498328	357
PFKP	uc001igg.1	chr10:3099819-3100819	264
ARHGAP21	uc009xkl.1	chr10:25052175-25053175	300
ITGB1	uc001iws.2	chr10:33286828-33287828	269
GLUD1	uc001keq.1	chr10:88843807-88844807	268
FAM35A	uc001kei.2	chr10:88843933-88844933	263
LIPA	uc001kgb.2,uc001kgc.2	chr10:91164362-91165362	241
GALNAC4S-6ST	uc001lhm.1	chr10:125841898-125842898	254
STK32C	uc001lld.1	chr10:133970706-133971706	368
PDDC1	uc001lrc.1,uc001lrd.1,uc001lrf.1,uc009ycg.1	chr11:767487-768487	227
PHLDA2	uc009yds.1,uc001lxa.1	chr11:2907170-2908170	263
TP53I11	uc001myi.1	chr11:44928287-44929287	325
SYT7	uc001nrv.1,uc009ynr.1	chr11:61104874-61105874	238
GPR137	uc001nzg.1	chr11:63808907-63809907	242
LRFN4	uc001ojs.2,uc001ojr.1	chr11:66380572-66381572	287
PITPNM1	uc001olx.1	chr11:67028416-67029416	245
APOLD1	uc001rau.2	chr12:12769130-12770130	229
DDX47	uc001rav.1	chr12:12769130-12770130	229
AEBP2	uc001req.1	chr12:19483782-19484782	238
TMTC1	uc001rjc.1	chr12:29827246-29828246	254
BICD1	uc001rku.1,uc001rkv.1	chr12:32150452-32151452	263

44

SCN8A	uc001ryw.1	chr12:50270287-50271287	277
DKFZp761C241	uc001tdw.1	chr12:94134972-94135972	238
APAF1	uc001tfy.1,uc001tfz.1,uc001tga.1,uc001tgb.1,uc001tgc.1	chr12:97562209-97563209	238
TMEM116	uc001ttc.1,uc001ttd.1,uc001tte.1,uc001ttf.1,uc001tth.1,uc001tti.1	chr12:110935318-110936318	240
RAB35	uc001txm.1,uc009zww.1	chr12:119038982-119039982	227
CCDC92	uc001ufw.1,uc001ufy.1,uc001ufx.1	chr12:123023116-123024116	340
ZNF664	uc001uga.1,uc001ugb.1	chr12:123022733-123023733	317
PXMP2	uc001ukt.1	chr12:131773265-131774265	261
TSC22D1	uc001uzo.1	chr13:44048392-44049392	288
PCDH9	uc001vik.1,uc001vil.1,uc001vin.2	chr13:66702464-66703464	237
KLF12	uc010aeq.1	chr13:73606395-73607395	245
ARHGAP5	uc001wrm.1,uc001wrn.1,uc001wro.1,uc001wrp.1	chr14:31615246-31616246	232
CCNK	uc001ygi.2,uc001ygg.2	chr14:99016492-99017492	324
EVL	uc001ygt.1	chr14:99506904-99507904	263
CDC42BPB	uc001ymi.1	chr14:102593495-102594495	289
EIF5	uc001yms.1	chr14:102869469-102870469	242
CKB	uc001yne.1	chr14:103057774-103058774	384
TMEM121	uc010axm.1	chr14:105065217-105066217	265
GABRB3	uc001zaz.1	chr15:24569344-24570344	259
PAK6	uc001zlb.2	chr15:38331671-38332671	242
LOC196993	uc002ata.2	chr15:69194893-69195893	250
PKM2	uc010bit.1	chr15:70309117-70310117	234
KIAA1024	uc002bew.1	chr15:77510913-77511913	295
AXIN1	uc002cgp.1,uc002cgq.1	chr16:342465-343465	322
SOX8	uc002ckn.1	chr16:970809-971809	314
CACNA1H	uc002cks.1,uc002ckt.1	chr16:1142242-1143242	298
ADCY9	uc002cvx.1	chr16:4106187-4107187	300
DKFZp564M1672	uc002cxu.1	chr16:4792164-4793164	235
UBN1	uc002cyb.1	chr16:4836913-4837913	396
MAP1LC3B	uc002fjx.1	chr16:85982320-85983320	250
SPG7	uc002fni.1,uc002fnj.1	chr16:88101306-88102306	271
ZNF276	uc002fos.2	chr16:88314453-88315453	322
ZMYND15	uc002fyt.1	chr17:4589086-4590086	270
KCTD11	uc002gge.2	chr17:7194932-7195932	232
MYH10	uc002gll.1,uc002glm.1,uc010cnx.1	chr17:8474761-8475761	239
ZSWIM7	uc002gpe.1,uc002gpf.1	chr17:15843731-15844731	263
CDK5R1	uc002hhn.1	chr17:27837218-27838218	392
SMARCD2	uc010dea.1	chr17:59273558-59274558	249
MFSD11	uc002jtd.2,uc010dhb.1,uc002jte.1	chr17:72244723-72245723	243
USP36	uc002jwa.1	chr17:74348204-74349204	266
CBX8	uc002jxd.1	chr17:75385485-75386485	250
RAC3	uc002kdf.1	chr17:77581821-77582821	334
RFNG	uc002kdj.1	chr17:77602939-77603939	234
ASXL3	uc010dmg.1,uc002kxq.2	chr18:29411539-29412539	247
SMAD2	uc002lcy.1	chr18:43710924-43711924	355
DOK6	uc002lkl.1	chr18:65218271-65219271	247
NETO1	uc002lkw.2,uc002lky.1	chr18:68685790-68686790	290
TXNL4A	uc010drg.1	chr18:75894903-75895903	309
ADAMTSL5	uc002ltd.1	chr19:1464019-1465019	247
PLK5	uc002ltq.1	chr19:1474958-1475958	332
SF3A2	uc002lvg.1	chr19:2186816-2187816	247
SAFB2	uc002mcd.1	chr19:5573938-5574938	289
NFIX	uc002mwd.1	chr19:12966584-12967584	230
CCNE1	uc002nsn.1	chr19:34993741-34994741	265
TSHZ3	uc002nsy.2	chr19:36532030-36533030	274
CEBPA	uc002nun.1	chr19:38485160-38486160	360
HNRNPUL1	uc010eho.1	chr19:46461826-46462826	239
SAPS1	uc002qjw.2	chr19:60461850-60462850	233
LOC284296	uc002qkz.2	chr19:60636112-60637112	234
SIRPA	uc002wfs.1,uc002wft.1	chr20:1822826-1823826	236
ADRA1D	uc002wkr.2	chr20:4177659-4178659	273
FOXA2	uc002wsm.1	chr20:22513101-22514101	238
C20orf24	uc002xfo.1	chr20:34635370-34636370	242
DBNDD2	uc002xoe.1	chr20:43467932-43468932	246
CTSA	uc002xqk.2	chr20:43952607-43953607	242
PARD6B	uc002xvo.1	chr20:48780488-48781488	266
LSM14B	uc010qjz.1	chr20:60130588-60131588	342
SLC2A4RG	uc002ygr.1	chr20:61841103-61842103	383
TIAM1	uc002yow.1	chr21:31853161-31854161	233
SFRS15	uc002ypd.1,uc002ype.1,uc010glu.1	chr21:32026133-32027133	249
KIAA1172	uc002ypg.1	chr21:32026133-32027133	249
AGPAT3	uc002zdv.1	chr21:44108544-44109544	332
MRPL40	uc002zpq.1	chr22:17799036-17800036	316
ZNRF3	uc003aeq.2	chr22:27608890-27609890	266
CARD10	uc003asx.1,uc003asy.1	chr22:36245170-36246170	285

Table S3.3

Top-1% mostly covered Saqqaq promoters virtually devoid of cytosine methylation.

Name	UCSC	Coordinates	Cov(PM)	M_s
TTLL10	uc001acv.1	chr1:1098146-1099146	76	0.105
IIP45	uc009vnj.1	chr1:12004477-12005477	53	0.113
HRIHFB2003	uc001avr.2	chr1:15158767-15159767	36	0.111
KIAA0445	uc001azv.2	chr1:17166285-17167285	49	0.122
ALPL	uc001beu.2	chr1:21749394-21750394	38	0.158
MAP3K6	uc001bnz.1	chr1:27560945-27561945	38	0.105
UBE2U	uc001dbn.1	chr1:64441078-64442078	58	0.138
GPR61	uc001dxy.2	chr1:109883017-109884017	27	0.148
HORMAD1	uc001evk.1,uc001evl.1,uc001evm.1	chr1:148959976-148960976	52	0.115
MUC1	uc001fhz.1	chr1:153428022-153429022	29	0.103
CACNA1E	uc001gox.1,uc009wxt.1	chr1:179967143-179968143	36	0.111
CNTN2	uc001hbs.1	chr1:203293356-203294356	44	0.114
SLC26A9	uc001hdo.1	chr1:204164137-204165137	27	0.185
OBSCN	uc001hsn.1,uc009xez.1	chr1:226461484-226462484	48	0.104
COLEC11	uc002qxz.1,uc002qya.1,uc002qyb.1,uc002qyc.1,uc010ewo.1	chr2:3619512-3620512	47	0.170
ALLC	uc010ewt.1	chr2:3682661-3683661	47	0.128
ADCY3	uc002rfr.2	chr2:24918040-24919040	55	0.109
ABCG5	uc002rtn.1,uc002rto.1,uc002rtp.1	chr2:43919462-43920462	29	0.207
ACTG2	uc010fex.1,uc002sjw.1,uc010fey.1	chr2:73972601-73973601	27	0.111
CYP27C1	uc002tod.2	chr2:127679813-127680813	58	0.103
RAB3GAP1	uc002tuj.1,uc010fnf.1	chr2:135525323-135526323	36	0.111
DKFZp434A012	uc010fnq.1	chr2:135525323-135526323	36	0.111
DGKD	uc010fyh.1,uc002vuj.1	chr2:233991626-233992626	26	0.115
KLHL30	uc002vxr.1	chr2:238713156-238714156	58	0.103
FLJ00133	uc002wak.2	chr2:241652388-241653388	60	0.117
C2orf85	uc010fzu.1	chr2:242459559-242460559	67	0.104
C3orf32	uc003bqz.1	chr3:8761726-8762726	34	0.118
VGLL4	uc010hdv.1	chr3:11585398-11586398	36	0.139
CAND2	uc003bxk.1	chr3:12831873-12832873	34	0.118
FGD5	uc003bzc.1	chr3:14834810-14835810	28	0.107
SLC6A20	uc003cow.1	chr3:45788835-45789835	28	0.179
TESSP2	uc003cqk.1	chr3:46850589-46851589	38	0.105
KIAA0540	uc010hjo.1	chr3:47020679-47021679	45	0.111
RNF123	uc010hky.1	chr3:49710905-49711905	25	0.120
NT5DC2	uc003dem.1	chr3:52537822-52538822	52	0.154
MORC1	uc003dxl.1	chr3:110319658-110320658	76	0.132
IGSF11	uc003eby.1,uc003ebz.1,uc010hqs.1	chr3:120347588-120348588	57	0.123
SLC41A3	uc003eii.1	chr3:127258308-127259308	39	0.154
C3orf56	uc003eji.1	chr3:128393664-128394664	76	0.118
ABTB1	uc010hsm.1	chr3:128877500-128878500	47	0.106
CPN2	uc003fts.1	chr3:195553350-195554350	40	0.125
TNK2	uc003fvv.1	chr3:197097073-197098073	27	0.148
KIAA1530	uc010ibv.1	chr4:1348865-1349865	32	0.156
TACC3	uc010ica.1	chr4:1701697-1702697	28	0.143
ZFYVE28	uc003gew.1	chr4:2312606-2313606	78	0.154
RGS12	uc003ghb.2	chr4:3383754-3384754	45	0.133
TBC1D14	uc003giu.2	chr4:7052696-7053696	41	0.122
ACOX3	uc003gle.1	chr4:8468626-8469626	48	0.125
GPRIN3	uc003hsm.1	chr4:90448184-90449184	28	0.107
C4orf32	uc003iah.2	chr4:113285002-113286002	38	0.105
SLC6A18	uc003jby.1	chr5:1277470-1278470	61	0.115
CLPTM1L	uc003jcg.1	chr5:1394975-1395975	51	0.137
ANKH	uc003jfl.2	chr5:14799111-14800111	29	0.172
LOC153328	uc003laz.1,uc003lba.2	chr5:135197264-135198264	28	0.107
CAMK2A	uc003lrt.1,uc003lru.1,uc010jhe.1	chr5:149649529-149650529	29	0.103
NAF1	uc010jhl.1,uc010jho.1	chr5:150423501-150424501	39	0.128
TNIP1	uc010jhp.1,uc010jhq.1	chr5:150423501-150424501	39	0.128
AK127224	uc003mil.1	chr5:177479705-177480705	83	0.108
KIAA0676	uc003mlk.1	chr5:179230101-179231101	38	0.105
IGnT2	uc010jol.1	chr6:10599442-10600442	68	0.103
AX747210	uc003nak.1	chr6:13403797-13404797	28	0.107
GRM4	uc003oio.1,uc010jvi.1	chr6:34138994-34139994	26	0.115
TCP11	uc003okd.2,uc003oka.2,uc003okb.2,uc003okc.2	chr6:35217165-35218165	118	0.127
TFEB	uc003oqu.1,uc003oqy.1	chr6:41811975-41812975	29	0.138
UNC84A	uc003sji.1	chr7:854683-855683	31	0.129
KIAA0810	uc003sjk.1	chr7:857410-858410	45	0.156
C7orf27	uc003smh.2	chr7:2548679-2549679	94	0.128
WIPI2	uc003snz.2,uc003soa.1	chr7:5219426-5220426	37	0.162
PON1	uc003uns.1	chr7:94791780-94792780	34	0.118
AZGP1	uc003ush.1	chr7:99411623-99412623	31	0.129
FLJ23834	uc003vdl.2,uc003vdm.2	chr7:105389921-105390921	27	0.148
ASZ1	uc003vjb.2	chr7:116854813-116855813	67	0.104
TCRBC2	uc003wbh.2	chr7:142203493-142204493	36	0.139
ARHGEF10	uc010lre.1	chr8:1817208-1818208	93	0.108
PTK2B	uc010luq.1,uc003xfr.1	chr8:27343310-27344310	42	0.119
CLU	uc003xfy.1	chr8:27525085-27526085	35	0.114
PTP4A3	uc003ywg.1,uc003ywh.1	chr8:142500189-142501189	85	0.106
hPRL-3	uc010met.1	chr8:142500189-142501189	85	0.106
SPATC1	uc003zaq.1	chr8:145157595-145158595	26	0.115
WNK2	uc004atl.1	chr9:95089891-95090891	26	0.115
CORO2A	uc004ayl.1	chr9:99974995-99975995	45	0.111
ANKS6	uc004ayv.1	chr9:100579557-100580557	29	0.241
GSN	uc004blg.1	chr9:123112570-123113570	26	0.115
UNQ6496	uc004brg.1	chr9:129296297-129297297	29	0.103
SH2D3C	uc004brz.2	chr9:129564518-129565518	37	0.216
ODF2	uc004bvf.1	chr9:130261650-130262650	26	0.115
KIAA0515	uc004cao.2	chr9:133334996-133335996	41	0.122
DBH	uc004cel.1	chr9:135490306-135491306	38	0.105
FLJ00280	uc004cfd.1	chr9:136447825-136448825	44	0.114
MAN1B1	uc004clf.1	chr9:139116621-139117621	41	0.122
LOC441476	uc004cmj.1,uc004cmk.1,uc004cml.1,uc010ncd.1	chr9:139264551-139265551	37	0.108
NELF	uc010nci.1	chr9:139470759-139471759	38	0.158

ERCC6	uc009xod.1	chr10:50354362-50355362	28	0.107
LDB3	uc001kdr.1,uc001kds.1,uc009xsy.1,uc001kdu.1,uc001kdv.1	chr10:88417301-88418301	31	0.161
KIAA0613	uc009xsz.1	chr10:88417301-88418301	31	0.161
KCNK18	uc001ldc.1	chr10:118945990-118946990	25	0.120
C10orf141	uc001lju.1,uc009yap.1	chr10:128864690-128865690	30	0.167
INPP5A	uc001llq.1	chr10:134270409-134271409	121	0.116
FLJ00228	uc001lov.2	chr11:282823-283823	48	0.104
EFCAB4A	uc001lrw.2	chr11:818297-819297	60	0.117
KIAA0899	uc009ycq.1	chr11:981000-982000	50	0.140
HCCA2	uc001lto.1,uc001ltp.1	chr11:1458691-1459691	65	0.108
TNNT3	uc001lun.2,uc001luu.2,uc001lup.2,uc001luw.2,uc001luo.2,uc001luq.2	chr11:1896375-1897375	53	0.113
C11orf21	uc009ydj.1,uc001lvv.1	chr11:2279866-2280866	53	0.113
APBB1	uc009yey.1,uc009yez.1,uc009yfa.1,uc001mdd.2,uc009yfb.1	chr11:6382668-6383668	49	0.122
NR1H3	uc001nej.1,uc001nek.1,uc001nel.1	chr11:47226043-47227043	46	0.109
TNKS1BP1	uc001njp.1,uc001njq.1	chr11:56827529-56828529	35	0.114
AB429224	uc009ypx.1	chr11:64417497-64418497	59	0.119
FLJ00043	uc001oep.1	chr11:65104695-65105695	34	0.118
CABP2	uc001omc.1,uc001omd.1,uc001ome.1	chr11:67047475-67048475	56	0.143
BC127192	uc009ysn.1	chr11:69995572-69996572	34	0.118
PLEKHB1	uc001ouc.1,uc001oud.1	chr11:73035330-73036330	37	0.108
CAPN5	uc001oya.1	chr11:76507101-76508101	45	0.156
FXYD2	uc001prj.2,uc001prl.2	chr11:117200669-117201669	50	0.120
IFFO1	uc001qoz.1,uc001qpa.1	chr12:6528535-6529535	50	0.120
HOM-TES-103	uc001qpg.1	chr12:6528535-6529535	50	0.120
LAG3	uc001qqs.2,uc001qqt.2,uc001qqu.2	chr12:6750931-6751931	30	0.133
C1RL	uc001qsn.1,uc009zft.1,uc001qso.1	chr12:7153069-7154069	25	0.120
STYK1	uc001qvs.1	chr12:10717906-10718906	64	0.125
COL2A1	uc009zkw.1	chr12:46668521-46669521	36	0.111
POU6F1	uc001rxy.1,uc001rxz.1	chr12:49877085-49878085	64	0.141
FIGNL2	uc001rzc.1	chr12:50502475-50503475	25	0.120
NR4A1	uc001rzr.2,uc009zmc.1	chr12:50722784-50723784	32	0.125
GEFT	uc009zpy.1,uc001soz.1	chr12:56289230-56290230	55	0.109
BC061638	uc009zrx.1	chr12:72972678-72973678	70	0.114
NOS1	uc001twm.1	chr12:116283965-116284965	26	0.192
MLXIP	uc001ubt.1	chr12:121182783-121183783	26	0.192
limkain	uc001ufv.1	chr12:122996061-122997061	35	0.114
KIAA1517	uc001ugz.1	chr12:124003069-124004069	27	0.222
PGAM5	uc001ukw.2	chr12:131800517-131801517	27	0.111
DNAJC15	uc001uyy.1	chr13:42494362-42495362	26	0.192
ADPRHL1	uc001vtq.1	chr13:113155840-113156840	47	0.106
ATP4B	uc001vtz.1,uc010agy.1	chr13:113360502-113361502	107	0.140
C14orf104	uc001wws.2,uc001wwt.2	chr14:49171698-49172698	26	0.154
PAPLN	uc010arm.1	chr14:72796170-72797170	38	0.105
KIAA1509	uc001xzi.1	chr14:90819720-90820720	25	0.120
SLC24A4	uc010auj.1	chr14:91974432-91975432	31	0.129
SLC25A29	uc010avv.1	chr14:99831824-99832824	26	0.115
PPP2R5C	uc001ykq.1	chr14:101417251-101418251	35	0.114
AHNAK2	uc001ypx.2	chr14:104507865-104508865	36	0.111
FAM82A2	uc001zmq.1	chr15:38835341-38836341	29	0.103
MAPKBP1	uc010bcl.1	chr15:39891006-39892006	28	0.107
CAPN3	uc001zpq.1	chr15:40480880-40481880	38	0.105
BCL2L10	uc002abq.1	chr15:50192264-50193264	38	0.105
TIPIN	uc002apr.1	chr15:64436108-64437108	57	0.105
NR2E3	uc002ath.1,uc002ati.1	chr15:69888948-69889948	30	0.133
SEMA4B	uc010bnv.1	chr15:88564850-88565850	34	0.118
CLCN7	uc002clu.2	chr16:1439799-1440799	140	0.107
IFT140	uc002clz.1	chr16:1548136-1549136	43	0.116
DNASE1	uc002cvr.1	chr16:3641941-3642941	34	0.118
JMJD5	uc010bxw.1	chr16:27127915-27128915	28	0.107
MYST1	uc002eba.2	chr16:31045011-31046011	33	0.121
ABCC12	uc002efe.1	chr16:46747430-46748430	54	0.130
SNX20	uc002egi.2,uc002egk.1	chr16:49272667-49273667	28	0.143
EDC4	uc002eus.1	chr16:66467893-66468893	38	0.105
BCAR1	uc002fdt.1	chr16:73827902-73828902	97	0.165
FAM38A	uc002flr.2	chr16:87330160-87331160	110	0.136
Mib	uc010cib.1	chr16:87330160-87331160	110	0.136
AK095785	uc002fpd.1	chr16:88496859-88497859	37	0.108
TUBB3	uc002fpk.1	chr16:88526821-88527821	47	0.149
DEF8	uc002fpq.1	chr16:88556749-88557749	53	0.132
ZZEF1	uc002fxq.1	chr17:3867623-3868623	27	0.111
ALOX15	uc002fyh.1	chr17:4491709-4492709	34	0.118
DHX33	uc002gbz.1	chr17:5305407-5306407	25	0.120
CACNB1	uc002hrl.1	chr17:34603662-34604662	34	0.118
CALCOCO2	uc002iof.1	chr17:44262371-44263371	66	0.106
CACNA1G	uc002irx.1,uc002iry.1,uc002irz.1,uc002isa.1,uc002isb.1,uc002isc.1,uc002isd.1 ,uc002ise.1,uc002isf.1,uc002isg.1,uc002ish.1,uc002isi.1	chr17:46000230-46001230	32	0.188
SDK2	uc010dfn.1	chr17:68945503-68946503	33	0.121
UBE2O	uc002jrl.2	chr17:71907597-71908597	25	0.120
AK125672	uc002jur.1	chr17:73644514-73645514	50	0.120
RNF213	uc002jyh.1	chr17:75927321-75928321	59	0.136
AK098403	uc010dhy.1	chr17:76701355-76702355	67	0.104
SLC38A10	uc002jzy.1	chr17:76874109-76875109	65	0.108
C17orf70	uc002kao.1	chr17:77125977-77126977	44	0.114
FN3K	uc002kfw.1	chr17:78287866-78288866	55	0.127
LOXHD1	uc002lcf.2	chr18:42393250-42394250	48	0.104
MOBKL2A	uc002luu.1	chr19:2029718-2030718	90	0.133
TMPRSS9	uc002lvv.1,uc002lvw.1	chr19:2339784-2340784	29	0.138
TBXA2R	uc002lye.1	chr19:3551665-3552665	27	0.111
FSD1	uc002maa.1	chr19:4261593-4262593	43	0.116
ALKBH7	uc002meo.1	chr19:6322444-6323444	29	0.172
KLF1	uc002mvo.1	chr19:12859017-12860017	28	0.143
CC2D1A	uc002mxq.1	chr19:13889317-13890317	41	0.146
INSL3	uc002nhm.1,uc010ebf.1,uc010ebg.1	chr19:17793320-17794320	31	0.129
KIAA0892	uc010ece.1	chr19:19318459-19319459	33	0.152
GRAMD1A	uc002nxi.1	chr19:40176528-40177528	35	0.114

47

NTF4	uc002pmf.2	chr19:54258936-54259936	31	0.161
KIAA0654	uc002pmt.1	chr19:54336873-54337873	29	0.103
ASPDH	uc002psr.2	chr19:55709759-55710759	72	0.222
KLK11	uc002pvf.1	chr19:56223102-56224102	27	0.148
SIGLEC12	uc002pwx.1	chr19:56696855-56697855	35	0.171
UNQ9215	uc010eoy.1	chr19:56696855-56697855	35	0.171
LILRA2	uc002qgg.2,uc010ero.1	chr19:59776110-59777110	26	0.154
KIAA1398	uc010qcm.1	chr20:17565374-17566374	42	0.119
SGK2	uc002xkv.1	chr20:41627151-41628151	33	0.121
TNFRSF6B	uc002yfy.1	chr20:61795538-61796538	45	0.178
AIRE	uc002zej.1,uc010gpr.1	chr21:44533576-44534576	62	0.113
APECED	uc010gpq.1	chr21:44533576-44534576	62	0.113
C21orf2	uc010gps.1	chr21:44578835-44579835	46	0.109
COL18A1	uc002zhk.1	chr21:45750842-45751842	45	0.111
ARVCF	uc002zra.1	chr22:18358335-18359335	31	0.129
KIAA0330	uc010gul.1	chr22:22896681-22897681	27	0.111
CRYBB1	uc003acy.1	chr22:25343991-25344991	33	0.121
CYB5R3	uc003bcw.2	chr22:41362840-41363840	27	0.111
PARVB	uc010gzn.1	chr22:42752517-42753517	37	0.108
TBC1D22A	uc003bif.1	chr22:45547488-45548488	26	0.115
C22orf34	uc003bit.2	chr22:48437194-48438194	45	0.111
TRABD	uc003bjv.1,uc003bjw.1	chr22:48972570-48973570	72	0.111
dJ402G11.5	uc010hap.1,uc003bjy.1	chr22:48985873-48986873	77	0.117
MAPK12	uc003bkp.1	chr22:49036709-49037709	41	0.146
FLNA	uc004fki.2	chrX:153235605-153236605	45	0.111

Table S3.4

Top-1% methylated Saqqaq promoters.

Ms(-2000;-1000)	Cov.	Ms(-1000;TSS)	Cov.	Ms(TSS;1000)	Cov.	Ms(1000;2000)	Cov.	TSS
0.1200	25	0.0000	25	0.0256	31	0.0909	36	chr16:66075217
0.0250	40	0.0000	26	0.0227	39	0.0800	54	chr11:60450096
0.0250	40	0.0000	26	0.0227	39	0.0800	54	chr11:60450096
0.0682	44	0.0000	27	0.0328	88	0.1111	34	chr12:107510288
0.0294	34	0.0000	29	0.0078	81	0.0588	29	chr11:62363707
0.0750	40	0.0000	29	0.0204	46	0.0784	44	chr14:23723809
0.0286	35	0.0000	29	0.0526	33	0.1111	33	chr1:203222168
0.0741	27	0.0000	29	0.0208	35	0.0667	37	chr2:96980965
0.0638	47	0.0000	30	0.0693	77	0.1429	33	chr1:202493620
0.0638	47	0.0000	30	0.0693	77	0.1429	33	chr1:202493620
0.0638	47	0.0000	30	0.0693	77	0.1429	33	chr1:202493620
0.0638	47	0.0000	30	0.0693	77	0.1429	33	chr1:202493620
0.0638	47	0.0000	30	0.0693	77	0.1429	33	chr1:202493620
0.0638	47	0.0000	30	0.0693	77	0.1429	33	chr1:202493620
0.0286	35	0.0000	31	0.0194	64	0.0758	85	chr19:4509501
0.0968	31	0.0000	31	0.0114	103	0.0513	27	chr19:53663423
0.1200	50	0.0000	31	0.0221	162	0.0579	76	chr7:719269
0.0571	35	0.0000	32	0.0208	45	0.0500	62	chr21:44612431
0.0278	36	0.0000	32	0.0278	38	0.0909	48	chr9:134850866
0.0364	55	0.0000	33	0.0213	53	0.0521	80	chr19:5885473
0.0492	61	0.0000	33	0.0323	40	0.0714	59	chr21:42578058
0.0714	42	0.0000	34	0.0222	43	0.1136	50	chr14:72797810
0.0667	45	0.0000	35	0.0120	56	0.0300	63	chr17:78634237
0.0294	34	0.0000	37	0.0417	34	0.1304	33	chr19:38052523
0.0294	34	0.0000	37	0.0417	34	0.1304	33	chr19:38052523
0.0294	34	0.0000	37	0.0417	34	0.1304	33	chr19:38052523
0.0200	50	0.0000	37	0.0146	185	0.0400	57	chr7:733813
0.0571	35	0.0000	38	0.0081	79	0.0588	27	chr10:103805922
0.0571	35	0.0000	38	0.0081	79	0.0588	27	chr10:103805922
0.0323	31	0.0000	38	0.0128	141	0.0400	31	chr15:81527110
0.0323	31	0.0000	38	0.0128	141	0.0400	31	chr15:81527110
0.1200	25	0.0000	38	0.0059	88	0.0339	49	chr19:1388362
0.0714	28	0.0303	38	0.0645	38	0.1321	50	chr8:21957128
0.0220	91	0.0000	40	0.0189	51	0.0758	65	chr16:66255916
0.0364	55	0.0000	41	0.0375	67	0.0851	80	chr11:1811609
0.0339	59	0.0000	41	0.0467	70	0.1071	59	chr9:129205140
0.0526	57	0.0000	42	0.1389	52	0.3333	28	chr19:56583009
0.0526	57	0.0000	42	0.1389	52	0.3333	28	chr19:56583009
0.0185	54	0.0000	44	0.0286	49	0.1000	39	chr12:50872051
0.0185	54	0.0000	44	0.0286	49	0.1000	39	chr12:50872051
0.0339	59	0.0000	44	0.0147	72	0.0625	45	chr12:54902002
0.0339	59	0.0000	44	0.0290	74	0.0845	67	chr19:5773817
0.0189	106	0.0000	44	0.0278	59	0.0909	44	chr1:55236204
0.0417	48	0.0000	44	0.0233	42	0.0789	48	chr21:43984728
0.0323	31	0.0000	44	0.0148	110	0.0536	74	chr2:45021540
0.0211	95	0.0000	45	0.0800	49	0.1818	47	chr11:125645049
0.0577	52	0.0000	45	0.0044	146	0.0449	80	chr19:19600739
0.0303	33	0.0000	46	0.0098	66	0.0714	27	chr11:62363552
0.0606	33	0.0192	46	0.0606	35	0.1538	25	chr3:127358216
0.0482	83	0.0000	47	0.0286	49	0.0814	71	chr2:128118155
0.0667	45	0.0000	47	0.0106	129	0.0959	69	chr4:969784
0.0667	45	0.0000	47	0.0106	129	0.0959	69	chr4:969784
0.0588	51	0.0000	48	0.0625	62	0.1553	81	chr16:73830059
0.1154	78	0.0000	48	0.0149	66	0.0645	48	chr16:88604030
0.0444	45	0.0000	48	0.0129	98	0.0380	64	chr19:1993675
0.0847	59	0.0204	48	0.0390	62	0.0833	30	chr7:150283848
0.0847	59	0.0204	48	0.0390	62	0.0833	30	chr7:150283848
0.0645	62	0.0000	49	0.0421	82	0.0860	82	chr22:16654433
0.0238	84	0.0000	50	0.0139	69	0.1000	49	chr12:120740759
0.0385	26	0.0000	50	0.0071	143	0.0588	34	chr16:630849
0.0814	86	0.0000	50	0.0077	106	0.0748	101	chr16:88510787
0.0909	55	0.0000	50	0.0303	48	0.2143	52	chr17:59361606
0.0625	32	0.0000	50	0.0035	156	0.0741	26	chr5:175724107
0.0625	32	0.0000	50	0.0035	156	0.0741	26	chr5:175724107
0.0345	58	0.0000	51	0.0175	63	0.0556	46	chr12:54901971
0.0345	58	0.0000	51	0.0175	63	0.0556	46	chr12:54901971
0.0345	58	0.0000	51	0.0175	63	0.0556	46	chr12:54901971
0.0345	58	0.0000	51	0.0175	63	0.0556	46	chr12:54901971
0.0789	38	0.0000	51	0.0056	115	0.1053	29	chr17:7078587
0.0789	38	0.0000	51	0.0056	115	0.1053	29	chr17:7078587
0.0233	86	0.0000	51	0.0278	57	0.0641	82	chr3:130760230
0.0385	26	0.0000	53	0.0066	169	0.0292	81	chr3:185579554
0.0385	26	0.0000	53	0.0066	169	0.0292	81	chr3:185579554
0.0385	26	0.0000	53	0.0066	169	0.0292	81	chr3:185579554
0.0385	26	0.0000	53	0.0066	169	0.0292	81	chr3:185579554
0.0800	25	0.0000	54	0.0155	118	0.0604	81	chr9:20612514
0.0800	25	0.0000	54	0.0155	118	0.0604	81	chr9:20612514
0.1304	69	0.0000	55	0.0128	75	0.1176	45	chr15:38848469
0.0976	41	0.0000	55	0.0252	106	0.0833	40	chr9:129517745
0.0606	33	0.0000	56	0.0083	185	0.0227	34	chr4:6961071
0.0233	86	0.0000	57	0.0079	175	0.0435	79	chr11:65413450
0.0625	32	0.0000	57	0.0056	108	0.0588	64	chr16:2194450
0.0625	32	0.0000	57	0.0056	108	0.0588	64	chr16:2194450
0.0625	32	0.0000	57	0.0056	108	0.0588	64	chr16:2194450
0.0625	32	0.0000	57	0.0056	108	0.0588	64	chr16:2194450
0.0625	32	0.0000	57	0.0056	108	0.0588	64	chr16:2194450
0.0759	79	0.0000	57	0.0244	67	0.0741	57	chr21:44596911
0.0759	79	0.0000	57	0.0244	67	0.0741	57	chr21:44596911
0.0625	64	0.0000	57	0.0069	124	0.4286	48	chr4:4279522
0.0270	37	0.0000	58	0.0070	83	0.0667	26	chr19:48791516

