

SCIENTIFIC REPORTS



OPEN

Transcriptional heterogeneity in the lactase gene within cell-type is linked to the epigenome

Received: 21 September 2016

Accepted: 04 January 2017

Published: 31 January 2017

Edward Oh^{1,*}, Richie Jeremian^{1,*}, Gabriel Oh¹, Daniel Groot¹, Miki Susic¹, KwangHo Lee², Kelly Foy², Peter W. Laird², Arturas Petronis^{1,†} & Viviane Labrie^{1,3,†}

Transcriptional variation in histologically- and genetically- identical cells is a widespread phenomenon in tissues, yet the processes conferring this heterogeneity are not well understood. To identify contributing factors, we analyzed epigenetic profiles associated with the *in vivo* transcriptional gradient of the mouse lactase gene (*Lct*), which occurs in enterocytes along the proximal-to-distal axis of the small intestine. We found that epigenetic signatures at enhancer and promoter elements aligns with transcriptional variation of *Lct* in enterocytes. Age and phenotype-specific environmental cues (lactose exposure after weaning) induced changes to epigenetic modifications and CTCF binding at select regulatory elements, which corresponded to the alterations in the intestinal *Lct* mRNA gradient. Thus, epigenetic modifications in combination with CTCF binding at regulatory elements account for the transcriptional gradient in *Lct* in cells of the same type. Epigenetic divergence within enterocytes may contribute to the functional specialization of intestinal subregions.

Within a tissue, seemingly identical cells have been found to exhibit a large degree of variation in their transcriptomes^{1,2}. Although this cell-to-cell variation can be partially stochastic³, there are coordinated transcriptional gradients in cells of the same type across whole organs. For example, prominent transcriptional gradients have been detected in intestinal epithelial cells (enterocytes) along the length of the small intestine^{4–6}, in hepatocytes across liver zones^{7–9}, in myocytes in different heart chambers^{10–13}, and in adipocytes located in bone marrow¹⁴. Gradients in transcriptional states enable cells that are histologically identical to perform biologically distinct roles, vary in their response to environmental stimuli, and exhibit differential disease vulnerability^{1,15}. Despite its evident biological significance, the molecular processes coordinating same cell-type transcriptional divergence across tissue subregions are not well understood.

Epigenetic mechanisms could contribute to transcriptional variation within the same cell type, as epigenetic modifications are important in regulating gene transcription¹⁶, determining cell identity^{17,18}, and affecting genomic functions in response to aging and environmental cues¹⁹. Recently, DNA modification analysis of single cells from *in vitro* cultures found significant heterogeneity^{20–22}, suggesting the potential for epigenetic marks to differ between cells of the same type in living organisms. In this study, we sought to examine the divergence of DNA modifications within a cell type *in vivo* and explored its potential in regulating same cell type transcriptional gradients in tissue. To answer this question, we examined transcriptional and epigenetic variation of the lactase gene (*Lct*) in enterocytes along the proximal-to-distal axis of the mouse small intestine. *Lct*, which is responsible for lactose metabolism, represents an ideal model to identify DNA modifications that can contribute to transcriptional heterogeneity for tissue subregion specialization, as it exhibits a distinguished transcriptional gradient in enterocytes; a gradual elevation from duodenum to jejunum, followed by a steady decline to the ileum^{23–25}. Recent studies have shown that *Lct* mRNA expression is dependent upon DNA modifications at key genomic regulatory elements^{18,26}, however, the extent to which DNA modifications diverge to regulate *Lct* transcription along the length of the intestine remains uninvestigated. In addition to this, we also examined how aging and the environment can modify *Lct* expression through DNA modifications at *Lct* genomic regulatory elements.

¹Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. ²Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI, USA. ³Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, MI, USA. *These authors contributed equally to this work. †These authors jointly supervised this work. Correspondence and requests for materials should be addressed to A.P. (email: Art.Petronis@camh.ca) or V.L. (email: Viviane.Labrie@vai.org)

Here, we found that in enterocytes isolated from different intestinal regions, DNA modifications at specific genomic regulatory elements were concordant with the transcriptional variation of *Lct*. Aging and environmental exposures (lactose feeding after weaning) resulted in DNA modification changes at these regulatory sites and changes in *Lct* transcriptional patterns. Overall, DNA modification patterns support the aging- and environmentally-induced gradients of *Lct* mRNA and, more broadly, could affect phenotypic outcome by modifying transcriptional programs within same cell types.

Results

Within cell-type differences in *Lct* mRNA are correlated to epigenetic alterations. Enterocytes exhibit differing *Lct* transcriptional patterns along the proximal-distal axis of the small intestine, which enables intestinal subregion specialization in lactose metabolism (Fig. 1a). We examined *Lct* mRNA levels in enterocytes from nine segments of the adult (postnatal day 60; P60) mouse small intestine; the proximal, middle and distal segments of the duodenum (segment 1–2), jejunum (segment 3–5), and ileum (segment 6–9), and observed a gradient in steady-state *Lct* mRNA levels (main effect of segment: $F_{8,46} = 23.4$, $p = 8.4 \times 10^{-14}$; Fig. 1b). *Lct* mRNA levels were highest in the proximal jejunum (segment 3), and then gradually declined toward the proximal duodenum ($p = 0.008$ compared to segment 3) and distal ileum ($p = 1.2 \times 10^{-10}$). In contrast, *Mcm6* mRNA levels did not change across the intestinal segments. This indicates that enterocytes from the mid-duodenum to mid-jejunum are most specialized for the digestion of lactose.

Using a targeted bisulfite sequencing approach²⁷, we sought to determine whether DNA modifications (methylation and other cytosine modifications) could account for this within cell-type transcriptional gradient. In enterocytes, selectively isolated from the villi of each of the nine intestinal segments, we characterized 7,580 cytosines (609 CpG sites & 6,971 CpH sites) along the *Lct* and its neighbouring gene, *Mcm6*. We included *Mcm6* in our investigation because DNA variation in this gene, particularly at *MCM6* intron 13, affects inter-individual differences in lactase levels in humans^{28–30}. We identified several clusters of modified cytosines (3 or more in <500 bp) which showed significant inverse correlations with *Lct* mRNA levels ($R^2 = 0.62–0.85$; $p < 0.01$ after Bonferroni correction; Fig. 1c). Most notably, at *Mcm6* exon 13–intron 13, we observed a large cluster modified cytosines associated with *Lct* mRNA variation across the intestine (18 CpGs and 1 CpH site, $p = 3.3 \times 10^{-3}$ to 1.4×10^{-11} after Bonferroni correction; $R^2 = 0.67–0.85$). DNA modification densities in this region were lowest in segment 3 (34.7%) and reached maximal levels in the distal ileum (77.1%; $p = 4.4 \times 10^{-21}$; Fig. 2). There were also DNA modification clusters strongly correlated with *Lct* mRNA at *Lct* exon 1 and *Lct* intron 2 (each genomic site containing 6 CpGs, $R^2 = 0.62–0.85$; $p = 1.8 \times 10^{-2}$ to 4.0×10^{-12} after Bonferroni correction; Figs 1c and 2). Thus, DNA modifications at specific regions are related to the *Lct* transcriptional gradient within enterocytes along the small intestine.

We next examined ENCODE data of adult mouse small intestine³¹ to determine whether the sites we associated with transcriptional variation in *Lct* could function as genomic regulatory elements (Fig. 1d). We found histone marks indicative of a poised element at *Mcm6* exon 13–intron 13 (H3Kme1), an active enhancer at *Lct* intron 2 (H3K4me1 & H3K27ac), and an active promoter at *Lct* exon 1 (H3K4me3 & H3K27ac). *Mcm6* exon 13–intron 13, *Lct* exon 1 and intron 2 showed binding of CTCF, a protein that facilitates interactions between transcription regulatory sequences by affecting chromatin architecture^{32,33}. In addition, *Lct* exon 1 and intron 2 overlapped GATA1 binding sites. GATA1 is a member of GATA transcription factor family that modulates *Lct* expression³⁴, and modifies epigenetic marks to promote subregion transcriptomic differences in the small intestine³⁵. Thus, sites in which DNA modifications were significantly predictive of regional differences in *Lct* contained chromatin signatures of enhancers and a promoter, along with an enrichment in transcriptional regulatory proteins.

