Large-scale dissection suggests that ultraconserved elements are dispensable for mouse embryonic stem cell survival and fitness

Aksana Schneider^{1,4}, Michael Hiller^{4,5,*}, Frank Buchholz^{1,2,3,4,*}

¹ Medical Faculty and University Hospital Carl Gustav Carus, UCC Section Medical Systems Biology, TU Dresden, 01307 Dresden, Germany

² German Cancer Research Center (DKFZ), Heidelberg and German Research Consortium (DKTK) Partner Site Dresden, 01307 Dresden, Germany

³ National Center for Tumor Diseases (NCT), University Hospital Carl Gustav Carus, TU Dresden, 01307 Dresden, Germany

⁴ Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany

⁵ Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany

*Correspondence: frank.buchholz@tu-dresden.de, hiller@mpi-cbg.de

Working title: Role of ultraconserved elements in mESCs

Keywords: noncoding; ultraconserved elements; embryonic stem cells; CRISPR/Cas9

ABSTRACT

Ultraconserved elements (UCEs) are genomic regions that are highly conserved among vertebrates and are under extreme purifying selection. Despite this extreme conservation, the biological importance of UCEs has remained elusive thus far. Here, we investigate the relevance of 93 UCEs for survival and fitness of mouse embryonic stem cells (mESCs) utilizing CRISPR/Cas9 technology. Surprisingly, systematic analyses of deletion clones revealed no measurable effects on mESC viability, proliferation or morphology. Furthermore, growth of homozygous UCE knockout clones under stress conditions or growth in direct competition with wild-type mESC revealed no phenotypic difference. Likewise, no difference in differentiation potential and on the expression level of key marker genes of differentiation was observed upon embryonic body formation. While functional overlap (redundancy) between several UCEs has been proposed to account for the absence of effects in single UCE knockouts, we obtained phenotypically normal mESC clones carrying up to five homozygous UCEs deletions, challenging this model. Finally, integration of additional UCE copies into the genome of mESCs also revealed no adverse effects. In summary, under all tested experimental conditions, UCEs appear to be dispensable for fitness, viability and differentiation potential of mESCs, despite their strong conservation.

INTRODUCTION

The human genome comprises approximately 3 billion base pairs, of which 5-10% are under evolutionary constraint (Lindblad-Toh et al. 2011). About a third of these constrained regions encode for proteins (Abascal et al. 2018), while the rest is allocated to non-coding regions, which include non-coding RNAs and regulatory elements such as promoters, enhancers, silencers, and insulators that control gene expression (Hoffman et al. 2013). The most extremely conserved genomic regions are called UltraConserved Elements (UCEs). These genomic regions were initially defined as being more than 200 bp long, with perfect sequence identity between human, rat, and mouse (Bejerano et al. 2004). According to this criterion, 481 UCEs have been identified in total. Many UCEs are also conserved in other mammals and non-mammalian vertebrates such as chicken, frog and fish (Bejerano et al. 2004), signifying that UCEs have been under strong purifying selection for more than 450 million years and implying that they have important biological functions (Dermitzakis et al. 2005). Furthermore, population genetics has revealed that polymorphisms inside UCEs have very low derived allele frequencies, suggesting that these elements are indeed under strong purifying selection and not simply mutational cold spots (Drake et al. 2005; Katzman et al. 2007; Sakuraba et al. 2008).

Of the 481 UCEs, 111 overlap exons and might have additional roles in mRNA processing and alternative splicing (Bejerano et al. 2004; Lareau et al. 2007; Ni et al. 2007). The remaining 370 UCEs are non-exonic elements located in intronic and intergenic regions (Bejerano et al. 2004). A number of these non-exonic UCEs has been proposed to function as gene-regulatory elements, as many are able to direct tissue-specific expression in transgenic mice (Nobrega et al. 2003; Woolfe et al. 2005; Pennacchio et al. 2006; Visel et al. 2008). In addition, these UCEs are often found near key developmental genes (Bejerano et al. 2004) and exhibit preserved synteny and distances between themselves and their flanking genes (Simons et al. 2006).