0.0893	56	0.0000	60	0.0156	56	0.1053	34	chr19:45773629
0.0893	56	0.0000	60	0.0156	56	0.1053	34	chr19:45773629
0.0893	56	0.0000	60	0.0156	56	0.1053	34	chr19:45773629
0.0103	97	0.0000	62	0.0118	95	0.0714	62	chr16:19440951
0.0103	97	0.0000	62	0.0118	95	0.0714	62	chr16:19440951
0.0465	43	0.0114	62	0.0294	40	0.0714	30	chr1:151966172
0.0465	43	0.0114	62	0.0294	40	0.0714	30	chr1:151966190
0.0465	43	0.0114	62	0.0294	40	0.0714	30	chr1:151966190
0.0465	43	0.0114	62	0.0294	40	0.0714	30	chr1:151966190
0.0588	34	0.0000	63	0.0060	100	0.0583	65	chr16:2194472
0.0189	53	0.0000	63	0.0068	94	0.0909	33	chr19:48791330
0.0400	25	0.0000	63	0.0118	94	0.0357	26	chr2:11211990
0.0250	40	0.0000	64	0.0171	100	0.0500	33	chr19:51941560
0.0250	40	0.0000	64	0.0171	100	0.0500	33	chr19:51941560
0.0323	31	0.0000	64	0.0189	47	0.0556	33	chr19:55062325
0.0323	31	0.0000	64	0.0189	47	0.0556	33	chr19:55062325
0.0093	107	0.0000	65	0.0244	88	0.1200	73	chr11:118796859
0.0429	70	0.0135	67	0.0400	46	0.1111	39	chr19:45774302
0.0517	58	0.0000	67	0.0227	56	0.0714	45	chr19:54094380
0.0517	58	0.0000	67	0.0227	56	0.0714	45	chr19:54094380
0.0128	78	0.0000	69	0.0073	184	0.0159	116	chr19:55524446
0.0256	78	0.0000	69	0.0151	164	0.0444	56	chr21:27260703
0.0313	160	0.0000	69	0.0291	153	0.1047	113	chr7:27152306
0.0208	48	0.0000	70	0.0079	83	0.0313	62	chr2:202948294
0.0208	48	0.0000	70	0.0079	83	0.0313	62	chr2:202948294
0.0208	48	0.0000	70	0.0079	83	0.0313	62	chr2:202948294
0.0789	38	0.0000	71	0.0036	149	0.0909	30	chr17:1412860
0.0789	38	0.0000	71	0.0036	149	0.0909	30	chr17:1412860
0.0789	38	0.0000	71	0.0036	149	0.0909	30	chr17:1412860
0.0698	43	0.0116	71	0.0263	56	0.0588	35	chr19:55012357
0.0698	43	0.0116	71	0.0263	56	0.0588	35	chr19:55012357
0.0213	47	0.0000	72	0.0050	126	0.0930	51	chr17:71648240
0.0649	77	0.0000	72	0.0333	55	0.1176	42	chr19:45773976
0.0342	117	0.0000	73	0.0053	153	0.0327	131	chr19:1221259
0.0230	87	0.0000	74	0.0123	124	0.1071	58	chr17:18102780
0.0230	87	0.0000	74	0.0123	124	0.1071	58	chr17:18102780
0.0323	31	0.0080	74	0.0179	65	0.0627	157	chr19:43570508
0.0323	31	0.0080	74	0.0179	65	0.0627	157	chr19:43570508
0.0286	35	0.0000	74	0.0213	40	0.0714	28	chr19:50517974
0.0667	45	0.0000	74	0.0034	164	0.0566	51	chr1:2451544
0.0645	31	0.0000	75	0.0083	70	0.0608	142	chr19:7650706
0.0238	42	0.0000	75	0.0125	155	0.0897	54	chr3:198241083
0.0238	42	0.0000	75	0.0125	155	0.0897	54	chr3:198241083
0.0667	45	0.0082	76	0.0263	60	0.0741	37	chr22:22423279
0.0806	62	0.0000	77	0.0050	228	0.0116	163	chr20:62151423
0.0217	92	0.0000	78	0.0073	177	0.0152	68	chr11:32413663
0.0217	92	0.0000	78	0.0073	177	0.0152	68	chr11:32413663
0.0217	92	0.0000	78	0.0073	177	0.0152	68	chr11:32413663
0.0217	92	0.0000	78	0.0073	177	0.0152	68	chr11:32413663
0.0217	92	0.0000	78	0.0073	177	0.0152	68	chr11:32413663
0.0690	58	0.0213	78	0.0441	91	0.1053	39	chr16:3008189
0.0625	80	0.0156	79	0.0253	92	0.0870	66	chr1:3596095
0.0625	80	0.0156	79	0.0253	92	0.0870	66	chr1:3596095
0.0625	80	0.0156	79	0.0253	92	0.0870	66	chr1:3596095
0.0091	110	0.0000	79	0.0143	78	0.0370	56	chr1:153412693
0.0091	110	0.0000	79	0.0143	78	0.0370	56	chr1:153412693
0.0091	110	0.0000	79	0.0143	78	0.0370	56	chr1:153412693
0.0250	40	0.0000	80	0.0085	79	0.0345	48	chr16:30569204
0.0606	33	0.0000	82	0.0091	61	0.0608	142	chr19:7650760
0.0606	33	0.0000	82	0.0091	61	0.0608	142	chr19:7650763
0.0400	25	0.0000	83	0.0028	176	0.0129	107	chr16:88313893
0.0400	25	0.0000	83	0.0028	176	0.0129	107	chr16:88313893
0.0400	25	0.0000	83	0.0028	176	0.0129	107	chr16:88313893
0.0400	25	0.0000	83	0.0028	176	0.0129	107	chr16:88313893
0.0308	65	0.0000	83	0.0056	102	0.0588	34	chr1:3763657
0.0308	65	0.0000	83	0.0056	102	0.0588	34	chr1:3763657
0.0455	44	0.0000	84	0.0033	153	0.0213	41	chr17:26741767
0.0370	27	0.0080	84	0.0128	134	0.0690	30	chr19:5671463
0.0385	52	0.0093	84	0.0287	126	0.1176	87	chr19:17978976
0.0071	141	0.0000	85	0.0227	89	0.0667	87	chr15:43192815
0.0071	141	0.0000	85	0.0227	89	0.0667	87	chr15:43192815
0.0233	86	0.0000	87	0.0090	139	0.0182	69	chr11:128068198
0.0233	86	0.0000	87	0.0090	139	0.0182	69	chr11:128068198
0.0215	93	0.0000	87	0.0056	144	0.0182	73	chr12:101874581
0.0286	70	0.0115	87	0.0244	59	0.0667	105	chr16:620012
0.0328	61	0.0064	87	0.0112	125	0.0909	30	chr16:2767298
0.0328	61	0.0064	87	0.0112	125	0.0909	30	chr16:2767298
0.0851	47	0.0177	87	0.0316	72	0.0690	30	chr16:87764663
0.0851	47	0.0177	87	0.0316	72	0.0690	30	chr16:87764663
0.0656	61	0.0094	87	0.0249	155	0.0645	41	chr7:44152238
0.0667	60	0.0093	88	0.0254	153	0.0556	43	chr7:44152253
0.0137	146	0.0000	88	0.0058	146	0.1081	106	chr8:70908179
0.0732	41	0.0000	89	0.0042	135	0.0112	68	chr19:50372842
0.0140	143	0.0000	89	0.0441	117	0.1250	86	chr5:141040593
0.0130	154	0.0000	89	0.0400	72	0.1395	84	chr5:176875428
0.0189	53	0.0000	90	0.0220	71	0.0882	47	chr17:18102595
0.0566	53	0.0068	90	0.0117	134	0.0417	40	chr18:55091605
0.0566	53	0.0068	90	0.0117	134	0.0417	40	chr18:55091605
0.1200	25	0.0000	91	0.0070	147	0.0417	32	chr5:179712921
0.0357	28	0.0000	92	0.0092	71	0.0244	33	chr16:4792455
0.0444	45	0.0000	93	0.0150	76	0.0370	25	chr16:2872318
0.0290	69	0.0000	94	0.0045	126	0.0300	74	chr11:549957
0.0087	115	0.0000	94	0.0769	64	0.1667	56	chr15:41449550
0.0087	115	0.0000	94	0.0769	64	0.1667	56	chr15:41449550
0.0087	115	0.0000	94	0.0769	64	0.1667	56	chr15:41449550
0.0087	115	0.0000	94	0.0769	64	0.1667	56	chr15:41449550
0.0909	44	0.0070	94	0.0212	145	0.0465	34	chr9:139152429

0.0909	44	0.0070	94	0.0212	145	0.0465	34	chr9:139152429
0.0909	44	0.0070	94	0.0212	145	0.0465	34	chr9:139152429
0.0909	44	0.0070	94	0.0212	145	0.0465	34	chr9:139152429
0.0909	44	0.0070	94	0.0212	145	0.0465	34	chr9:139152429
0.0652	46	0.0000	95	0.0127	56	0.1000	28	chr12:8075625
0.0652	46	0.0000	95	0.0127	56	0.1000	28	chr12:8075625
0.0270	37	0.0068	95	0.0192	97	0.0517	41	chr16:2194742
0.0270	37	0.0068	95	0.0192	97	0.0517	41	chr16:2194742
0.0417	48	0.0064	95	0.0139	110	0.0800	48	chr17:4656377
0.0417	48	0.0064	95	0.0139	110	0.0800	48	chr17:4656377
0.0110	91	0.0000	95	0.0058	223	0.0417	158	chr19:539892
0.0323	124	0.0000	95	0.0089	125	0.0270	87	chr6:10527783
0.0345	29	0.0121	96	0.0211	68	0.0423	53	chr10:134606054
0.0345	29	0.0121	96	0.0211	68	0.0423	53	chr10:134606054
0.0400	75	0.0182	96	0.0476	73	0.1429	50	chr2:27158968
0.0656	61	0.0000	96	0.0143	199	0.1053	45	chr2:233926891
0.0606	33	0.0051	98	0.0185	84	0.0789	30	chr16:88250710
0.0714	28	0.0000	98	0.0189	58	0.0787	64	chr17:7405042
0.0244	82	0.0000	99	0.0074	175	0.0212	139	chr20:22978301
0.0057	176	0.0000	99	0.0093	208	0.0684	155	chr22:49049005
0.0056	179	0.0000	99	0.0163	154	0.0330	137	chr4:5762176
0.0143	140	0.0000	99	0.0188	137	0.1000	99	chr8:70907958
0.0093	108	0.0000	100	0.0432	130	0.1270	85	chr11:75595222
0.0625	32	0.0000	100	0.0103	175	0.0217	33	chr18:490685
0.0333	30	0.0054	100	0.0114	139	0.0247	59	chr20:60228707
0.0333	30	0.0054	100	0.0114	139	0.0247	59	chr20:60228707
0.0333	30	0.0054	100	0.0114	139	0.0247	59	chr20:60228707
0.0333	30	0.0054	100	0.0114	139	0.0247	59	chr20:60228707
0.0625	48	0.0000	101	0.0109	61	0.1081	33	chr11:8058484
0.0769	26	0.0062	101	0.0205	89	0.0469	41	chr11:66914998
0.0250	40	0.0057	101	0.0107	200	0.0370	29	chr19:4861019
0.1143	35	0.0000	101	0.0063	147	0.0222	33	chr20:60743241
0.1023	88	0.0000	101	0.0095	163	0.1000	79	chr7:1510544
0.1023	88	0.0000	101	0.0095	163	0.1000	79	chr7:1510544
0.0392	51	0.0123	102	0.0248	136	0.0968	39	chr9:139068326
0.0392	51	0.0123	102	0.0248	136	0.0968	39	chr9:139068326
0.0244	41	0.0093	103	0.0166	267	0.0390	59	chr11:1366704
0.0244	41	0.0093	103	0.0166	267	0.0390	59	chr11:1366704
0.0244	41	0.0093	103	0.0166	267	0.0390	59	chr11:1366704
0.0244	41	0.0093	103	0.0166	267	0.0390	59	chr11:1366704
0.0244	41	0.0093	103	0.0166	267	0.0390	59	chr11:1366704
0.0227	44	0.0000	103	0.0080	167	0.0301	105	chr2:10359490
0.0227	44	0.0000	103	0.0080	167	0.0301	105	chr2:10359490
0.0185	162	0.0000	105	0.0333	99	0.0813	151	chr17:73634634
0.0313	32	0.0000	106	0.0247	49	0.0610	50	chr19:60845797
0.0303	33	0.0000	106	0.0385	33	0.1077	50	chr9:129536664
0.0426	47	0.0047	107	0.0221	82	0.0909	33	chr16:88250779
0.0233	86	0.0000	109	0.0086	159	0.0571	70	chr19:50373394
0.0103	97	0.0000	109	0.0179	70	0.0769	66	chr19:60588529
0.0476	42	0.0192	110	0.0309	106	0.0811	60	chr7:718659
0.0560	125	0.0084	110	0.0149	85	0.0400	66	chr9:135283855
0.0455	44	0.0000	111	0.0139	59	0.1429	50	chr17:2153563
0.0741	27	0.0000	111	0.0091	158	0.0685	62	chr8:145714008
0.0123	81	0.0000	113	0.0143	223	0.1250	43	chr2:100802044
0.0231	173	0.0000	114	0.0238	123	0.1176	114	chr6:34963797
0.0588	34	0.0049	115	0.0154	100	0.0488	37	chr16:87450875
0.0588	34	0.0049	115	0.0154	100	0.0488	37	chr16:87450875
0.0588	34	0.0049	115	0.0154	100	0.0488	37	chr16:87450875
0.0588	34	0.0049	115	0.0154	100	0.0488	37	chr16:87450875
0.0127	79	0.0059	116	0.0357	79	1.0000	34	chr13:26231964
0.0769	78	0.0000	116	0.0053	120	0.0323	50	chr16:87405843
0.0769	78	0.0000	116	0.0053	120	0.0323	50	chr16:87405843
0.0364	55	0.0053	117	0.0280	79	0.0882	46	chr16:29818647
0.0152	66	0.0068	117	0.0333	78	0.0769	42	chr19:40740589
0.0208	48	0.0060	118	0.0122	98	0.0357	41	chr19:1441164
0.0882	34	0.0099	121	0.0488	42	0.2000	25	chr19:63589467
0.0882	34	0.0099	121	0.0476	42	0.2222	25	chr19:63589467
0.0097	103	0.0000	121	0.0075	189	0.0255	133	chr20:22978141
0.0238	42	0.0052	123	0.0085	147	0.0263	42	chr5:577447
0.0169	59	0.0061	123	0.0168	132	0.1875	53	chr9:139565129
0.1000	50	0.0050	124	0.0126	147	0.0667	39	chr2:232499203
0.0244	82	0.0000	125	0.0048	152	0.1111	67	chr11:47164534
0.0244	82	0.0000	125	0.0048	152	0.1111	67	chr11:47164534
0.1200	25	0.0119	125	0.0223	99	0.1304	28	chr9:135140451
0.1200	25	0.0119	125	0.0223	99	0.1304	28	chr9:135140451
0.1200	25	0.0119	125	0.0223	99	0.1304	28	chr9:135140451
0.0159	126	0.0000	126	0.0160	168	0.0727	113	chr1:50661729
0.0244	82	0.0000	127	0.0181	154	0.0513	85	chr16:638363
0.0244	82	0.0000	127	0.0181	154	0.0513	85	chr16:638363
0.0244	82	0.0000	127	0.0181	154	0.0513	85	chr16:638363
0.0408	98	0.0000	129	0.0088	165	0.0526	65	chr11:65442303
0.0141	71	0.0053	129	0.0172	60	0.0690	51	chr14:74815098
0.0192	52	0.0000	130	0.0104	179	0.0541	67	chr11:72696316
0.0364	55	0.0000	130	0.0062	192	0.0488	45	chr22:48980534
0.0465	86	0.0059	131	0.0122	135	0.0704	82	chr11:617173
0.0667	30	0.0094	131	0.0181	135	0.0465	36	chr8:144887902
0.0612	49	0.0050	132	0.0274	61	0.1000	38	chr16:30680130
0.0612	49	0.0050	132	0.0274	61	0.1000	38	chr16:30680130
0.0612	49	0.0050	132	0.0274	61	0.1000	38	chr16:30680130
0.0612	49	0.0049	132	0.0290	61	0.0833	39	chr16:30680143
0.0102	197	0.0000	134	0.0040	240	0.0283	162	chr2:14690554
0.0200	50	0.0051	135	0.0207	131	0.0526	41	chr8:143482444
0.0200	50	0.0051	135	0.0207	131	0.0526	41	chr8:143482444
0.0196	51	0.0042	139	0.0180	128	0.0385	35	chr17:2186555
0.0196	51	0.0042	139	0.0180	128	0.0385	35	chr17:2186555
0.0192	52	0.0000	139	0.0043	159	0.2000	35	chr7:19123469

0.0148	135	0.0068	142	0.0420	136	0.0943	99	chr11:64878855
0.0342	146	0.0104	142	0.0556	85	0.1333	85	chr17:58111550
0.0210	143	0.0054	143	0.0244	83	0.0625	77	chr16:3036416
0.0248	121	0.0075	144	0.0202	165	0.0679	147	chr22:49332327
0.0349	86	0.0041	144	0.0385	77	0.1143	61	chr6:36915198
0.0568	88	0.0000	145	0.0051	134	0.1099	93	chr17:77602939
0.0566	106	0.0000	146	0.0064	121	0.1481	76	chr19:41296727
0.0690	29	0.0080	147	0.0168	74	0.0513	28	chr17:959074
0.0055	183	0.0000	148	0.0154	123	0.1429	105	chr19:47079285
0.0055	183	0.0000	148	0.0154	123	0.1429	105	chr19:47079285
0.0278	72	0.0047	149	0.0112	140	0.0278	53	chr16:626348
0.0278	72	0.0047	149	0.0112	140	0.0278	53	chr16:626348
0.0278	72	0.0047	149	0.0112	140	0.0278	53	chr16:626348
0.0278	72	0.0047	149	0.0112	140	0.0278	53	chr16:626348
0.0278	72	0.0047	149	0.0112	140	0.0278	53	chr16:626348
0.0296	135	0.0000	150	0.0441	99	0.1176	72	chr19:18196303
0.0392	102	0.0106	150	0.0161	154	0.0357	55	chr4:7095165
0.0050	201	0.0000	152	0.0066	177	0.0213	121	chr14:74814283
0.0149	67	0.0000	154	0.0069	110	0.0526	72	chr17:7194931
0.0357	56	0.0000	155	0.0162	142	0.0625	49	chr11:2907226
0.0632	95	0.0093	155	0.0198	98	0.1098	84	chr19:2046867
0.0632	95	0.0093	155	0.0198	98	0.1098	84	chr19:2046867
0.0732	41	0.0000	155	0.0065	99	0.0303	34	chr1:229539304
0.0732	41	0.0000	155	0.0065	99	0.0303	34	chr1:229539304
0.0732	41	0.0000	155	0.0065	99	0.0303	34	chr1:229539304
0.0625	64	0.0000	156	0.0053	124	0.0135	64	chr1:53565659
0.0263	38	0.0000	157	0.0385	27	0.0833	30	chr17:4589085
0.0167	299	0.0000	157	0.0278	159	0.0714	168	chr2:238737972
0.0306	98	0.0000	157	0.0127	159	0.0323	55	chr8:145661545
0.0164	61	0.0081	158	0.0179	66	0.0476	43	chr11:65781593
0.0659	91	0.0139	158	0.0280	116	0.0851	72	chr19:2047269
0.1023	88	0.0045	159	0.0157	150	0.0333	87	chr16:549422
0.0194	103	0.0050	159	0.0212	272	0.0522	109	chr16:88168676
0.0194	103	0.0050	159	0.0212	272	0.0522	109	chr16:88168676
0.0185	54	0.0000	159	0.0041	146	0.0465	63	chr4:3263552
0.0185	54	0.0000	159	0.0041	146	0.0465	63	chr4:3263552
0.1000	100	0.0052	160	0.0159	92	0.0417	72	chr14:104522660
0.0611	131	0.0057	160	0.0155	172	0.1020	87	chr8:10625432
0.0220	91	0.0047	162	0.0081	116	0.0260	92	chr1:163591641
0.0076	131	0.0000	162	0.0070	136	0.0769	78	chr5:172687330
0.0233	43	0.0079	162	0.0326	75	0.2143	44	chr9:139565354
0.0600	100	0.0082	164	0.0137	163	0.0357	107	chr16:778384
0.0600	100	0.0082	164	0.0137	163	0.0357	107	chr16:778384
0.0600	100	0.0082	164	0.0137	163	0.0357	107	chr16:778384
0.0288	104	0.0041	167	0.0089	121	0.0779	97	chr11:550779
0.0173	173	0.0063	167	0.0110	174	0.0256	103	chr7:126679664
0.0273	110	0.0127	168	0.0435	105	0.1071	66	chr11:63440892
0.0110	273	0.0000	172	0.0056	314	0.0252	180	chr18:75255759
0.0110	273	0.0000	172	0.0056	314	0.0252	180	chr18:75255759
0.0110	273	0.0000	172	0.0056	314	0.0252	180	chr18:75255759
0.0571	35	0.0034	172	0.0131	128	0.0303	91	chr1:37272431
0.0571	35	0.0034	172	0.0131	128	0.0303	91	chr1:37272431
0.0250	80	0.0039	176	0.0087	101	0.0882	91	chr17:76623096
0.0328	61	0.0000	177	0.0191	130	0.0612	54	chr11:2907170
0.0095	105	0.0041	177	0.0069	131	0.2222	48	chr11:34335378
0.0286	140	0.0108	177	0.0216	175	0.1000	74	chr16:579107
0.0313	64	0.0122	178	0.0184	107	0.0448	73	chr4:7095629
0.0313	288	0.0000	179	0.0254	323	0.0784	186	chr19:59666706
0.0556	72	0.0033	184	0.0069	264	0.0245	119	chr4:1232908
0.0556	72	0.0033	184	0.0069	264	0.0245	119	chr4:1232908
0.0093	216	0.0000	184	0.0046	201	0.0187	158	chr7:100331477
0.0093	216	0.0000	184	0.0046	201	0.0187	158	chr7:100331477
0.0093	216	0.0000	184	0.0046	201	0.0187	158	chr7:100331477
0.0093	216	0.0000	184	0.0046	201	0.0187	158	chr7:100331477
0.0565	124	0.0042	185	0.0179	159	0.0400	96	chr16:777622
0.0565	124	0.0042	185	0.0179	159	0.0400	96	chr16:777622
0.0565	124	0.0042	185	0.0179	159	0.0400	96	chr16:777622
0.0565	124	0.0042	185	0.0179	159	0.0400	96	chr16:777622
0.0293	205	0.0066	185	0.0789	125	0.3333	109	chr18:3001945
0.0167	60	0.0031	188	0.0455	55	0.3333	26	chr11:63815674
0.0690	29	0.0000	188	0.0128	194	0.0521	70	chr16:4106187
0.0042	236	0.0000	188	0.0063	186	0.0164	169	chr3:129689454
0.0385	26	0.0067	189	0.0221	130	0.0588	68	chr11:74818881
0.0079	253	0.0000	189	0.0250	168	0.0571	142	chr19:50373335
0.0029	339	0.0000	190	0.0723	201	0.2000	155	chr15:71830869
0.0078	256	0.0000	191	0.0260	169	0.0571	144	chr19:50373325
0.0087	230	0.0000	192	0.0408	164	0.1250	118	chr11:65442261
0.0101	99	0.0000	193	0.0068	194	0.0316	99	chr2:236741391
0.0980	51	0.0079	200	0.0186	104	0.1212	36	chr12:122201238
0.0032	308	0.0000	200	0.0435	198	0.0909	159	chr6:109882346
0.0299	67	0.0028	202	0.0417	84	0.1429	48	chr10:80672430
0.0125	321	0.0000	205	0.0189	180	0.0400	169	chr11:62251527
0.0260	77	0.0031	206	0.0526	63	0.1538	49	chr2:10505357
0.0345	87	0.0089	208	0.0175	118	0.0678	81	chr12:131193945
0.0194	206	0.0000	209	0.0145	143	0.0333	128	chr11:62251159
0.0274	73	0.0000	210	0.0112	80	0.0294	70	chr8:144969537
0.0274	73	0.0000	210	0.0112	80	0.0294	70	chr8:144969537
0.0076	131	0.0035	211	0.0147	129	0.0690	97	chr14:105027748
0.0076	131	0.0035	211	0.0147	129	0.0690	97	chr14:105027748
0.0076	131	0.0035	211	0.0147	129	0.0690	97	chr14:105027748
0.0076	131	0.0035	211	0.0147	129	0.0690	97	chr14:105027748
0.0104	289	0.0000	213	0.0726	211	0.2857	142	chr8:29262241
0.0194	103	0.0000	213	0.0087	179	0.0526	75	chr9:137938826
0.0769	26	0.0078	216	0.0149	44	0.1000	26	chr13:113286713
0.0182	165	0.0040	216	0.0115	130	0.0455	91	chr9:137531582
0.0182	165	0.0040	216	0.0115	130	0.0455	91	chr9:137531582
0.0426	235	0.0000	218	0.0040	250	0.0278	140	chr16:2958342

0.0426	235	0.0000	218	0.0040	250	0.0278	140	chr16:2958342
0.0442	113	0.0031	223	0.0079	226	0.0263	55	chr10:725606
0.0314	159	0.0077	225	0.0286	101	0.1515	105	chr16:30577743
0.0071	280	0.0000	229	0.0108	265	0.0244	179	chr6:108986718
0.0071	280	0.0000	229	0.0108	265	0.0244	179	chr6:108986718
0.0500	80	0.0089	234	0.0182	75	0.0671	275	chr14:102045325
0.0161	62	0.0026	238	0.0076	109	0.0204	58	chr16:57054705
0.0198	101	0.0047	240	0.0286	103	0.1250	37	chr16:1816225
0.0192	104	0.0047	243	0.0286	106	0.1250	40	chr16:1816230
0.0192	104	0.0047	243	0.0286	106	0.1250	40	chr16:1816230
0.0071	141	0.0031	244	0.0088	205	0.0769	80	chr8:140784481
0.0071	141	0.0031	244	0.0088	205	0.0769	80	chr8:140784481
0.0198	202	0.0000	250	0.0060	270	0.0313	160	chr16:970808
0.0160	125	0.0069	301	0.0179	196	0.0700	113	chr17:77649395
0.0132	228	0.0000	307	0.0122	152	0.0625	128	chr10:133970706

Table S3.5

List of promoters presenting a methylation profile matching that of GC skew promoters.

Cov. = Coverage = N1+N2+M1+M2 (see Figure S3.3). M_s (-2000;-1000): 1kb windows starting 2kb upstream of TSS. M_s (-1000;TSS): 1kb windows ending at TSS. M_s (TSS;1000): 1kb windows starting at TSS. M_s (1000;2000): 1kb windows starting 1kb downstream of TSS.

Gene	Entry	Coordinates	M_s	Cov(1st Exon)
MORN3	uc001uax.1	chr12:120591628-120591943	0.240	25
HNRNPK	uc004ank.2	chr9:85784854-85784917	0.240	25
PINK1	uc001bdn.1	chr1:20844588-20844803	0.231	26
THSD4	uc002atg.1	chr15:69837226-69837468	0.200	35
KRT13	uc002hwu.1	chr17:36914834-36915391	0.200	25
KRT13	uc002hwv.1	chr17:36914834-36915391	0.200	25
KRT13	uc010cxo.1	chr17:36914834-36915391	0.200	25
HIP1R	uc001udk.1	chr12:121906426-121906663	0.194	31
CST7	uc002wtx.1	chr20:24877866-24878211	0.194	31
FLJ00251	uc001mdw.2	chr11:6524472-6525472	0.186	43
ATP8B3	uc002lty.1	chr19:1756887-1757159	0.185	27
9-Sep	uc002jtw.1	chr17:72883760-72884075	0.182	33
TP53	uc002gii.1	chr17:7519096-7519536	0.179	28
TP53	uc010cnf.1	chr17:7519096-7519536	0.179	28
TP53	uc010cng.1	chr17:7519096-7519536	0.179	28
UBXN6	uc010dty.1	chr19:4404927-4405090	0.179	28
C10orf108	uc001ifr.2	chr10:686017-686382	0.175	40
CDC42BPB	uc001ymj.1	chr14:102500592-102500740	0.172	29
PTPRS	uc002mbz.1	chr19:5237061-5237399	0.167	42
BAGE2	uc002yit.1	chr21:10120574-10120796	0.167	36
CR617046	uc001myp.2	chr11:45187118-45187422	0.162	74
KIAA1335	uc002xjz.1	chr20:39475378-39477119	0.161	31
CPNE4	uc003eom.1	chr3:133486774-133486944	0.161	31
MYO1H	uc009zvh.1	chr12:108310871-108311032	0.161	31
DNAJC14	uc001shu.1	chr12:54507303-54508765	0.160	50
KIAA0676	uc003mlk.1	chr5:179229861-179230101	0.160	25
BAGE5	uc002yiu.1	chr21:10120594-10120808	0.159	44
BAGE	uc002yiv.1	chr21:10120594-10120808	0.159	44
GRM2	uc003dbp.1	chr3:51717904-51718489	0.158	57
SH2D4B	uc001kck.1	chr10:82287638-82288251	0.157	70
MICAL2PV1	uc001mkb.2	chr11:12140202-12140542	0.156	32
MICAL2PV2	uc001mkc.2	chr11:12140202-12140542	0.156	32
AKT2	uc002one.1	chr19:45440281-45441049	0.156	32
KIAA0676	uc003mlf.1	chr5:179225444-179226478	0.156	32
MED15	uc002zst.1	chr22:19257965-19259519	0.155	71
NOC2L	uc001aby.2	chr1:882342-883781	0.154	26
KANK4	uc001daf.2	chr1:62501352-62501684	0.154	26
DPH1	uc002ftv.1	chr17:1889009-1889867	0.154	26
LZTR1	uc002ztp.1	chr22:19680252-19680401	0.154	26
MOV10L1	uc003bjj.1	chr22:48870621-48870741	0.154	26
MOV10L1	uc003bjk.2	chr22:48870621-48870741	0.154	26
CNTN2	uc009xbi.1	chr1:203293932-203294414	0.154	26
LZTS1	uc003wzr.1	chr8:20156628-20157083	0.151	53
LZTS1	uc010ltg.1	chr8:20156628-20157083	0.151	53
APBB1	uc001mda.2	chr11:6383193-6383465	0.150	40
APBB1	uc001mde.2	chr11:6383193-6383465	0.150	40
ACACB	uc001toc.1	chr12:108061585-108062246	0.148	27
GRK6	uc003mgs.1	chr5:176791318-176791714	0.148	27
MID1	uc004ctm.1	chrX:10494928-10495643	0.148	27
MID1	uc004ctn.1	chrX:10494928-10495643	0.148	27
MID1	uc004ctt.2	chrX:10494928-10495643	0.148	27
MID1	uc004ctu.2	chrX:10494928-10495643	0.148	27
MID1	uc004cub.1	chrX:10494928-10495643	0.148	27
MID1	uc010ndy.1	chrX:10494928-10495643	0.148	27
TGM2	uc002xhq.1	chr20:36194156-36195576	0.147	68
ATP13A1	uc002nnf.2	chr19:19626928-19627260	0.146	41
KIAA1640	uc002jxn.2	chr17:75669837-75670452	0.146	48
SBSN	uc002oad.1	chr19:40709386-40711052	0.146	48
pp9964	uc001any.1	chr1:6567949-6569434	0.143	49
SLC24A4	uc001yan.1	chr14:91988863-91990166	0.143	28
RFNG	uc002kdh.1	chr17:77601083-77601679	0.143	28
TSKS	uc002ppm.1	chr19:54958147-54958327	0.143	28
ZC3H12D	uc003qmn.1	chr6:149837068-149837442	0.143	63
BC075797	uc003ssl.1	chr7:12562712-12562881	0.143	28
GRIK4	uc009zax.1	chr11:120036188-120036319	0.143	28
CACNG5	uc010deq.1	chr17:62303810-62304108	0.143	28
NBEAL2	uc010hjm.1	chr3:47011526-47012127	0.143	63
CYP8B1	uc010hif.1	chr3:42890966-42892394	0.140	50
GPRC5D	uc001rbc.1	chr12:12993691-12994585	0.139	36
CNTN2	uc001hbs.1	chr1:203294356-203295044	0.138	29
MYST1	uc002eaz.1	chr16:31044456-31045910	0.138	29
HDGF2	uc002maq.1	chr19:4450488-4450716	0.138	29
NFAM1	uc003bcn.2	chr22:41158187-41158345	0.138	29
C7orf46	uc003swo.2	chr7:23686274-23686427	0.138	29
C7orf46	uc003swq.2	chr7:23686274-23686427	0.138	29
C7orf46	uc003swr.2	chr7:23686274-23686427	0.138	29
HDGF2	uc010dua.1	chr19:4450488-4450716	0.138	29
TG	uc010mdw.1	chr8:133979076-133979213	0.138	29
ESPNL	uc002vxq.2	chr2:238673690-238674093	0.137	102
ETV6	uc001raa.1	chr12:11913625-11914170	0.136	44
HNRNPK	uc004anf.2	chr9:85784888-85785004	0.136	44
HNRNPK	uc004ang.2	chr9:85784888-85785004	0.136	44
HNRNPK	uc004anh.2	chr9:85784888-85785004	0.136	44