The *Lct* transcriptional and epigenetic gradients in enterocytes are age-dependent. We investigated whether transcriptional gradients could be established from age-associated epigenetic changes. To do this, we first examined *Lct* mRNA levels in enterocytes across the small intestine of infant mice (P6) and adult mice (P60). *Lct* mRNA levels in infants were mostly similar across intestinal segments, and were 5–15 fold higher than adults (main effect of age: $F_{8,85} = 13.23$, $p = 3.12 \times 10^{-12}$; Fig. 3a). This signifies that transcriptional differences in *Lct* along the small intestine manifests with age, where over time *Lct* becomes mainly suppressed in enterocytes of the proximal duodenum and ileum relative to medial segments.

Next, the extent to which DNA modifications diverge with age at the *Lct-Mcm6* locus was examined in villi enterocytes from different intestinal regions. In adults, segments 1, 3 and 7 (duodenal, jejunal and ileal segments) differed substantially in *Lct* abundance, and consistently, DNA modifications differed significantly between adult segments ($p < 0.01$ after Bonferroni correction; Fig. 3b, x-axis). In infants, no significant DNA modification differences were observed between the intestinal segments (Fig. 3b, y-axis), which reflects their lack of an *Lct* transcriptional gradient.

We then searched for epigenetically-controlled DNA sites that could contribute to the age-dependent establishment of the *Lct* transcriptional gradient. For this, we investigated DNA modifications at the *Lct-Mcm6* locus in isolated villi enterocytes of adult and infant mice, comparing only matched segments between the age groups (Fig. 3c). We found differences in CpG (but not CpH) modifications with age in enterocytes along the intestine. In particular, *Mcm6* exon 13–intron 13 exhibited the highest localized gain of DNA modifications in adulthood (Fig. 3c). The age-related DNA modification increase at *Mcm6* exon 13–intron 13 was most prevalent in the segment with the largest *Lct* mRNA loss (segment 7; age-segment interaction: $F_{7,78} = 5.78$, $p = 4.0 \times 10^{-4}$; Fig. 3d). DNA modification densities in *Mcm6* exon 13–intron 13 increased by 30% in segment 7 in P60 adults compared to infants ($p = 1.3 \times 10^{-4}$), while there was an 18% increase segment 1 ($p = 1.3 \times 10^{-4}$) and a 12% increase in segment 3 ($p = 0.0013$; Fig. 3d). In older adults (P90), DNA modification levels continued to rise (by ~10%) in segment 1 and 3 ($p = 0.025$; Fig. 3d), indicating continued epigenetic aging at this *Mcm6* site. Thus, in addition to enabling lactase persistence in certain human populations^{28,29}, epigenetic regulation of *Mcm6* exon 13–intron 13

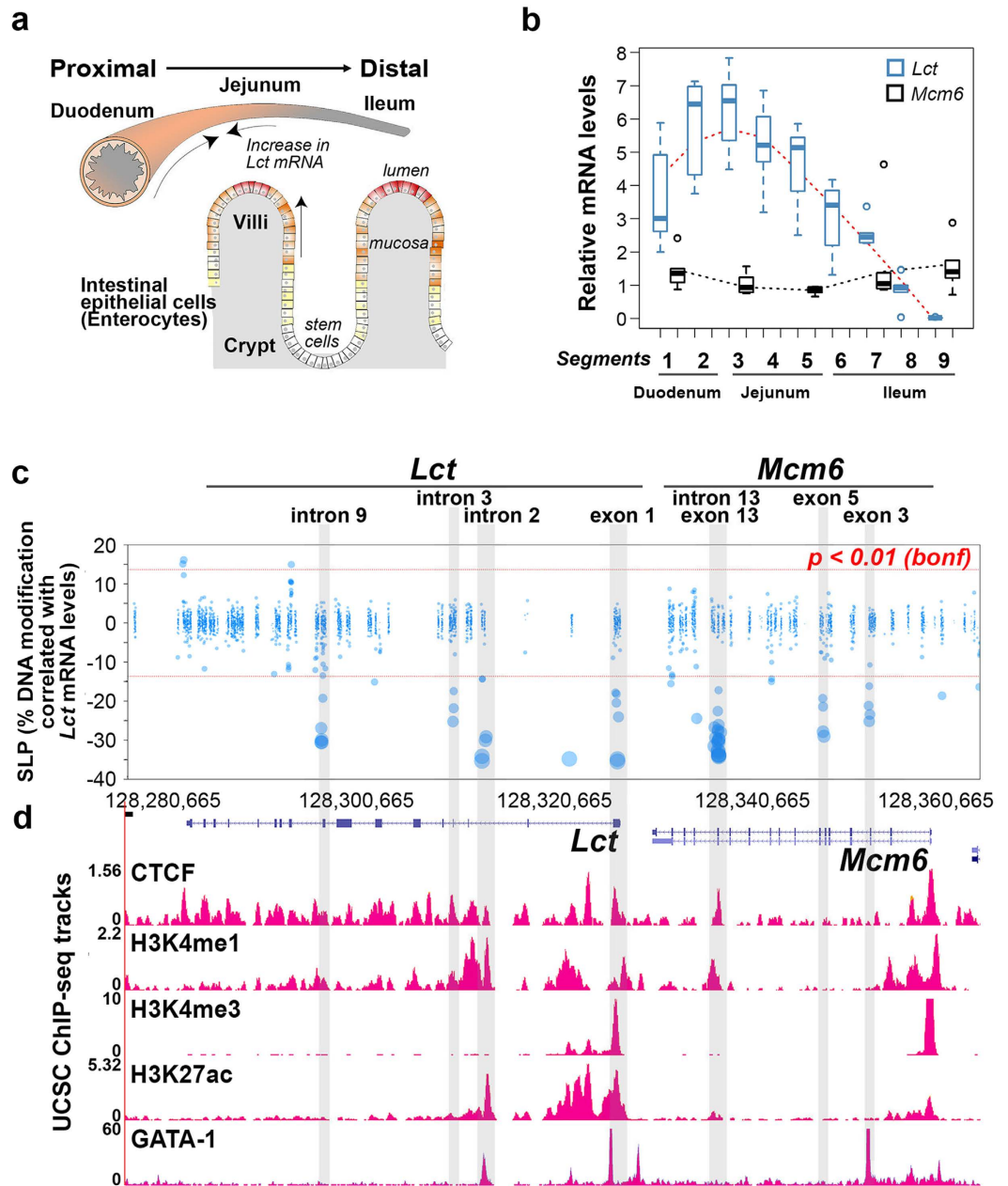


Figure 1. Within cell-type variation in *Lct* mRNA along the proximal-to-distal axis of small intestine is correlated to DNA modification differences at genomic regulatory elements. **(a)** Illustration of the proximal-to-distal axis of the small intestine and villi. **(b)** There is a segment-specific gradient in steady-state *Lct* mRNA levels. There is no transcriptional gradient in *Mcm6* along the intestine. Villin-1 was used as an endogenous control. Data are represented as mean \pm SEM. $n = 5-7$ mice. $**p < 0.01$, $***p < 10^{-3}$ by Tukey's HSD post hoc test in comparison to segment 3. **(c)** Enterocyte-specific DNA modification clusters (>3 cytosine modifications) within the *Lct*-*Mcm6* locus significantly correlated with *Lct* mRNA levels ($n = 53$; 9 segments \times 5-6 mice per segment). Bonferroni correction for multiple testing ($p < 0.01$) is indicated by dash lines. SLP refers to the $-\log p$ -value of the Pearson's correlation R^2 (negative values signify an inverse correlation). **(d)** DNA modification clusters associated with the *Lct* transcriptional gradient (highlighted by grey vertical bars) overlap histone marks characteristic of promoters and enhancers, as well as the transcription factor GATA1 and the chromatin architectural protein CTCF (ChIP-seq data from ENCODE, $n = 2$).

could be central to the age-dependent decline of *Lct* in proximal and distal portions of the small intestine, leading to their inability to metabolize lactose. Thus, cells of the same type can exhibit a divergence in DNA modifications with age across tissue regions, which in turn may facilitate transcriptional gradients and the functional specialization of tissue subregions.