Strikingly, despite the fact that the extraordinary degree of constraint on UCE sequences strongly implies that they are functionally important (Woolfe et al. 2005), previous

studies found that some UCEs are dispensable for the viability of mice. Specifically, transgenic mice with four UCEs deleted individually (Ahituv et al. 2007) or in pairs (Dickel et al. 2018) have been generated. The selected UCEs drive tissue-specific expression of a reporter gene in transgenic mice and are located near protein-coding genes that cause marked phenotypes when deleted. Contrary to expectations, heterozygous as well as homozygous mice of both sexes with individual or paired deletions of the UCEs were viable and fertile. Only simultaneous deletion of two UCEs adjacent to the *Arx* gene resulted in body mass reduction and a decrease in the density of cholinergic neurons in the striatum, although without a change of overall size or structure of the organs (Dickel et al. 2018). The results suggested that either the functions of chosen UCEs are only relevant in a non-laboratory environment or that the elements are highly redundant. Alternatively, the tested UCEs may not be representative of non-exonic UCEs in general.

In addition, it is possible that the effect of UCE deletions in relevant cells or organs is masked at the whole-organism level by compensatory mechanisms—(Juarez-Carreño et al. 2018). If this is the case, UCE knock-out (KO) in cell lines, where these compensatory mechanisms are not expected, could result in measurable phenotypes. We therefore sought to test this hypothesis in large-scale by deleting UCEs using CRISPR/Cas9 technology. Since many UCEs are predicted to be developmental enhancers, mouse embryonic stem cells (mESCs) were chosen for these experiments since these cells are the closest approximation to the developing embryo. Moreover, mESCs preserve their potency to differentiate (Poh et al. 2014), providing an opportunity to analyze the differentiation potential of modified cells.

RESULTS

Hetero- and homozygous deletion of 93 UCEs

To investigate the role of a large number of UCEs for mESC viability, we selected non-exonic UCEs that likely have regulatory activity in mESCs. To assign putative regulatory functions to each UCE, we made use of regulatory annotations generated by the human and mouse Encode projects (Consortium 2012; The et al. 2012). Specifically, we overlapped UCEs with regulatory annotations generated by ChromHMM, a method that integrates a variety of DNase-seq and ChIP-seq data of various histone modifications and assigns regulatory classes to genomic regions (Ernst and Kellis 2010; Ernst and Kellis 2012). We selected 54 UCEs that overlapped enhancers and/or binding sites for important pluripotency transcription factors (Oct4, Sox2, Nanog) in human or mouse ESCs (Supplemental_Table_S1). Therefore, these UCEs may have a role in ESC maintenance. In addition, we selected 26 UCEs that are classified as enhancers active in the brain and in endothelial cells, allowing us to test the effect of UCE deletion during the differentiation of mESCs into other cell types. We also included 13 non-enhancer UCEs as controls, some of which are located near predicted enhancer UCEs, allowing us to perform redundancy tests. Altogether, 93 non-exonic UCEs were selected for knockout in mESC (Supplemental_Table_S1). Of them, 50 showed enhancer activity in mouse embryos at e11.5 (Visel et al. 2008). To control for transfection efficiency and toxicity in our experiments, we finally included a genomic control region that does not contain genes or other conserved sequences.

In order to delete individual UCEs from the genome we utilized the power of CRISPR/Cas9 technology. For this we generated a mESC cell line stably expressing Cas9 and designed two sgRNAs flanking each target UCE. Pairs of sgRNAs for individual UCEs were cloned into a piggyBac vector (Supplemental_Figure_S1) and transfected into mES-Cas9 cells in an arrayed format together with a transposase expression plasmid (Figure 1). To investigate the efficacy of the employed gRNAs, we performed PCR analyses to detect UCE-overlapping deletions in the genome of cell pools. For all pools, we detected PCR bands signifying the deletion of the targeted UCEs (data not shown), demonstrating that