HNRNPK	uc004ani.2	chr9:85784888-85785004	0.136	44
HNRNPK	uc004anj.2	chr9:85784888-85785004	0.136	44
MST066	uc002dmw.1	chr16:24849606-24850226	0.135	37
ARHGAP17	uc002dmx.1	chr16:24849606-24850226	0.135	37
MST066	uc002dne.1	chr16:24849606-24850226	0.135	37
TACC1	uc003xmh.2	chr8:38796197-38797310	0.135	37
TACC1	uc010lwq.1	chr8:38796197-38797310	0.135	37
DKFZp761P19121	uc003ejh.1	chr3:128227918-128229619	0.135	89
BCR-ABL	uc010gtx.1	chr22:21925986-21926167	0.135	52
C13orf35	uc001vsh.1	chr13:112349359-112349913	0.133	30
INPP4A	uc002syy.1	chr2:98502841-98503049	0.133	30
SLC22A7	uc003ous.2	chr6:43373976-43374467	0.133	30
SLC22A7	uc003out.1	chr6:43373976-43374467	0.133	30
CAMTA2	uc010ckv.1	chr17:4823703-4824640	0.133	30
SLC22A7	uc010jyl.1	chr6:43373976-43374467	0.133	30
SLC22A7	uc010jym.1	chr6:43373976-43374467	0.133	30
ARHGEF10L	uc001baq.1	chr1:17814170-17815284	0.132	53
FLJ00133	uc002wak.2	chr2:241653388-241653684	0.132	38
hCG_31249	uc004bzl.2	chr9:132295490-132295812	0.131	61
GPSM1	uc004che.1	chr9:138369803-138370119	0.130	46
AMPD2	uc001dyd.1	chr1:109969572-109969937	0.129	31
PPFIA4	uc001gyz.1	chr1:201286934-201287628	0.129	31
BLK	uc003wua.1	chr8:11455251-11456370	0.129	31
PRKAG2	uc003wkj.1	chr7:151142435-151142890	0.128	39
HBE269	uc010ncg.1	chr9:139409576-139409677	0.128	39
CRTAC1	uc001kot.1	chr10:99685917-99686113	0.127	55
limkain	uc001ufv.1	chr12:122993847-122996061	0.126	95
FLAD1	uc001fgh.1	chr1:153227157-153227949	0.125	32
ACP2	uc001nei.1	chr11:47226803-47226939	0.125	32
MYO7A	uc001oyd.1	chr11:76563292-76563608	0.125	32
BEAN	uc002eoq.1	chr16:65061006-65061269	0.125	40
CBFA2T3	uc002fmm.1	chr16:87570566-87570902	0.125	32
CBFA2T3	uc002fmn.1	chr16:87570566-87570902	0.125	32
SERPINB6	uc003muk.1	chr6:2904401-2906560	0.125	40
AGAP3	uc003wjj.1	chr7:150462420-150462588	0.125	32
ACP2	uc009ylj.1	chr11:47226803-47226939	0.125	32
ACP2	uc009ylk.1	chr11:47226803-47226939	0.125	32
dJ402G11.5	uc010hap.1	chr22:48986873-48987076	0.125	48
MYT1	uc002yij.1	chr20:62309391-62310284	0.123	57
MYO7A	uc009yut.1	chr11:76568429-76568647	0.123	57
TCF3	uc002lto.1	chr19:1573053-1573222	0.122	41
MEGF5	uc003mac.1	chr5:168204167-168204434	0.122	41
ZNF687	uc009wmp.1	chr1:149525375-149527506	0.122	82
SMARCA4	uc010dxt.1	chr19:10984438-10984788	0.122	41
PAX8	uc002tjo.1	chr2:113710649-113711555	0.121	33
TESSP2	uc003cqj.1	chr3:46850205-46850317	0.120	25
FGF17	uc003xai.1	chr8:21958129-21959748	0.120	75
KIAA1542	uc009ybz.1	chr11:582035-582674	0.119	42
KRT74	uc001sap.1	chr12:51253358-51253876	0.118	34
CATSPER2	uc001zsh.1	chr15:41728119-41728331	0.118	34
CATSPER2	uc001zsi.1	chr15:41728119-41728331	0.118	34
CATSPER2	uc001zsk.1	chr15:41728119-41728331	0.118	34
SMARCA4	uc002mqg.1	chr19:10955800-10956049	0.118	34
SMARCA4	uc002mqj.2	chr19:10955800-10956049	0.118	34
MED15	uc002zss.1	chr22:19238526-19239435	0.118	51
LYPD2	uc003ywz.1	chr8:143830814-143830954	0.118	34
SMARCA4	uc010dxq.1	chr19:10955800-10956049	0.118	34
SMARCA4	uc010dxr.1	chr19:10955800-10956049	0.118	34
SMARCA4	uc010dxs.1	chr19:10955800-10956049	0.118	34
pp9320	uc003yzv.2	chr8:144993669-144994190	0.117	120
CTBP1	uc003gct.1	chr4:1221971-1222132	0.116	43
SLC6A19	uc003jbw.2	chr5:1254710-1254967	0.116	43
RALGDS	uc004ccn.1	chr9:134966728-134967125	0.116	43
TMEM132A	uc001nqm.1	chr11:60458536-60458812	0.116	95
FBLIM1	uc001axi.1	chr1:15964046-15964315	0.115	26
DISC1	uc001hvc.2	chr1:229896199-229897174	0.115	78
MCF2L	uc001vst.1	chr13:112746970-112747203	0.115	26
CLK3	uc002ayl.2	chr15:72703014-72704412	0.115	26
SPATA22	uc002fvo.1	chr17:3321726-3321831	0.115	26
ZNF335	uc002xqv.1	chr20:44015735-44016750	0.115	26
TRIM7	uc003mmv.1	chr5:180559457-180560536	0.115	26
IRF4	uc003mtc.1	chr6:342056-342252	0.115	26
ANKRD6	uc003pnh.2	chr6:90369431-90369565	0.115	26
GLIS3	uc003zic.1	chr9:4276038-4276996	0.115	26
HPCAL1	uc010exf.1	chr2:10426341-10426794	0.115	26
NM_001127438	uc010nmy.1	chrX:99081047-99081497	0.115	26
SLC22A1	uc003qtc.1	chr6:160462853-160463368	0.114	35
SLC22A1	uc003qtd.1	chr6:160462853-160463368	0.114	35
EP400NL	uc001ujv.2	chr12:131154495-131155802	0.114	44
UNQ830	uc002vtr.1	chr2:233443238-233443368	0.114	44
KIF1A	uc002vzx.1	chr2:241313361-241314088	0.114	44
TBC1D24	uc002cqm.1	chr16:2486036-2487115	0.112	188
PIWIL4	uc001pfa.1	chr11:93940182-93940419	0.111	36
AQP2	uc001rvn.1	chr12:48630796-48631240	0.111	63
MBD6	uc001sok.1	chr12:56205398-56206438	0.111	36
ACADS	uc001tzb.2	chr12:119659157-119659321	0.111	27

RHBDL1	uc002cis.1	chr16:666076-666498	0.111	27
ARHGAP17	uc002dna.1	chr16:24873425-24876358	0.111	54
EFCBP2	uc002fhe.1	chr16:82563017-82563281	0.111	27
RHBDF2	uc002jrs.1	chr17:71989052-71989222	0.111	27
MAP4K1	uc002oix.1	chr19:43800277-43800483	0.111	27
MAP4K1	uc002oiy.1	chr19:43800277-43800483	0.111	27
SGK2	uc002xkq.1	chr20:41602205-41602362	0.111	27
CARMA3	uc003asv.1	chr22:36220039-36220529	0.111	27
RHO	uc003emt.1	chr3:130730172-130730627	0.111	27
AMOTL2	uc003eqf.1	chr3:135572232-135573026	0.111	72
SLC26A1	uc003gbx.1	chr4:977089-977224	0.111	27
SLC26A1	uc003gcb.1	chr4:977089-977224	0.111	27
SLC26A1	uc003gcc.1	chr4:977089-977224	0.111	27
KIAA0294	uc003wpu.2	chr8:1795470-1795757	0.111	36
ATP2B3	uc004fhs.1	chrX:152454774-152455107	0.111	27
ATP2B3	uc004fht.1	chrX:152454774-152455107	0.111	27
KIAA1933	uc010cfd.1	chr16:66943484-66943813	0.111	27
ASPDH	uc010enz.1	chr19:55708841-55708954	0.111	36
PIWIL2	uc010ltv.1	chr8:22189025-22189487	0.111	63
AK123118	uc001uke.1	chr12:131362142-131362495	0.109	55
CALML4	uc002arb.1	chr15:66284637-66285502	0.109	46
CALML4	uc002arc.1	chr15:66284637-66285502	0.109	46
SH3BP2	uc003gfm.2	chr4:2797850-2798843	0.109	46
NCOR2	uc001ugh.2	chr12:123545646-123545867	0.108	37
MAP3K10	uc002onb.1	chr19:45396122-45396302	0.108	37
CTCFL	uc002xym.2	chr20:55533490-55533589	0.108	37
CTCFL	uc010giw.1	chr20:55533490-55533589	0.108	37
CTCFL	uc010giy.1	chr20:55533490-55533589	0.108	37
CTCFL	uc010giz.1	chr20:55533490-55533589	0.108	37
CTCFL	uc010gjf.1	chr20:55533490-55533589	0.108	37
CTCFL	uc010gjg.1	chr20:55533490-55533589	0.108	37
CTCFL	uc010gji.1	chr20:55533490-55533589	0.108	37
TNKS1BP1	uc001njp.1	chr11:56826509-56827529	0.108	65
TNKS1BP1	uc001njq.1	chr11:56826509-56827529	0.108	65
SDK1	uc003smy.1	chr7:4135843-4136255	0.108	65
PMPCA	uc004chm.1	chr9:138430340-138431487	0.108	65
C1orf167	uc001ata.2	chr1:11762446-11762654	0.107	28
C18orf22	uc002lnt.1	chr18:75902290-75903207	0.107	28
MCM5	uc003anw.1	chr22:34136402-34136903	0.107	28
GOLGA2	uc004buh.1	chr9:130062565-130062902	0.107	28
NLRC3	uc010btn.1	chr16:3567151-3567393	0.107	28
IGSF11	uc003eby.1	chr3:120347350-120347588	0.106	47
IGSF11	uc003ebz.1	chr3:120347350-120347588	0.106	47
UBAP2	uc003ztn.1	chr9:33917795-33918833	0.106	47
IGSF11	uc010hqs.1	chr3:120347350-120347588	0.106	47

Table S3.6

Top-1% Saqqaq genes showing highest methylation levels in the first exon.

3.6. Predicting the Saqqaq age at death using age-dependent CpG signatures

DNA methylation levels have been shown to change with age in human tissues [Alish et al. 2012; Day et al. 2013; Johansson et al. 2013]. In this section, we explored whether ancient methylation levels as measured through M_s in the Saqqaq sequence data could be used to propose an age at death for the Saqqaq individual. Recently, Koch and Wagner [2011] proposed to use methylation levels at four CpG sites associated with the genes TRIM58, KCNQ1DN, NPTX2 and GRIA2 (cg07533148, cg01530101, cg1279989, cg25148589, respectively) and at one additional hypomethylated CpG site (BIRC4BP, cg23571857) to predict the age of modern human individuals. This framework enabled relatively precise predictions, with average differences between predicted and real age of about 11 years and could potentially be used to propose a range for the age at death of the Saqqaq individual. As cg1279989 showed no overlap with the Illumina 450k array data available from [Siekler et al. 2013], further analyses were restricted to the other four loci. As a positive control, we first used the linear regressions provided in [Koch and Wagner 2011] in order to predict the age of five donor individuals (PT1, PT2, PT3, PT4 and PT5) from their methylation levels at each four loci. Such linear models have been built using methylation levels from two types of tissues (dermis and epidermis) and three cell lines (cervical smear cells, monocytes and T cells) and are likely robust across a large range of somatic tissues. We therefore hypothesized that the linear model was correct for hairs. Predicted ages are plotted in Figure S3.14 together with the known age for each donor individual. Overall, predictions based on observed methylation levels at cg07533148 and cg01530101 were consistent with each other, providing age estimates within 1.7-12.4 years (standard deviation = 5.7 years) around the real age of donor individuals. This is in agreement with the accuracy levels reported in Koch and Wagner [2011]. Inconsistent values showing large differences between predicted and real age values (16.1-39.4 years, standard deviation = 9.4 and 12.4 years, respectively) were found for the other two loci. Consequently, we decided to base our predictions for the age at death of the Saqqaq individual on the first two CpGs only (cg07533148 and cg0153010).
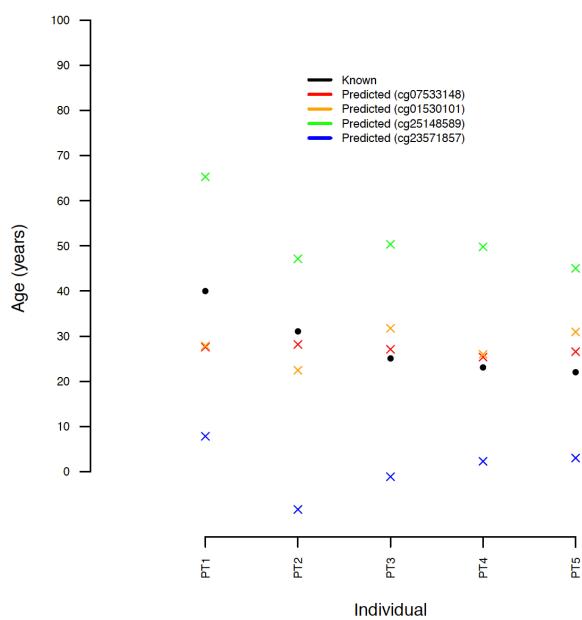
Figure S3.14

Predictions for the age of five modern human individuals based on methylation levels at four CpG sites.
Four CpG sites showing age-dependent methylation levels have been considered (cg0733148, red; cg01530101, orange; cg25148589, green, and cg23571857, blue; [Koch and Wagner 2011]). We predicted (coloured crosses) the age of five modern human individuals (PT1, PT2, PT3, PT4 and PT5) for which methylation information was available and compared our predictions to the known age of each individual (black).

We then used the procedure described in Supplemental Section SI3.4 in order to calculate absolute methylation levels for the Saqqaq at these two loci. Briefly, we relied on a linear model relating M_s values (as calculated within a 2,000 bp-wide region centered on each CpG site) and hair methylation levels measured in any of the five modern human individuals. In this first linear model, different CpG subsets were considered depending on their coverage (a full range of possible coverage was considered, spanning every 50 counts from 100 to 500). Methylation levels were predicted for the Saqqaq individual using the predict() function in R [R Development Core Team 2012] and a confidence level of 0.999. Saqqaq methylation values as predicted from the first linear model were further converted into absolute age using a second linear model, namely the linear regression provided in Koch and Wagner [2011]. Predictions are provided in Figure S3.15 for the full range of coverage investigated. The age at death is predicted at around 44.1-69.4 and 52.1-64.2 years for both CpG considered. Given error margins estimated with modern human donors (shifted by up to 12.4 years), these values should not be taken as precise point estimates but rather as an indication that the Saqqaq individual was most likely old when he died. We should also insist that our estimate depend on a number of strong assumptions that remain to be investigated with further studies: 1) the age-dependency of methylation levels at cg07533148 and cg0153010 has remained constant in recent modern human evolution; 2) the age-dependency of methylation levels at cg07533148 and cg0153010 as described in Koch and Wagner [2011] is robust across a range of human populations and environments.
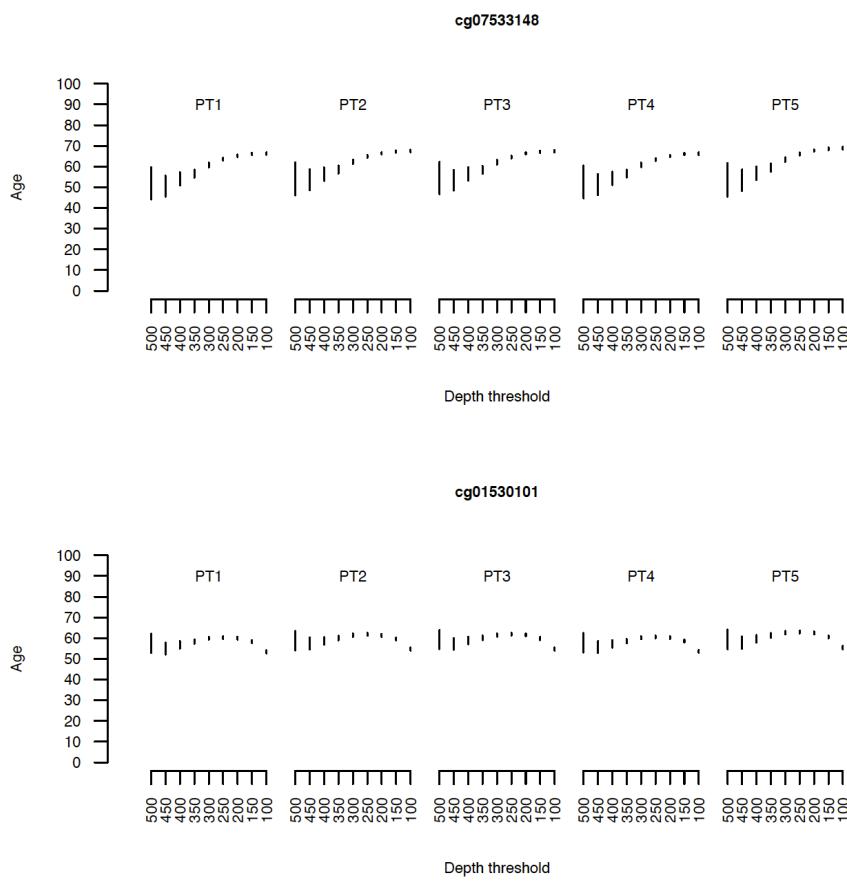
Figure S3.15

Predictions for the age at death of the Saqqaq individual based on inferred methylation levels at two CpG sites. We predicted the age at death of the Saqqaq individual using on one hand the methylation levels estimated at two CpG sites (cg07533148 and cg0153010) and on the other hand the linear relationship between methylation levels and age as reported in [Koch and Wagner 2011]. Different subsets of CpGs showing a range of minimal coverage were considered to estimate the Saqqaq methylation levels based on M_s values.

3.7. Methylation profiles across nucleosomes

CTCF (CCCTC-binding Factor) sites play a critical role in shaping the 3D architecture of our genome [Bell and Felsenfeld 2000]. An overall number of 15,000-40,000 of CTCF sites showing quite permissive consensus sequences of 20 nucleotides are interspersed in the human genome [Kim et al. 2007]. Recruitment of the CTCF protein at CTCF sites induces strong chromatin remodelling, positionning arrays of 10 nucleosomes over the ~2kb flanking genomic regions. Within those regions, nucleosomes are strongly phased, showing footprints of ca. 110-170 nucleotides (depending on the nature of histone modifications; 118 bp for mononucleosome data) and are separated by ca. 0-60 bp long spacers (33 bp for mononucleosome data) [Fu et al. 2008].

We retrieved hg18 genomic coordinates of 12,864 CTCF 20-mer binding sites from the Supplemental Table 1 provided in Fu and colleagues [Fu et. al 2008]. We next calculated average M_s values for each 25bp window (1bp increment) within 1kb upstream and downstream of the CTCF binding site and compared these to nucleosome positioning represented by GC-corrected read depth (Figure 4.d). Overall, peaks in methylation levels coincided with sites of low nucleosome occupancy in agreement with independent empirical evidence in modern human cells [Kelly et al. 2012]. We also recovered a 63bp large footprint for the CTCF protein using the central region with lowest methylation levels (<0.024), as methylation is known to suppress CTCF DNA-binding [Bell et al. 2001].

We investigated whether the greater power to detect methylation in positions with greater read depth could be responsible for the pattern observed at CTCF sites. We therefore performed the same analysis by down-sampling sites across the full region. Down-sampling was performed for each position located 1,000 nucleotides upstream and downstream of CTCF binding sites in order to represent each position with similar depth-of-coverage (from 20 to 100, iterating every 10). We generated 100 down-sampled datasets (per depth-of-coverage investigated), calculated M_s at each position then tested for correlation with nucleosome positioning. The results are provided in Table S3.7. While stronger correlation were found for datasets down-sampled at higher depth, we found a significantly negative correlation between M_s and nucleosome occupancy for all dataset investigated (-0.0896 ≤ Pearson correlation coefficient ≤ -0.254 ; p-value ≤ 6.05x10^{-5}). This clearly demonstrates that the greater power to detect methylation in positions with greater read depth does not preclude our ability to detect correct methylation and nucleosome positioning signal at CTCF sites.

We then profiled methylation levels at non-CTCF nucleosomes (Figure 4.e). We defined CTCF-nucleosomes as those nucleosomes located within a 1kb of a CTCF binding sites. Conversely, non-CTCF nucleosomes consist of any nucleosome call not overlapping these regions.

We then defined any position within 73bp of a nucleosome center as inside (labeled "In"), while sites located within 85-110bp from the nucleosome centers were considered flanking regions (labeled "Out"; Figure 3.a). We required two consecutive nucleosomes centers to be separated by at least 183bp in order to avoid overlap between In and Out. We calculated M_s for the In and Out group and for each position relative to the nucleosome center (-200;+200) following the procedure described in Figure S3.4. Finally, we plotted raw M_s (without smoothing) as a function of the distance to nucleosome centers.

We found that on average non-CTCF nucleosomes showed similar methylation as that of nucleosomal DNA ("In") and spacer DNA ("Out"; data not shown). We also identified that the 20 nucleotides flanking nucleosome centers were enriched in strong dinucleotides and were strongly depleted of methylation, suggesting a possible functional role for this region in relation with correct nucleosome positioning (Figure 4.e, Figure S2.12).

Depth	20	30	40	50	60	70	80	90	100
Mean cor.	-0.11831	-0.13805	-0.1576	-0.1714	-0.1871	-0.1986	-0.2094	-0.2191	-0.2306
Max cor.	-0.1656	-0.18501	-0.215	-0.22	-0.2203	-0.2381	-0.2401	-0.2605	-0.254
Min cor.	-0.08958	-0.08654	-0.1213	-0.1216	-0.1468	-0.158	-0.1772	-0.1883	-0.1911
Mean p-value	3.01E-06	1.85E-06	1.89E-09	5.74E-10	5.15E-13	1.53E-14	3.06E-17	3.45E-19	6.97E-20
Max p-value	6.05E-05	1.07E-04	5.29E-08	5.06E-08	4.32E-11	1.25E-12	1.51E-15	2.18E-17	6.92E-18
%Data lost	0.1	0.2	0.2	0.25	0.3	0.4	1.9	6.8	18.94

Table S3.7

Nucleosome and methylation patterns at CTCF sites are robust to down-sampling.

Each position located within 1,000 upstream and downstream of CTCF sites was down-sampled to a given depth-of-coverage. We considered a range of possible coverage from 20 to 100, where coverage = N1+N2+M1+M2 (Figure S3.3). Mean (Min) cor. = average (minimum) Pearson correlation coefficient across 100 correlation tests performed between M_s and nucleosome positioning, where M_s has been calculated following down-sampling. Mean (Max) p-value = average (maximum) p-value of Pearson correlation analyses performed across 100 correlation tests performed between M_s and nucleosome positioning, where M_s has been calculated following down-sampling. %Data lost indicates the fraction of the positions showing depth-of-coverage inferior to the threshold considered.

Section SI4. Expression analysis

We retrieved Saqqaq methylation levels at promoters and gene body regions from 280 genes whose proteins have been identified following proteomic analysis of human hair shaft (an extra number of 65 other proteins described in the original publication could not be identified without ambiguity and were therefore disregarded) [Lee et al. 2006]. Those represented a total number of 567 possible transcripts. We then calculated the level of gene body to promoter methylation as a proxy for gene expression, following the procedure described by Ball et al. [2009]. Promoters (PM) were defined as spanning -500 to +2000 around the TSS and gene bodies (GB) as the region from +2000 to the termination site. We then calculated the M_s statistics for PM and GB as indicated in Figure S3.3. We finally calculated the ratio of methylation $M_s(GB)$ over $M_s(PM)$ that we called called R_s. At promoters showing null M_s values, we arbitrarily set R_s to an infinite number, suggesting high expression levels [Ball et al. 2009]. Following a procedure identical to that described in Supplemental Section SI3.4, we first selected all genes showing a minimum coverage of 25 and 75 over their promoter and gene body, respectively. We then selected genes where coverage at promoters was at least equivalent to that observed within gene bodies, correcting for length differences between PM and RE regions. A total number of 140 genes (271 transcripts) from the original list fulfilled those criteria (Table S4.1). Unfortunately, genes for Keratin Associated Proteins, which represent together with keratins major structural components of the hair shaft [Gong et al. 2012], showed not sufficient coverage and could not be investigated for expression patterns.

We compared the distribution of R_s for that list of 140 genes known to be expressed at the protein level in hair shaft and for the full list of genes annotated in the human reference genome (hg18; using similar filtering criteria, this represented a total number of 29,070 genes). We found the former to be significantly greater than the latter (Kolmogorov-Smirnov test, p-value = 0.00152), in agreement with the fact that it consists of known expressed genes in contrast to the whole gene dataset that is a mixture of silenced and expressed genes. This suggests that R_s captures genuine expression information. Of note, we identified amongst the genes showing high to extremely high R_s values a range of transcripts for keratin proteins such as keratins 71 and 85 [Moll et al. 2008] (Table S4.1). Those proteins were likely expressed in the Saqqaq hairs, in agreement with functional studies of modern hairs. Keratin 85 has been detected at high levels in the hair shaft cortex and medulla [Langbein et al. 2010]. Keratin 71 is also a hair-specific keratin expressed in the inner root sheath [Langbein et al. 2010]. This contrasts with keratin 82, one member of group C type II hair keratins, that is absent from the medulla and restricted to a thin external cuticle layer [Langbein et al. 2010] and for which we also found relatively low R_s value (Table S4.1). Similarly, we found a low R_s value for keratin 79, suggesting low expression levels, if any, in agreement with its classification amongst non hair-specific epithelial keratins [Moll et al. 2008]. In addition to hair-specific keratins, we confirmed the presence of a number of proteins involved in cellular adhesion and cytoskeleton organization. In particular, we predicted the presence of Plakophilin 1 and 3 and, to a lesser extent, Plectin and Desmoplakin (Table S4.1), two proteins associated with cell-cell adhesion via desmosomes, a feature known to be key for the structural organization of hairs [Bazzi et al. 2009]. We also predicted high levels of trichohyalin (TCHH), a protein known to confer mechanical strength to the hair follicle inner root sheath. Altogether, this suggests that the level of gene body over promoter methylation, R_s, could be used as a genuine proxy for predicting gene expression. We therefore stratified genes according to R_s quartiles and we used the first and the top quartiles as provisional lists of genes (7,263-7,264 UCSC entries per quartile considered, representing 4,174 and 3,700 non-redundant entries respectively). In

Supplemental Section SI3.5, we identified a subset of 214 genes showing highest (top-1%) methylation levels in their first exon (159 are not redundant; Table S3.5). This represents a list of candidates for low (null) expression levels given the documented tight association between methylation at first exon and silencing [Brenet et al. 2011]. Therefore, this subset of genes is expected to vastly coincide with genes present in the first R_s quartile (ie. with minimal R_s values), as the latter defines a provisional list of genes whose expression was down-regulated in the Saqqaq hairs. Interestingly, 100.0% of the genes identified as showing highest methylation levels in their first exon belong to the first R_s quartile. No genes identified as showing highest methylation levels in their first exon belong to other quartiles, which demonstrates the strong consistency in our expression predictions. Extending the list of genes showing highest methylation levels in their first exon to the top-5% (instead of the top-1%; for a total number of 1,089 genes) does not affect those conclusions, as 88.3% and 9.0% intersect the first and second R_s quartiles, respectively, leaving only a number of 20 (2.35%) and 3 (0.35%) genes in the third and fourth quartiles. In tables S4.2 and S4.3, we provide the bottom-1,000 and top-1,000 genes ranked by R_s values, together with their respective coverage in PM and RE regions. Those lists provide candidate genes whose expression was respectively down-regulated and up-regulated in the Saqqaq hairs. We performed functional enrichment analyses of those candidates in DAVID [Huang et al. 2009], considering only those categories with enrichment scores (ES) superior or equal to 1.2 and within each category, the term showing lowest Benjamini-Hochberg p-values (terms with p-values superior to 0.05 were disregarded). Down-regulated candidates were enriched in the following functional categories: Signal (ES = 3.95; Counts = 203; BH p-value = 7.77×10^{-6}), Cell adhesion (ES = 3.87; Counts = 37; BH p-value = 0.0052) Peptidase S1/S6 – Chymotrypsin/Hap (ES = 3.66; Counts = 20; BH p-value = 6.78×10^{-4}), Plasma membrane part (ES = 3.10; Counts = 134; BH p-value = 0.0433), Ionic channel (ES = 2.98; Counts = 35; BH p-value = 1.60×10^{-4}), Homeobox (ES = 2.51; Counts = 23; BH p-value = 0.0432), Glycoprotein (ES = 2.08; Counts = 239; BH p-value = 0.0016), and Muscle protein (ES = 1.27; Counts = 14; BH p-value = 1.34×10^{-4}). Up-regulated candidates were enriched in the following functional categories: Ubiquitin ligase complex (ES = 3.97; Counts = 19; BH p-value = 6.24×10^{-5}), Phosphorus metabolic process (ES = 2.36; Counts = 78; BH p-value = 0.0231), Ligase (ES = 2.16; Counts = 30; BH p-value = 0.0079), Metal-binding (ES = 2.11; Counts = 173; BH p-value = 0.0110) and Inorganic anion transport (ES = 1.72; Counts = 15; BH p-value = 0.0383).

In order to further assess the validity of our expression predictions, we contrasted microarray expression data from modern hair and R_s values. Mean expression of 10 hair microarray samples from [Kim et al. 2006] (normalized expression data downloaded from GEO (id:GSE3058) was used to partition the represented genes into 10, 20 and 50 quantiles by expression. In addition to R_s (Expression Proxy 1) two other proxies for gene expression were defined based on the literature:

Expression Proxy 2: The read depth over the +1 nucleosome. Calculated as the mean GC-corrected read depth in the region 0 to 200 downstream of the TSS, following [Valouev et al. 2011].

Expression Proxy 3: The strength of the phasing. Signal strength of regularly spaced nucleosomes downstream of the TSS. The mean spectral density was computed over the TSS ± 1000 for the genes of each quantile by the Welch method and the signal strength at the maximum frequency estimated for each quantile, following [Schones et al. 2008].

The correlation for the proxy and the mean expression of these groups was then calculated using Spearman correlation and significance was determined based on Spearman's rho statistic. Importantly, we found significant correlation between mean expression levels and all our three proxys. This supports the validity of our expression predictions and consequently the quality of both our nucleosome and methylation maps. Spearman's rank correlation coefficient for each of the measures was for 10 groups: **1)** 0.99 ($p < 2.2e-16$), **2)** 0.85 ($p = 0.00171$), **3)** 0.79 ($p = 0.009844$); for 20 groups **1)** 0.95 ($p = 6.217e-06$) **2)** 0.75 ($p = 0.0002104$) **3)** 0.71 ($p = 0.0006549$) and for 50 groups: **1)** 0.87 ($p < 2.2e-16$) **2)** 0.62 ($p = 3.242e-06$) **3)** 0.53 ($p = 8.968e-05$).