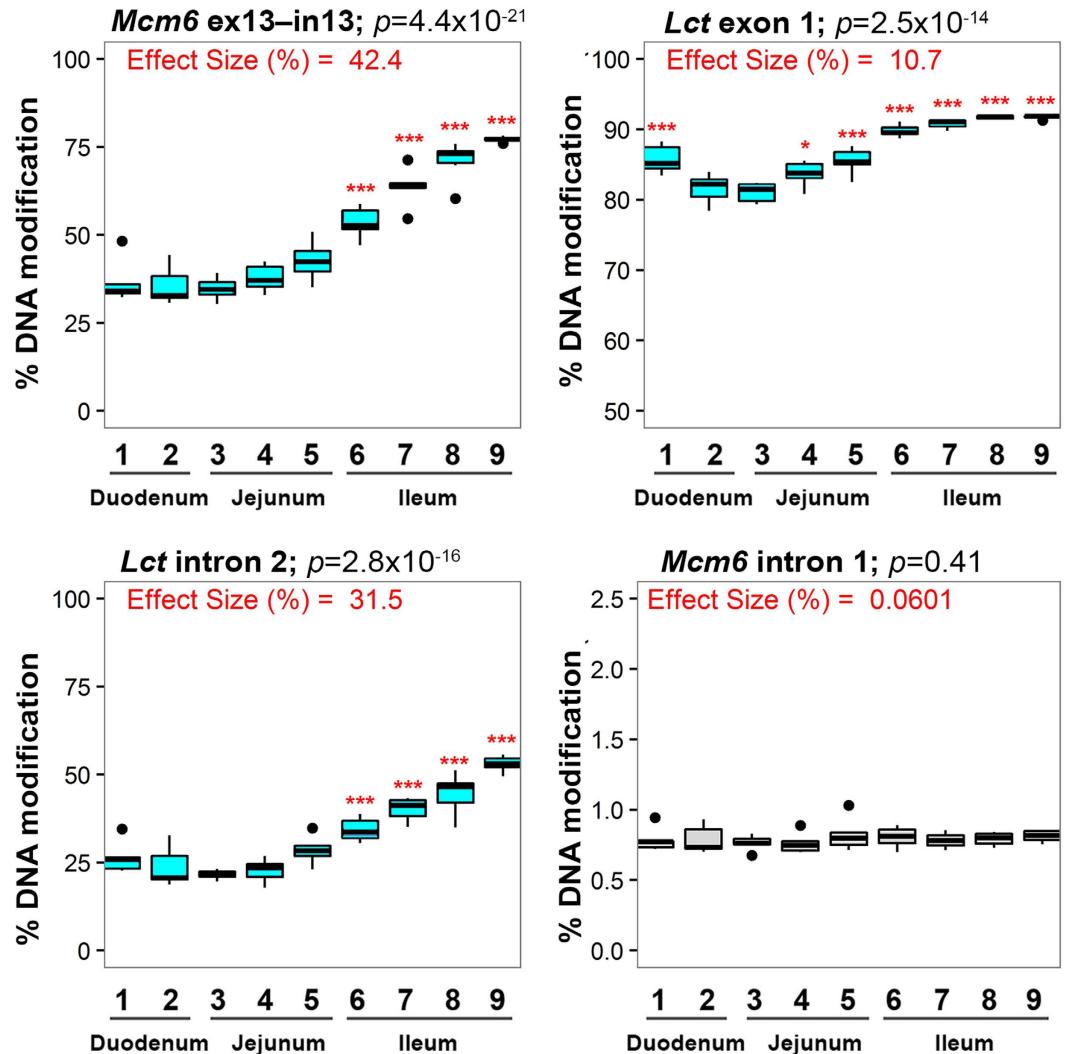


Figure 2. Divergence of DNA modification at *Lct* regulatory elements in enterocytes across intestinal segments. Average % DNA modification (y-axis) at *Mcm6* exon 13-intron 13, *Lct* intron 2 and *Lct* exon 1 is significantly associated with *Lct* transcriptional variation along the proximal-to-distal axis of the intestine (x-axis). *Mcm6* exon 13-intron 13, *Lct* intron 2 and *Lct* exon 1 regions selected based on findings of Fig. 1c. Average % DNA modification of *Mcm6* intron 1 included as a negative control. P-value obtained by one-way ANOVA ($n = 53$; 9 segments \times 5–6 mice per segment). * $p < 0.05$, *** $p < 10^{-3}$ by Tukey's HSD post hoc test in comparison to segment 3.

The environment modifies the *Lct* transcriptional gradient and induces epigenetic alterations.

We determined whether a phenotype-related environmental signal (i.e. lactose exposure) could modify transcriptional patterns and induce epigenetic changes in enterocytes along the small intestine. For this experiment, we fed mice either lactose-containing milk (LAC⁺) or lactose-free milk (lac⁻) for 60 days. In response to lactose treatment, *Lct* mRNA levels increased in the enterocytes of distal intestinal segments (segment-diet interaction $F_{8,88} = 5.2$, $p = 2.0 \times 10^{-5}$; Fig. 4a). Specifically, lactose exposure increased *Lct* mRNA levels in enterocytes of segments 7 ($p = 0.007$) and 8 by 1.3- to 2- fold ($p = 0.023$) compared to mice in the lactose-free group (Fig. 4a).

To test whether the epigenome could be involved in the environmentally-induced changes in *Lct* mRNA across the intestine, we investigated DNA modifications in isolated villi enterocytes at the *Lct*-*Mcm6* locus in LAC⁺ and lac⁻ treated mice. We evaluated intestinal segments in which *Lct* was induced (segments 7 and 8) or showed no statistically significant change (segments 1, 3 and 4). Distal intestinal segments exhibited a significant decrease in CpG modification densities at *Lct* exon 12 (segment 7 = 5.2%, $p = 4.7 \times 10^{-4}$; segment 8 = 4.7%, $p = 0.010$), *Lct* intron 8 (segment 8 = 3.6%, $p = 3.6 \times 10^{-3}$), and *Lct* intron 2 (segment 8 = 6.4%; $p = 0.012$; Fig. 4b,c). CpG modifications were not changed significantly in segments 1, 3 and 4 (Fig. 4b,c). Although epigenetic modifications at *Mcm6* exon 13-intron 13 were found to be important for aging and regional differences in *Lct*, this genomic site was unchanged by milk-feeding, which could partly explain why *Lct* mRNA levels were not returned to that of pre-weaned infants after prolonged lactose exposure.