CRISPR/Cas9 genome editing is suitable for the generation of genomic deletions of UCEs. However, for all 93 PCR assays performed, the wild-type (WT) band was also visible (Figure 1), raising the possibility that homozygous UCE deletions might be lethal. To investigate this possibility and to quantitatively assess frequencies of homo- and heterozygous deletions, we single cell sorted multiple clones from each pool and analyzed deletions in a total of 3478 single clones (Figure 1), with an average of 37 clones per UCE. Surprisingly, for all 93 target UCEs, we identified clones harboring homozygous deletions, validated with PCR and sequencing (Supplemental_Figure_S2). Of all the clones, about 10% were homozygous for the desired deletions, and about 13% had a heterozygous status. The rest of the analyzed clones had small indels in the sgRNA target sites but no full-length deletions. Detailed information about all evaluated clones is presented in Supplemental_Table_S2. All acquired clones were stable in culture, although they had slightly different growth rates and morphology. Since we observed the same growth rate and morphology differences in sorted WT clones, these effects were likely not the consequence of the deletions but instead reflect natural clonal variability. Therefore, we conclude that homozygous deletions of 93 different UCEs can be produced and have no adverse effects on mESCs growth and viability.

Figure 1. Experimental workflow. For each of the 93 selected UCEs we designed two sgRNAs up- and downstream of the element, which were then cloned in pairs in a piggyBac vector. Then mES cells stably expressing Cas9 protein were transfected with pairs of sgRNAs for different UCEs in arrayed format. PCR analysis with primers flanking the targeted deletion was used to confirm the deletion of all 93 UCEs, exemplified by the electrophoretogram where the amplicons expected after deletion are marked with an asterisk. Since UCE deletion was incomplete, as shown by the visible wild-type band, we sorted single cells and performed a

total of 3478 PCRs to detect clones with a heterozygous and homozygous deletion. Portion of the analyzed electrophoretograms is presented. Homozygous UCE KO clones were used in different assays, including test for differences in fitness, stress resistance, and developmental potential. HeZ – heterozygous clones, HoZ – homozygous clones.

UCE knock-out and wild-type clones exhibit similar growth in direct competition

Although analysis of clones with homo- and heterozygous UCE deletion did not reveal severe effects, subtle growth defects may not be visible in isolation. To account for this possibility, we used a competitive cell growth assay to compare fitness of wild-type mESC and UCE deletion clones directly in the same well under the same conditions. A similar assay has been informative in *Drosophila melanogaster* embryos (Morata and Ripoll 1974), where a mixed cell population contained ‘winner cells’ with a higher fitness that gradually eliminate ‘loser cells’ with a lower fitness via competition-induced apoptosis, although loser cells grown alone were viable (Tamori and Deng 2011; Merino et al. 2016). Importantly, this competition assay was shown to be sensitive in detecting even subtle growth differences between two populations of cultured mammalian cells, which can be due to differences in proliferation rate, signaling, metabolism, or genome stability (Eekels et al. 2012; Sancho et al. 2013; Penzo-Mendez et al. 2015).

To assess cellular fitness of KO clones in a competition assay, we selected the three UCEs for which the lowest proportion of homozygous UCE deletion clones was detected (uce42, uce118 and uce291). We reasoned that if the low percentage of detected KOs reflects the importance of the deleted UCE, tested clones should be more vulnerable in the direct competition with wild-type cells. To directly compare fitness, we mixed wild-type mES cells expressing GFP with clones carrying homozygous deletions of uce42, uce118 and uce291 and measured the percentage of GFP-positive cells over the course of 18 days by FACS (Figure 2A and B). If deletion of a UCE confers a subtle disadvantage compared to wild-type clones, we expected that wild-type clones will outcompete UCE knockout clones, resulting in an increased proportion of GFP-positive cells. As a control experiment, we mixed two wild-

type cell populations that just differ in GFP expression or cells with KO of the control region containing no genes or conserved elements. Testing 20 different control clones showed variations in the proportion of GFP-positive cells that is consistent with the existence of “winner” and “loser” cells in the population but showed no systematic trend that one cell population always wins or loses, as expected. Next, for each of the UCE KOs, we tested 4-6 different clones individually to allow statistical analyses. Surprisingly, the results showed that 9 out of 15 UCE KO clones outperformed the wild-type cells, whereas in only one experiment the wild-type cells won the competition assay against the KO clone (Figure 2C). Hence, many KO clones showed a tendency to overgrow mES-GFP cells, but clearly rejected the hypothesis that UCE KO negatively affects cell growth and survival. Overall, we conclude that clones with a deletion of the three UCes that were most likely to induce growth defects behaved indistinguishably from their wild-type counterparts in the direct competition assay.