Gene	UCSC	R_s	Cov(GB)	Cov(PM)	M_s(GB)	M_s(PM)
GARS	uc003tbm.1	NaN	106	146	0.000	0.000
RBM14	uc001oiz.1	Inf	80	64	0.088	0.000
RBM14	uc001oiy.1	Inf	81	80	0.086	0.000
RBM14	uc001oiw.1	Inf	81	84	0.086	0.000
RBM14	uc001oix.1	Inf	81	84	0.086	0.000
RBM14	uc001oiz.1	Inf	80	64	0.088	0.000
RBM14	uc001oiy.1	Inf	81	80	0.086	0.000
RBM14	uc001oiw.1	Inf	81	84	0.086	0.000
RBM14	uc001oix.1	Inf	81	84	0.086	0.000
UNC84B	uc010gxr.1	Inf	2013	99	0.039	0.000
RTN4	uc002ryf.1	Inf	117	246	0.017	0.000
RTN4	uc002ryg.1	Inf	117	246	0.017	0.000
RTN4	uc002ryd.1	Inf	114	66	0.018	0.000
RTN4	uc002rye.1	Inf	117	246	0.017	0.000
2-Sep	uc002wbf.1	Inf	260	233	0.035	0.000
2-Sep	uc002wbg.1	Inf	260	233	0.035	0.000
2-Sep	uc002wbh.1	Inf	260	233	0.035	0.000
SERPINB5	uc002liz.2	Inf	141	30	0.014	0.000
TSPAN7	uc004deg.2	Inf	154	96	0.039	0.000
ACOT7	uc001ams.1	Inf	2372	34	0.055	0.000
PADI3	uc001bai.1	Inf	371	27	0.065	0.000
AIM1	uc003prh.1	Inf	260	317	0.054	0.000
SELENBP1	uc001exx.1	Inf	135	32	0.022	0.000
TCHH	uc001ezp.2	Inf	356	61	0.006	0.000
LMNA	uc001fnk.2	Inf	303	35	0.053	0.000
LMNA	uc001fnj.1	Inf	292	38	0.051	0.000
PKP1	uc001gwe.1	Inf	536	250	0.076	0.000
PKP1	uc001gwd.1	Inf	536	250	0.076	0.000
VCP	uc003zvy.2	Inf	108	206	0.028	0.000
SND1	uc003vmi.1	Inf	2821	178	0.045	0.000
SEC24C	uc001jux.1	Inf	193	59	0.052	0.000
SEC24C	uc001juw.1	Inf	193	59	0.052	0.000
EZR	uc003qrt.2	Inf	607	137	0.026	0.000
NME2	uc002itk.1	Inf	448	127	0.016	0.000
NME2	uc002itj.1	Inf	448	127	0.016	0.000
NME2	uc002iti.1	Inf	87	127	0.046	0.000
NME2	uc002ith.1	Inf	87	127	0.046	0.000
NME2	uc002itk.1	Inf	448	127	0.016	0.000
NME2	uc002itj.1	Inf	448	127	0.016	0.000
NME2	uc002iti.1	Inf	87	127	0.046	0.000
NME2	uc002ith.1	Inf	87	127	0.046	0.000
PKP3	uc001lpc.1	Inf	601	125	0.048	0.000
P4HB	uc002kbn.1	Inf	196	27	0.036	0.000
EEF1G	uc001ntm.1	Inf	115	85	0.061	0.000
CRAT	uc004bxg.1	Inf	399	49	0.025	0.000
MDH2	uc003ueo.1	Inf	420	125	0.048	0.000
RPS9P4	uc002qed.1	Inf	117	119	0.026	0.000
RPS9P4	uc002qea.1	Inf	117	119	0.026	0.000
RPS9P4	uc002qed.1	Inf	117	119	0.026	0.000
RPS9P4	uc002qea.1	Inf	117	119	0.026	0.000
ATP6V1B2	uc003wzp.1	Inf	87	124	0.069	0.000
ATP6V1A	uc003eao.1	Inf	331	51	0.048	0.000
PGD	uc001arc.1	Inf	157	180	0.038	0.000
NPC1	uc002kum.2	Inf	488	182	0.055	0.000
CLTC	uc002ixq.1	Inf	250	152	0.056	0.000
CLTC	uc002ixr.1	Inf	250	152	0.056	0.000
CLTC	uc002ixp.2	Inf	211	152	0.052	0.000
UQCRC2	uc002djx.1	Inf	108	85	0.019	0.000
KRT71	uc001sao.1	Inf	130	46	0.062	0.000
TARS	uc003jhy.1	Inf	84	103	0.036	0.000
YWHAE	uc002fsj.1	Inf	527	187	0.040	0.000
NOP58	uc002uzb.1	Inf	277	74	0.029	0.000
BLMH	uc002hez.1	Inf	134	201	0.045	0.000
LPCAT3	uc001qsi.1	Inf	201	155	0.050	0.000
KRT85	uc001sag.1	Inf	85	34	0.012	0.000
SARS	uc001dwv.1	Inf	146	93	0.048	0.000
SARS	uc001dwu.1	Inf	146	93	0.048	0.000
USP5	uc001qri.2	Inf	232	118	0.047	0.000
USP5	uc001qrh.2	Inf	232	118	0.047	0.000
EIF3E	uc003ymu.1	Inf	125	59	0.056	0.000
OTUB2	uc001yci.1	Inf	417	189	0.041	0.000
PGLS	uc002ngw.1	18.016	189	227	0.079	0.004
HIP1R	uc001udj.1	17.893	1304	307	0.058	0.003
EFHD1	uc002vtc.1	16.748	548	353	0.047	0.003
LMNB2	uc002lvy.1	15.878	1522	281	0.057	0.004
EEF2	uc002lze.1	14.509	611	197	0.074	0.005
CSNK1E	uc003avj.1	13.821	776	275	0.050	0.004
MAP7	uc003qgz.1	13.778	1019	312	0.044	0.003
YWHAG	uc003uez.1	13.264	318	222	0.060	0.005
CSNK1E	uc003avk.1	13.230	793	269	0.049	0.004
YWHAZ	uc003yjv.1	13.071	127	332	0.039	0.003
2-Sep	uc002wbc.1	12.607	262	367	0.034	0.003
2-Sep	uc002wbd.1	12.607	262	367	0.034	0.003

CSNK1E	uc003avp.1	12.343	557	275	0.045	0.004
CSNK1E	uc003avq.1	12.343	557	275	0.045	0.004
VDAC2	uc001jwz.1	12.333	90	185	0.067	0.005
VDAC2	uc001jxa.1	12.333	90	185	0.067	0.005
PEBP1	uc001twu.1	12.194	93	189	0.065	0.005
PLD3	uc002onj.2	11.667	621	207	0.056	0.005
PLD3	uc002onm.2	11.667	621	207	0.056	0.005
PLD3	uc002onl.2	11.667	621	207	0.056	0.005
GDI2	uc001iim.2	10.600	225	265	0.040	0.004
GDI2	uc009xid.1	10.600	225	265	0.040	0.004
GDI2	uc001iil.2	10.600	225	265	0.040	0.004
UNC84B	uc003awh.1	10.286	741	206	0.050	0.005
ACTN1	uc001xkm.1	9.721	1581	327	0.059	0.006
ACTN1	uc001xkl.1	9.721	1581	327	0.059	0.006
CDH1	uc002ewg.1	9.385	797	220	0.043	0.005
CDH1	uc010cfg.1	9.385	797	220	0.043	0.005
EZR	uc003qru.2	9.344	613	358	0.026	0.003
DYNC1H1	uc001yks.1	8.824	1771	329	0.054	0.006
TXNL1	uc002lgg.1	8.777	130	163	0.054	0.006
YWHAB	uc002xmt.1	8.718	78	170	0.103	0.012
YWHAB	uc002xmu.1	8.718	78	170	0.103	0.012
VPS35	uc002eef.2	8.684	76	220	0.079	0.009
CALM3	uc002pew.1	8.681	235	255	0.068	0.008
CTNNA1	uc003ldh.1	8.679	949	289	0.060	0.007
HNRNPD	uc003hmn.1	8.625	80	230	0.038	0.004
HNRNPD	uc003hmp.1	8.625	80	230	0.038	0.004
HNRNPD	uc003hmo.1	8.625	80	230	0.038	0.004
HNRNPD	uc003hmm.1	8.625	80	230	0.038	0.004
UNC84B	uc010gxq.1	8.195	745	330	0.050	0.006
UNC84B	uc003awi.1	8.195	745	330	0.050	0.006
ATP5O	uc002ytl.1	8.155	103	120	0.068	0.008
YWHAZ	uc003yjw.1	8.051	127	409	0.039	0.005
ACOX1	uc002jqe.1	7.924	301	159	0.050	0.006
ACOX1	uc002jqf.1	7.924	301	159	0.050	0.006
ALDH2	uc001tst.1	7.652	615	181	0.042	0.006
YWHAZ	uc010mbr.1	7.469	192	478	0.031	0.004
LAMP1	uc001vtm.1	7.295	757	597	0.049	0.007
PPL	uc002cyd.1	7.263	1376	253	0.057	0.008
CTSD	uc001luc.1	7.111	486	216	0.066	0.009
LMNA	uc001fni.2	7.000	441	147	0.048	0.007
LMNA	uc001fnh.2	7.000	441	147	0.048	0.007
LMNA	uc009wro.1	7.000	441	147	0.048	0.007
LMNA	uc001fng.2	6.853	429	147	0.047	0.007
HEXB	uc003kdf.2	6.722	133	149	0.045	0.007
GCN1L1	uc001txo.1	6.715	1311	142	0.047	0.007
SEC23B	uc002wqz.1	6.583	319	105	0.063	0.010
SEC23B	uc002wra.1	6.583	319	105	0.063	0.010
CNDP2	uc002llm.1	6.319	771	252	0.075	0.012
ACOT7	uc001amt.1	6.151	2454	227	0.054	0.009
LMNB1	uc003kud.1	6.112	349	474	0.026	0.004
PDIA6	uc002rau.1	5.840	162	172	0.068	0.012
PLEC1	uc003zac.1	5.783	3979	195	0.059	0.010
AHNAK	uc001ntk.1	5.752	1413	301	0.038	0.007
UBE2L3	uc002zva.1	5.649	539	203	0.056	0.010
HADHA	uc002rgy.1	5.494	172	105	0.052	0.010
CANX	uc003mkk.1	5.484	366	223	0.025	0.004
CANX	uc003mkl.1	5.484	366	223	0.025	0.004
GPNMB	uc003swb.1	5.412	153	138	0.039	0.007
GPNMB	uc003swc.1	5.412	153	138	0.039	0.007
DUSP14	uc002hnx.2	5.409	281	304	0.036	0.007
HSD17B12	uc001mxq.2	5.385	390	105	0.051	0.010
DSP	uc003mxp.1	5.291	326	345	0.015	0.003
CPT1A	uc001oof.2	5.278	1766	329	0.048	0.009
CPT1A	uc001oog.2	5.278	1766	329	0.048	0.009
CRAT	uc004bxh.1	5.034	445	224	0.022	0.004
PCBP2	uc001sdb.2	4.972	181	150	0.033	0.007
PCBP2	uc001sdi.2	4.972	181	150	0.033	0.007
PCBP2	uc001sdc.2	4.972	181	150	0.033	0.007
LAP3	uc003gph.1	4.935	207	227	0.043	0.009
CANX	uc010jlb.1	4.678	286	223	0.021	0.004
PLEC1	uc003zab.1	4.496	3749	217	0.062	0.014
HSP90AA2	uc001ykv.2	4.471	1054	152	0.029	0.007
HSP90AA2	uc001ykv.2	4.471	1054	152	0.029	0.007
AHNAK	uc001ntl.1	4.470	404	301	0.030	0.007
PREP	uc003prc.1	4.262	566	193	0.044	0.010
CTNNBIP1	uc001aqk.1	4.138	622	198	0.042	0.010
CTNNBIP1	uc001aql.1	4.138	622	198	0.042	0.010
ECHS1	uc001lmu.1	4.016	310	249	0.048	0.012
YWHAZ	uc003yjx.1	3.997	289	385	0.021	0.005
PPIB	uc002and.1	3.900	100	156	0.050	0.013
SNRPD1	uc002ktj.1	3.892	116	129	0.060	0.016
PLEC1	uc003zad.1	3.837	4044	263	0.058	0.015
RPS9P4	uc002qdz.1	3.737	118	147	0.025	0.007
RPS9P4	uc002qdy.1	3.737	118	147	0.025	0.007
RPS9P4	uc002qdx.1	3.737	118	147	0.025	0.007
RPS9P4	uc002qdz.1	3.737	118	147	0.025	0.007

RPS9P4	uc002qdy.1	3.737	118	147	0.025	0.007
RPS9P4	uc002qdx.1	3.737	118	147	0.025	0.007
PDIA3	uc001zsu.1	3.684	152	112	0.033	0.009
PLEC1	uc003zai.1	3.633	5847	292	0.050	0.014
DYNLL1	uc001tyj.1	3.604	323	291	0.037	0.010
FAM83H	uc003yzk.1	3.505	1200	437	0.064	0.018
LMNB1	uc003kuc.1	3.410	278	474	0.014	0.004
RBM14	uc009yrk.1	3.311	474	327	0.051	0.015
RBM14	uc009yrk.1	3.311	474	327	0.051	0.015
PLEC1	uc003zaf.1	3.310	4532	240	0.055	0.017
STX12	uc001bou.2	3.245	208	75	0.043	0.013
ALDOA	uc002dwc.1	3.224	85	274	0.035	0.011
CKAP4	uc001tlk.1	3.202	162	389	0.049	0.015
RAB1B	uc001ohf.1	3.167	90	171	0.056	0.018
FBP1	uc004auw.2	3.109	499	179	0.052	0.017
METAP2	uc001tec.1	2.927	193	113	0.026	0.009
METAP2	uc001tef.1	2.927	193	113	0.026	0.009
PPP1CB	uc002rmg.1	2.902	164	272	0.043	0.015
PPP1CB	uc002rmh.1	2.902	164	272	0.043	0.015
DSC3	uc002kwi.2	2.869	107	307	0.009	0.003
DSC3	uc002kwj.2	2.869	107	307	0.009	0.003
PTBP1	uc002lpp.1	2.829	1069	432	0.039	0.014
PTBP1	uc002lps.1	2.829	1069	432	0.039	0.014
PTBP1	uc002lpr.1	2.829	1069	432	0.039	0.014
RTN3	uc001nxm.1	2.812	448	189	0.045	0.016
RTN3	uc001nxq.1	2.812	448	189	0.045	0.016
RTN3	uc001nxp.1	2.812	448	189	0.045	0.016
RTN3	uc001nxo.1	2.812	448	189	0.045	0.016
RTN3	uc001nxn.1	2.812	448	189	0.045	0.016
PLEC1	uc003zah.1	2.784	5688	331	0.050	0.018
PKM2	uc002atw.1	2.707	238	451	0.042	0.016
PKM2	uc002atx.1	2.707	238	451	0.042	0.016
PKM2	uc002aty.1	2.707	238	451	0.042	0.016
PPA1	uc001jqv.1	2.675	106	189	0.028	0.011
SEC23B	uc002wrc.1	2.671	313	88	0.061	0.023
FASN	uc002kdu.1	2.577	2175	568	0.068	0.026
CAPN12	uc002ojd.1	2.476	458	54	0.046	0.019
TOLLIP	uc001lte.1	2.476	1703	278	0.053	0.022
TOLLIP	uc009ycu.1	2.476	1703	278	0.053	0.022
PHGDH	uc001ehz.1	2.438	273	121	0.040	0.017
G6PD	uc004fly.1	2.409	274	165	0.044	0.018
C3	uc002mfm.1	2.404	892	64	0.075	0.031
ANXA2	uc002agn.1	2.373	236	140	0.017	0.007
ANXA2	uc002agl.1	2.373	236	140	0.017	0.007
ANXA2	uc002agm.1	2.373	236	140	0.017	0.007
ANXA2	uc002agn.1	2.373	236	140	0.017	0.007
ANXA2	uc002agl.1	2.373	236	140	0.017	0.007
ANXA2	uc002agm.1	2.373	236	140	0.017	0.007
LMNA	uc001fnf.1	2.358	966	134	0.035	0.015
G6PD	uc004flx.1	2.256	266	150	0.045	0.020
KRT32	uc002hwr.1	2.247	89	40	0.056	0.025
ABCD1	uc004fif.1	2.234	366	218	0.041	0.018
ALDOA	uc002dvx.1	2.204	599	120	0.018	0.008
DHX9	uc001gpr.1	2.203	240	235	0.038	0.017
YWHAQ	uc002qzx.1	2.200	153	202	0.033	0.015
RBM14	uc001oit.1	1.972	199	327	0.030	0.015
RBM14	uc009yri.1	1.972	199	327	0.030	0.015
RBM14	uc009yrh.1	1.972	199	327	0.030	0.015
RBM14	uc001oit.1	1.972	199	327	0.030	0.015
RBM14	uc009yri.1	1.972	199	327	0.030	0.015
RBM14	uc009yrh.1	1.972	199	327	0.030	0.015
KRT80	uc001rzw.1	1.972	289	57	0.035	0.018
IQGAP1	uc002bpl.1	1.936	550	213	0.036	0.019
VCL	uc001jwe.1	1.934	547	230	0.042	0.022
VCL	uc001jwd.1	1.934	547	230	0.042	0.022
PLEC1	uc003zag.1	1.868	4793	282	0.053	0.028
CSNK1E	uc003avm.1	1.827	1943	150	0.037	0.020
ACAA1	uc003chu.1	1.748	90	118	0.044	0.025
ACAA1	uc003cht.1	1.748	90	118	0.044	0.025
ENO1	uc001apj.1	1.722	151	208	0.033	0.019
YWHAQ	uc002qzw.1	1.710	153	157	0.033	0.019
ME1	uc003pjy.1	1.691	510	138	0.049	0.029
ARF4	uc003dix.2	1.640	150	123	0.013	0.008
EFHD1	uc010fyf.1	1.631	1156	138	0.035	0.022
ATP5A1	uc002lbr.1	1.558	107	100	0.047	0.030
RAB1A	uc002sdn.1	1.553	229	97	0.048	0.031
RAB1A	uc002sdm.1	1.553	229	97	0.048	0.031
RAB1A	uc002sdo.1	1.553	229	97	0.048	0.031
EFHD1	uc002vtd.1	1.497	191	26	0.058	0.038
PLEC1	uc003zae.1	1.477	4153	233	0.057	0.039
CD9	uc001qnq.1	1.337	449	191	0.049	0.037
CD9	uc001qnp.1	1.265	450	207	0.049	0.039
TAGLN2	uc001fun.1	1.232	79	146	0.025	0.021
PDIA6	uc002rav.1	1.223	485	113	0.043	0.035
C1orf204	uc001fuh.1	1.198	248	33	0.036	0.030
C1orf204	uc001fuh.1	1.198	248	33	0.036	0.030

LRRC15	uc003ftu.1	1.179	311	55	0.064	0.055
ACOT7	uc001amr.1	1.090	2144	38	0.057	0.053
PHGDH	uc001eib.1	1.078	232	25	0.043	0.040
ZWILCH	uc002aqb.1	0.798	188	50	0.032	0.040
KRT80	uc001rzx.1	0.794	423	48	0.033	0.042
KRT80	uc001rzy.1	0.794	423	48	0.033	0.042
ATP5A1	uc002lbt.1	0.766	248	95	0.032	0.042
ACOT7	uc001amq.1	0.719	2115	38	0.057	0.079
TGM3	uc002wfx.2	0.680	441	25	0.054	0.080
EIF2S3	uc004dbc.1	0.660	97	32	0.021	0.031
TGM1	uc001wod.1	0.592	224	106	0.045	0.075
KRT82	uc001sai.1	0.553	152	42	0.079	0.143
PLD3	uc002onn.1	0.476	455	50	0.057	0.120
ATG9B	uc010lpv.1	0.297	768	48	0.025	0.083
PDIA6	uc002raw.1	0.277	682	27	0.041	0.148
ACTA1	uc001htm.1	0.000	81	298	0.000	0.007

Table S4.1

R_s values for a list of genes whose proteins have been detected in modern hair shaft.

We selected original gene accessions from Lee et al. [2006], filtering for genes showing sufficient sequence coverage.

Gene	UCSC	Cov(G	Cov(PM)	Rs
MAP2K3	uc002gvu.1	1029	95	0.708
CAPN9	uc009xfg.1,uc001hua.1,uc001htz.1	697	85	0.707
IGSF9B	uc001qgy.1	747	132	0.707
MYH3	uc002gmq.1	433	36	0.707
RAB3IL1	uc001nsp.1	784	41	0.706
NFIC	uc002lxq.1	1471	225	0.706
KIAA0802	uc010dkw.1	95	268	0.705
LYSMD2	uc002abj.1	335	236	0.704
ATP6V1B1	uc002shi.1,uc010fdv.1,uc002shj.1	265	35	0.704
FHAD1	uc001awe.1	560	58	0.704
SQSTM1	uc003mkx.1,uc003mku.1	269	102	0.704
CTBP1	uc003gct.1	1674	152	0.704
ACTN4	uc002ojb.1	173	152	0.703
PDE6B	uc003gao.2,uc003gap.1	1649	132	0.703
SLC11A1	uc002vhw.1,uc002vhv.1	428	41	0.702
PLEKHA7	uc001mmn.1	363	34	0.702
OSBPL5	uc001lxj.1,uc009ydv.1	231	146	0.702
ANKRD13D	uc001okh.1,uc001okg.1	188	113	0.701
SCN3B	uc001pzb.1,uc001pza.1	190	213	0.701
TNR	uc001gkp.1	458	37	0.700
CST7	uc002wtx.1	228	87	0.700
GRIK4	uc009zax.1	3407	86	0.698
BNPI	uc002pno.1	216	181	0.698
TCP10L	uc002ypw.2	147	57	0.698
KIAA1855	uc003our.2	143	133	0.698
CNGA3	uc010fij.1	323	25	0.697
ZNF714	uc002npl.2,uc010ecp.1,uc002npo.2	106	59	0.696
SIGLEC1	uc002wiz.2,uc002wja.1	537	103	0.695
MB	uc003anz.1,uc003aoa.1	159	63	0.693
PRG-3	uc010mtc.1	150	26	0.693
VAV2	uc004cet.1	662	200	0.693
DKFZp762	uc002jvn.1	686	104	0.693
NFKBID	uc002och.1	104	108	0.692
AMOTL2	uc003eqe.1,uc003eqf.1	294	37	0.692
vWF-CP	uc004cdp.2	1536	43	0.691
UPK1B	uc003ecc.1	113	26	0.690
SLC38A10	uc002jzy.1	2553	165	0.689
SLC22A2	uc003qte.1,uc003qtf.1	272	125	0.689
GPR56	uc002emg.2	585	52	0.689
MYH7B	uc002xbi.1	1234	25	0.689
CNKSR1	uc009vsd.1,uc001bln.2,uc009vse.1,uc001blm.2	196	45	0.689
ESF1	uc002woj.1	213	110	0.689
AACS	uc009zyi.1	1081	60	0.688
GSG1L	uc010bxz.1	993	41	0.688
TMPRSS13	uc009yzr.1,uc001prt.1	288	54	0.687
FAM60A	uc001rke.1,uc001rkd.1,uc001rkb.1	193	398	0.687
RAB19	uc010lni.1,uc003vvr.1	163	28	0.687
FAM125B	uc010mxd.1	2603	49	0.687
ZNF57	uc002lwr.1	151	166	0.687
TDRD12	uc002ntr.2,uc002ntq.2	483	300	0.686
KBTBD5	uc003clv.1	114	266	0.686
FAS/ER	uc010kit.1	341	26	0.686
FAM107B	uc001ina.1,uc001imx.1	1750	45	0.686
ATCAY	uc010dts.1	299	64	0.685
PSMA8	uc002kvr.1,uc002kvq.1,uc002kvp.1,uc002kvo.1	168	69	0.685
ZFHX2	uc010akq.1	104	61	0.684
HEATR7B2	uc003jmi.2	117	32	0.684
XHRIP110	uc003mff.1	169	33	0.683
DLGAP3	uc001byc.1	734	76	0.683
NKX3-2	uc003gmx.2	165	338	0.683
C1orf177	uc001cyb.2,uc001cya.2	221	96	0.683
ZNF506	uc002noh.2,uc010eci.1	152	83	0.683
SLC39A4	uc003zcq.1,uc003zco.1,uc003zcp.1	334	243	0.682
SLC22A7	uc003ous.2,uc010jvm.1,uc003out.1	88	50	0.682
C11orf52	uc001pmh.1	113	77	0.681
SMURF1	uc003upt.1	364	62	0.681
DPY30	uc002rob.1,uc002roa.1	91	62	0.681
AGT	uc009xff.1,uc001hty.2,uc009xfe.1	279	95	0.681
TGM3	uc002wfx.2	441	25	0.680
NLRP3	uc001ics.1,uc001icr.1,uc001icv.1,uc001icw.1,uc001icu.1	412	28	0.680
PHLDB3	uc010eit.1	287	65	0.679
CEP192	uc002krw.1	318	27	0.679
GRK6	uc003mgs.1	316	143	0.679
hBSSP-4	uc002crz.1	178	161	0.678
PRSS22	uc002cry.1	178	161	0.678
KIAA1486	uc002vof.1	326	182	0.678
CDH26	uc002ybg.1,uc002ybf.1	155	35	0.677
DNMT3L	uc002zeg.1,uc002zeh.1	676	109	0.677
ZC3H14	uc001xwz.1	193	28	0.677
PRICKLE4	uc003ore.1	263	89	0.677
ABCB9	uc001udr.2	461	120	0.677
FLJ00133	uc002wak.2	437	180	0.677
KIAA0654	uc002pmt.1	343	87	0.676
RNF17	uc001ups.1,uc001upr.1	302	84	0.675
C17orf64	uc002iyq.1	120	45	0.675
FLJ00108	uc003gcm.1	144	340	0.675
LZTS1	uc003wzr.1,uc010ltq.1	276	121	0.674
MYT1	uc002yih.2,uc002yij.1	1943	66	0.674
BPIL3	uc002wyl.1,uc002wyk.1	95	32	0.674
PLXNB3	uc004fii.1	668	72	0.674
CACNA1H	uc002ckv.1,uc010brj.1,uc002cku.1	1058	344	0.674
C1orf14	uc001gpw.1,uc001gpv.1,uc001gpu.1	201	203	0.673
BMI1	uc001irh.1	112	377	0.673

OBSL1	uc010fwl.1	485	147	0.671
AZU1	uc002lpz.1	124	104	0.671
HNMT	uc002tvc.1	97	26	0.670
SLC34A1	uc003mgk.2	264	101	0.670
BRD1	uc003biv.1	2446	186	0.669
STK17B	uc010fsh.1	417	31	0.669
GRHL3	uc001biz.1,uc001biy.1	286	41	0.669
SLC30A8	uc010mcz.1	615	121	0.669
DKFZp586B	uc010dzy.1	423	165	0.669
CALCOCO2	uc002iof.1	169	113	0.669
EIF2C1	uc001bzk.1	399	47	0.668
KLHL31	uc003pcb.2	189	42	0.667
DKFZp686E	uc001cpz.1	225	25	0.667
C4orf44	uc003ggs.1,uc003ggt.1	266	228	0.667
TMPRSS4	uc001psd.2,uc009yzt.1	418	38	0.667
PHKG1	uc003trz.1	257	38	0.665
BC042092	uc002bcv.2	115	51	0.665
IL34	uc002ezi.1,uc002ezh.1	262	134	0.665
ARRDC2	uc002nhv.1	115	344	0.665
RUNX3	uc001bjr.1,uc009vrj.1,uc009vrk.1	1720	98	0.665
TSNARE1	uc003ywj.1	5621	136	0.664
ADRA1A	uc010lum.1,uc003xfe.1,uc010lul.1,uc003xfc.1,uc003xfh.1,uc003xfg.1	338	204	0.664
AK096230	uc001ukq.1	136	203	0.663
NAV2	uc009yhz.1	407	30	0.663
TOX2	uc002xlg.1	991	34	0.663
INPP5D	uc002vtw.1,uc002vtv.1	2060	115	0.663
KIAA1075	uc001sbo.1,uc001sbm.2	178	59	0.663
FBXO46	uc002pcy.1	513	34	0.663
GUCY1B2	uc001vfd.2	312	31	0.662
BC132948	uc002vym.1	214	62	0.662
ALS2CL	uc003cpx.1	139	46	0.662
TP73	uc001aks.2,uc009vlk.1,uc001akr.2	1315	105	0.662
ANKRD24	uc002lzt.2,uc002lzs.2	567	210	0.661
WDR5	uc004cez.1	876	81	0.660
UNC84B	uc010gxs.1	546	173	0.660
EIF2S3	uc004dbc.1	97	32	0.660
ACHE	uc003uxd.1	128	232	0.659
DNAH6	uc002soo.1	88	29	0.659
HSPA12B	uc002wje.1,uc002wjd.1	781	196	0.659
ITGA2B	uc002igu.1	186	49	0.659
LOC285501	uc010iru.1	719	33	0.658
PKD1L2	uc002fgi.1,uc002fgh.1	753	45	0.657
SV2B	uc002bqv.1	426	30	0.657
GRK1	uc001vua.1	190	222	0.657
EVL	uc001ygv.1,uc001ygu.1	369	97	0.657
NTN3	uc002cqj.1	140	368	0.657
PRPF40B	uc001rus.1,uc001rur.1	118	62	0.657
CEL	uc010naa.1	265	29	0.657
BIRC7	uc002yei.1,uc002yej.1,uc010gkc.1	192	162	0.656
P2RX1	uc002fww.1	425	82	0.656
ATXN7L1	uc003vdg.2,uc003vdf.1	187	92	0.656
PGLYRP2	uc002nbg.2,uc002nbf.2	313	41	0.655
AQR	uc001ziv.1	379	73	0.655
ACRC	uc004eae.1	134	117	0.655
VRL	uc002gpz.1	280	50	0.655
TRPV2	uc002gpy.1	280	50	0.655
LGI4	uc002nxx.2,uc002nxz.1	381	95	0.655
TEX14	uc010dda.1,uc010dcz.1,uc002iws.1,uc002iwr.1	1079	72	0.654
SERPINA3	uc001ydo.2	286	34	0.654
ARHGEF10	uc010lre.1,uc003wpt.2	1828	74	0.653
CAV3	uc003brb.1,uc003bra.1	98	64	0.653
EIF4G1	uc003fny.2,uc003fnx.2	292	52	0.653
S100Z	uc003keq.2,uc003kep.1	291	38	0.653
ACSL1	uc003iww.1	326	56	0.653
SUT-1	uc003vtc.1	239	117	0.653
RGS14	uc003mgf.1,uc003mgi.1,uc003mgh.1	609	106	0.653
DAGK1	uc001sio.1	124	97	0.652
KIAA1549	uc003vuk.2	1053	70	0.651
EYA2	uc002xsn.1,uc002xso.1	1765	58	0.651
FLJ00369	uc002osc.2	1461	100	0.650
OAT4L	uc009ypt.1	154	89	0.650
ADCK4	uc002oop.1	145	47	0.648
TRXR2A	uc002zqq.1	211	117	0.647
SLC27A5	uc002qtc.1,uc010eus.1	295	106	0.647
RHBDL2	uc001ccu.1,uc001ccv.2	368	28	0.647
MAP3K14	uc002iiv.1,uc002iiu.1	533	47	0.647
WISP1	uc010med.1,uc010mec.1,uc010meb.1,uc003yuc.1,uc003yub.1	448	79	0.647
TRAPPC9	uc010mel.1	6978	45	0.646
ALDH1L1	uc003eio.2,uc003ein.1	445	46	0.646
LPHN3	uc010ihh.1	1115	36	0.646
PTPRE	uc001lkd.1	819	47	0.646
SHANK1	uc002psx.1,uc002psw.1	2072	175	0.645
EML2	uc002pcq.1	802	141	0.645
ARC41	uc003uqc.1	495	58	0.644
C14orf4	uc001xsy.1	239	462	0.644
WFIKKN2	uc002isv.2,uc010dbu.1	238	184	0.644
PTK2B	uc003xfr.1	170	73	0.644
FBXO47	uc002hrc.1	118	57	0.644
CRYBB1	uc003acy.1	205	36	0.644
TDRD1	uc001lbf.2,uc001lbi.1,uc001lbh.1,uc001lbg.1	254	104	0.643
TRPV4	uc001tpj.1,uc001tpi.1,uc001tph.1,uc001tpg.1	700	72	0.643
TSSC6	uc009ydk.1	336	81	0.643
PABPN1	uc001wjh.2	327	70	0.642
FLJ00059	uc003uwv.1	221	81	0.641
DKFZp434	uc002qkq.1	149	191	0.641
ARPC4	uc003btb.1,uc003bta.1,uc003bsz.1	126	121	0.640