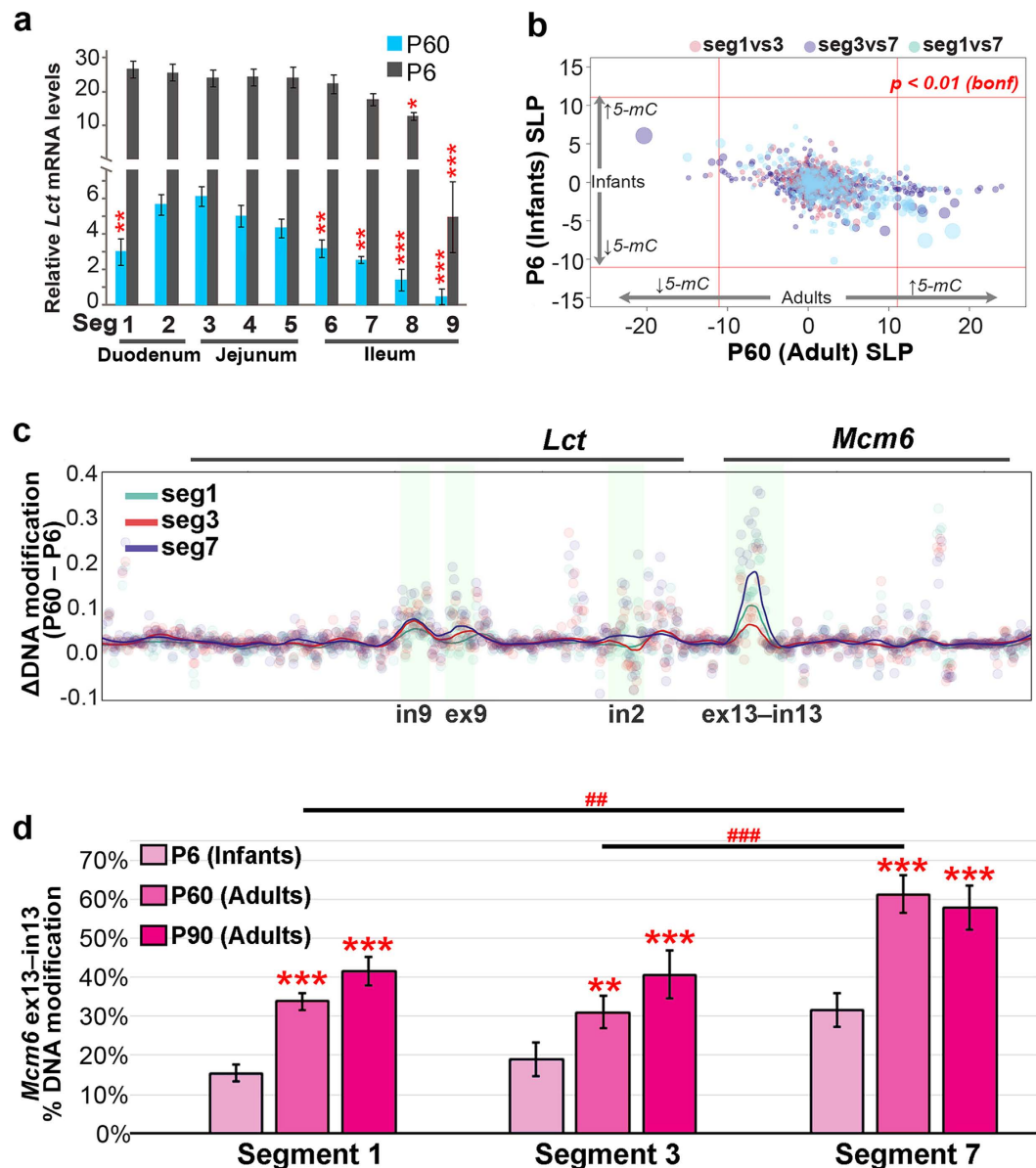


Figure 3. Regional differences in *Lct* transcription arises with age and the divergence of DNA modification patterns within enterocytes. (a) Infant mice (P6; $n = 6$) had higher *Lct* mRNA levels than adults (P60; $n = 4-6$), and had an attenuated transcriptional gradient. Data are represented as mean \pm SEM. $*p < 0.05$, $**p < 0.01$, $***p < 10^{-3}$ by Tukey's HSD post hoc test compared to age-matched segment 3. (b) Comparison of individual DNA modifications between intestinal segments (within age group) revealed a significant divergence in DNA modifications in villi enterocytes along the intestine of adults (x-axis), but not infants (y-axis). SLP refers to signed log p -value. Significant comparisons exceeded the dashed red line ($p < 0.01$ after Bonferroni correction). (c) Identification of genomic sites exhibiting differential DNA modification with age. The *Mcm6* exon 13-intron 13 site in segment 7 (blue) displayed the largest increase in % DNA modification with age, with segment 1 (red) and segment 3 (turquoise) showing smaller DNA modification increases. There was a smaller gain in DNA modifications at *Lct* exon 9-intron 9 in adults. Modification density differences y-axis; average % DNA modification in adults minus in infants by individual CpGs ($n = 632$) fitted with LOESS curve for each segment. (d) Average % DNA modification densities at *Mcm6* exon 13-intron 13 site (Data are represented as mean \pm SEM; $n = 14$ CpGs) in enterocytes along the intestine of infants at P6 and adults at P60 and P90. DNA modification levels increased with age at *Mcm6* exon 13-intron 13, particularly in segment 7 enterocytes. Data are represented as mean \pm SEM. $**p < 0.01$, $***p < 10^{-3}$ by Tukey's HSD post hoc test compared to same segment of infant mice. $##p < 0.01$, $###p < 10^{-3}$ by Tukey's HSD post hoc test for between segments of P60 adults. (b-d) $n = 3-7$ per age group for each segment.