Figure 2. UCE deletion clones have no growth defects in direct competition with WT cells. A) Scheme of the experiment. A pool of GFP positive wild-type cells was mixed with a pool of homozygous UCE KO cells in equal proportions. Cells were followed for 18 days by monitoring GFP percentage. B) Example of FACS plots from the GFP measurement for control region that is devoid of genes or conserved sequences for days 2, 10 and 18 shows a stable percentage of GFP positive cells. FSC is forward scatter used to determine size of the cells.

C) Heatmap shows the percentage of GFP positive cells in the cell competition assay across nine time points. Orange color indicates an increase in GFP positive cells (up to 100% positive cells), blue color indicates an increase in UCE deletion cells (down to 0% GFP positive cells). Hollow triangle marks the control clone shown in panel B. To assess changes in GFP percentage over time, a linear regression analysis was performed, asterisks (*) mark $p < 0.01$.

UCE knock-outs have no effect on stress tolerance

While many developmental genes have multiple, seemingly-redundant enhancers that can compensate for the deletion of a single enhancer (Lagha et al. 2012), phenotypic effects of single enhancer KO can be revealed under stress conditions, where buffering by the remaining enhancers is less efficient. For example, while shadow enhancers of *Drosophila shavenbaby* buffer the effect of single enhancer deletions on larval trichome development at optimal temperatures, a significant loss of trichomes occurs under stress induced by low or high temperatures (Frankel et al. 2010). To test whether UCE deletion clones have a lower capacity to tolerate stress, we measured mESC viability under non-standard culturing conditions (Figure 3A). To this end, we subjected cells to temperature stress, oxidative stress, acidification and treatment with ethanol, using a stress intensity that impairs growth of wild-type cells but does not abolish it. For these experiments we employed UCE KOs with multiple available clones to account for potential clonal variability.

Figure 3. UCE knock-outs do not affect cellular resistance to stress conditions. A) Scheme of the experiment. UCE KO and WT cells were seeded in 96-well plates and divided into two groups. The first group was kept under standard culture conditions while the second group was subjected to stress. To test whether UCE deletion affects cell viability during stress conditions, we analyzed the number of viable cells by flow cytometry based on PI staining. B)

The difference in the number of viable cells between stress and normal conditions are shown on the Y-axis. Mean and standard variation was calculated from 3 replicates. All tested clones are displayed along the X-axis. We tested multiple clones with the same UCE deletion to control for clonal variation and pipetting error. Alternating colors mark groups of the clones with the deletion of the same UCE. Wild-type clones or cells with KO of the control region are shown on the right (dashed line). Multiple t-test was performed to compare the difference between cell number in control and experimental conditions. Significantly different clones are marked with a red asterisk, FDR 1%.

First, since various “cold shock” treatments have been shown to influence viability (Wang et al. 2017), differentiation potential (Belinsky and Antic 2013; Saito et al. 2013), and DNA integrity (Guo et al. 2018) of human and mouse stem cells, we cultured UCE-KO clones at a lower temperature of 35°C. Of the 122 clones examined, only one clone with a deletion of uce283 showed a significant decrease in viable cells numbers under stress conditions. However, three independent clones with the same UCE-KO were not affected (Figure 3B), suggesting that the effect is caused by clonal variability rather than a phenotype of the deleted UCE. As a second stress condition we grew the cells at low pH. Medium acidification is known to cause DNA damage due to lactate accumulation (Jacobs et al. 2016), and allows to test for metabolic, genome instability and DNA repair deficiencies among knock-out clones. We kept experimental plates with all available homo- and heterozygous clones without medium change for a week, resulting in acidification, while the medium in control plates was changed every day. Of all 122 clones, none was significantly affected by acidification (Figure 3B). Third, ethanol treatment was applied, which has adverse effects on the developing embryo (Arzumanyan et al. 2009). After cells were subjected to treatment with 1% ethanol for 48h, one clone with a uce470 KO showed a significant growth defect, but the second clone with the same deletion grew normally (Figure 3B), again arguing against a reproducible phenotypic consequence of the UCE deletion. Fourth, we subjected cells to oxidative stress, which affects many cellular activities. We started by seeding cells in a medium with a low dosage of H₂O₂

(12 μ M). Of the 122 clones, one clone with a uce10 KO exhibited a reduced number of viable cells comparing to the growth of the same clone under normal conditions (Figure 3B). However, this clone was one out of three clones with independent deletion of the same UCE, again most likely indicating clonal variability instead of real stress susceptibility.