SERPINA6	uc001ycv.1	146	28	0.639
ZNF507	uc002nte.1,uc002ntd.1	169	108	0.639
CES7	uc002eip.1,uc002eio.1	155	33	0.639
NLRP1	uc010clh.1,uc002gcg.1,uc002gch.2,uc002gcl.1,uc002gcj.1,uc002gck.1,uc002gci.1	337	43	0.638
PPFIA1	uc001opr.1	207	44	0.638
RHBDF2	uc002jrs.1	255	133	0.637
PPP1R14D	uc001zmz.1,uc001zmy.1	91	29	0.637
BTBD2	uc002luo.1	290	231	0.637
NBR1	uc002idj.1,uc002idk.1	113	72	0.637
WNK4	uc002ibk.1	117	205	0.637
SSH1	uc001tno.1	201	48	0.637
SERPINE2	uc002vnt.1	438	44	0.636
KLK15	uc002pto.1,uc002ptm.1,uc002ptn.1,uc002ptl.1	170	48	0.635
MUC17	uc003uxp.1	255	54	0.635
ANO1	uc001opk.1	2126	98	0.635
IL2RA	uc009xih.1,uc001iiz.1	358	31	0.635
RAPGEF1	uc010mzm.1,uc010mzl.1,uc010mzs.1,uc010mzr.1,uc010mzq.1,uc010mzp.1,uc010mzo.1,uc010mzn.1,uc004cbc.1	540	75	0.635
CERCAM	uc004buy.1,uc010mya.1	706	56	0.635
C21orf56	uc002zii.1	763	175	0.634
TIMM50	uc002olv.1	101	64	0.634
DPEP2	uc002eve.1	232	49	0.634
INPP5A	uc001llq.1	8836	173	0.633
ZNF236	uc002lmi.1	1840	96	0.633
HYDIN2	uc002ezs.1	294	143	0.632
BC061638	uc009zrx.1	399	63	0.632
TJP3	uc002lyk.1	1065	82	0.631
CAMLG	uc003kzu.1,uc003kzt.1	99	125	0.631
ERCC1	uc002pbu.1	2160	47	0.631
KIAA1618	uc002jyg.1	831	121	0.631
SMCR7L	uc003axw.2	333	45	0.631
ZFHX4	uc003yaw.1	434	57	0.630
MYOC	uc001ghu.1	108	34	0.630
ZMIZ2	uc003tls.1,uc003tlt.1	365	53	0.629
TMEM111	uc003buo.2,uc003bun.1	97	61	0.629
GPR77	uc010ela.1	210	48	0.629
PDE4A	uc002moo.2	258	54	0.628
TINAGL1	uc001bta.1	228	91	0.627
TGFBRAP1	uc010fjc.1	287	36	0.627
HNRNPA1L	uc001vgz.1,uc001vgy.1,uc001vgx.1	83	52	0.627
ELFN1	uc010ksq.1	224	421	0.626
DKFZp434	uc002kdy.2	895	58	0.626
KIAA2034	uc010env.1	623	65	0.626
ZBTB46	uc002ygv.1	2998	113	0.626
HDLBP	uc002wbb.1	597	70	0.625
F7	uc001vsw.1,uc001vsv.1,uc010agp.1	1037	108	0.625
PPP1R16A	uc003zdf.1,uc003zdd.1	429	119	0.624
ILDR1	uc003eeq.1	152	43	0.622
CDH23	uc001jsc.1,uc001jsi.2,uc001jsh.2,uc001jsg.2	617	36	0.622
LRFN1	uc002okw.2	423	374	0.622
MYL6B	uc001sjt.1,uc001sjs.1	111	69	0.622
FAM30A	uc001ysr.2,uc001ysq.2	656	94	0.621
FHOD3	uc010dmz.1	1242	34	0.621
TFAP2C	uc002xya.1	139	431	0.620
CUX2	uc001tsb.1	2520	62	0.620
KIAA1669	uc003bka.1	283	193	0.620
UNC84A	uc003sji.1	836	88	0.620
GALNAC4S	uc001lhl.1	850	86	0.620
TRIM63	uc001bli.1	92	38	0.620
FLJ00074	uc001bbd.1	143	59	0.619
hSK4	uc010eiz.1	160	173	0.618
KCNE3	uc001ovc.1	75	139	0.618
UNQ830	uc002vtr.1	291	147	0.617
GIP	uc002iol.1	81	30	0.617
KCNIP3	uc002sup.1	2094	34	0.617
HPN	uc002nxt.1	386	35	0.617
C5orf38	uc003jdc.1	132	244	0.616
SRCRB4D	uc003ufb.1	607	34	0.616
AK128833	uc001acx.1	190	156	0.616
LRIT1	uc001kcz.1	156	32	0.615
IGnT2	uc010jol.1	547	42	0.614
GMEB2	uc002yfp.1	1133	114	0.613
DBN1	uc003mgz.1	297	238	0.613
FGFR2	uc009xzs.1,uc009xzr.1	809	55	0.612
C11orf41	uc001mun.1,uc001mup.2	352	38	0.612
C7orf25	uc010kxr.1	242	74	0.612
LRRC24	uc003zdn.1,uc003zdm.1	495	528	0.610
C19orf54	uc002oou.1,uc002ooy.1	186	34	0.609
CRYBA4	uc003acz.2	104	38	0.609
KCNIP2	uc009xwv.1,uc009xwu.1,uc001kuf.1,uc001kue.1,uc001kud.1,uc001kuc.1,uc001kub.1,uc001kug.1	479	136	0.608
KCHIP2	uc001kua.1	479	136	0.608
KLF17	uc001clp.1	102	62	0.608
RPL3L	uc002cnh.1	264	126	0.607
C14orf73	uc001ymk.1	851	86	0.606
BC026998	uc002qte.2	179	31	0.606
SERPINB6	uc003muk.1	245	27	0.606
ACTG2	uc010fey.1,uc002sjw.1	132	40	0.606
MED1	uc002hru.2,uc002hrv.2	352	64	0.606
ZCCHC8	uc009zxq.1,uc009zxp.1	165	25	0.606
FOSL1	uc001ogg.1	85	309	0.606
SLC22A6	uc001nwm.1,uc001nwl.1,uc001nwj.1,uc001nwk.1	97	47	0.606
MMP9	uc002xqz.1	366	133	0.606
UROC1	uc010hsi.1,uc003eiz.1	814	56	0.605
LOC645545	uc010ddb.1	333	72	0.605
KLHL30	uc002vxr.1	566	224	0.604
AMBP	uc004bie.2	149	72	0.604
GLIS3	uc003zhy.1,uc003zic.1,uc003zhz.1	782	217	0.604

70

Gene	Transcript			
GCNT7	uc002xxw.2	418	28	0.603
ZNF709	uc002mtx.2,uc002mtw.2	777	104	0.602
PLOD3	uc010lhs.1	269	162	0.602
FOXRED1	uc001qdk.1	123	37	0.602
HAL	uc001tem.1	131	90	0.601
GALNTL2	uc003car.2	366	33	0.601
PCDHGB4	uc003lkc.1	2845	115	0.601
P/OKcl.13	uc010mrd.1	429	106	0.600
C1orf167	uc001ata.2	170	51	0.600
MUC20	uc010hzo.1	316	54	0.598
NOD27	uc002ekp.1	328	49	0.598
LOXL3	uc010ffm.1,uc010ffn.1,uc002smq.1,uc002smp.1	375	64	0.597
BARHL1	uc004cbp.1	447	160	0.597
NLGN3	uc004dzd.1	114	51	0.596
KIAA0134	uc002pgm.1	150	161	0.596
SOHLH2	uc001uvj.1	108	161	0.596
KIAA0999	uc001ppw.1	143	31	0.596
EML1	uc010avt.1	2293	135	0.596
FLJ00009	uc003uwu.1,uc003uws.1	292	87	0.596
FLJ00100	uc003uwt.1	292	87	0.596
NLRP5	uc002qmi.1,uc002qmj.1	698	43	0.596
GLTSCR1	uc002phi.2	1133	428	0.595
GABRB3	uc001zbb.1	3074	29	0.594
TET3	uc002skb.2,uc010fez.1	691	112	0.594
PHF12	uc002hdh.1	175	26	0.594
C12orf34	uc001tpf.1	702	61	0.594
SYDE1	uc002naj.1	221	223	0.594
HIP1R	uc001udk.1	252	184	0.593
SLC41A3	uc003eii.1	492	56	0.592
SEMA5B	uc003egb.1	1301	56	0.592
P2X	uc010ckm.1	194	82	0.592
PLA2G2F	uc009vpp.1	142	36	0.592
TGM1	uc001wod.1	224	106	0.592
PCDHGA5	uc003lju.1	3468	144	0.591
ABCA13	uc010kyu.1	430	186	0.590
STK24	uc001vnm.1	1201	59	0.590
DHX33	uc002gbz.1	190	42	0.589
AZGP1	uc003ush.1	112	33	0.589
SERPIND1	uc002ztc.1	84	33	0.589
DKFZp313K	uc002jew.2	280	33	0.589
ATXN7L3	uc002iga.1,uc002ifz.1	87	64	0.589
DUSP22	uc003msy.1	1047	88	0.588
IL27	uc002dqc.1	153	54	0.588
TBL1X	uc004css.1	398	26	0.588
FLJ00010	uc003uwr.1	326	87	0.587
PCDHGB7	uc003lkn.1	1961	157	0.587
KIAA1274	uc001jre.2	486	95	0.586
TCRBV2S1	uc003vzp.2	765	46	0.586
TNFRSF18	uc001add.1,uc001adc.1,uc001adb.1	95	167	0.586
HRIHFB200	uc001avr.2	1494	35	0.586
BEND7	uc001imo.2,uc001imm.2	563	26	0.585
SIX3	uc002run.1	103	241	0.585
TCL6	uc001yev.1,uc001yeu.1	106	31	0.585
KIAA0447	uc001ahe.2	118	69	0.585
SLC24A6	uc001tvb.1	344	86	0.583
RILPL1	uc001ufd.1	944	55	0.583
RDH10	uc003xzj.1	103	30	0.583
BV13S6J2.	uc010lnu.1	103	30	0.583
NSUN4	uc009vyg.1,uc009vyf.1,uc001cpr.1	146	85	0.582
SP6	uc002imq.1	232	135	0.582
Cap43	uc003yue.1	165	48	0.582
ADCY6	uc001rsh.2	252	213	0.581
SLC23A1	uc003leh.1,uc003leg.1	269	78	0.580
IIP45	uc009vnj.1	161	35	0.580
MID1	uc004cub.1,uc010ndy.1,uc004ctw.2,uc004ctv.2,uc004ctu.2,uc004ctt.2,uc004cto.1,uc004ctn.1,uc004ctm.1	88	34	0.580
SPDYA	uc002rmk.1,uc002rmj.1	99	172	0.579
WDR85	uc004cnj.1,uc010ncl.1	188	79	0.578
C10orf71	uc001jhp.1	395	57	0.577
CSK	uc010bkc.1	156	120	0.577
RLTPR	uc002etn.1,uc010cel.1	831	213	0.577
GALNT6	uc001ryk.1	336	31	0.577
SULT4A1	uc003bed.1	642	167	0.576
AMH	uc002lvh.2	288	199	0.576
JSRP1	uc002lvi.1,uc002lvj.1	211	182	0.575
KCTD1	uc002kvv.1	1706	40	0.574
RNF165	uc002lby.1	1805	37	0.574
CLDN10	uc001vmq.1	921	48	0.573
FXY	uc004cty.2	89	34	0.573
LOC389827	uc004ceh.1	288	60	0.573
R3HDML	uc002xls.1	110	63	0.573
PHAX	uc003kua.1	154	147	0.573
PRSS35	uc010kbm.1,uc003pjz.1	77	44	0.571
ITIH3	uc003dfv.2	217	62	0.571
KLK9	uc010eol.1,uc002pux.1	168	60	0.571
FGF6	uc001qmr.1	103	98	0.571
OVCH1	uc001rix.1	137	65	0.569
CACNA1A	uc002mwx.2	1480	41	0.568
CHRNE	uc002fzk.1	295	195	0.567
C11orf2	uc001ocs.1	139	328	0.566
CCDC144N	uc002qyf.1	106	36	0.566
GPRC5C	uc002jkt.2,uc002jks.1	244	113	0.566
DYRK4	uc009zeh.1	277	57	0.566
RHOBTB2	uc003xcp.1	792	28	0.566
DMRTC2	uc002orr.1,uc002ors.1	87	131	0.565
C8orf82	uc003zdq.1	163	674	0.564
MGC70857	uc003zdp.1	163	674	0.564

PTCHD2	uc001asi.1	235	176	0.562
SLC6A18	uc003jby.1	792	176	0.561
UNQ5783	uc010clb.1	209	44	0.561
C17orf87	uc002gbh.2,uc002gbi.2	209	44	0.561
GPLD1	uc010jpr.1	181	29	0.561
ZNF556	uc002lwq.2,uc002lwp.1	107	96	0.561
AK308669	uc010jvk.1	778	64	0.559
SCN5A	uc010hhj.1	595	38	0.559
LOC126520	uc010dsm.1	153	122	0.558
LDLRAD1	uc001cwm.1	121	27	0.558
CLDN18	uc003ero.1	212	59	0.557
RBP3	uc001jez.1	179	224	0.556
RBMY1F	uc004fvd.1	96	32	0.556
DKFZp434P	uc001jry.2	1120	68	0.555
ABCC5	uc010hxl.1	1014	25	0.555
MYL2	uc001try.2	92	51	0.554
PTPRH	uc002qjs.1,uc010esv.1,uc002qjq.1	199	49	0.554
OLFML2A	uc004bow.1	295	35	0.554
MYO7B	uc002top.1	1949	49	0.553
ASB2	uc001ycc.1	660	73	0.553
KIAA1812	uc001jrz.2	1669	68	0.553
KRT82	uc001sai.1	152	42	0.553
MGAT5B	uc002jtj.1	77	117	0.553
KvLQT1	uc009ydp.1	6354	97	0.552
IP6K3	uc010jvf.1,uc003ofb.1	450	54	0.552
KIAA0772	uc002ycm.1	212	117	0.552
VWF	uc001qno.1,uc001qnn.1	237	28	0.551
C13orf26	uc001uti.1	118	78	0.551
GRK7	uc003euf.1	358	91	0.551
LRIG1	uc003dmw.1	361	106	0.551
MAGIX	uc004dmt.2,uc010nio.1,uc010nin.1,uc004dmu.2	114	94	0.550
LAIR1	uc002qfn.1,uc002qfm.1,uc002qfl.1,uc002qfk.1	141	43	0.549
C14orf115	uc001xpw.2	257	78	0.546
C2orf84	uc002rfc.1	76	166	0.546
NAIF1	uc004bta.1	100	150	0.545
ADCYAP1R	uc003tce.1,uc003tcd.1,uc003tcc.1	470	32	0.545
C10orf141	uc009yap.1,uc001lju.1	316	132	0.543
A1BG	uc002qsd.2	490	133	0.543
TRIM7	uc003mmv.1	291	42	0.541
CATSPER2	uc001zsj.1,uc001zsi.1,uc001zsh.1	91	160	0.541
TBR1	uc002ubw.1	416	75	0.541
CLDN14	uc002yvo.1	1535	30	0.541
MORC1	uc003dxl.1	272	105	0.540
PCDHA8	uc003lhs.1	2194	60	0.540
ITGAE	uc002fwo.2	1567	65	0.539
CD247	uc001gej.2,uc001gei.2,uc001gek.2	1062	88	0.539
KIAA1267	uc002ikb.1	672	35	0.538
DNAJB13	uc001ouo.1	302	25	0.538
MEGF5	uc003mac.1	679	76	0.537
PLA2G5	uc001bcx.1	252	29	0.537
EBI3	uc002lzu.1	136	73	0.537
IQCA1	uc002vwb.1	988	50	0.536
GTF2A1L	uc002rwt.1,uc002rws.1	214	51	0.536
MX2	uc002yzf.1,uc002yzg.1,uc010gop.1	784	72	0.536
POLR1A	uc002sqt.1	127	34	0.535
NRG1	uc010lvt.1,uc003xiy.1	188	67	0.535
TUSC5	uc002fsi.1	449	120	0.535
TM9SF4	uc010gdz.1	292	104	0.534
AMFR	uc002eix.1	278	33	0.534
PAX1	uc002wsj.2	299	319	0.533
TM4SF5	uc002fyw.1	197	30	0.533
TKT	uc003dqp.1	84	67	0.532
ATP2C2	uc002fhy.1	1357	48	0.531
ABHD12B	uc001wys.1,uc001wyr.1,uc001wyq.1	133	141	0.530
UBE2O	uc002jrl.2	316	134	0.530
ANKS6	uc004ayw.1	416	42	0.530
GCLC	uc003pbx.1	117	248	0.530
DKFZp686	uc003mjq.1	102	126	0.529
DUS1L	uc002kdp.1	241	153	0.529
SLC25A31	uc003ifl.1	84	211	0.529
LRFN4	uc001ojr.1,uc001ojs.2	108	533	0.529
BEST2	uc002mux.1	275	67	0.528
TRAF3IP3	uc009xcr.1	108	38	0.528
C11orf9	uc001nse.1	1002	94	0.528
KIAA1234	uc003jaz.1,uc003jax.1	173	73	0.527
TNPO2	uc002muo.1	402	53	0.527
JMJD5	uc010bxw.1	202	76	0.527
SASH3	uc004euu.1	142	32	0.526
TOP3A	uc002gsw.1,uc010cpz.1	331	29	0.526
ZHX3	uc010gqg.1	165	52	0.525
PCDHA5	uc003lhk.1,uc003lhl.1	2686	173	0.525
SAMM50	uc003bek.1	445	112	0.524
SERPINF1	uc010cjw.1,uc002ftl.1	316	69	0.524
DAO	uc001tnr.2	215	45	0.523
CAPN8	uc009xee.1	474	32	0.523
SLC25A24	uc001dvm.1	112	78	0.522
VESPR	uc009zsv.1	272	71	0.522
LOC91948	uc002buh.1	431	27	0.522
KIAA1835	uc010gyk.1,uc003azr.1	307	59	0.522
AK125726	uc003elh.1	326	34	0.521
KCNQ1	uc001lwo.1	8533	97	0.521
SPNS3	uc002fxv.1	613	33	0.520
ZNF564	uc002mty.1	150	104	0.520
CAPZB	uc001bcd.1	1143	27	0.520
TSGA10IP	uc001ogk.1	179	93	0.520
LLGL2	uc010dgg.1	1449	43	0.519

AK310432	uc009vll.1	259	96	0.519
TUB	uc001mfy.1	871	43	0.518
CMIP	uc002fqq.1	4964	63	0.518
SPTBN5	uc001zos.1	1589	95	0.518
NKX6-3	uc010lxa.1	3067	62	0.518
FLJ00275	uc002qfa.1	118	61	0.517
CD4	uc009zfc.1,uc001qqv.1,uc009zez.1	260	48	0.517
PDCD7	uc002aol.1	89	276	0.517
AML2	uc009vrl.1	946	118	0.517
TMEM108	uc003epm.1	91	94	0.516
KIAA0702	uc002apg.1,uc010bhm.1	194	25	0.515
TMEM110	uc003dgb.2	114	47	0.515
SLC26A9	uc001hdo.1	198	68	0.515
C7orf27	uc003smh.2	188	172	0.515
EPN3	uc002irb.2	273	158	0.514
UHRF1BP1	uc001tgp.1	177	26	0.514
PEBP4	uc003xcn.1	2579	32	0.513
PHOX2A	uc001osh.2	285	219	0.512
RASGRP4	uc010egb.1,uc002oir.2	293	50	0.512
SLC18A1	uc003wzl.1	127	26	0.512
PARP14	uc003efr.1	127	26	0.512
CHRND	uc010fyc.1,uc002vsw.1	191	76	0.512
FKSG71	uc009xqt.1	88	30	0.511
PLA2G12B	uc001jtf.1	88	30	0.511
NLRP7	uc010esk.1,uc002qih.2,uc002qig.2	356	39	0.511
VGLL2	uc003pxo.1,uc003pxn.1	270	138	0.511
AFAP1L1	uc003lgi.1	184	47	0.511
C4orf29	uc010inz.1,uc003ifu.1,uc003ifv.2,uc003ift.2	94	32	0.511
CTCFL	uc010gjg.1,uc010gjf.1,uc010gje.1,uc010giy.1,uc010giz.1,uc002xym.2,uc010giw.1,uc010gja.1,uc010gjd.1,uc010gjb.1,uc010gjc.1,uc010gix.1,uc010gji.1,uc010gjj.1,uc010gjh.1	300	181	0.511
FAM107A	uc003dko.1	635	54	0.510
TRIM2	uc003inh.1	961	28	0.510
CARM1	uc002mqa.1	357	156	0.510
CD226	uc002lkm.2	365	31	0.510
BPI	uc002xib.2	242	46	0.507
PCDHGA1	uc003lji.1	4289	54	0.507
ASAP3	uc001bhb.2	150	57	0.507
SPATA16	uc003fin.2	545	46	0.506
MEGF2	uc010hkf.1	177	64	0.506
FSCN2	uc002kan.1,uc002kam.1	220	289	0.505
AK094715	uc003pta.1	126	53	0.505
BIN2	uc001ryh.1,uc009zlz.1,uc001ryg.1	248	143	0.505
CAMTA2	uc010ckv.1	111	35	0.505
LILRB5	uc002qez.1,uc002qey.1,uc002qex.1	119	60	0.504
GATA4	uc003wub.1	2101	45	0.503
PRSS16	uc003njb.1,uc003nja.1	163	41	0.503
NCF4	uc003apz.2,uc003apy.2	281	77	0.502
FCGBP	uc002omp.2	1130	36	0.502
NKX6-1	uc003hpa.1	75	301	0.502
BAZ2A	uc009zov.1	101	38	0.502
DKFZp434B	uc003gcx.1	224	159	0.501
KIAA0467	uc001cjk.1,uc001cjl.1	499	40	0.501
HRIHFB200	uc010kyj.1	263	94	0.500
MS4A10	uc001npz.1	135	45	0.500
PTPRU	uc001brx.1	88	44	0.500
NM_001001	uc002yxh.1	390	30	0.500
EPAS1	uc002ruw.1	78	39	0.500
BCR-ABL	uc010qtx.1	900	88	0.500
BSPH1	uc002phs.1	107	32	0.498
SH2D3C	uc004bsc.1,uc004brz.2	1213	93	0.498
IFNGR2	uc002yrr.2	259	43	0.498
ZNF718	uc003fzt.2	621	581	0.497
SLC6A19	uc003jbw.2	1026	144	0.496
MEFV	uc002cun.1	236	65	0.496
COLQ	uc003bzv.1	353	25	0.496
KCNN4	uc002oxl.1	386	51	0.495
SLC25A19	uc002jns.2,uc010dge.1	231	52	0.495
SPATA20	uc002irf.1,uc002ire.1,uc002ird.1	248	135	0.495
HOXC9	uc001sep.1	330	136	0.495
ARID1A	uc009vsm.1	91	45	0.495
CEACAM20	uc010ejq.1,uc010ejp.1,uc010ejo.1,uc010ejn.1	151	32	0.494
CASS4	uc002xxp.1,uc002xxr.1,uc002xxq.2	506	30	0.494
RHOBTB1	uc001jll.1	75	37	0.493
XKR9	uc010lze.1,uc010lzd.1,uc003xyq.1	141	139	0.493
BC111724	uc002cuf.1	141	139	0.493
LETM1	uc003qdw.1	164	127	0.493
ZNF717	uc010hoq.1	207	34	0.493
CACNA1B	uc004coh.1	2334	60	0.493
PROZ	uc010agr.1,uc001vta.1	326	111	0.492
MUC3A	uc003uxl.1	470	33	0.491
CNGB1	uc010cdh.1,uc002emt.1	1276	44	0.491
KIAA1123	uc002isj.2	149	122	0.491
TUBA8	uc002znx.1	226	37	0.491
FRG1	uc003izs.1	89	240	0.490
DOCK4	uc003vfv.1	204	25	0.490
ZNF232	uc002gat.1	403	158	0.490
BOC	uc003eac.2	121	83	0.490
DKFZp547P	uc001zbi.1	212	135	0.490
SLC7A9	uc002ntw.2,uc002ntu.2,uc002ntv.2	564	48	0.489
ATN1	uc001qrw.1	804	28	0.488
PPP1R14A	uc010efv.1,uc002ohq.1	80	234	0.488
MAN1B1	uc004clf.1,uc004cle.1	382	62	0.487
PASK	uc002wap.2	512	83	0.486
SPG20	uc001uvq.1	305	37	0.485
C20orf107	uc002xxy.1	99	32	0.485
ARRDC5	uc002mbm.1	181	39	0.485

DYNC2LI1	uc002rtl.1,uc002rtk.1,uc002rtj.1	95	46	0.484
dJ402G11.5	uc010hap.1,uc003bjy.1	537	195	0.484
GRM2	uc003dbp.1	204	69	0.483
PEAR1	uc001fqj.1	546	111	0.483
HOXD3	uc002ukt.1	232	112	0.483
CD300LF	uc002jlf.1,uc002jlj.1	140	27	0.482
PRKAG2	uc003wkj.1	5405	75	0.482
FRMD4A	uc001imu.1	2489	38	0.481
IL7	uc003ybq.1	131	126	0.481
ZFAT	uc010meh.1,uc003yuo.1,uc003yun.1,uc003yur.1	1968	187	0.480
MYLK2	uc002wwq.2	313	50	0.479
ARNT2	uc002bfs.1	1259	36	0.479
NPHS2	uc009wxi.1,uc001qmq.2	82	157	0.479
LDB3	uc009xsy.1,uc001kds.1,uc001kdr.1,uc001kdv.1,uc001kdu.1	604	51	0.478
LIG4	uc001vqp.1	136	195	0.478
SNCG	uc001keb.1	76	121	0.478
HEFL	uc010qio.1	506	29	0.478
NKX2-5	uc003mcm.1	181	259	0.477
PAQR5	uc002asa.2	839	39	0.476
PLD3	uc002onn.1	455	50	0.476
PTPRS	uc002mbz.1	216	57	0.475
TXNDC3	uc003tfn.1	93	44	0.473
RNB6	uc010avu.1	1598	28	0.473
LOC201175	uc002iiz.1	444	105	0.473
ZNF284	uc002oyq.1	86	61	0.473
DDX1	uc002rce.1	134	95	0.473
ABAT	uc010buh.1,uc002czd.2	730	69	0.473
TAF15	uc002hkd.1,uc002hkc.1	218	103	0.472
PVALB	uc003apx.1	392	37	0.472
DDC	uc010kza.1	677	69	0.471
ACSBG1	uc002bdh.1	821	43	0.471
KIAA1569	uc002krv.1	397	28	0.470
GLT25D1	uc010eax.1	375	47	0.470
C10orf110	uc001ifx.2	147	46	0.469
PCDHGB5	uc003lkf.1	2529	138	0.469
CCDC79	uc002eqd.1	109	102	0.468
BAG4	uc003xkz.1,uc003xky.1	202	157	0.466
UBE3B	uc001tos.1	530	38	0.466
CLCN2	uc003foh.2	260	69	0.464
ZNF442	uc002mtr.1	96	89	0.464
BTN2A2	uc003nht.1,uc003nhs.2,uc003nhr.1,uc003nhq.1	82	38	0.463
ZC3H12D	uc003qmn.1	247	88	0.463
EPB42	uc001zqz.2	81	75	0.463
ABCD1	uc004fiq.1	128	79	0.463
NINJ2	uc001qil.1	1395	28	0.462
PECAM1	uc002jeq.1	585	40	0.462
SGOL2	uc002uvv.2,uc002uvw.1	78	72	0.462
SAP130	uc002tpn.1	506	25	0.461
CD300A	uc010dfs.1,uc010dfr.1,uc002jkw.1,uc002jkv.1	254	91	0.461
TH	uc001lvt.1,uc001lvs.1,uc001lvr.1,uc001lvq.1,uc001lvp.1	365	96	0.460
SFTPB	uc002sqj.1,uc002sqi.1,uc002sqh.1	174	32	0.460
ZNF595	uc003fzv.1	158	581	0.460
RAB11FIP1	uc003xkp.1	99	26	0.460
COBL	uc003tpp.2	1371	40	0.460
TEAD1	uc009yql.1,uc001mkk.2	293	31	0.458
LRRC37B	uc010csu.1,uc002hgu.1	102	35	0.458
CAPN3	uc001zpq.1	125	80	0.457
CPN2	uc003fts.1	259	93	0.457
IL22RA2	uc003qhn.1,uc003qhm.1,uc003qhl.1	124	34	0.457
CR617046	uc001myp.2	148	146	0.455
TEPP	uc002emw.2,uc002emv.2	346	35	0.455
SSR2	uc009wrm.1,uc001fmx.1	77	70	0.455
RTN1	uc001xem.1	397	219	0.454
PIP5KL1	uc004bsu.1	175	159	0.454
GPIHBP1	uc003yxu.1	116	171	0.454
C7orf51	uc003uve.1	222	181	0.453
ACVRL1	uc001rzk.1	217	86	0.453
TC2N	uc001xzv.2	460	32	0.452
PIK3R6	uc002glq.1	640	35	0.451
SELO	uc003bjz.1	292	117	0.451
ATP1A4	uc001fve.2	234	35	0.449
CYP27C1	uc002tod.2	284	100	0.448
MST091	uc002vqq.2	186	50	0.448
KIAA1132	uc001pri.1	4245	98	0.448
CTRC	uc001awj.1,uc001awi.1	105	47	0.448
C20orf197	uc002ybj.1	226	56	0.446
CRHR2	uc003tbp.2	655	46	0.445
PPP1R1B	uc002hsc.1,uc002hsb.1	90	32	0.444
MRP6	uc002dem.1	90	60	0.444
TUBB3	uc002fpf.2	1043	240	0.444
ARVCF	uc002zra.1	504	71	0.443
KIF25	uc003qwl.1,uc003qwk.1	686	85	0.443
DUSP18	uc003aiu.2	194	143	0.442
LPIN2	uc002klo.1	643	42	0.441
ASZ1	uc003vjb.2	113	149	0.440
TOP3b	uc002zvr.2,uc010qtm.1	332	31	0.436
PRSS36	uc002ebd.1	405	197	0.432
PSAPL1	uc010idj.1	82	124	0.432
RPS6KA1	uc009vsl.1,uc001bms.1	345	62	0.431
grb7V	uc002hst.1	164	101	0.431
EDC4	uc002eus.1	202	87	0.431
CELSR3	uc010hkq.1	405	58	0.430
KAT2A	uc002hyx.2	270	116	0.430
KIAA1501	uc002hqc.1	599	147	0.429
NR4A1	uc001rzr.2	331	71	0.429
EIF3L	uc003auh.1	169	29	0.429

Gene	Transcript	Col1	Col2	Col3
KIAA1787	uc002ggc.1	173	84	0.425
ADAMTS17	uc002bvx.1	3868	33	0.424
BAIAP2L2	uc003auw.1	748	152	0.423
NM_001127	uc010nmy.1	319	45	0.423
PCNA	uc002wlq.1	188	159	0.423
ENTPD1	uc001kle.1,uc001kli.2	447	28	0.423
C1orf186	uc001hdt.1	278	47	0.423
MCM5	uc003anw.1	354	52	0.422
PHKA1	uc004eay.2,uc004eax.2	178	25	0.421
BOLL	uc002uut.2,uc002uus.2,uc002uuu.1,uc002uur.2	91	460	0.421
C6orf146	uc003mvy.1	270	31	0.421
CASQ1	uc001fvi.1	99	25	0.421
CNTN2	uc001hbs.1	486	71	0.420
AX721097	uc002fji.2	86	48	0.419
LRRC17	uc003vau.1,uc003vat.1	104	29	0.418
C21orf63	uc002ypu.1	558	35	0.418
SCML4	uc003pry.2	192	37	0.418
LOC283867	uc002eol.1	766	29	0.416
KLK8	uc002puu.1	92	134	0.416
FAM83E	uc002pjn.2	481	153	0.416
RAB11FIP5	uc002sit.2	202	84	0.416
ALPL	uc001beu.2	453	25	0.414
RBM47	uc003qvd.2	790	28	0.414
PLD5	uc001hzm.2,uc001hzl.2	1198	30	0.413
HSF5	uc002iwi.1	324	401	0.413
DNAJC14	uc001shu.1	461	57	0.412
pp6037	uc009yue.1	306	56	0.412
WDR59	uc002fdq.1	447	35	0.411
NUP210L	uc001fdw.1	849	74	0.411
LGALS4	uc002ojq.1	186	87	0.409
EXOC3L	uc002erx.1	348	79	0.409
PCDH8	uc001vhj.1,uc001vhi.1	285	582	0.408
HEATR5A	uc010ami.1,uc001wrg.1	459	28	0.407
ETV6	uc001raa.1	1387	49	0.406
CCDC64B	uc002ctf.2,uc002cte.2	386	67	0.405
TRAK1	uc003ckz.2	494	25	0.405
TCF4	uc002lgb.1	267	36	0.404
VAX1	uc009xyx.1,uc001ldb.1	220	400	0.404
RPA3	uc003sri.1	312	37	0.403
ARX	uc004dbp.2	413	83	0.402
ST3GAL2	uc002eyw.2	627	84	0.402
MYL5	uc003qav.1	196	105	0.402
graf-2	uc003ili.1	337	27	0.401
ADAM2	uc003xnl.1,uc003xnk.1,uc003xnj.1	162	54	0.400
gklp	uc001oee.1	245	49	0.400
XAF1	uc002gdo.1,uc002gdn.1,uc002gdr.1,uc002gdq.1,uc002gdp.1	85	68	0.400
Insp3ri	uc010hcc.1	317	38	0.400
CGB8	uc002pmc.1	89	105	0.393
CETP	uc002ekj.2,uc002eki.2	372	64	0.393
TACC1	uc010lwq.1,uc003xmh.2	121	37	0.393
FOXN1	uc002hbj.1	208	109	0.393
CHRFAM7A	uc010baj.1,uc010bak.1	229	30	0.393
BC127913	uc010axk.1	275	90	0.393
PCDHGB1	uc003ljo.1	3925	85	0.393
KIAA1218	uc003vdh.2	786	41	0.391
WDR88	uc002nui.1	545	213	0.391
SEPX1	uc002cnq.1	118	184	0.390
LYST	uc001hxi.1	308	40	0.390
FAM13C	uc001jkp.2	230	57	0.389
KIAA0148	uc001tpr.1	207	46	0.389
WNT5B	uc009zdq.1	1291	35	0.389
IL1F8	uc002tiq.1	76	59	0.388
LAX1	uc001haa.1	80	31	0.388
RIMKLB	uc001quu.2	792	34	0.386
MYST1	uc002eaz.1	90	52	0.385
AGAP3	uc003wjj.1	417	117	0.383
HOM-TES-	uc001qpg.2	159	73	0.383
SDC3	uc001bsd.2	131	60	0.382
ABCC3	uc002isn.1	228	26	0.380
GRIK2	uc003pqn.2,uc010kcw.1,uc003pqp.2,uc003pqo.2	734	117	0.380
C10orf108	uc001ifr.2	211	89	0.380
ZNF580	uc002qlm.1	506	64	0.379
DKFZp564A	uc002evs.1	91	69	0.379
CA5A	uc002fkn.1	1016	81	0.379
CYP4F12	uc002nbl.2	111	30	0.378
NWD1	uc002nev.2	866	28	0.377
DUXA	uc002qoa.1	116	29	0.375
TP53	uc002gih.1,uc002gig.1	98	55	0.374
OSTCL	uc003qrw.1	102	38	0.373
SPAG4L	uc002wyi.1	242	45	0.372
SLC6A12	uc001qhy.2,uc001qhx.2	147	41	0.372
PTGIR	uc002pex.1	111	299	0.372
TNNT1	uc002qjd.2,uc002qje.2,uc002qjc.2,uc002qjb.2,uc002qjf.2	400	27	0.371
FAM151A	uc001cxn.1	158	35	0.369
DKFZp564A	uc004ead.2	95	28	0.368
OGT	uc004eac.2	95	28	0.368
MIOX	uc003bln.1,uc003blm.1,uc003bll.1	125	138	0.368
SNX20	uc002egk.1,uc002egi.2	272	70	0.368
ADAT1	uc002fep.1,uc002feo.1	164	120	0.366
BEAN	uc002eoq.1	241	80	0.365
SIGIRR	uc001lpd.1	343	48	0.364
PIWIL2	uc010ltv.1,uc003xbn.1	440	160	0.364
SLC5A11	uc010bxt.1,uc002dmu.1,uc002dmt.1,uc002dms.1	473	43	0.364
CATSPER4	uc001blp.1	110	40	0.364
GSG1	uc001rbp.1,uc001rbo.1,uc001rbn.1,uc001rbq.1	235	32	0.363
VSTM1	uc002qcx.2,uc002qcw.2	144	87	0.362