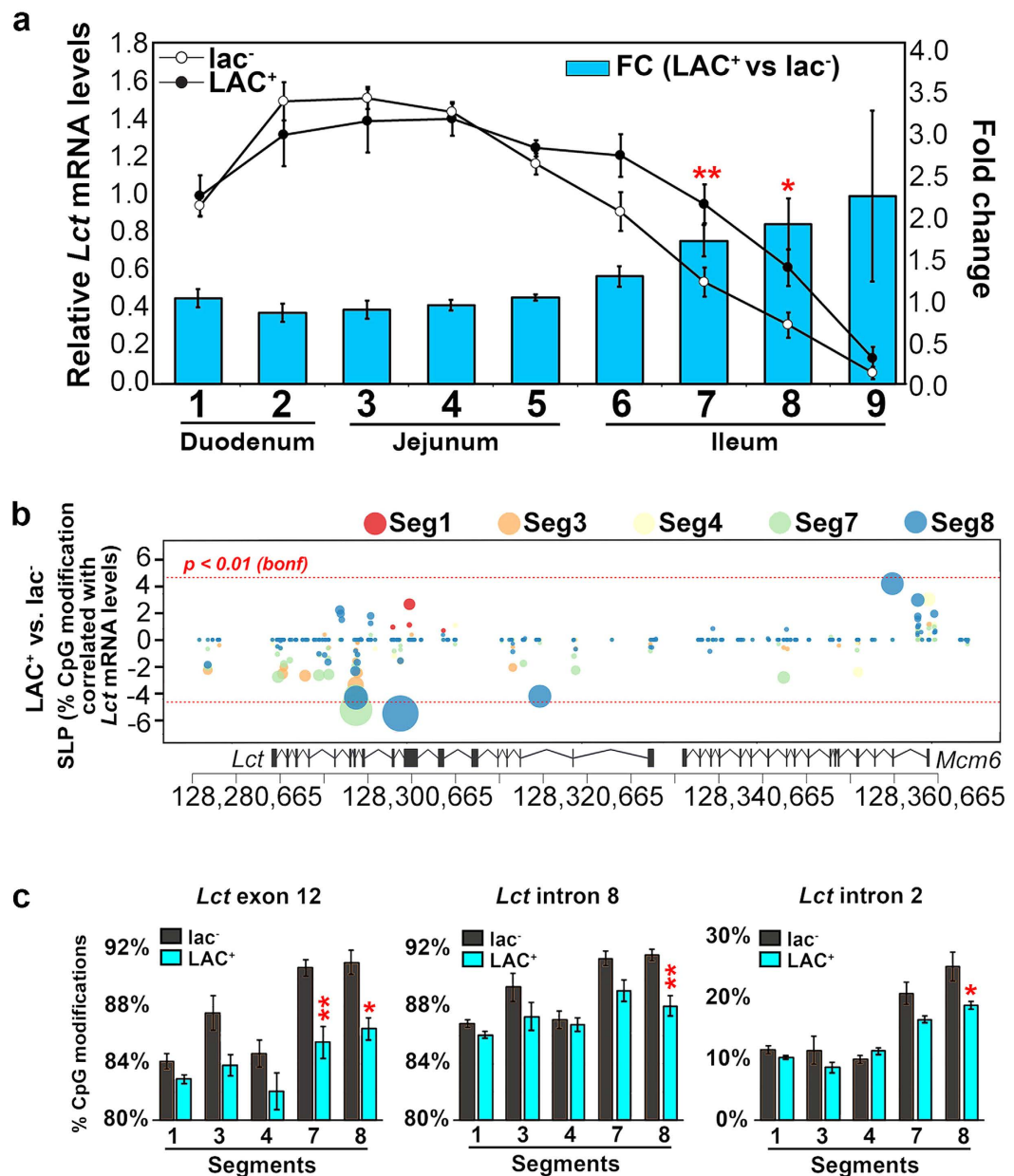


Figure 4. Lactose-treatment induces DNA modification changes in intestinal subregions that correspond to changes in the *Lct* transcriptional gradient. (a) Mice given lactose-containing milk (*LAC*⁺, $n = 6$) for 60-days had segment-specific increases in *Lct* mRNA in comparison to mice given lactose-free milk (*lac*⁻, $n = 7$). *Lct* mRNA following lactose exposure was increased in the ileum (seg7–8). * $p < 0.05$, ** $p < 0.01$ by one-way ANOVA. (b) Changes in CpG modification were detected in intestinal segments that displayed *Lct* mRNA changes in response to lactose treatment ($p < 0.01$ after Bonferroni correction; dashed red line; *LAC*⁺ $n = 6-7$ mice per segment; *lac*⁻ $n = 5-7$ mice per segment). (c) Reductions in CpG modifications occurred in response to lactose treatment at *Lct* exon 12 (Factorial ANOVA; $F_{4,55} = 2.3$, $p = 0.074$), intron 8 ($F_{4,55} = 2.3$, $p = 0.074$) and intron 2 ($F_{4,55} = 3.1$, $p = 0.023$) in distal intestinal segments. Data are represented as average % CpG modification density \pm SEM. Effect of lactose treatment by intestinal segment was calculated by factorial ANOVA. * $p < 0.05$ ** $p < 0.01$, and *** $p < 0.01$ by Tukey's HSD post hoc test for between lactose treatment within intestinal segment.

CTCF binding along the intestine targets epigenetically-controlled sites associated with *Lct* transcriptional variation. We wanted to examine the potential mechanism behind how DNA modification status at the *Mcm6* exon 13–intron 13 region located ~13 kb upstream of the *Lct* gene was able to exert such strong effects on the *Lct* promoter. We chose to examine the contribution of the chromatin architectural protein CTCF because of its ability to mediate architectural DNA interactions that facilitate transcription³⁶. Furthermore, sequence-dependent CTCF occupancy can be regulated by DNA methylation status^{37–41} which leads to differential splicing and gene transcription. The contribution of CTCF binding in the *Lct-Mcm6* locus in mediating transcriptional gradients in

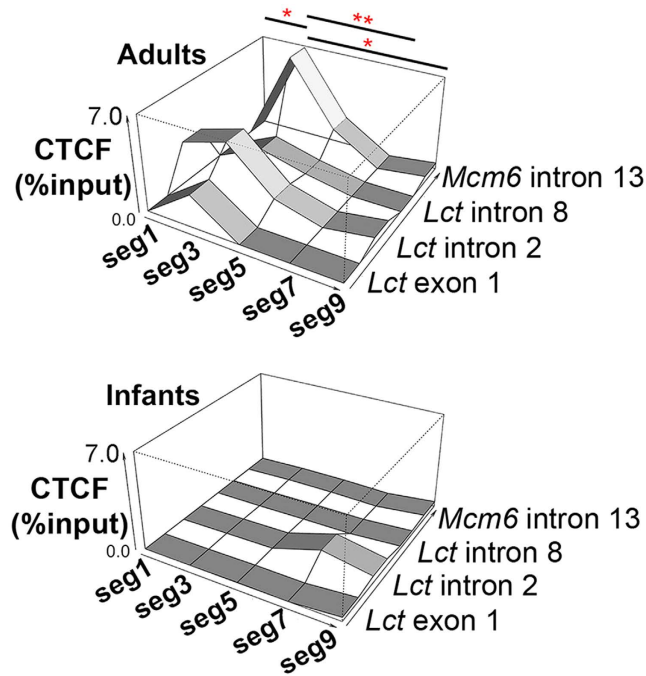


Figure 5. CTCF binding at epigenetically-controlled sites in the *Lct-Mcm6* locus corresponds to the *Lct* transcriptional gradient. ChIP-qPCR analysis of CTCF binding at *Mcm6* intron 13, *Lct* intron 8, *Lct* intron 2 and *Lct* exon 1 in adults (P60; top) and infants (P6; bottom) by segment (Data are represented as mean; $n = 3$ per group). Data presented as percent input after background normalization. * $p < 0.05$ and ** $p < 0.01$ by Tukey's HSD post hoc test for one-way ANOVA by effects of intestinal segment.

cells of the same type across the adult and infant mouse small intestine was investigated by ChIP-qPCR (Fig. 5). In adults, CTCF binding was highest in the proximal jejunum (segment 3) and gradually declined in more proximal and distal intestinal regions (main effect segment: $F_{4,19} = 5.12$, $p = 0.0083$). Segment differences in CTCF binding were most apparent at *Mcm6* exon 13–intron 13, and to a lesser extent at *Lct* intron 2. In infant mice, on the other hand, we found that CTCF was absent from *Lct-Mcm6* regulatory elements (Fig. 5).

Discussion

Our results indicate that divergent epigenetic programming enables aging- and environmentally-induced changes in gene transcription occurring in cells of the same type (Fig. 6). In particular, the *Mcm6* exon 13–intron 13 site was found to be a key modulator of the age-dependent establishment and maintenance of the *Lct* transcriptional gradient. These findings are consistent with reports that this upstream region is an enhancer of *Lct* transcription based on genetic (lactase persistence SNPs in various populations)^{28,30} and molecular evidence that demonstrate its ability to regulate *Lct* transcription *in vitro*^{42–44}. Our new observation that epigenetic control of *Mcm6* exon 13–intron 13 is important to *Lct* regulation in both mice and humans^{26,28,29} adds an epigenetic layer to these previous findings. Our study also suggests an evolutionary conservation of epigenetic regulation, which appears to precede genetic polymorphisms for the evolutionarily favorable lactase persistence trait^{45,46}. Furthermore, the *Mcm6* exon 13–intron 13 site was not sensitive to environmental cues (i.e. milk), indicative that age-dependent epigenetic programming of *Mcm6* exon 13–intron 13 is not malleable. By contrast, other regulatory elements affecting the *Lct* gradient (i.e. *Lct* intron 2 and intron 8) remained epigenetically adaptive to environmental signals, enabling a partial recovery of *Lct* expression in the adult intestine after weaning. Therefore, localized DNA modification changes accumulating with age may facilitate the intestinal *Lct* transcriptional gradient, while some remain partially dynamic to environmental signals.