Gene	Transcript			
EPX	uc002ivq.2	162	70	0.360
DISC1	uc001hvc.2	95	111	0.360
C9	uc003jlv.2	115	31	0.359
MRAP	uc002ypl.1	200	154	0.359
ZBTB7B	uc001fql.2	141	126	0.357
FN3K	uc002kfw.1	142	81	0.357
FAT1	uc010isn.1,uc003ize.1	101	36	0.356
COL8A1	uc003dth.1,uc003dtg.1	337	30	0.356
RP11-	uc003oqk.1	213	126	0.355
RASAL3	uc002nbe.2,uc010eaa.1	564	75	0.355
TMEFF1	uc004bay.1	799	53	0.354
FLJ00251	uc001mdw.2	389	60	0.353
PPP1R7	uc002wav.1	394	26	0.352
OGDH	uc003tlp.1,uc003tlo.1	446	28	0.352
PILRA	uc003uuq.1,uc003uup.1,uc003uuo.1	243	64	0.351
DEF6	uc010jvt.1,uc010jvs.1,uc003okk.1	361	39	0.351
SYNGAP1	uc010juz.1	295	47	0.351
CLCNKA	uc001axv.1,uc001axu.1	117	41	0.350
ARSA	uc003bmz.2,uc003bnd.2,uc003bnc.2,uc003bnb.2,uc003bna.2	80	280	0.350
CORO6	uc002hen.2,uc002hem.2,uc002hel.1	154	269	0.349
PRMT2	uc002zjz.1	491	52	0.348
CACNA1S	uc001qvv.1	1019	40	0.348
UNQ385	uc003gbb.1	256	216	0.347
SAMD14	uc002iqe.1	312	48	0.346
ERCC6L	uc004eap.1,uc004eaq.1	110	76	0.345
AP2M1	uc003fmy.1,uc010hxu.1	116	50	0.345
APBA2	uc001zcl.1,uc001zck.1,uc010azj.1	3028	32	0.345
ARRDC3	uc003kjz.1	180	62	0.344
C3orf22	uc003ejb.1	214	55	0.343
ARID3C	uc003zuy.1	251	43	0.343
SERINC3	uc010ggs.1,uc002xmf.1,uc002xme.1	127	87	0.343
SCIN	uc003sso.2	219	25	0.342
CCDC81	uc001pbx.1,uc001pbw.1	148	76	0.342
GRIP2	uc003byt.1	943	28	0.341
NEUROD2	uc002hry.1	218	297	0.341
SAG	uc002vuh.2	652	38	0.340
KIAA0613	uc009xsz.1	1276	51	0.340
ACCN3	uc003wip.2,uc003wio.1,uc003win.2	196	244	0.340
ACACB	uc001toc.1	1736	33	0.338
HTR3A	uc001pon.1,uc001pom.1	202	34	0.337
LGALS12	uc001nxc.1,uc001nxb.1,uc001nxa.1	121	27	0.335
CAPN11	uc003owt.1	314	30	0.334
ZNF687	uc009wmp.1	128	71	0.333
TMEM82	uc001axc.1	81	121	0.332
FCHSD1	uc010jgg.1	127	42	0.331
limkain	uc001ufv.1	361	94	0.330
LIM2	uc002pwm.1,uc002pwl.1	91	60	0.330
CPT1A	uc009ysj.1	1186	60	0.329
CTNND1	uc001nlh.1	378	148	0.326
TBC1D24	uc002cqm.1	1153	232	0.326
SLC9A3R1	uc002jlp.1	120	97	0.323
POU2F3	uc001pxe.1	99	48	0.323
KLK14	uc002pvs.1	171	46	0.323
NKX2-3	uc009xwj.1	222	143	0.322
SYTL1	uc009vsv.1	392	56	0.321
C19orf51	uc002qjl.1,uc002qjk.1,uc002qjj.1,uc002qji.1	374	144	0.321
FUT3	uc002mdl.2,uc002mdj.2,uc002mdm.2,uc002mdk.2	97	31	0.320
KPL1	uc009ytr.1	188	30	0.319
NRXN2	uc001oao.1	384	49	0.319
PAQR6	uc001fny.1,uc001fnx.1,uc001fnw.1,uc001fnv.1,uc001fnu.1	147	78	0.318
DQ866763	uc001vue.2	150	167	0.318
DPH1	uc002ftu.1	215	51	0.316
IGLON5	uc002pwc.2	399	98	0.316
OPN4	uc001kdq.1,uc001kdp.1	317	50	0.315
FBXO32	uc003yqq.1	180	34	0.315
MICAL2PV2	uc001mkc.2	884	51	0.313
MICAL2PV1	uc001mkb.2	884	51	0.313
GHDC	uc002hzd.1	80	75	0.313
MEOX1	uc002iea.1,uc002idz.1,uc002ieb.1	405	72	0.311
FANCL	uc010fcf.1,uc010fce.1,uc002rzw.2,uc002rzx.2	127	79	0.311
SCAND2	uc002bks.2	253	78	0.308
BTNL9	uc003mmt.1	832	45	0.306
SDPR	uc002utb.1	93	57	0.306
HCCA2	uc001ltp.1,uc001lto.1	301	122	0.304
COL8A2	uc001bzw.1,uc001bzv.1	112	68	0.304
NM_207423	uc001ijx.1	133	281	0.302
ALOX12B	uc002qjy.1	421	79	0.300
IRF1	uc003kxb.1	394	47	0.298
SNRPA	uc002opa.1	166	33	0.298
KIR3DX1	uc002qge.1	131	39	0.298
ATG9B	uc010lpv.1	768	48	0.297
NRBP2	uc003yzw.1	158	131	0.296
TPSG1	uc002ckw.1	100	128	0.295
FLJ00180	uc010bto.1	112	33	0.295
TRIM61	uc003iqw.1	188	41	0.291
ATP13A3	uc003ftx.2	112	26	0.290
AK308580	uc010ekl.1	216	141	0.290
TBC1D4	uc010aer.1	83	36	0.289
CYP8B1	uc010hif.1	119	57	0.287
TEKT4	uc002stw.1	216	134	0.286
MYO18A	uc002hds.1	1635	39	0.286
DHH	uc001rtf.1	326	93	0.285
C15orf39	uc002azp.2	589	42	0.285
AQP4	uc002kwa.1	102	29	0.284
KANK4	uc001daf.2	158	28	0.284
DGUOK	uc002sjy.1,uc002sjx.1	122	69	0.283

F12	uc003mgo.2	518	36	0.278
ABI3	uc002ioq.1,uc002iop.1	244	45	0.277
NR4A3	uc004bah.1	304	63	0.276
HCA58	uc002xel.1	145	32	0.276
BTBD17	uc002jkn.1	396	135	0.273
KIAA1335	uc002xjz.1	172	39	0.272
PDX1	uc001urt.2	209	282	0.270
RCCD1	uc002bql.1,uc002bqk.1,uc002bqj.1	288	272	0.270
CAPN5	uc001oya.1	75	111	0.269
CIB3	uc010eaq.1,uc010eae.1,uc002nds.1	123	33	0.268
AK123582	uc002fhs.1	89	167	0.268
ART1	uc009yeb.1,uc001lye.1	126	27	0.268
NCAN	uc010ecc.1	136	45	0.265
LDLR	uc010dxu.1	91	72	0.264
UNQ6496	uc004brq.1	179	60	0.261
PNLIPRP2	uc001lcq.1,uc009xyt.1	223	58	0.260
MASP2	uc001aru.1	161	97	0.258
C22orf28	uc003amm.2	130	67	0.258
MORN3	uc001uax.1	402	69	0.257
BATF2	uc001ocf.1	149	38	0.255
TP53INP2	uc002xau.1	304	229	0.251
PRTN3	uc002lqa.1	287	72	0.251
DKFZp686B	uc010els.1	206	71	0.251
U2AF1	uc002zcy.1	250	25	0.250
ISLR2	uc002axd.1,uc010bjf.1,uc002axe.1,uc002axf.1	666	248	0.248
HPCA	uc001bwh.1	199	147	0.246
IL15RA	uc001iiu.1	114	37	0.243
UNQ3098	uc002orc.1	263	32	0.243
DBX1	uc001mpw.1	144	137	0.238
SLC29A2	uc001ohs.1	97	69	0.237
HYLS1	uc001qcw.1	297	35	0.236
FXYD1	uc002nyc.1	134	125	0.233
IGSF11	uc010hqs.1,uc003ebz.1,uc003eby.1	667	81	0.233
DNAH8	uc003ooe.1	776	28	0.231
AX747263	uc003syp.1	224	187	0.228
EIF4A1	uc002qho.1	347	45	0.227
tMDC	uc003xnf.1,uc003xne.1	81	55	0.226
AB209692	uc001nwn.1	84	38	0.226
MYO1H	uc009zvh.1	469	39	0.226
TRPC3	uc003ief.1,uc010inr.1	76	166	0.218
DNM1	uc004buc.1,uc004bub.1	492	43	0.210
RUSC1	uc001fkk.2,uc001fkj.2	342	188	0.206
BC075797	uc003ssl.1	192	79	0.206
NLRP13	uc002qmq.1	213	50	0.205
RINL	uc002ojq.1	284	38	0.201
ABCG8	uc002rtq.1	258	62	0.200
NBLA00301	uc003itl.2	196	157	0.200
C20orf78	uc002wrj.1	130	26	0.200
KRT79	uc001sbb.1	124	62	0.200
PCOLCE	uc003uvo.2	231	92	0.199
NHLH2	uc001efy.1,uc009wqz.1	272	81	0.199
RENBP	uc004fjo.1	101	99	0.196
ZNF19	uc002fal.1,uc010cqc.1,uc002fam.1	94	70	0.186
MAPKBP1	uc010bcl.1	352	36	0.184
PNMAL2	uc002pes.2	280	405	0.181
HOXB3	uc010dbg.1,uc010dbf.1,uc002inn.1	1207	39	0.178
PHOSPHO	uc002ior.1	281	149	0.177
MPL	uc009vwr.1,uc001ciw.1,uc001civ.2	204	27	0.176
TTLL4	uc010fvx.1	81	28	0.173
SLAMF1	uc001fwl.2	120	26	0.163
MYOZ3	uc003lsr.2,uc003lss.2	309	28	0.159
NRTN	uc002mde.1	292	130	0.148
GPS2	uc002qfz.1	258	113	0.146
AQP2	uc001rvn.1	132	103	0.142
C14orf93	uc001wib.1	91	51	0.140
EVX2	uc002uke.1	201	247	0.137
KIAA0985	uc001tua.1	174	26	0.120
PON1	uc003uns.1	77	60	0.111
CTCF	uc002etm.1	130	55	0.106
NR1I2	uc003edl.1	82	43	0.105
PRRG2	uc002pon.2	179	61	0.097
ZNF335	uc002xqv.1	94	54	0.096
hCG_31249	uc004bzl.2	269	165	0.094
KCNC3	uc002prt.1	364	195	0.067
SERPINC1	uc001qjt.1	83	30	0.060
SPRED3	uc002oim.1	311	68	0.044
EFNB1	uc004dxe.2,uc004dxd.2	106	173	0.000
GPR172A	uc003zcc.1,uc003zce.1,uc003zcd.1	76	339	0.000
AP3M2	uc003xop.1,uc003xoo.1	76	243	0.000
ZNF703	uc003xjy.1	130	458	0.000
PP838	uc003uss.1	94	174	0.000
EPHA7	uc003poq.2	148	293	0.000
AMD1	uc003pul.1,uc003puk.1	89	323	0.000
C6orf57	uc003pfq.1	83	83	0.000
SOX4	uc003ndi.1	112	336	0.000
SHROOM1	uc003kxy.1	186	306	0.000
GFM2	uc010izj.1,uc003kdi.1,uc003kdh.1	85	101	0.000
NPM1	uc003mbh.1,uc003mbj.1,uc003mbi.1	93	193	0.000
TNFAIP8	uc003ksh.1	110	88	0.000
HAND2	uc003ith.1	83	332	0.000
DUSP7	uc003dct.1	110	325	0.000
CCK	uc003cld.1	95	139	0.000
RPS19BP1	uc003ayb.1	83	126	0.000
HNRPLL	uc002rqx.1	87	209	0.000
SOX11	uc002qyj.1	245	293	0.000
FLJ36070	uc002pkg.1	268	132	0.000

KISS1R	uc002lqk.2	262	352	0.000
TOB1	uc010dbv.1	126	295	0.000
HOXB13	uc002ioa.1	85	128	0.000
GALR2	uc002jqm.1	161	401	0.000
HEXIM1	uc002iiq.1	122	153	0.000
BCL6B	uc002geq.2,uc010clt.1	100	179	0.000
IRX5	uc002ehv.1	131	422	0.000
TMEM159	uc002dih.2,uc002dif.2	137	122	0.000
IMP3	uc002bat.2	219	221	0.000
AEN	uc010bnl.1	75	142	0.000
ZFP36L1	uc001xki.1,uc001xkh.1	83	102	0.000
ATP6V1D	uc001xjf.1	107	143	0.000
KLF5	uc001vje.1	95	404	0.000
C13orf36	uc001uvt.2	83	240	0.000
GPN3	uc001tqs.1,uc001tqr.1	88	82	0.000
SILV	uc001siq.1,uc001sip.1	75	27	0.000
ELA1	uc001ryi.1	89	32	0.000
PRPH	uc001rtu.1	100	357	0.000
WEE1	uc001mhs.1	95	404	0.000
AL157440	uc001kec.1	89	147	0.000
C10orf114	uc001iqn.2	98	312	0.000
KIAA0335	uc001klr.1,uc001klq.1	94	121	0.000
ZNF518A	uc001klo.1,uc001klp.1	94	121	0.000
ACTA1	uc001htm.1	81	298	0.000
UBE2U	uc001dbn.1	76	42	0.000

Table S4.2

List of candidate genes whose expression was down-regulated in the Saqqaq hairs.

This list corresponds to the first 1,000 non-redundant genes showing the lowest R_s values (transcripts are collapsed into a single Gene name, providing the maximal R_s value observed in this quartile). R_s values and respective coverage in PM and RE regions are provided.

Gene	UCSC	Cov(GB)	Cov(PM)	Rs
UTY	uc004fsv.1,uc004fsx.1	400	49	Inf
PRKY	uc004fre.1	395	163	Inf
MPP1	uc010nvg.1,uc004fmq.1,uc004fmp.1	86	101	Inf
FAM3A	uc004flu.1,uc004flt.1,uc004flw.1,uc004fls.1	161	98	Inf
DNASE1L1	uc004fkw.1,uc004fkv.1,uc004fku.1	127	140	Inf
IRAK	uc010nur.1	192	164	Inf
IRAK1	uc004fju.1,uc004fjt.1,uc004fjs.1,uc004fjr.1	96	164	Inf
HCFC1	uc004fjp.1	710	148	Inf
ARD1A	uc004fjn.1,uc004fjm.1	89	141	Inf
L1CAM	uc004fje.1,uc010nuo.1,uc004fjc.1,uc004fjb.1	213	156	Inf
UCHL5IP	uc004fho.1,uc004fhn.1	341	95	Inf
GABRE	uc004ffi.1,uc004ffh.1,uc004fff.1,uc004ffg.1	83	123	Inf
CD99L2	uc004fen.1,uc004fem.1,uc004fel.1	564	157	Inf
IDS	uc004fcv.2,uc004fcu.2,uc004fct.2	94	59	Inf
ZNF75D	uc004eyp.1	124	51	Inf
MOSPD1	uc004evb.1	103	66	Inf
ZNF280C	uc010nrf.1,uc004evm.1	119	48	Inf
APLN	uc004eus.1	78	152	Inf
SMARCA1	uc004eup.2,uc004eun.2	101	74	Inf
CUL4B	uc004esw.1	149	30	Inf
FAM70A	uc010nqo.1,uc004esp.2,uc004eso.2	144	99	Inf
KIAA1210	uc004era.2	172	44	Inf
KLHL13	uc004eqm.1	182	95	Inf
AMOT	uc004eps.1	158	56	Inf
TRPC5	uc004epm.1,uc004epl.1	400	36	Inf
AMMECR1	uc004eop.1,uc004eoo.1	110	103	Inf
TSC22D3	uc004enj.1,uc004eni.1,uc004enh.1	329	102	Inf
NUP62CL	uc004ena.1	128	44	Inf
MORC4	uc004emw.2,uc004emv.2,uc004emu.2	77	124	Inf
MCART6	uc004elu.1	209	39	Inf
BTK	uc010nno.1	152	38	Inf
TRMT2B	uc004egq.1,uc004eqv.1,uc004eqt.1,uc004egr.1	144	39	Inf
DKFZp667H0	uc004equ.1	144	39	Inf
CXorf34	uc004eqs.1	144	39	Inf
HDX	uc004eel.1,uc004eek.1	103	35	Inf
RPS6KA6	uc004eej.1	120	92	Inf
BRWD3	uc004edu.2,uc004eeb.2,uc004eea.2,uc004edz.2,uc004edy.2,uc004edx.2,uc004edw.2,uc004edv.2,uc004edt.2	194	56	Inf
ZDHHC15	uc004ech.1,uc004ecg.1	141	41	Inf
ZMYM3	uc004dzj.1,uc004dzi.1,uc004dzh.1	178	76	Inf
LAS1L	uc004dwd.1,uc004dwc.1,uc004dwa.1	80	44	Inf
KIAA0424	uc004dvj.1,uc004dvk.1	117	48	Inf
ARHGEF9	uc004dvm.1,uc004dvl.2	117	48	Inf
PHF8	uc004dsy.2,uc004dsw.1,uc004dsu.1,uc004dst.1	307	33	Inf
HUWE1	uc004dsp.1	522	71	Inf
JARID1C	uc004dsa.1,uc004drz.1	196	75	Inf
SYP	uc004dna.1,uc004dmz.1	79	89	Inf
WDR45	uc004dmo.1,uc004dmp.1	190	60	Inf
DKFZp761J18	uc004dmc.2	99	92	Inf
TFE3	uc004dmb.2,uc004dmd.1	99	92	Inf
SLC35A2	uc004dlo.1	76	54	Inf
ELK1	uc010nhw.1,uc010nhv.1,uc004dik.2	107	64	Inf
SYN1	uc004die.1,uc004did.1	371	135	Inf
ZNF41	uc004dhx.2,uc004dhy.2	112	74	Inf
SLC9A7	uc004dgv.1,uc004dgu.1	146	140	Inf
BCOR	uc004deq.2,uc004deo.2,uc004dep.2,uc004dem.2,uc004den.2	571	236	Inf
SRPX	uc004ddz.1,uc004ddv.1	118	80	Inf
KLHL15	uc004dbb.1	196	101	Inf
PHKA2	uc010nfe.1	173	27	Inf
DKFZp434M1	uc010nfd.1	247	144	Inf
SCML2	uc004cyl.2	250	144	Inf
RAI2	uc004cyg.1,uc004cyf.1,uc004cyh.2,uc010nfa.1	191	135	Inf
ARHGAP6	uc004cun.1	453	91	Inf
MID1	uc004cth.2,uc004ctg.2	320	44	Inf
HDHD1A	uc010ndl.1,uc004crv.1	403	166	Inf
NLGN4X	uc010ndh.1,uc010ndj.1,uc010ndi.1,uc004crr.1,uc004crq.1	600	110	Inf
MXRA5	uc004crg.2	158	158	Inf
BRCC3	uc004fnb.1,uc004fna.1	119	114	Inf
TAZ	uc004flc.2,uc010nuy.1,uc004flb.1,uc004fla.1,uc004fky.1,uc004fkx.1	93	133	Inf
SLC6A8	uc004fic.2,uc004fib.2	114	125	Inf
BGN	uc004fhr.1	273	75	Inf
NSDHL	uc004fgs.1,uc004fgt.1	197	70	Inf
MTMR1	uc004feh.1,uc004fei.1	312	97	Inf
FHL1	uc010nrz.1,uc004ezl.1	198	172	Inf
FAM122C	uc010nru.1	86	64	Inf
HPRT1	uc004exl.2	84	117	Inf
PHF6	uc004exi.2,uc004exk.1,uc004exj.1	88	54	Inf
FLJ30058	uc004ewa.1,uc004evz.1	238	75	Inf
XPNPEP2	uc004eut.1	169	31	Inf
XIAP	uc010nqv.1,uc010nqu.1,uc004etx.1	227	73	Inf
GRIA3	uc004etr.2,uc004etq.2	301	67	Inf
SLC25A43	uc004erd.1	171	82	Inf
LONRF3	uc004eqz.1,uc004eqx.1,uc004eqw.1	100	103	Inf
IL13RA1	uc004eqt.1,uc004eqs.1	214	126	Inf
PLS3	uc010nqf.1,uc010nqg.1,uc004eqd.1	198	84	Inf
ZCCHC16	uc004epo.1	392	30	Inf
PAK3	uc010npu.1,uc010npt.1,uc004eoz.2	307	36	Inf
ATG4A	uc004ens.1,uc004ent.1,uc004enr.1	76	52	Inf
IL1RAPL2	uc004elz.1	896	106	Inf
DACH2	uc010nmq.1,uc004eex.1,uc004eew.1	459	63	Inf
CHIC1	uc004ebl.2,uc004ebk.1	92	66	Inf
OGT	uc004eab.1,uc004eaa.1	142	53	Inf
NLGN3	uc010nlb.1,uc004dzc.1,uc004dzb.1	169	42	Inf
MED12	uc004dyz.1,uc004dyy.1	166	41	Inf
DLG3	uc004dyj.1,uc004dyi.1	192	35	Inf
KIF4A	uc004dyf.1,uc010nkw.1,uc004dyg.1	180	43	Inf

EDA	uc004dxm.1,uc004dxn.1,uc004dxq.1,uc004dxp.1,uc004dxs.1,uc004dxr.1	223	98	Inf
STARD8	uc004dxc.2	235	85	Inf
MSN	uc004dwf.1	162	58	Inf
GNL3L	uc004dth.1	178	47	Inf
CCNB3	uc004dox.2	140	33	Inf
CLCN5	uc004doq.1,uc004dor.1	203	92	Inf
CCDC22	uc004dnd.1	158	31	Inf
ZNF81	uc010nhv.1	75	26	Inf
DKFZp434P0	uc004dhe.1	328	67	Inf
RBM10	uc004dhi.1,uc010nhq.1,uc004dhh.1,uc004dhg.1,uc004dhf.1	500	60	Inf
RP2	uc004dgw.1	121	74	Inf
TSPAN7	uc004deq.2	154	96	Inf
LANCL3	uc004ddp.1	89	55	Inf
PRRG1	uc004ddo.1,uc004ddn.1	128	94	Inf
PRGP1	uc004ddm.2	113	94	Inf
PDK3	uc004dbh.1,uc004dbq.1	185	93	Inf
PTCHD1	uc010nfu.1,uc004dal.2	85	163	Inf
YY2	uc010nfq.1	129	60	Inf
MBTPS2	uc004dab.1,uc004dac.1	177	60	Inf
CDKL5	uc004cym.1	637	100	Inf
REPS2	uc004cxw.1,uc004cxv.1	404	75	Inf
SYAP1	uc004cxo.1,uc004cxp.1	165	58	Inf
CA5B	uc004cxe.1	180	54	Inf
OFD1	uc004cvv.2,uc004cvu.2,uc010nen.1,uc004cvq.2,uc004cvr.2,uc004cvp.2	97	63	Inf
PRPS2	uc004cvb.1,uc004cva.1	99	144	Inf
FRMPD4	uc004cuz.1	769	101	Inf
KIAA1280	uc010nds.1	991	171	Inf
WWC3	uc004csx.2	991	171	Inf
GYG2	uc004cqw.1	257	50	Inf
SURF6	uc004cdb.2	288	149	Inf
TOR1A	uc004bvl.1	174	153	Inf
CRAT	uc004bxq.1	399	49	Inf
ANGPTL2	uc010mxq.1,uc004bqr.1	447	118	Inf
PPP6C	uc010mww.1,uc004bpq.2,uc010mwv.1	232	220	Inf
RBM18	uc004bma.1	109	106	Inf
NDUFA8	uc004blv.1	76	86	Inf
ASTN2	uc004bjv.1,uc004bjp.1,uc004bjq.1,uc004bjt.1,uc004bjs.1,uc004bjr.1	150	111	Inf
DFNB31	uc004bix.1	424	59	Inf
FKBP15	uc004bgt.2,uc004bgs.2	144	99	Inf
KIAA0674	uc010mut.1	171	99	Inf
C9orf80	uc004bgg.1	80	91	Inf
PTPN3	uc010mtv.1	344	34	Inf
ABCA1	uc004bcm.2,uc004bcl.1	408	225	Inf
KIAA0573	uc010msz.1	350	131	Inf
ERP44	uc004bam.1	350	131	Inf
ANKS6	uc004ayt.1	484	41	Inf
HIATL2	uc004aws.1	117	31	Inf
CDC14B	uc004awi.1	446	120	Inf
hfrc	uc004awe.1,uc004awd.1	263	163	Inf
SLC35D2	uc010msf.1,uc010mse.1,uc004awc.1	488	163	Inf
FAM120AOS	uc004atu.2	144	258	Inf
IPPK	uc004asl.1	863	175	Inf
CCRK	uc004apu.1,uc004apt.1,uc004aps.1,uc004apr.1	113	90	Inf
C9orf64	uc004anc.1,uc004anb.1	145	141	Inf
GKAP1	uc004amz.1,uc004amy.1	236	76	Inf
UBQLN1	uc004amw.1,uc004amv.1	175	217	Inf
PRUNE2	uc004akn.1,uc010mpk.1	343	146	Inf
TRPM6	uc004ajn.1,uc010mpe.1,uc010mpd.1,uc010mpc.1,uc004ajk.1,uc004ajl.1	235	106	Inf
FAM108B1	uc004aim.1,uc004ail.1	151	270	Inf
MCART1	uc004aau.1,uc004aav.1	81	194	Inf
FBXO10	uc004aab.1	584	163	Inf
PAX5	uc010mls.1,uc010mlr.1,uc010mlp.1,uc010mlo.1,uc003zzo.1	2969	136	Inf
RNF38	uc003zzl.1,uc003zzi.1,uc003zzk.1,uc003zzm.1,uc003zzh.1,uc003zzj.1	271	322	Inf
KIAA1539	uc003zwm.1,uc003zwl.1	102	84	Inf
FLJ00135	uc003zwg.1	113	104	Inf
FLJ00350	uc003zwc.1	113	104	Inf
PIGO	uc003zwe.1,uc003zwd.1,uc003zwf.1	113	104	Inf
VCP	uc010mki.1,uc010mkh.1,uc003zvy.2	95	128	Inf
UBAP2	uc003ztr.1,uc003ztq.1	862	89	Inf
APTX	uc003zrr.1,uc003zru.1,uc003zrq.1,uc003zrp.1,uc003zrt.1,uc003zro.1,uc003zry.1,uc003zrw.1,uc003zrx.1,uc010	107	63	Inf
NDUFB6	uc003zrf.1,uc003zre.1	158	53	Inf
TOPORS	uc003zrc.1,uc003zrb.1	82	217	Inf
DDX58	uc010mjk.1,uc003zra.1	154	58	Inf
ELAVL2	uc003zpw.1,uc003zpv.1,uc003zpu.1	685	111	Inf
CDKN2A	uc003zpj.1,uc003zpk.1	165	123	Inf
PSIP1	uc003zlz.2	136	139	Inf
TTC39B	uc010mif.1,uc010mie.1,uc003zlr.1	466	171	Inf
FREM1	uc003zll.1,uc003zlm.1	106	25	Inf
ZDHHC21	uc003zlq.1,uc003zli.1	218	142	Inf
NFIB	uc003zlf.1,uc003zle.1	507	248	Inf
PTPRD	uc003zkn.1,uc003zko.1,uc003zkm.1,uc003zkl.1,uc003zkk.1	3789	75	Inf
KIAA2026	uc003ziq.2	152	205	Inf
ERMP1	uc003zjn.1,uc010mhs.1,uc003zjm.1	80	194	Inf
C9orf68	uc003zim.2,uc010mhj.1	223	47	Inf
RFX3	uc010mhe.1,uc010mhd.1,uc003zhs.1,uc003zhr.1	427	140	Inf
C9orf7	uc004cec.1,uc010nan.1,uc004ced.1	332	194	Inf
GTF3C5	uc010mzz.1,uc004cci.2,uc004ccj.2	419	149	Inf
GFI1B	uc004ccg.1	348	51	Inf
BAT2L	uc004can.2	1535	30	Inf
EXOSC2	uc004bzu.2	77	67	Inf
USP20	uc004byt.1,uc004byr.2,uc004bys.2	1285	192	Inf
DKFZp781M1	uc010myr.1	568	198	Inf
PPP2R4	uc004bxo.1,uc004bxn.1,uc004bxl.1,uc004bxm.1	568	198	Inf
PHYHD1	uc004bwp.2	221	80	Inf
SET	uc004bvt.2	622	28	Inf
SPTAN1	uc004bvn.2,uc004bvm.2,uc004bvl.2	1282	311	Inf
ODF2	uc004bve.1,uc004bvc.1,uc004bvb.1,uc004bvd.2	460	144	Inf
URM1	uc004buv.1	319	78	Inf

SLC25A25	uc004btb.1	837	187	Inf
STXBP1	uc004brl.1,uc004brk.1	822	298	Inf
MRRF	uc010mwa.1,uc004bmc.1,uc004bmb.1	201	108	Inf
TRIM32	uc004bjw.2,uc004bjx.2	108	122	Inf
ZNF618	uc004bib.1,uc004bid.1,uc004bic.1	1395	206	Inf
RGS3	uc010muz.1,uc010mva.1,uc004bhw.1	399	69	Inf
SLC31A1	uc004bgv.2,uc004bgu.1	192	101	Inf
HSDL2	uc004bgc.1,uc004bga.1	399	190	Inf
DKFZp666L09	uc004bgb.1	399	190	Inf
AKAP2	uc004bem.1	650	162	Inf
FKTN	uc004bcs.1,uc004bcr.1	124	74	Inf
SMC2	uc004bbu.1,uc004bbw.1,uc004bbv.1,uc004bbx.1	80	94	Inf
RNF20	uc004bbn.1	89	60	Inf
PRG-3	uc004bbb.1	505	205	Inf
INVS	uc010mta.1,uc010mtb.1,uc004bar.1,uc004baq.1,uc004bap.1,uc004bao.1	522	114	Inf
NR4A3	uc004bae.1,uc004bai.1,uc004bag.1,uc004baf.1	439	288	Inf
TGFBR1	uc004azd.1,uc004azc.1	110	315	Inf
ANP32B	uc004aya.1	191	631	Inf
TDRD7	uc004axj.1	197	294	Inf
HABP4	uc010msh.1,uc010msg.1	351	290	Inf
SYK	uc004arc.1	664	36	Inf
S1PR3	uc004aqe.1	267	174	Inf
CR613032	uc004aoh.1	82	35	Inf
GCNT1	uc004akf.2,uc010mph.1	247	141	Inf
OSTF1	uc004aix.2,uc004aiw.2,uc004aiv.2	222	189	Inf
C9orf61	uc010moo.1,uc010mon.1	102	34	Inf
RG9MTD3	uc004aak.1,uc004aai.1	113	69	Inf
FRMPD1	uc004aag.1	631	195	Inf
UNC13B	uc010mkl.1	612	128	Inf
UNC13B	uc003zwr.1,uc003zwq.1	799	128	Inf
KIAA1045	uc003zvg.1	79	238	Inf
IL11RA	uc003zvi.1	108	54	Inf
NFX1	uc010mjr.1,uc003zsr.1,uc003zso.1,uc003zsp.1,uc003zsq.1	303	90	Inf
MTAP	uc003zpi.1,uc003zph.1	1683	134	Inf
JMJD2C	uc010mhv.1,uc003zkg.1,uc003zkh.1	390	277	Inf
KIAA0780	uc010mhw.1	792	277	Inf
KANK1	uc003zgs.1	304	96	Inf
TRAPPC9	uc003yvi.1	9100	86	Inf
LRRC6	uc010mdu.1,uc003ytk.1	265	143	Inf
TATDN1	uc010mdm.1,uc003yrf.1,uc003yrd.1	219	103	Inf
DERL1	uc003ypn.1,uc003ypm.1,uc003ypl.1	126	155	Inf
MRPL13	uc003ypa.1	89	51	Inf
SAMD12	uc003yom.2,uc010mda.1	606	112	Inf
EIF3H	uc003yoa.1	290	48	Inf
CSMD3	uc003ynx.2,uc010mcx.1,uc003ynv.1,uc003ynu.1	396	55	Inf
NUDCD1	uc010mcl.1,uc003ynb.2	224	111	Inf
EIF3E	uc010mci.1,uc003ymv.1,uc003ymu.1	125	59	Inf
PABPC1	uc003yjt.1,uc003yjs.1	100	275	Inf
SNX31	uc003yjr.1	360	146	Inf
ANKRD46	uc003yjm.2,uc003yjp.1,uc003yjn.1,uc003yjo.1	166	140	Inf
STK3	uc003yio.1	1837	300	Inf
RPL30	uc010mbk.1	82	81	Inf
KIAA1429	uc003ygp.1,uc003ygo.1	199	69	Inf
FAM82B	uc003ydu.1	117	103	Inf
SNX16	uc003yco.1,uc003ycn.1	76	130	Inf
MRPS28	uc003ybo.1,uc003ybp.1	232	106	Inf
TPD52	uc003ybt.1,uc003ybs.1	510	193	Inf
MSC	uc003xyx.1	124	197	Inf
EYA1	uc003xyv.1,uc003xyu.1	438	39	Inf
COPS5	uc003xxd.1,uc003xxf.1,uc003xxe.1	86	37	Inf
PDE7A	uc003xvq.1,uc003xvr.1	322	141	Inf
ARMC1	uc003xvl.1	190	85	Inf
PLAG1	uc010lyj.1,uc010lyi.1,uc003xsr.2	234	202	Inf
TCEA1	uc003xrv.1,uc003xru.1	248	153	Inf
PLAT	uc003xot.2,uc003xos.2	581	37	Inf
ZMAT4	uc003xns.1,uc003xnr.1	1324	56	Inf
FLJ43582	uc003xlx.1	142	33	Inf
FGFR1	uc003xlt.2	566	31	Inf
DUSP26	uc003xjq.1,uc003xjp.1	110	93	Inf
C8orf41	uc003xjl.2,uc003xjm.2	75	40	Inf
GTF2E2	uc003xiq.1	363	193	Inf
INTS9	uc003xhb.1,uc003xha.1	316	51	Inf
PNMA2	uc003xez.2	77	186	Inf
TNFRSF10A	uc003xda.1	241	158	Inf
CSGALNACT	uc003wzf.1	1289	25	Inf
ASAH1	uc003wym.2,uc003wyl.2,uc003wyo.2,uc003wyn.2	137	155	Inf
MTMR7	uc003wxm.1	575	120	Inf
SGCZ	uc003wwq.1	2273	171	Inf
DLC1	uc003wwl.1	1245	118	Inf
LOC649305	uc003wwc.2	223	67	Inf
PINX1	uc003wti.2,uc003wth.2	630	119	Inf
ERICH1	uc003wpe.1,uc003wph.1	3194	151	Inf
ZNF707	uc003yzh.2,uc003yzf.2,uc010mfi.1,uc010mfh.1,uc003yze.2	468	109	Inf
PHF20L1	uc003yts.1,uc003ytt.1	76	159	Inf
EFR3A	uc003yte.1	166	276	Inf
MTBP	uc003ypc.1	153	39	Inf
TTC35	uc003ymw.1	103	134	Inf
TM7SF4	uc003vlx.1	91	29	Inf
C8orf47	uc003yih.1	115	153	Inf
CPNE3	uc003ydv.1	109	126	Inf
WWP1	uc003ydt.1	394	221	Inf
CHMP4C	uc003ycl.1	83	75	Inf
PKIA	uc003ybb.1,uc003vba.1	120	140	Inf
CRISPLD1	uc003yan.1	83	131	Inf
AK056720	uc003xzl.2	237	43	Inf
SULF1	uc003xyg.2,uc003xyf.2	771	48	Inf
C8orf46	uc003xwg.1	253	33	Inf
DKFZp451J08	uc003xuz.1,uc003xva.1,uc010lys.1	144	197	Inf

YTHDF3	uc003xuy.1	144	200	Inf
RLBP1L1	uc003xug.1,uc010lyp.1,uc003xuh.1	185	63	Inf
RP1	uc003xsc.1	249	30	Inf
SNTG1	uc010lxz.1,uc003xqs.1	1715	29	Inf
KIAA0146	uc010lxs.1,uc003xqe.1,uc003xqd.1	2158	186	Inf
FNTA	uc003xpu.1,uc003xpt.1,uc003xps.1	126	228	Inf
POLB	uc003xpa.1,uc003xoz.1	149	69	Inf
LETM2	uc003xlo.1,uc003xln.1,uc003xlm.1,uc003xll.1	122	124	Inf
DCTN6	uc003xhy.1	122	76	Inf
HMBOX1	uc003xhc.2,uc010lvd.1,uc003xhd.2,uc003xhe.1	421	114	Inf
ELP3	uc003xgo.2	347	54	Inf
PTK2B	uc003xfq.1,uc003xfp.1	1017	95	Inf
DKFZp762K0	uc003xeq.1	175	116	Inf
CDCA2	uc003xep.1	175	116	Inf
TNFRSF10C	uc003xcy.1	88	71	Inf
XPO7	uc003xaa.2,uc010lti.1	328	255	Inf
ATP6V1B2	uc003wzp.1	87	124	Inf
INTS10	uc003wzj.1	182	119	Inf
PCM1	uc003wyg.2,uc003wyh.2,uc010lta.1,uc003wyi.2	82	212	Inf
TUSC3	uc003wwr.1,uc003wwy.1,uc003wws.1,uc003www.1,uc003wwv.1,uc003wwu.1,uc003wwt.1	429	129	Inf
FDFT1	uc003wuh.1	481	29	Inf
MTMR9	uc010lrx.1,uc003wtm.1	118	128	Inf
AGPAT5	uc003wqo.1	220	245	Inf
CLN8	uc003wpo.2	422	194	Inf
FBXO25	uc003woz.1,uc003wov.1,uc003wox.1	264	97	Inf
FAM62B	uc003woc.1,uc003wob.1,uc003wod.1	1117	31	Inf
SMARCD3	uc003wju.1	1228	124	Inf
FAM115A	uc003wdp.1,uc003wdo.1	253	109	Inf
SLC37A3	uc010lnh.1,uc003vvp.1,uc003vvo.1	543	152	Inf
CREB3L2	uc003vty.2,uc003vtx.1,uc003vtw.1	487	116	Inf
PTN	uc003vtr.1,uc010lmx.1,uc003vtq.1	287	32	Inf
DKFZp586B0	uc010lmq.1	251	26	Inf
WDR91	uc003vsp.1	289	214	Inf
C7orf49	uc003vsh.1	280	184	Inf
UBE2H	uc003vpg.1,uc003vpf.1	683	149	Inf
TNPO3	uc010llz.1,uc010llv.1,uc003vol.1	372	90	Inf
TRN-SR	uc003vom.1	372	90	Inf
RBM28	uc003vmp.2	181	78	Inf
POT1	uc003vlo.1,uc003vlm.1	189	68	Inf
IQUB	uc003vko.1	100	69	Inf
CTTNBP2	uc003vjf.1	478	333	Inf
WNT2	uc003vja.1,uc003viz.1	276	187	Inf
COG5	uc003vee.1,uc003ved.1,uc003vec.1	1066	106	Inf
SLC26A5	uc003vbv.1,uc003vbu.1,uc003vbz.1,uc003vbt.1	349	193	Inf
ZRF1	uc003vbp.1	190	106	Inf
DNAJC2	uc010lix.1,uc003vbo.1	190	106	Inf
C7orf18	uc003vbf.2	90	173	Inf
NAPEPLD	uc003vbd.2,uc003vbc.2	180	173	Inf
ALKBH4	uc003uzm.1,uc003uzl.1	231	200	Inf
RABL5	uc003uym.1,uc003uyl.1	90	139	Inf
LRCH4	uc003uvj.1	558	147	Inf
STAG3OS	uc003uua.2,uc003uuc.1	723	96	Inf
DKFZp761F1	uc010lqq.1	121	56	Inf
GAL3ST4	uc003utu.1	121	56	Inf
BAIAP2L1	uc003upj.1	1665	181	Inf
TECPR1	uc003uph.1,uc003upg.1	958	158	Inf
CALCR	uc003umw.2,uc003umv.1,uc003umu.1	179	96	Inf
CDK6	uc003ume.1	485	253	Inf
PEX1	uc010ley.1,uc003uly.1	106	184	Inf
KRIT1	uc003ulq.1	135	40	Inf
SEMA3E	uc003uhy.1	502	26	Inf
STYXL1	uc003uek.2,uc010ldh.1,uc003uej.2,uc003uem.1,uc003uel.1	516	132	Inf
MKSTYX	uc010ldg.1	508	132	Inf
HIP1	uc003uds.1	2401	163	Inf
POM121C	uc003udk.2	192	26	Inf
MLXIPL	uc003typ.1,uc003tyn.1,uc003tym.1,uc003tyl.1,uc003tyk.1	586	251	Inf
BAZ1B	uc003tyc.1	525	266	Inf
LOC401365	uc010lac.1	858	177	Inf
TNS3	uc010kyo.1,uc003tnv.1	2054	40	Inf
DDX56	uc003tlf.1	86	119	Inf
NPC1L1	uc003tla.1	102	44	Inf
POLR2J2	uc003tjd.1	153	40	Inf
C7orf44	uc003tip.1,uc003tiq.1,uc003tio.1,uc003tin.1	311	91	Inf
DPY19L2P1	uc003teq.1	124	57	Inf
CRH2R	uc010kvy.1	288	213	Inf
CRHR2	uc010kvx.1,uc010kvw.1,uc003tbn.1	288	213	Inf
KIAA0644	uc003szt.1	227	378	Inf
SKAP2	uc003syc.1	324	69	Inf
IGF2BP3	uc003swq.1	879	265	Inf
DRCTNNB1A	uc003svn.2	100	158	Inf
FAM126A	uc003svm.2	100	158	Inf
MGC87042	uc010kum.1,uc003svh.2	197	30	Inf
DKFZp761E0	uc003sty.2	150	152	Inf
KIAA0713	uc003stx.1,uc003stw.1,uc010kuc.1	164	152	Inf
SNX13	uc003stv.1	344	152	Inf
ETV1	uc003ssw.2	264	33	Inf
THSD7A	uc003ssf.2	686	99	Inf
ICA1	uc003srs.1,uc003srp.2,uc003srr.1,uc003sro.2,uc003srn.2,uc003srq.1,uc003srm.1	281	202	Inf
DAGLB	uc003sqd.2	1291	194	Inf
KDELR2	uc003sqf.2,uc003sqe.2	182	194	Inf
RNF216	uc003soz.1,uc003soy.1,uc003sox.1	1255	162	Inf
C7orf27	uc003smj.1,uc003smi.1	298	157	Inf
KIAA0010	uc003wni.2	627	30	Inf
NUB1	uc003wjx.1,uc003wjw.1	497	161	Inf
DKFZp761E2	uc003wjk.1	167	74	Inf
ZNF282	uc003wfm.1	910	230	Inf
ZNF398	uc003wfk.1	1050	184	Inf
EPHB6	uc003wbt.1	322	29	Inf

Gene	Transcripts			
CHRM2	uc003vtm.1,uc003vto.1,uc003vtl.1,uc003vtg.1,uc003vtn.1,uc003vtk.1,uc003vtj.1,uc003vtf.1,uc003vti.1	316	262	Inf
CALD1	uc003vsd.1,uc003vsc.1	375	36	Inf
BPGM	uc003vrw.1,uc003vrv.1	85	41	Inf
EXOC4	uc003vri.2,uc003vrj.1,uc003vrk.1	605	25	Inf
FAM40B	uc003vow.1,uc003vox.1	181	198	Inf
CCDC136	uc003vny.1	114	27	Inf
CALU	uc003vns.1,uc003vnr.1,uc003vnq.1	132	145	Inf
METTL2B	uc003vng.1,uc003vnf.1	131	46	Inf
SND1	uc003vmi.1	2821	178	Inf
CFTR	uc003vjd.1	331	77	Inf
HBP1	uc003vdy.1	124	187	Inf
RINT1	uc010ljj.1,uc003vda.1	260	96	Inf
DKFZp586I12	uc010liw.1,uc010liv.1,uc003vbm.1	77	93	Inf
PMPCB	uc003vbk.1,uc003vbl.1	76	93	Inf
PRKRIP1	uc003uzh.2	434	72	Inf
SERPINE1	uc003uxt.2	175	52	Inf
TRIM56	uc003uxr.2,uc003uxq.1	317	86	Inf
FBX24	uc003uvn.1	228	178	Inf
FBXO24	uc003uvl.1,uc003uvm.1	154	178	Inf
ZNF789	uc010lfw.1,uc003uqq.1	122	94	Inf
MYH16	uc003upw.1	361	27	Inf
LMTK2	uc003upd.1	1385	183	Inf
KIAA1861	uc003ump.1	145	38	Inf
CCDC132	uc003umo.1	145	38	Inf
PFTK1	uc003ukv.1	1274	294	Inf
GTPBP10	uc003uko.1,uc003ukn.1,uc003ukm.1	99	64	Inf
ZNF804B	uc003uju.1	791	160	Inf
ADAM22	uc003uji.1,uc003ujl.1,uc003ujp.1,uc003ujk.1,uc003ujo.1,uc003ujm.1,uc003ujn.1	387	194	Inf
PHTF2	uc003ugp.2,uc010ldv.1,uc003ugo.2,uc003ugq.2	480	206	Inf
MDH2	uc003uep.1,uc003ueo.1	420	125	Inf
RHBDD2	uc003udw.1,uc003udv.1	155	90	Inf
ELN	uc003tzq.1,uc003tzp.1,uc003tzo.1,uc003tzz.1,uc003tzy.1,uc003tzx.1,uc003tzw.1,uc003tzv.1,uc003tzu.1,uc003tz	757	115	Inf
POM121	uc010lam.1,uc003twj.1	399	25	Inf
AUTS2	uc003tvv.2,uc003tvx.2,uc003tvw.2	1754	258	Inf
STAG3L4	uc003tvt.2	89	151	Inf
TYW1	uc003tvn.1	715	29	Inf
C7orf42	uc003tvl.1,uc003tvk.1	367	32	Inf
TPST1	uc010laa.1,uc010kzz.1,uc003tuw.1	548	197	Inf
VKORC1L1	uc003tul.1	352	160	Inf
ZNF273	uc003tto.1,uc003ttn.1	85	53	Inf
ZNF138	uc003tth.1,uc010kzs.1,uc003ttg.1	96	52	Inf
ZNF107	uc003tte.1,uc003ttd.1	223	41	Inf
DKFZP566I10	uc003trv.1	190	101	Inf
SUMF2	uc003trt.1,uc003trs.1,uc003trr.1,uc003trq.1,uc003trp.1	190	101	Inf
CMAP	uc003tjn.2,uc003tjr.2	171	117	Inf
DBNL	uc003tjp.2,uc003tjq.2,uc003tjo.2	238	117	Inf
RALA	uc003thd.1	261	232	Inf
BMPER	uc003tdw.1	828	196	Inf
BBS9	uc003tdq.1,uc003tdp.1,uc003tdo.1,uc003tdn.1	784	74	Inf
ADCYAP1R1	uc003tcf.1	125	29	Inf
GHRHR	uc003tbw.1,uc003tbx.1	114	36	Inf
FAPP2	uc003tao.1	226	197	Inf
PLEKHA8	uc003tam.1,uc003tap.1,uc003tan.1	115	197	Inf
TAX1BP1	uc003szl.1,uc003szk.1	161	127	Inf
CCDC126	uc003swn.1,uc003swm.1,uc003swl.1	145	57	Inf
KLHL7	uc003svs.2,uc003svr.2	234	122	Inf
TSPAN13	uc003stq.1	105	168	Inf
ZDHHC4	uc003sqm.1,uc003sql.1,uc003sqj.1,uc003sqi.1,uc003sqh.1,uc003sqk.1	87	97	Inf
JTV1	uc003spo.1	138	135	Inf
EIF3B	uc003slz.1,uc003sly.1,uc003slx.1	109	225	Inf
UNCX	uc003skq.1	386	265	Inf
GPER	uc003ska.1,uc003sjz.1	383	78	Inf
UNC84A	uc010ksb.1,uc003sjf.1	1388	46	Inf
THBS2	uc003qwt.1	1388	57	Inf
PARK2	uc010kke.1,uc003qtz.2,uc003qtv.2,uc003qtx.2	7523	196	Inf
RSPH3	uc010kju.1,uc003qrx.1	103	165	Inf
EZR	uc003qrt.2	607	137	Inf
SERAC1	uc003qrc.1	303	180	Inf
CNKSR3	uc003qpy.1	803	297	Inf
SYNE1	uc010kiy.1	102	75	Inf
ZC3H12D	uc010kid.1,uc003qmm.2	915	56	Inf
HIVEP2	uc003qjd.1	856	211	Inf
AX747618	uc003qid.1	226	39	Inf
MAP	uc010kqv.1	585	26	Inf
MAP7	uc010kgr.1,uc010kgs.1,uc010kgt.1,uc010kgq.1,uc003qha.1,uc003qgz.1	664	26	Inf
DKFZp686H1	uc003qep.2	357	43	Inf
SGK1	uc003qeo.2	1386	43	Inf
SLC2A12	uc003qem.1	271	43	Inf
STX7	uc003qdq.2	160	103	Inf
ARHGAP18	uc003qbr.1	440	44	Inf
C6orf174	uc003qbq.2,uc003qbf.1	246	61	Inf
ECHDC1	uc010kez.1,uc003qav.2,uc010kev.1,uc003qaz.2	87	115	Inf
ZUFSP	uc003pxf.1	76	51	Inf
LAMA4	uc003pvw.2,uc003pvu.2,uc003pvt.2,uc003pvv.2	102	113	Inf
TRAF3IP2	uc010kdx.1,uc010kdw.1,uc003pvg.1,uc003pvf.1,uc003pve.1	220	34	Inf
DDO	uc003pud.1,uc003puc.1	280	25	Inf
RTN4IP1	uc003prk.1	251	80	Inf
ASCC3	uc003pql.2,uc003pqk.1	298	153	Inf
USP45	uc003ppz.1,uc003pqa.1,uc010kcq.1,uc003ppx.1	139	168	Inf
SFRS18	uc003ppq.2,uc003ppp.2,uc003ppo.2	116	55	Inf
AKIRIN2	uc003pmk.1	135	200	Inf
KIAA1009	uc003pkk.2,uc003pkj.2,uc010kbp.1	85	61	Inf
UBE2CBP	uc003pjp.1	370	84	Inf
IBTK	uc003pjm.1,uc003pjl.1	206	116	Inf
KIAA1417	uc010kbi.1	246	116	Inf
LCA5	uc003pix.1,uc003piy.1	92	80	Inf
SLC17A5	uc003phn.2	393	163	Inf
C6orf150	uc003pgx.1	128	191	Inf

83

B3GAT2	uc003pfw.2,uc003pfv.1	318	408	Inf
COL9A1	uc003pfe.2	217	34	Inf
LMBRD1	uc010kal.1,uc003pfa.1	200	64	Inf
HMGCLL1	uc010jzx.1,uc003pco.1,uc003pcn.1	137	70	Inf
C6orf138	uc003ozf.2,uc003oze.2	253	102	Inf
GPR116	uc003oyr.2,uc003oyp.2,uc003oyo.2	166	42	Inf
RCAN2	uc003oyb.1	195	28	Inf
NFKBIE	uc003oxe.1	77	152	Inf
TRERF1	uc003osc.1,uc003osb.1,uc003ose.1,uc003osd.1	1689	34	Inf
USP49	uc003ori.1	620	161	Inf
FRS3	uc003orc.1	171	143	Inf
MOCS1	uc003ope.2,uc003opc.2,uc003opb.2,uc003opa.2,uc003opd.2	300	74	Inf
KCNK5	uc003oon.1	444	207	Inf
BTBD9	uc010jwx.1,uc003ooa.2	1895	145	Inf
C6orf129	uc003ont.1	237	38	Inf
TMEM217	uc003onm.2,uc010jws.1,uc010jwr.1,uc003onl.1	169	78	Inf
PPIL1	uc003omu.1	122	77	Inf
STK38	uc003omi.1,uc003omh.1	258	97	Inf
AK309286	uc010jqw.1	226	63	Inf
DCDC2	uc003ndx.1	424	87	Inf
DTNBP1	uc010jph.1,uc003nbp.1,uc003nbm.1,uc003nbl.1	579	226	Inf
CCDC90A	uc003nbd.1	106	249	Inf
GFOD1	uc003nas.1	781	45	Inf
NEDD9	uc003mzx.1,uc003mzv.1	126	30	Inf
MAK	uc003mzm.1	435	102	Inf
F13A1	uc003mwv.1	530	57	Inf
TBP	uc003qxu.1,uc003qxt.1	77	86	Inf
PACRG	uc003quc.1,uc003qub.1,uc003qua.1	4487	137	Inf
GTF2H5	uc003qrd.1	109	183	Inf
OPRM1	uc010kjf.1,uc003qpq.1,uc003qpt.1,uc003qpo.1,uc003qpr.1,uc003qpn.1	235	70	Inf
C6orf97	uc003qol.1	594	219	Inf
C6orf211	uc003qok.1	89	171	Inf
C6orf103	uc003qlp.1,uc010khx.1	475	64	Inf
DREG	uc003qix.2	107	250	Inf
GPR126	uc010khf.1,uc010khe.1,uc010khd.1,uc010khc.1	230	250	Inf
VTA1	uc003qiw.1	148	69	Inf
TBPL1	uc003qel.1	112	137	Inf
ENPP1	uc003qcx.2	267	203	Inf
ENPP3	uc003qcu.2	295	73	Inf
LAMA2	uc010kfe.1,uc003qbo.1,uc003qbn.1	1397	53	Inf
RSPO3	uc003qas.1,uc003qar.1	106	336	Inf
TRMT11	uc010kev.1,uc003qam.1	97	95	Inf
NCOA7	uc003qae.2,uc010kes.1	773	27	Inf
IBRDC1	uc003pzr.1	273	297	Inf
NKAIN2	uc003pzn.1,uc003pzo.1	1378	248	Inf
TCBA1	uc010keq.1,uc003pzp.1	1624	248	Inf
PKIB	uc003pyz.1	758	79	Inf
CDC40	uc003pua.1	92	50	Inf
FIG4	uc003ptt.2	370	74	Inf
QRSL1	uc003prl.1,uc003prm.1	119	80	Inf
AIM1	uc003prh.1	260	317	Inf
PRDM1	uc003prd.2	252	78	Inf
KLHL32	uc003poy.2,uc010kcm.1	533	163	Inf
KIAA1900	uc003poz.2	579	163	Inf
ORC3L	uc003pmi.1,uc003pmh.1,uc003pmg.1	274	95	Inf
DOPEY1	uc003pjs.1	195	144	Inf
CD109	uc010kaz.1,uc010kba.1,uc003phq.1,uc003php.1	331	161	Inf
KCNQ5	uc003pgj.2,uc010kat.1,uc003pgk.1	936	285	Inf
SMAP1	uc010kap.1	557	79	Inf
FAM135A	uc003pfh.2,uc003pfi.2	394	189	Inf
COL19A1	uc003pfc.1	629	170	Inf
BAI3	uc010kak.1,uc003pev.2	1087	70	Inf
ZNF451	uc003pdk.1,uc003pdl.2,uc003pdn.1,uc003pdm.1	146	88	Inf
EFHC1	uc003pap.2	132	95	Inf
PAQR8	uc003pao.2	239	246	Inf
CENPQ	uc003ozh.1	90	54	Inf
CDC5L	uc003oxl.1	142	76	Inf
DKFZp547K1	uc010jyy.1	111	184	Inf
TMEM63B	uc003owq.1,uc003owr.1	178	188	Inf
POLR1C	uc003ovo.1	75	83	Inf
TTBK1	uc003ouq.1	1329	130	Inf
PARC	uc003ouj.1	105	90	Inf
KIAA0708	uc010jyk.1	595	90	Inf
CUL9	uc003oul.1,uc003ouk.1	595	90	Inf
PTK7	uc010jyj.1	427	55	Inf
KLC4	uc003otu.1,uc003otv.1	84	93	Inf
FLJ00338	uc003otb.2,uc003otc.2	102	120	Inf
UNQ1934	uc003osz.2	102	120	Inf
CNPY3	uc003ota.2	102	120	Inf
UBR2	uc003osf.2,uc003osg.1	159	67	Inf
BYSL	uc003orl.1	123	92	Inf
RNF8	uc003onr.2,uc003onq.2	117	143	Inf
TBC1D22B	uc003onn.1	356	74	Inf
FLJ00276	uc003onq.2	432	27	Inf
FGD2	uc010jwp.1	432	27	Inf
C6orf89	uc003omx.1,uc003omw.1	203	58	Inf
AK125083	uc010jwl.1	158	41	Inf
MAPK14	uc003olo.1,uc003olr.1,uc003olq.1,uc003olp.1	216	160	Inf
ZKSCAN3	uc003nlf.2,uc010jrc.1,uc003nle.2	105	40	Inf
SCGN	uc010jpz.1,uc003nfb.1	123	49	Inf
LRRC16A	uc010jpy.1,uc010jpx.1	971	177	Inf
THEM2	uc010jpv.1,uc003nek.1	129	131	Inf
NRSN1	uc010jpq.1	85	62	Inf
E2F3	uc003ncz.2,uc003nda.2	439	364	Inf
RNF144B	uc003ncs.1	192	42	Inf
GMPR	uc003nbs.1	593	152	Inf
JARID2	uc003nbj.1	2373	104	Inf
HIVEP1	uc003nac.1	662	199	Inf

Gene	Transcript			
RREB1	uc003mxe.1	1483	78	Inf
PRPF4B	uc003mvv.1	143	103	Inf
BPHL	uc003mva.1,uc003muy.1	421	197	Inf
RIPK1	uc003muw.2,uc010jni.1,uc003muv.2,uc003mux.1	512	128	Inf
AX748230	uc003mnb.1	252	43	Inf
MGAT1	uc003mmi.2,uc010jlg.1,uc010jlf.1,uc003mmh.2,uc003mmg.2,uc010jlh.1	855	29	Inf
HK3	uc003mfa.1	476	32	Inf
FBXW11	uc003mbn.1,uc003mbl.1,uc003mbm.1	510	180	Inf
GABRB2	uc010jiu.1,uc003lyr.1,uc003lyt.1,uc003lys.1	442	182	Inf
RNF145	uc010jiq.1,uc003lxp.2	113	267	Inf
EBF1	uc003lxl.2,uc010jip.1	1588	138	Inf
HAVCR1	uc003lwi.2,uc010jij.1	141	29	Inf
DKFZp434C1	uc010jhu.1	289	100	Inf
CCDC69	uc003ltq.1	289	100	Inf
DCTN4	uc003lsu.1,uc010jhi.1,uc003lsv.1	186	34	Inf
CD74	uc003lsf.1,uc003lse.1,uc003lsd.1,uc003lsc.1	174	33	Inf
ARSI	uc003lrv.2	190	162	Inf
CSF1R	uc003lrm.1	699	32	Inf
HTR4	uc003lpo.1,uc003lpk.1,uc003lpj.1,uc003lpn.1,uc003lpi.1,uc003lpm.1,uc003lpl.1	263	143	Inf
PCDH1	uc003llq.1,uc003llp.1	394	281	Inf
HARS	uc010jfu.1,uc003lgw.1,uc003lgv.1	116	93	Inf
PFDN1	uc003lff.1	250	31	Inf
NRG2	uc003ley.1,uc003lex.1,uc003lew.1,uc003lev.1	1968	223	Inf
DNAJC18	uc010jff.1,uc003len.1	153	77	Inf
SIL1	uc003ldp.1,uc003ldo.1	1357	98	Inf
CDC25C	uc003lcs.1	364	124	Inf
CDC23	uc003lcl.1	135	56	Inf
BRD8	uc010jer.1	89	25	Inf
FAM13B	uc003lca.2,uc003lcb.2,uc003lbz.2	332	213	Inf
KLHL3	uc010jem.1,uc010jek.1	318	25	Inf
NKIAMRE	uc010jdw.1	444	88	Inf
SKP1	uc010jdv.1,uc003kzd.2,uc003kzc.2	90	199	Inf
VDAC1	uc003kyr.1,uc003kyq.1,uc003kyp.1	221	446	Inf
ZCCHC10	uc003kyh.1,uc003kyg.1	180	85	Inf
AFF4	uc003kyf.2,uc003kye.1,uc003kyd.1	305	202	Inf
KIF3A	uc003kxp.1,uc003kxo.1	141	124	Inf
ZNF608	uc003kts.1,uc003ktq.1	579	37	Inf
PGGT1B	uc010jch.1,uc003kqx.2,uc003kqw.2	85	84	Inf
PJA2	uc003kos.2	192	135	Inf
GIN1	uc003koc.1,uc003kob.1,uc003koa.1	84	42	Inf
SLCO4C1	uc003knm.1	95	121	Inf
ELL2	uc003klr.2	191	258	Inf
FAM172A	uc003kkm.2,uc010jbd.1	463	75	Inf
TMEM161B	uc003kjc.1	142	46	Inf
DHFR	uc003kgx.1,uc003kgv.1	126	159	Inf
TBCA	uc003kfi.1,uc003kfh.1	194	153	Inf
C5orf37	uc003keh.2	101	83	Inf
CENPK	uc003jtu.1,uc003jtt.1,uc003jts.1	152	80	Inf
ADAMTS6	uc003jtp.1	715	27	Inf
SFRS12IP1	uc003jtk.1	89	62	Inf
ERCC8	uc003jsm.2,uc003jsl.2	161	49	Inf
PDE4D	uc003jsc.2,uc003jrz.2,uc003jry.2	956	39	Inf
SLC38A9	uc010ivy.1,uc003jqf.1	218	68	Inf
FGF10	uc003joq.1	83	30	Inf
C5orf28	uc003jny.1,uc003jnv.2,uc003jnx.1	93	115	Inf
OXCT1	uc003jmn.1	205	233	Inf
NUP155	uc010iuz.1,uc003jku.1,uc003jkt.1	507	55	Inf
LMBRD2	uc003jkb.1	95	135	Inf
MTMR12	uc010iul.1,uc010iuk.1,uc003jhq.1	511	235	Inf
CDH12	uc003jgk.1	1715	26	Inf
CDH18	uc003jgd.1	777	44	Inf
FAM134B	uc003jfs.1	794	187	Inf
FAM173B	uc003jeo.2	151	95	Inf
FLJ33360	uc003jdn.1	157	25	Inf
TPPP	uc003jbh.2	1242	141	Inf
CCDC127	uc003jam.1	140	121	Inf
RMND5B	uc003miq.1	119	43	Inf
PRR7	uc003mgw.1	671	47	Inf
FAF2	uc003mej.2	429	103	Inf
C5orf25	uc003mdt.2,uc003mds.2,uc010jka.1	335	26	Inf
HMP19	uc003mcx.1	548	32	Inf
BNIP1	uc003mcl.2,uc003mck.2,uc003mcj.2,uc003mci.2	193	93	Inf
C5orf41	uc003mcg.2,uc003mcf.2,uc003mch.1	289	265	Inf
ATP6V0E1	uc003mcd.1	418	117	Inf
ERGIC1	uc003mbw.2	2172	159	Inf
GABRP	uc003mau.1	167	25	Inf
KCNIP1	uc003map.1	3485	47	Inf
RARS	uc003lzx.1	137	78	Inf
HMMR	uc003lzh.1,uc003lzg.1,uc003lzf.1	113	58	Inf
GABRG2	uc003lyz.2,uc003lyy.2,uc010jic.1	122	26	Inf
FABP6	uc003lya.1	83	27	Inf
CNOT8	uc003lvu.1	109	190	Inf
GRIA1	uc003luz.2,uc003lva.2,uc003luy.2	955	25	Inf
GM2A	uc003ltr.2	135	111	Inf
SLC6A7	uc003lrr.1	414	143	Inf
SLC26A2	uc003lrh.1	85	218	Inf
FLJ41603	uc003lra.1	421	123	Inf
AFAP1L1	uc003lqg.2,uc010jqv.1,uc003lqh.1	400	247	Inf
FBXO38	uc003lph.2,uc003lpg.1,uc003lpf.1	149	50	Inf
STK32A	uc003lom.2,uc003lol.2,uc010jgn.1	370	95	Inf
SH3RF2	uc003lnt.1	500	92	Inf
KCTD16	uc003lnm.1	667	60	Inf
PCDHGB2	uc003lis.1	3625	114	Inf
PCDHGC3	uc003lkw.1,uc003lkv.1	954	179	Inf
PAIP2	uc003led.1	204	154	Inf
FAM53C	uc003lcx.1,uc003lcv.1,uc003lcw.1	147	111	Inf
STSG4523	uc010jev.1	147	111	Inf
SEC24A	uc003kzs.1	641	160	Inf