CTCF occupancy was concordant with the *Lct* transcriptional gradient along the small intestine, and inversely correlated to DNA modification profiles at these genomic regulatory elements, particularly at *Mcm6* exon 13–intron 13. CTCF is a multifunctional protein that participates in many epigenetic regulatory functions, including insulation via enhancer blocking, imprinting, X chromosome inactivation, and both transcriptional activation and repression^{47,48}. CTCF can also influence DNA modification distribution both locally, through binding to chromatin boundaries, and distally, through effects on chromatin architecture^{37,49,50}. Here, our findings indicate that CTCF binding at the *Lct-Mcm6* locus functions as an intestinal region-specific transcriptional activator in adult mice, potentially by enabling DNA looping of distal enhancers, such as the *Mcm6* exon 13–intron 13 locus³². CTCF binding may limit the accumulation of cell-type specific DNA modifications with age at middle regions of the intestine relative to distal regions during development, whereas the absence of CTCF could promote down-regulation of *Lct* through unobstructed accumulation of DNA modifications and epigenetic silencing. On the other hand, the lack of CTCF at *Lct-Mcm6* sites in infant mice signifies that CTCF binding is not required for the expression of high *Lct* mRNA per se. Rather, CTCF occupancy (and its putative effects on chromatin structure)

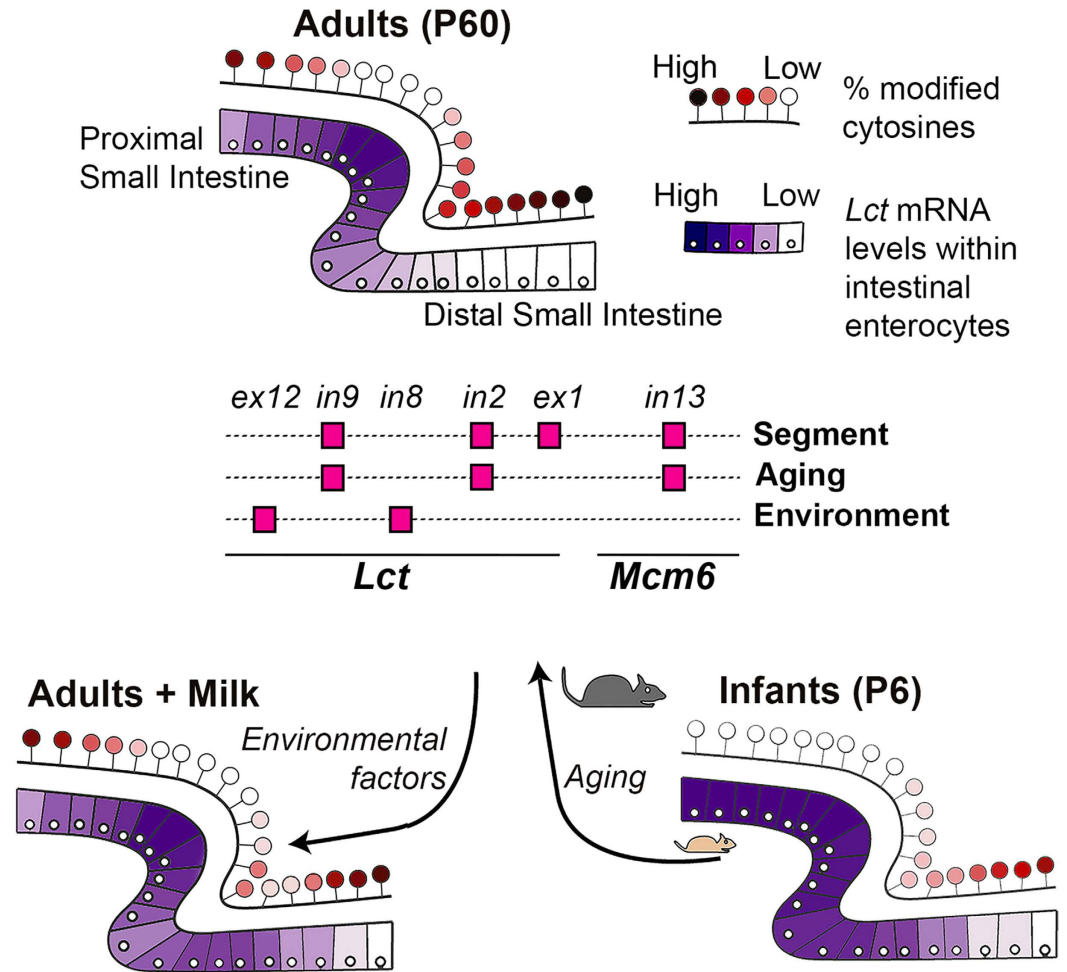


Figure 6. Relationship between *Lct* transcriptional gradient and DNA modification status. In enterocytes along the small intestine, transcriptional variation in the *Lct* gene corresponds to variation in DNA modification profiles at regulatory elements. The *Lct* transcriptional gradient is established with age and is modifiable by environmental signals; changes which co-occur with DNA modification density alterations in enterocytes at the same intestinal regions. The key genomic sites exhibiting DNA modification differences predictive of *Lct* transcriptional heterogeneity across intestinal segments and in response to aging and environmental signals are summarized.

occurs in tandem with DNA modification changes following weaning, to prevent uniform epigenetic silencing of *Lct* with age. The resulting effect in adulthood is that CTCF and the opposing DNA modification landscape work in concert to facilitate a transcriptional gradient in *Lct* across the intestine.

DNA modifications and CTCF may work in tandem with transcription factors to create and maintain age-dependent transcriptional gradients in cells of the same type. Transcription factors can interact with gene enhancers, including those affecting *Lct*^{25,44}. Transcription factors have also been shown to play a role in enterocyte differentiation along the crypt-to-villi axis (i.e., HNF-1a) and in intestinal subregion specialization (i.e., GATA4)^{35,51–53}. For example, GATA4 has been shown to be upregulated in the proximal small intestine and downregulated in the distal small intestine to enable the development and divergence of duodenal and ileal portions of the intestine^{5,6}. Together, our findings suggest an interplay between molecular and epigenetic factors that facilitate the biological specialization of cell of the same type. The molecular mechanism behind this concerted effort of CTCF, DNA modifications, and transcription factors at enhancer region(s) should be investigated in future studies.

Profiling epigenetic modifications of individual cell types in a tissue region-specific manner could offer insights into tissue specialization. Our findings emphasize the future studies should examine epigenetic contributions to the transcriptional divergence of numerous genes within cells of the same type, as these could help explain why tissue subregions can perform diverse biological functions^{4,10,14} and vary widely in disease susceptibility and treatment^{54,55}. In addition, our findings provide an important lesson for epigenetic studies of phenotypes, as failure to consider within cell-type transcriptional variation and epigenetic divergence limits the detection of biologically significant effects. Future epigenetic and gene regulation studies in health and disease will be greatly refined by not only isolating the cell type of interest, but by sampling a single cell type across tissue subregions, and across aging and environmental parameters.

Methods

Mouse intestinal samples and milk treatments. Infant C57BL/6NCrl mice at postnatal day 6 (P6) and adult mice at postnatal day 60 (P60) and 90 (P90) were used to investigate epigenetic changes across the small intestine (segments 1–9). To investigate diet/environmental associated epigenetic changes, P30 mice were supplied *ad libitum* sterile food and either 2% lactose-containing milk or lactose-free milk in the place of water for 60 days. All animal procedures were approved by the Institutional Animal Care Committee of the Toronto Centre for Phenogenomics (TCP) and compiled per the requirements of the Canadian Council on Animal Care and Province of Ontario Animals for Research Act.

RNA extraction and quantitative PCR. Enterocyte *Lct* mRNA levels were examined in each of the nine intestinal segments. For each segment, a small proximal portion (~30 mg) was homogenized with a ceramic bead-based homogenizer. Total RNA was extracted using Qiagen RNeasy Mini Kit with Qiagen RNase-free DNase I. RNA yield was quantified using NanoDrop ND-1000 (Thermo Fisher Scientific), and RNA integrity was verified via the Agilent Bioanalyzer 2100 system (Agilent Technologies). Purified RNA was converted to cDNA using High Capacity RNA-to-cDNA Kit (Life Technologies). *Lct* (Mm01285112_m1) and *Mcm6* (Mm00484848_m1) mRNA levels were quantified using TaqMan Gene Expression Master Mix (Life Technologies) using Applied Biosystems ViiA 7 real-time PCR system. The enterocyte marker *Villin-1* (Mm00494146_m1) mRNA was used as endogenous control for both *Lct* and *Mcm6* gene expression. $\Delta\Delta C_t$ was used to calculate the relative steady-state mRNA levels of each sample. Analysis was performed using repeated-measures (RM) ANOVA, and significant interactions were analyzed by Tukey's honest significant difference (HSD) post hoc comparisons. *Mcm6* mRNA levels did not display tissue subregion variations, indicating that within cell-type transcriptional gradients can be gene specific.