Gene	Transcripts			
UBE2B	uc003kzh.1	80	169	Inf
TCF7	uc003kzb.1,uc003kyt.1	85	49	Inf
CDC42SE2	uc003kvh.2,uc003kvj.1,uc003kvi.1	546	249	Inf
CSNK1G3	uc003kto.1,uc003ktn.1,uc003ktm.1,uc003ktl.1	163	244	Inf
SNX24	uc010jcy.1,uc003ktf.1,uc003ktg.1	558	172	Inf
SNCAIP	uc010jct.1,uc003ksz.1,uc003ksy.1,uc003ksx.1,uc003ksw.1	287	252	Inf
SRFBP1	uc003kst.1	112	48	Inf
DCP2	uc003kqh.1	137	127	Inf
ZRSR1	uc003kqd.2,uc010jcb.1	151	74	Inf
APC	uc010jbz.1,uc003kpy.2,uc003kpz.2	325	54	Inf
FER	uc003koq.1,uc003kop.1	289	205	Inf
HISPPD1	uc010jbo.1,uc003kod.2	117	53	Inf
CAST	uc003klt.1,uc003kly.1,uc003klx.1,uc003klw.1,uc003klv.1,uc003klu.1	453	283	Inf
RHOBTB3	uc003klm.1	225	280	Inf
XRCC4	uc003kia.1,uc003kie.1,uc003kic.1,uc003kid.1,uc003kib.1	449	53	Inf
MSH3	uc003kqz.1	636	160	Inf
FAM151B	uc003kqv.1	193	179	Inf
ZFYVE16	uc003kgp.2,uc003kgs.2,uc003kgq.2,uc003kgr.2	94	179	Inf
THBS4	uc003kqh.1	324	164	Inf
CMYA5	uc003kqc.1	371	104	Inf
PAPD4	uc003kfz.2,uc003kga.2,uc010jaf.1,uc003kgb.2,uc010jae.1	234	42	Inf
JMY	uc003kfv.1	358	114	Inf
HMGCR	uc003kdq.1,uc003kdp.1	88	163	Inf
TNPO1	uc003kck.2	459	187	Inf
PTCD2	uc003kcc.1,uc003kcb.1	123	73	Inf
MAP1B	uc010iyy.1	167	49	Inf
CENPH	uc010ixc.1,uc003jvp.1	116	78	Inf
SLC30A5	uc003jvh.1	142	150	Inf
KIAA0303	uc003juv.1	311	50	Inf
SFRS12	uc010iwy.1,uc003jun.1,uc003juo.1	105	201	Inf
C5orf44	uc010iwv.1,uc003juc.2,uc003jua.2,uc003itz.2	140	62	Inf
SDCCAG10	uc003jtm.2,uc010iwt.1,uc003jtn.1	98	56	Inf
RGS7BP	uc003jtj.1	198	146	Inf
IPO11	uc003jtc.1	781	127	Inf
NDUFAF2	uc003jsp.2	294	98	Inf
GPBP1	uc003jri.2,uc003jrh.2,uc010iwq.1	479	178	Inf
NDUFS4	uc003jpe.2	241	33	Inf
PELO	uc003jos.1	117	81	Inf
ITGA1	uc003jou.1	394	81	Inf
NNT	uc003joe.1,uc003jof.1	248	141	Inf
MGC42105	uc003jno.1	380	177	Inf
GHR	uc003jmt.1	498	206	Inf
DKFZp547K2	uc010iva.1	858	38	Inf
WDR70	uc003jkv.1	975	38	Inf
DNAJC21	uc003jic.1,uc003jjb.1	105	211	Inf
KIAA1334	uc010ius.1	538	28	Inf
RAI14	uc010iur.1,uc003jir.1	742	277	Inf
TARS	uc003jhz.1,uc010iup.1,uc003jhy.1	84	103	Inf
PDZD2	uc003jhm.1	2714	36	Inf
CDH6	uc003jhd.1,uc003jhe.1	321	136	Inf
GUSBP1	uc010iub.1	99	25	Inf
FBXL7	uc003jfn.1	1405	219	Inf
NDUFS6	uc003jcy.1	292	110	Inf
KIAA1234	uc010isy.1	3954	25	Inf
AHRR	uc003jaw.1,uc003jav.1	3954	25	Inf
PDLIM3	uc003ixy.2,uc003ixx.2,uc003ixw.2	133	186	Inf
ACSL1	uc003iwt.1	565	29	Inf
MLF1IP	uc003iwr.1,uc003iwq.1	152	179	Inf
CASP3	uc003iwg.1	116	135	Inf
ENPP6	uc003iwc.1	1542	27	Inf
GPM6A	uc003iuh.1,uc003iug.1	977	150	Inf
GLRA3	uc003itz.1,uc003ity.1	361	54	Inf
FBXO8	uc003itq.1,uc003itp.1	84	68	Inf
C4orf27	uc003isl.2	96	116	Inf
SPOCK3	uc003irk.2,uc003irj.1,uc003iri.1	907	122	Inf
FSTL5	uc010iqv.1,uc003iqh.1,uc003iqi.1	1572	33	Inf
C4orf18	uc003ipp.2	82	83	Inf
PDGFC	uc003ipi.1,uc003iph.1	457	196	Inf
PET112L	uc003imm.2,uc003iml.1	323	36	Inf
SH3D19	uc003imc.2	318	29	Inf
LRBA	uc003ilu.2	2468	104	Inf
DKFZp761C0	uc003ild.1	161	138	Inf
LOC90826	uc003ilc.1	161	138	Inf
OTUD4	uc003ikb.2,uc003ijz.2,uc003ika.2	116	76	Inf
ANAPC10	uc003ijv.2,uc003iju.2,uc003ijw.2	233	96	Inf
AK095554	uc003iic.1	390	97	Inf
SCLT1	uc003igt.2,uc003igr.2,uc010iob.1,uc003igq.2,uc003igp.2	86	138	Inf
BBS7	uc003iee.1,uc003ied.1	122	91	Inf
PDE5A	uc003idg.1,uc003idf.1	185	217	Inf
PRSS12	uc003ica.1	206	289	Inf
PITX2	uc003iaf.1	1181	31	Inf
AGXT2L1	uc010imc.1,uc003hzc.1	91	134	Inf
PAPSS1	uc003hyk.1	226	232	Inf
TBCKL	uc003hye.1,uc010ilv.1,uc003hyf.1,uc003hyc.1	472	78	Inf
TBCK	uc003hyd.1	472	78	Inf
CENPE	uc003hxc.1,uc003hxb.1	135	82	Inf
NHEDC1	uc010ilm.1,uc003hwu.1,uc003hww.1	242	71	Inf
UBE2D3	uc003hwl.1,uc003hwk.1,uc003hwi.1	83	206	Inf
MANBA	uc003hwg.1	398	141	Inf
PPP3CA	uc010ilk.1,uc010ilj.1,uc003hvu.1,uc003hvs.1,uc003hvv.1	528	236	Inf
DKFZp761L05	uc003hvt.1	528	236	Inf
SNCA	uc003hsq.1,uc003hsr.1,uc003hsp.1	247	170	Inf
HEL308	uc010ikb.1.uc003hom.1	234	61	Inf
HPSE	uc010ika.1,uc003hoj.2,uc003hok.2	182	125	Inf
KIAA2037	uc010iju.1,uc003hny.2	398	78	Inf
LIN54	uc010ijt.1,uc003hnz.2,uc003hnx.2	398	78	Inf
SEC31A	uc003hno.2,uc003hnn.1,uc003hnl.1,uc003hnk.1,uc003hnj.1,uc003hni.1,uc003hnh.1,uc003hng.1,uc003hnf.1	166	167	Inf
SCARB2	uc003hju.1	207	234	Inf

Gene	Transcripts			
NUP54	uc003hjt.1,uc010ije.1,uc003hjs.1	84	132	Inf
SDAD1	uc003hjf.2,uc003hje.2	110	54	Inf
PPEF-2	uc003hiz.1	233	25	Inf
G3BP2	uc003hit.1,uc003his.1,uc003hir.1	124	211	Inf
CDKL2	uc003hiq.1	230	102	Inf
BTC	uc003hig.2	107	110	Inf
RASSF6	uc010iil.1,uc010iik.1,uc003hhd.1,uc003hhc.1	93	40	Inf
ANKRD17	uc003hqr.1,uc003hqq.1,uc003hqp.1	562	192	Inf
ADAMTS3	uc003hqk.1	394	95	Inf
CENPC1	uc010ihm.1,uc003hdd.1	124	81	Inf
EPHA5	uc003hda.1,uc003hcx.1,uc003hcz.1,uc003hcy.1	414	128	Inf
PPAT	uc003hbr.1	127	177	Inf
CLOCK	uc003haz.1	327	143	Inf
SCFD2	uc010iqm.1,uc003qzu.1	1317	60	Inf
TEC	uc003gxz.1	439	197	Inf
CRN	uc003gxn.2	461	144	Inf
CORIN	uc003gxm.1	692	144	Inf
GABRA2	uc003gxc.2	144	34	Inf
PHOX2B	uc003gwf.2	211	102	Inf
RBM47	uc003gvc.2	1588	96	Inf
C4orf34	uc003guo.1	409	114	Inf
RFC1	uc003gtz.1,uc003gty.1,uc003gtx.1	160	89	Inf
KIAA0746	uc003gru.2	718	259	Inf
PPARGC1A	uc003gqs.1	241	25	Inf
KCNIP4	uc003gqe.1	465	38	Inf
LCORL	uc003gpq.1	343	322	Inf
FBXL5	uc010idy.1,uc003god.1,uc010idx.1,uc003goc.1,uc003gob.1	101	240	Inf
BC035722	uc003gmw.1	172	37	Inf
RAB28	uc003gmu.1,uc003gmt.1	260	166	Inf
ACOX3	uc003gld.2,uc003glc.2	2045	99	Inf
AFAP1	uc003gkf.1	1778	27	Inf
JAKMIP1	uc003giu.2	1725	35	Inf
STX18	uc003gic.1	590	118	Inf
WHSC2	uc003gen.1,uc003gem.1	842	248	Inf
FAM53A	uc010ibw.1,uc003gdh.1	1454	244	Inf
FLJ14297	uc010ibd.1,uc003gae.2,uc003gad.2,uc003gac.2	207	55	Inf
ZNF721	uc003gaf.2	146	55	Inf
NEIL3	uc010irs.1,uc003iut.2	159	104	Inf
CLCN3	uc003isi.1,uc003ish.1	357	113	Inf
PALLD	uc003irv.1	1937	113	Inf
ETFDH	uc010iqr.1,uc003iqb.1	115	142	Inf
TMEM144	uc003ipx.1	112	36	Inf
TRIM2	uc003ing.1	1322	180	Inf
ARFIP1	uc003inc.1,uc003inb.1,uc003ina.1,uc003imz.1	394	106	Inf
IL15	uc003iit.1,uc003iis.1	214	157	Inf
MGST2	uc003ihy.1	128	32	Inf
INTU	uc003ifk.1	107	32	Inf
SPATA5	uc003iey.2,uc003iez.2	333	85	Inf
FGF2	uc003iev.1	158	296	Inf
KIAA1627	uc003icg.1,uc003icf.1	96	59	Inf
NDST3	uc003ibx.1	325	26	Inf
C4orf16	uc010imm.1,uc003iak.2,uc003iaj.2	78	182	Inf
SCYE1	uc003hyg.1,uc003hyh.1	75	75	Inf
GSTCD	uc003hxx.2,uc010ils.1,uc003hxy.2	180	90	Inf
CISD2	uc003hwt.2	95	132	Inf
KIAA1122	uc010ila.1	219	60	Inf
SMARCAD1	uc003htb.2,uc003htc.2,uc003htd.2	219	89	Inf
ARHGAP24	uc003hpi.1,uc003hpj.2,uc003hpk.1	459	81	Inf
COPS4	uc010ijx.1,uc010ijw.1,uc003hob.1,uc003hoa.1	169	55	Inf
C4orf22	uc010ijp.1,uc003hmf.1	986	75	Inf
ANXA3	uc010ijk.1,uc003hle.1,uc003hld.1	134	137	Inf
MRPL1	uc003hku.2	292	36	Inf
AK310457	uc010ijf.1	89	122	Inf
CR627383	uc003hjw.1	89	122	Inf
FAM47D	uc003hiv.1,uc003hix.1,uc003hiv.1	315	203	Inf
MOBKL1A	uc003hfv.1,uc003hfw.1	394	168	Inf
ARL9	uc003hbv.1	84	128	Inf
KIAA1211	uc010iha.1	595	28	Inf
FIP1L1	uc003gzx.2,uc003hab.1,uc003gzz.1,uc003gzy.1	92	167	Inf
PDGFRA	uc003haa.1	4610	167	Inf
SPATA18	uc003gzk.1,uc003gzl.1	124	165	Inf
SLAIN2	uc003gya.2	186	238	Inf
GABRB1	uc003gxh.1	725	101	Inf
LIMCH1	uc003gwd.2,uc003gwc.2	365	26	Inf
PGM2	uc003gta.1	125	182	Inf
TBC1D19	uc010iew.1,uc003gsf.2	292	49	Inf
ZCCHC4	uc003grl.2	203	66	Inf
CD38	uc003goj.1,uc003goi.1	194	130	Inf
C4orf23	uc003glf.1	458	49	Inf
KIAA0232	uc003giq.2,uc003gir.2	587	346	Inf
DKFZp762M1	uc010ida.1	2091	240	Inf
STK32B	uc003gih.1	2091	240	Inf
AK056081	uc003gie.1	221	50	Inf
RGS12	uc010icu.1,uc010icv.1,uc003gha.1	1366	73	Inf
C4orf8	uc010ick.1	922	27	Inf
ZNF141	uc003fzz.2,uc003gab.2,uc003gaa.2	100	127	Inf
IQCG	uc003fyq.2,uc003fyn.1,uc003fyo.1	361	55	Inf
RNF168	uc010iah.1,uc003fwq.1	265	181	Inf
PCYT1A	uc003fwh.1,uc003fwg.1	527	177	Inf
MUC4	uc010hzu.1	597	29	Inf
FGF12	uc003fsx.1	564	104	Inf
BCL6	uc003frp.2,uc010hza.1	189	35	Inf
EHHADH	uc003fpf.1	235	73	Inf
ABCC5	uc010hxo.1,uc010hxn.1,uc003fmh.1,uc003fmg.1	283	192	Inf
PARL	uc003fme.1,uc003fmd.1	388	94	Inf
MCC-B	uc003flf.1	459	139	Inf
MCCC1	uc003fle.1	459	139	Inf
ZMAT3	uc003fji.1,uc010hxa.1,uc003fjg.1	195	205	Inf

87

AADACL1	uc003fiq.1	429	80	Inf
EVI1	uc003ffk.2,uc003ffn.2,uc010hwm.1,uc003ffi.2,uc010hwl.1	109	95	Inf
PDCD10	uc003fez.1,uc003fex.1,uc003fey.1	97	103	Inf
ZBBX	uc003feq.1,uc003fep.1	235	60	Inf
KPNA4	uc003fdn.1	213	358	Inf
TRIM59	uc003fdm.1	81	170	Inf
SHOX2	uc010hvw.1,uc003fbs.1,uc003fbr.1	261	189	Inf
DHX36	uc003ezz.2,uc003ezy.2,uc010hvq.1	118	46	Inf
WWTR1	uc003exe.1,uc003exf.1	666	165	Inf
HLTF	uc003ews.1,uc010hve.1,uc003ewr.1,uc003ewq.1	198	124	Inf
PLSCR4	uc003evu.2,uc010hva.1,uc010huy.1,uc003evt.2,uc010huz.1	96	147	Inf
PLOD2	uc003evs.1,uc003evr.1	154	223	Inf
AMOTL2	uc003eqh.1,uc003eqg.1	452	230	Inf
TOPBP1	uc003eps.1	163	113	Inf
ACAD11	uc003eoy.2,uc003eov.2	94	114	Inf
CPNE4	uc003eok.1	819	114	Inf
PLXND1	uc003emx.2	1971	550	Inf
C3orf25	uc003emg.1	324	83	Inf
KIAA1160	uc010hsy.1	300	250	Inf
RAB43	uc003eln.1	300	250	Inf
KIAA1257	uc003elg.1,uc003elj.2	653	135	Inf
RPN1	uc003ekr.1	259	196	Inf
ZNF148	uc003ehy.2,uc003eia.2,uc010hsa.1,uc003ehz.2,uc003ehx.2	521	157	Inf
SLC12A8	uc010hrz.1,uc003ehv.2	1002	154	Inf
POLQ	uc003eee.2	380	62	Inf
FSTL1	uc010hrb.1,uc003eds.1	151	252	Inf
GSK3B	uc003edo.1,uc003edn.1	753	53	Inf
LSAMP	uc003ebt.1	1207	59	Inf
ZBTB20	uc003ebn.1	1722	121	Inf
KIAA1407	uc003eax.1	211	76	Inf
CCDC52	uc003eaq.2	185	72	Inf
CBLB	uc003dwe.1,uc003dwd.1,uc003dwc.1	369	165	Inf
ZBTB11	uc003dve.2	136	205	Inf
PROS1	uc003drb.2	258	122	Inf
FAM86D	uc003dpq.2,uc003dpp.2	91	45	Inf
PDZRN3	uc003dpl.1	1247	388	Inf
FRMD4B	uc003dny.2,uc003dnv.2	320	117	Inf
MAGI1	uc003dmr.2,uc003dmp.1,uc003dmo.1,uc003dmn.1,uc003dmm.1	2118	222	Inf
THOC7	uc003dlu.2,uc003dlt.2	85	250	Inf
FHIT	uc010hnn.1,uc003dky.2,uc003dkx.2	3032	100	Inf
DNASE1L3	uc003djo.1	130	29	Inf
FAM116A	uc003dja.1	232	160	Inf
DNAH12	uc003diu.1	201	30	Inf
IL17RD	uc003dik.1	373	25	Inf
ARHGEF3	uc003dif.2,uc003dih.2	234	27	Inf
WNT5A	uc003dhm.1,uc010hmw.1	208	35	Inf
NEK4	uc003dfr.2,uc003dfq.2	144	101	Inf
TLR9	uc010hmc.1	282	194	Inf
TWF2	uc003ddd.1	282	194	Inf
WDR51A	uc003dcw.1,uc003dcu.1	878	113	Inf
DKFZp434C2	uc003dcv.1	878	113	Inf
PCBP4	uc003dcl.1	92	75	Inf
RRP9	uc003dbw.1	159	106	Inf
RASSF1	uc003dab.1,uc003dac.2	81	238	Inf
HYAL3	uc003czg.1,uc003czf.1,uc003cze.1,uc003czd.1	119	114	Inf
DKFZp564E0	uc003cxx.1,uc003cxw.1	123	138	Inf
CAMKV	uc003cxv.1,uc003cxu.1,uc003cxt.1	123	138	Inf
IHPK1	uc003cxo.1	444	142	Inf
IP6K1	uc003cxn.1,uc003cxm.1	500	142	Inf
RHOA	uc010hku.1	439	236	Inf
USP4	uc003cwp.1	225	25	Inf
SLC25A20	uc010hkj.1,uc003cva.2	263	72	Inf
COL7A1	uc003ctz.2	900	159	Inf
SHISA5	uc003cto.1,uc003ctq.1,uc003ctp.1	338	123	Inf
DKFZp547H1	uc003crw.1	297	25	Inf
MAP4	uc003csf.2,uc003csc.2,uc003csq.1,uc003csb.1	839	171	Inf
KIF9	uc010hjp.1,uc003cqz.1,uc003cqy.1,uc003cqx.1	347	165	Inf
LTF	uc010hjh.1,uc003cpq.1	104	103	Inf
FYCO1	uc003cpb.2	502	108	Inf
ZDHHC3	uc003coq.1,uc003cof.1	352	158	Inf
ZNF445	uc003cnf.2	193	106	Inf
SEC22C	uc003cli.1	364	141	Inf
ULK4	uc003ckx.1,uc003ckw.2,uc003ckv.2	395	113	Inf
GLB1	uc003cfj.1,uc003cfi.1,uc003cfh.1	625	163	Inf
FLJ45032	uc003cfk.1	84	159	Inf
SBC2	uc010hfm.1	330	304	Inf
SLC4A7	uc003cdu.2	419	304	Inf
NGLY1	uc003cdm.1,uc010hfg.1,uc003cdl.1	136	129	Inf
THRB	uc003cdd.2,uc003cde.1,uc003cdc.2,uc003ccz.2,uc003ccy.2,uc003ccx.2,uc010hfe.1	587	204	Inf
NKIRAS1	uc003ccm.1,uc003ccl.1,uc003cck.1,uc003ccj.1	105	173	Inf
ZNF385D	uc003cce.1	520	53	Inf
TBC1D5	uc003cbf.1	1197	29	Inf
SH3BP5	uc003bzp.1	585	217	Inf
RAF1	uc003bxf.2	363	150	Inf
C3orf31	uc003bwj.1,uc003bwh.1	238	74	Inf
SEC13	uc003bvq.1,uc003bvm.1,uc003bvp.1,uc003bvo.1,uc003bvn.1	166	73	Inf
PRRT3	uc003bul.2	399	89	Inf
KIAA0411	uc003brk.2	1342	48	Inf
SRGAP3	uc003brg.1,uc003brf.1	2039	48	Inf
RAD18	uc003brd.1	154	56	Inf
SUMF1	uc010hby.1,uc003bpz.1	315	84	Inf
LMLN	uc010ias.1,uc003fyu.1,uc003fyt.1,uc010iar.1,uc003fyv.1	468	151	Inf
OPA1	uc003ftn.1,uc003ftm.1,uc003ftl.1,uc003ftk.1,uc003ftj.1,uc003fti.1,uc003fth.1,uc003ftg.1	319	112	Inf
IL1RAP	uc010hzf.1,uc003fsl.1,uc003fso.1,uc010hzg.1,uc003fsq.2,uc003fsk.1,uc003fsm.1	375	198	Inf
LPP	uc003frs.1	1974	25	Inf
MAP3K13	uc010hyf.1	895	86	Inf
VPS8	uc010hyd.1,uc003fpb.1	781	96	Inf
FAM131A	uc003fog.1	125	51	Inf

88

PSMD2	uc003fnn.1	97	165	Inf
ECE2	uc003fni.2	1132	107	Inf
FXR1	uc003fkr.1,uc003fkq.1,uc003fkp.1	370	100	Inf
NDUFB5	uc003fke.1,uc003fkc.1	97	80	Inf
NLGN1	uc010hww.1,uc003fip.1	1309	52	Inf
FNDC3B	uc003fhz.2,uc010hwt.1	1593	241	Inf
SERPINI1	uc003ffb.2,uc003ffa.2	178	53	Inf
SCHIP1	uc010hvz.1	535	298	Inf
GFM1	uc003fcd.2,uc003fce.1	112	66	Inf
RSRC1	uc003fbu.1,uc003fbv.1,uc003fbt.1	652	111	Inf
KCNAB1	uc010hvt.1,uc003fat.2,uc003far.2	728	126	Inf
MME	uc003fad.1,uc003fac.1,uc003fab.1,uc003fae.1	206	132	Inf
MED12L	uc003eyo.2,uc003eyn.2,uc003eyp.1,uc003eym.1	428	232	Inf
SELT	uc003evf.1	189	143	Inf
FAM62C	uc003esk.1	435	283	Inf
ARMC8	uc003esb.1	376	30	Inf
BFSP2	uc003epn.1	535	52	Inf
DNAJC13	uc010htq.1,uc003eor.1	131	129	Inf
NEK11	uc003enw.1,uc003eoa.1,uc003enx.1,uc003eny.1	201	84	Inf
nek11L	uc003enz.1	544	84	Inf
C3orf37	uc003elu.1,uc003elw.1,uc003elv.1,uc003elt.1	137	234	Inf
ACAD9	uc010hsw.1,uc003elb.1,uc003ela.2	372	134	Inf
RAB7	uc010hsv.1	622	170	Inf

Table S4.3

List of candidate genes whose expression was up-regulated in the Saqqaq hairs.

This list corresponds to the top 1,000 non-redundant genes showing the highest R_s values (transcripts are collapsed into a single Gene name, providing the maximal R_s value observed in this quartile). R_s values and respective coverage in PM and RE regions are provided.

89

Section SI5. References.

R.S. Alisch, B.G. Barwick, P. Chopra, L.K. Myrick, G.A. Satten, K.N. Conneely, S.T. Warren. Age-associated DNA methylation in pediatric populations. *Genome Res.* **22**, 623-632 (2012).

P.J. Aruscavage, S. Hellwig, B.L. Bass. Small DNA pieces in *C. elegans* are intermediates of DNA fragmentation during apoptosis. *PLoS One* **5**, e11217 (2010).

M.P. Ball *et al*. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27**, 485 (2009).

H. Bazzi, S. Demehri, C.S. Potter, A.G. Barber, A. Awgulewitsch, R. Kopan, A.M. Christiano. Desmoglein 4 is regulated by transcription factors implicated in hair shaft differentiation. *Differentiation* **78**, 292-300 (2009).

A.C. Bell, G. Felsenfeld. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-485 (2000).

A.C. Bell, A.G. West, G. Felsenfeld. Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**, 447-450 (2001).

Y. Benjamini, T.P. Speed, TP. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acid Res 40, e72 (2012).

T. Beissbarth, T.P. Speed. GOstat: Find statistically overrepresented Gene Ontologies within a group of genes *Bioinformatics* **20**, 1464-1465 (2004).

N.V. Botchkavera, G. Ahluwalia, D. Shander. Apoptosis in the hair follicle. *J Invest Dermatol* **126**, 258-264 (2006).

F. Brenet, M. Moh, P. Funk, E. Feierstein, A.J. Viale, N.D. Socci, S.M. Scandura. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One* **6**, e14524 (2011).

A.W. Briggs *et al*. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* **38**, e87 (2010).

A.W. Briggs *et al*. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci.* **104**, 14616–14621 (2007).

P. Brotherton *et al*. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.* **35**, 5717–5728 (2007).

C.K. Collings, A.G. Fernandez, C.G. Pitschka, T.B. Hawkins, J.N. Anderson. Oligonucleotide sequence motifs as nucleosome positioning signals. *PLoS One* **5**, e10933 (2010).

A.M. Deaton, A. Bird. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010-1022 (2011).

S. Dedeurwaerder, M. Defrance, E. Calonne E, H. Denis, C. Sotiriou, F. Fuks. Evaluation of the Infinium Methylation 450K technology. *Epigenomics* 3, 771–784 (2011).

K. Day, L.L. Waite, A. Thalacker-Mercer, A. West, M.M. Bamman, J.D. Brooks, R.M. Myers, D. Absher. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol.* 14, R102 (2013).

J.H. Dennis *et al*. Independent and complementary methods for large-scale structural analysis of mammalian chromatin. *Genome Res.* **17**, 928-939 (2007).

J.C. Dohm, C. Lottaz, T. Borodina, H. Himmelbauer. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, e105 (2008).

T.A. Down, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* **26**, 779-785 (2008).

F. Eckardt *et al*. DNA methylation profiling of human chromosomes 6, 20, 22. *Nat. Genet.* **38**, 1378-85.

J. Ernst *et al*. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).

M.J. Fogg, L.H. Pearl, B.A. Connolly. Structural basis for uracil recognition by archaeal family b DNA polymerases. *Nature Struct. Biol.* **9**, 922–927 (2002).

Y. Fu, M. Sinha, C.L. Peterson, Z. Weng. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4**, e1000138 (2008).

P.A. Fujita *et al*. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876-82 (2010).

D.J. Gaffney *et al*. Controls of Nucleosome Positioning in the Human Genome. *PLoS Genetics* **8**, e1003036 (2012).

M.T. Gilbert *et al*. Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res.* **35**, 1-10 (2007).

M.T. Gilbert *et al*. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**, 1787-1789 (2008).

P.A. Ginno, P.L. Lott, H.C. Christensen, I. Korf, F. Chedin. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular Cell* **45**, 1-12 (2012).

A. Ginolhac, M. Rasmussen, M.T. Gilbert, E. Willerslev, L. Orlando. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**, 2153-2155 (2011).

H. Gong et al. An updated nomenclature for Keratin-Associated Proteins (KAPs). *Int. J. Biol. Sci.* **8**, 258-264 (2012).

R.E. Green *et al*. A draft sequence of the Neandertal genome. *Science* **328**, 710-722 (2010).

A. Johansson, S. Enroth, U. Gyllensten. Continuous aging of the human DNA methylome throughout the human lifespan. PLoS One 8, e67378 (2013).

T.K. Kelly, Y. Liu, F.D. Lay, G. Liang, B.P. Berman, P. Jones. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497-2506 (2012).

S.J. Kim *et al*. Gallagher and John C. Rockett. Gene Expression in Head Hair Follicles Plucked from Men and Women. *Ann. Clin. Lab. Sci. Spring* **36**, 115-126 (2006).

J.Y. Kim, S. Tavare, D. Shibata. Human hair genealogies and stem cell latency. Human hair genealogies and stem cell latency. *BMC Biology* 4, 2 (2006).

T.H. Kim *et al*. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).

C.M. Koch, W. Wagner. Epigenetic-aging-signature to determine age in different tissues. AGING 3, 1018-27 (2011).

J. Krause *et al*. A complete mtDNA genome of an early modern human from Kostenki Russia. *Curr. Biol.* **20**, 231-236 (2010).

L. Langbein, H. Yoshida, S.Praetzel-Wunder, D.A. Parry, J. Schweizer. The keratins of the human beard hair medulla: the riddle in the middle. *J. Invest. Derm.* **130**, 55-73 (2010).

L. Laurent *et al*. in the human methylome during differentiation. *Genome Res. 20*, 320-331 (2010).

Y.J. Lee, R.H. Rice, Y.M. Lee. Proteome analysis of human hair shaft. *Mol. Cell. Proteomics* **5**, 789-800 (2006).

H. Li, R. Durbin. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).

H. Li *et al*. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

S. Lindgreen. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012).

R. Lister, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **464**, 315-322 (2009).

C.Y. McLean et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**:495-501.

L.R. Meyer *et al*. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**(Database issue), D64-D69 (2012).

W. Miller *et al*. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc Natl Acad Sci U S A 109, E2382-E2390 (2012).

R. Moll, M. Divo, L. Langbein. The human keratins: biology and pathology. *Histochem. Cell. Biol.* **129**, 705-733 (2008).

S. Nagata, M. Enari, H. Sakahira, H. Yokoyama, K. Okawa, A. Iwamatsu. A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43–50 (1998).

A. Noer, A.L. Sorensen, A.C. Boquest, P. Collas. Stable CpG hypomethylation of adipogenic promoters in freshly isolated, cultured, and differentiated mesenchymal stem cells from adipose tissue. *Mol. Biol. Cell.* **17**, 3543-3556 (2006).

L. Orlando *et al*. Revising the recent evolutionary history of equids using ancient DNA. *Proc. Natl. Acad. Sci. USA* **106**, 21754-21759 (2009).

L. Orlando *et al*. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Res.* **21**, 1705-1719 (2011).

L. Orlando *et al*. Recalibrating *Equus* evolution using the genome sequence of an Early Middle Pleistocene horse. *Nature* **499**, 74-78 (2013).

F. Ozsolak, J.S. Song, X.S. Liu, D.E. Fisher. High-throughput mapping of the chromatin structure of human promoters. Nat. Biotechnol. 25:244-248 (2007).

H. Pages, P. Aboyoun, M. Lawrence. IRanges: Infrastructures for manipulating intervals on sequences. R packages version 1.8.9.

R Development Core Team 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

V.A. Randall, N.V. Botchkareva NV. *The biology of hair growth* In: Ahluwalia GA, ed. Cosmetic applications of laser and light based systems.Ch 1, pp.3-35 William Andrew Inc., USA (2009).

M. Rasmussen *et al*. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762 (2010).

M. Rasmussen *et al*. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94-98 (2011).

A.P. Rebello *et al*. *In vivo* methylation of mtDNA reveals the dynamics of protein-mtDNA interactions. *Nucleic Acids Res.* **37**, 6701-6715 (2009).

D. Reich *et al*. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060 (2010).

J. Roessler, O. Ammerpohl, J. Gutwein, B. Hasemeier, S.L. Anwar, H. Kreipe, U. Lehmann. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina. Inc. *BMC Res Notes* 5, 210 (2012).

N. Rohland, M. Hofreiter. Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2**, 1756-1762 (2007)

D.E. Schones *et al*. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-898 (2008).

M Schubert *et al*. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).

R.C. Slieker *et al.* Identification and systematic annotation of tissue-speciic differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* **6**, 26 (2013).

A.F.A. Smit,, R. Hubley, P. Green. RepeatMasker Open-3.0. http://www.repeatmasker.org. 1996-2010.Stiller, M et al. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc. Natl. Acad. Sci. USA* **103**, 13578-13584 (2006).

R. Straussman et al. Developmental programming of CpG island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.* **16**, 564-571 (2009).

A. Valouev *et al*. Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516-520 (2011).

M. Weber *et al*. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genetics* **39**, 457-466 (2007).

A.O. Zalensky *et al*. Human testis/sperm-specific histone H2B (hTSH2B): molecular cloning and characterization. *J. Biol. Chem.* **277**, 43474-43480 (2002).