Fine-mapping of DNA modifications in enterocytes. Enterocytes were isolated from only the villi of the small intestine segments (see Supplementary methods). The purity and specificity of the villi enterocyte isolation was confirmed to be $84.8\% \pm 4.4$ (median \pm SEM) across intestinal segments through histological analysis²⁶ and fluorescence-activated cell sorting (FACS) using mice expressing fluorescent protein, mKate2, under the control of the enterocyte-specific *Villin-1* promoter (see Supplementary methods and Supplementary Fig. 1). DNA modification state in the villi enterocytes was examined with single nucleotide resolution in the *Lct-Mcm6* locus of mice using the bisulfite padlock probe technique²⁷. Padlock probes ($n = 314$) were designed to target non-repetitive genomic sequences on both DNA strands (GRCm38/mm10; Supplementary File 1). Enrichment of targeted bisulfite-converted DNA was done as described²⁷ (see Supplementary methods), and sequencing was performed on an Illumina HiSeq 2500. Preprocessed reads were mapped onto a bisulfite-converted mouse genome (GRCm38/mm10) using Bismark⁵⁶. Modification estimates were included only for cytosines with > 30 reads. DNA modification density was interrogated at 7,580 modified cytosines (609 CpG sites) across 148 unique samples, and an additional 19 replicates. All data are available from the NCBI Gene Expression Omnibus (GEO) database under accession number GSE76373.

Statistical analysis of DNA modification data. All statistical analyses were performed with R statistical software. To investigate whether DNA modifications of isolated villi enterocytes were associated with transcriptional variation in *Lct* in enterocytes across the small intestine, we checked for normality at individual cytosines using a QQ plot, and performed a Pearson's correlation between DNA modification density and steady-state *Lct* mRNA levels for each individual cytosine, using the 9 segments of the small intestine ($n = 53$, 5–6 mice per segment). Significance was set at $p < 0.01$ after Bonferroni correction for multiple testing. Clusters of significantly associated modified cytosines contained 3 or more cytosines within 500 bp. *P*-values are expressed as the $-\log p$ -value of the correlation coefficient, with the sign (+/−) representing the direction of Pearson's correlation (SLP). Genomic sites with DNA modification clusters significantly associated with *Lct* transcriptional variation were further investigated. At these sites, significant changes in the average % DNA modification status in enterocytes across the intestine was determined by one-way ANOVA. *Mcm6* mRNA (Fig. 1b) and DNA modifications at negative control regions in *Mcm6* (*Mcm6* intron 1; Fig. 2) did not show significant differences between intestinal segments. No significant sex differences were observed as determined by repeated measures (RM)-ANOVA. The sequences were aligned to small intestine (adult 8 weeks) ENCODE tracks for CTCF (ENCSR000CED), H3K4me1 (ENCSR000CCR), H3K4me3 (ENCSR000CCS), H3K27ac (ENCSR000CCQ), and MEL cell line, GATA-1(ENCSR000EUG).

For testing age-associated change in DNA modifications, segment 1, 3 and 7 of P6 infants were compared against corresponding segments in P60 adults ($n = 3–6$ per segment per age group). We determined which genomic sites showed significant age-associated differences by analyzing change in DNA modification (adults P60 minus infants P6). Diet-associated DNA modification changes were identified by comparing LAC⁺ ($n = 34$; 6–7 per segment) and lac[−] ($n = 31$; 5–7 per segment) P90 mice. A group-wise variance matrix showed that heavily modified CpGs (>90%) lacked deviation and these were removed from analysis in this experiment (Supplementary Table S1). Significant age- and diet-induced DNA modification changes were identified after Bonferroni correction.

Chromatin immunoprecipitation of CTCF. Chromatin immunoprecipitation (ChIP) was performed to investigate CTCF binding in intestinal segments 1, 3, 5, 7 and 9 of both P6 and P60 mice ($n = 3$ per group). Tissue homogenization and ChIP were performed using the MAGnify ChIP kit (Life Technologies). Immunoprecipitation was performed overnight, using 3 μ l monoclonal CTCF antibody (Pierce G.758.4), and negative control reactions used 1 μ g of mouse IgG antibody (Life Technologies). Input controls were also taken for each sample. qPCR was performed with Universal SYBR Green Supermix (Bio-Rad) in triplicate for four *Lct-Mcm6* loci and two negative control locations up/downstream of this site (primers listed in Supplementary Table S2).

References

1. Dueck, H. *et al.* Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. *Genome Biol* **16**, 122, doi: 10.1186/s13059-015-0683-4 (2015).
2. Lovatt, D. *et al.* Transcriptome *in vivo* analysis (TIVA) of spatially defined single cells in live tissue. *Nat Methods* **11**, 190–196, doi: 10.1038/nmeth.2804 (2014).
3. Munsky, B., Neuert, G. & van Oudenaarden, A. Using gene expression noise to understand gene regulation. *Science* **336**, 183–187, doi: 10.1126/science.1216379 (2012).
4. Anderle, P. *et al.* Changes in the transcriptional profile of transporters in the intestine along the anterior-posterior and crypt-villus axes. *BMC Genomics* **6**, 69, doi: 10.1186/1471-2164-6-69 (2005).
5. Bosse, T. *et al.* Gata4 is essential for the maintenance of jejunal-ileal identities in the adult mouse small intestine. *Molecular and cellular biology* **26**, 9060–9070, doi: 10.1128/MCB.00124-06 (2006).
6. Kohlnhofer, B. M., Thompson, C. A., Walker, E. M. & Battle, M. A. GATA4 regulates epithelial cell proliferation to control intestinal growth and development in mice. *Cellular and molecular gastroenterology and hepatology* **2**, 189–209, doi: 10.1016/j.jcmgh.2015.11.010 (2016).
7. Braeuning, A. *et al.* Differential gene expression in periportal and perivenous mouse hepatocytes. *FEBS J* **273**, 5051–5061, doi: 10.1111/j.1742-4658.2006.05503.x (2006).
8. Gebhardt, R. & Matz-Soja, M. Liver zonation: Novel aspects of its regulation and its impact on homeostasis. *World journal of gastroenterology* **20**, 8491–8504, doi: 10.3748/wjg.v20.i26.8491 (2014).
9. Schleicher, J. *et al.* Zonation of hepatic fatty acid metabolism - The diversity of its regulation and the benefit of modeling. *Biochim Biophys Acta* **1851**, 641–656, doi: 10.1016/j.bbali.2015.02.004 (2015).
10. Barth, A. S. *et al.* Functional profiling of human atrial and ventricular gene expression. *Pflugers Arch* **450**, 201–208, doi: 10.1007/s00424-005-1404-8 (2005).
11. Muller-Ehmsen, J., Wang, J., Schwinger, R. H. & McDonough, A. A. Region specific regulation of sodium pump isoform and Na,Ca-exchanger expression in the failing human heart—right atrium vs left ventricle. *Cellular and molecular biology* **47**, 373–381 (2001).
12. Small, E. M. & Krieg, P. A. Molecular regulation of cardiac chamber-specific gene expression. *Trends in cardiovascular medicine* **14**, 13–18 (2004).
13. Wu, S. P. *et al.* Atrial identity is determined by a COUP-TFII regulatory network. *Developmental cell* **25**, 417–426, doi: 10.1016/j.devcel.2013.04.017 (2013).
14. Scheller, E. L. *et al.* Region-specific variation in the properties of skeletal adipocytes reveals regulated and constitutive marrow adipose tissues. *Nat Commun* **6**, 7808, doi: 10.1038/ncomms8808 (2015).
15. Avraham, R. *et al.* Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell* **162**, 1309–1321, doi: 10.1016/j.cell.2015.08.027 (2015).
16. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330, doi: 10.1038/nature14248 (2015).
17. Barrero, M. J., Boue, S. & Izpisua Belmonte, J. C. Epigenetic mechanisms that regulate cell identity. *Cell Stem Cell* **7**, 565–570, doi: 10.1016/j.stem.2010.10.009 (2010).
18. Sheaffer, K. L. *et al.* DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes Dev* **28**, 652–664, doi: 10.1101/gad.230318.113 (2014).
19. Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Biol* **16**, 593–610, doi: 10.1038/nrm4048 (2015).
20. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* **13**, 229–232, doi: 10.1038/nmeth.3728 (2016).
21. Farlik, M. *et al.* Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep* **10**, 1386–1397, doi: 10.1016/j.celrep.2015.02.001 (2015).
22. Stelzer, Y., Shivalila, C. S., Soldner, F., Markoulaki, S. & Jaenisch, R. Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell* **163**, 218–229, doi: 10.1016/j.cell.2015.08.046 (2015).
23. Lee, M. F., Russell, R. M., Montgomery, R. K. & Krasinski, S. D. Total intestinal lactase and sucrase activities are reduced in aged rats. *The Journal of nutrition* **127**, 1382–1387 (1997).
24. Rings, E. H. *et al.* Restriction of lactase gene expression along the proximal-to-distal axis of rat small intestine occurs during postnatal development. *Gastroenterology* **106**, 1223–1232 (1994).
25. Fang, R., Olds, L. C. & Sibley, E. Spatio-temporal patterns of intestine-specific transcription factor expression during postnatal mouse gut development. *Gene Expr Patterns* **6**, 426–432, doi: 10.1016/j.modgep.2005.09.003 (2006).
26. Labrie, V. *et al.* Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nat Struct Mol Biol* **23**, 566–573, doi: 10.1038/nsmb.3227 (2016).
27. Diep, D. *et al.* Library-free methylation sequencing with bisulfite padlock probes. *Nat Methods* **9**, 270–272, doi: 10.1038/nmeth.1871 (2012).
28. Tishkoff, S. A. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* **39**, 31–40, doi: 10.1038/ng1946 (2007).
29. Enattah, N. S. *et al.* Identification of a variant associated with adult-type hypolactasia. *Nat Genet* **30**, 233–237, doi: 10.1038/ng826 (2002).
30. Enattah, N. S. *et al.* Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *American journal of human genetics* **82**, 57–72, doi: 10.1016/j.ajhg.2007.09.012 (2008).
31. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74, doi: 10.1038/nature11247 (2012).
32. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–910, doi: 10.1016/j.cell.2015.07.038 (2015).
33. Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**, 234–246, doi: 10.1038/nrg3663 (2014).
34. van Wering, H. M. *et al.* Complex regulation of the lactase-phlorizin hydrolase promoter by GATA-4. *Am J Physiol Gastrointest Liver Physiol* **287**, G899–909, doi: 10.1152/ajpgi.00150.2004 (2004).
35. Aronson, B. E. *et al.* GATA4 represses an ileal program of gene expression in the proximal small intestine by inhibiting the acetylation of histone H3, lysine 27. *Biochim Biophys Acta* **1839**, 1273–1282, doi: 10.1016/j.bbagr.2014.05.018 (2014).
36. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev* **30**, 881–891, doi: 10.1101/gad.277863.116 (2016).
37. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* **22**, 1680–1688, doi: 10.1101/gr.136101.111 (2012).
38. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495, doi: 10.1038/nature10716 (2011).
39. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79, doi: 10.1038/nature10442 (2011).
40. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484–492, doi: 10.1038/nrg3230 (2012).

41. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485, doi: 10.1038/35013100 (2000).
42. Liebert, A. *et al.* *In Vitro* Functional Analyses of Infrequent Nucleotide Variants in the Lactase Enhancer Reveal Different Molecular Routes to Increased Lactase Promoter Activity and Lactase Persistence. *Annals of human genetics*, doi: 10.1111/ahg.12167 (2016).
43. Troelsen, J. T., Olsen, J., Moller, J. & Sjoström, H. An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology* **125**, 1686–1694 (2003).
44. Jensen, T. G. *et al.* The –14010°C variant associated with lactase persistence is located between an Oct-1 and HNF1alpha binding site and increases lactase promoter activity. *Human genetics* **130**, 483–493, doi: 10.1007/s00439-011-0966-0 (2011).
45. You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same coin? *Cancer cell* **22**, 9–20, doi: 10.1016/j.ccr.2012.06.008 (2012).
46. Heyn, H. *et al.* DNA methylation contributes to natural human variation. *Genome Res* **23**, 1363–1372, doi: 10.1101/gr.154187.112 (2013).
47. Kim, S., Yu, N. K. & Kaang, B. K. CTCF as a multifunctional protein in genome regulation and gene expression. *Experimental & molecular medicine* **47**, e166, doi: 10.1038/emmm.2015.33 (2015).
48. Wallace, J. A. & Felsenfeld, G. We gather together: insulators and genome organization. *Current opinion in genetics & development* **17**, 400–407, doi: 10.1016/j.gde.2007.08.005 (2007).
49. Kemp, C. J. *et al.* CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep* **7**, 1020–1029, doi: 10.1016/j.celrep.2014.04.004 (2014).
50. Hou, C., Zhao, H., Tanimoto, K. & Dean, A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20398–20403, doi: 10.1073/pnas.0808506106 (2008).
51. Boudreau, F. *et al.* Hepatocyte nuclear factor-1 alpha, GATA-4, and caudal related homeodomain protein Cdx2 interact functionally to modulate intestinal gene transcription. Implication for the developmental regulation of the sucrase-isomaltase gene. *The Journal of biological chemistry* **277**, 31909–31917, doi: 10.1074/jbc.M204622200 (2002).
52. Walker, E. M., Thompson, C. A., Kohlnhofer, B. M., Faber, M. L. & Battle, M. A. Characterization of the developing small intestine in the absence of either GATA4 or GATA6. *BMC research notes* **7**, 902, doi: 10.1186/1756-0500-7-902 (2014).
53. Middendorp, S. *et al.* Adult stem cells in the small intestine are intrinsically programmed with their location-specific function. *Stem cells* **32**, 1083–1091, doi: 10.1002/stem.1655 (2014).
54. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* **29**, 1120–1127, doi: 10.1038/nbt.2038 (2011).
55. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401, doi: 10.1126/science.1254257 (2014).
56. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572, doi: 10.1093/bioinformatics/btr167 (2011).

Acknowledgements

We thank Aiping Zhang, Miki Susic and Sasha Ebrahimi for technical assistance, Matt Carlucci and Orion Buske for statistical recommendations, as well as Akhil Nair for comments on the manuscript. We also thank Rachael Sheridan from Van Andel Research Institute Flow Cytometry Facility for thorough technical assistance and analysis. This work was supported in part by the Canadian Institutes of Health Research (MOP-199170, MOP-119451 and MOP-77689), the US National Institutes of Health (MH088413, DK085698), the Krembil Foundation and Brain Canada to A.P. A.P. is supported as the Tapscott Chair in Schizophrenia Studies, University of Toronto, Canada. V.L. is supported by grants from the Alzheimer's Society of Canada (16 15) and Scottish Rite Charitable Foundation of Canada (15110).

Author Contributions

E.O. and V.L. designed and coordinated the study. E.O. and R.J. performed the experimental work. E.O. performed the computational analysis. G.O. and D.G. designed the bisulfite padlock probes and contributed to the computational analysis of bisulfite padlock sequencing reads. K.L. generated transgenic mice for FACS, and K.L. and K.F. performed the FACS experiments under the supervision of P.W.L. The manuscript was written by E.O., V.L. and A.P. The manuscript was commented on and reviewed by all authors.

Additional Information

Accession codes: Bisulfite sequencing data can be access on GSE76373.

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Oh, E. *et al.* Transcriptional heterogeneity in the lactase gene within cell-type is linked to the epigenome. *Sci. Rep.* **7**, 41843; doi: 10.1038/srep41843 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017