In summary, exposure of the clones with UCE KOs to various sub-lethal stress conditions showed no significant decrease in cellular resistance to the environmental fluctuations. From this data we conclude that UCE deletions do not affect resistance and robustness of mESCs to the tested stress conditions.

UCE knock-outs have no effect on stem cell differentiation

Based on the observation that UCEs are often located in proximity to developmental genes, it has been hypothesized that they may function during embryo development and cell differentiation (Nobrega et al. 2003; Bejerano et al. 2004; Woolfe et al. 2005). In addition to the 54 UCEs overlapping putative enhancers in ESC, we selected 26 UCEs that overlap potential enhancers in more differentiated cell types. Knock-outs of these UCEs might not influence mESC viability under self-renewal culture conditions but may result in different phenotypes upon differentiation. Therefore, we initiated differentiation of all mESC clones through embryoid body (EB) formation, which gives rise to derivatives of all three germ layers (Doetschman et al. 1985; Leahy et al. 1999; Wang and Yang 2008; Poh et al. 2014). We reasoned that an effect on differentiation would be detectable via altered expression of key genes involved in establishment of different cell types. To assess the fidelity of differentiation and detect any potential deviations from this process, we used the NanoString platform to measure the expression of 95 pluripotency and germ layer marker genes as well as 122 neighboring genes (Supplemental_Table_S3) that could be regulated by UCEs after initiation of differentiation (Figure 4A).

Figure 4. UCE knock-outs have little effects on the expression of neighboring and marker genes during stem cell differentiation. A) Scheme: UCE KO and WT clones were used for EB formation induced by the hanging drop method and then allowed to undergo spontaneous differentiation. After 12 days all cells were analyzed for expression of selected genes that comprise 96 pluripotency and germ layer marker genes and 122 genes adjacent to the deleted UCEs. B) Expression levels of germ layer marker and pluripotency genes. For each gene, the expression level in each tested EB normalized to the mean of the expression of all EBs is shown on the Y-axis. Each dot represents the measurement of one gene in one of the EBs. Genes are placed along the X-axis. Outliers of the expression of every gene in every EB were identified using FDR cutoff of 1%. C) Expression of genes associated with UCEs. Genes adjacent to the deleted UCEs are shown in red. Multiple dots for one gene

represent multiple tested clones or replicas of experiments. The only element that showed a statistically significant change in expression of the corresponding gene is boxed (uce283).

First, we analyzed 12 samples from non-differentiated cells and 15 EBs derived from WT or heterozygous clones. The results showed that pluripotency genes were reproducibly downregulated, while differentiation marker genes for all lineages were consistently upregulated (Supplemental_Figure_S3), confirming that the differentiation protocol worked as expected. Next, we analyzed a total of 88 EBs representing clones with deletions of 66 UCEs, using expression of pluripotency and differentiation marker genes to detect possible deviations from the controls (Supplemental_Table_S3). Unexpectedly, we found that EBs from different UCE knock-outs showed no consistent variation in gene expression (Figure 4B), even for UCEs that overlap enhancers and/or binding sites for important pluripotency transcription factors.

We then analyzed the 122 genes located in the vicinity of the UCEs, as they may be regulated by these elements. We found that 80 of these 122 genes are expressed in mESC and EBs, but only one gene (*Prrxl1*) was confirmed to have changed expression in clones where the respective UCE (uce283) was deleted (Figure 4C). Uce283 is located 650 bp downstream of the transcription start site, and thus overlaps with the proximal promoter of *Prrxl1*, a gene that is required for the establishment of the neuronal sensory relay and nociceptive circuits in the spinal cord (Rebelo et al. 2007). Deletion of uce283 caused an upregulation of *Prrxl1*, which is normally not expressed in mESCs or differentiated EBs. Upregulation upon UCE deletion can be explained by the fact that the proximal promoter of *Prrxl1* mediates self-repression (Monteiro et al. 2014), suggesting that the deletion likely removed binding sites required for *Prrxl1* repression.

In summary, none of the tested UCE KOs showed a prominent difference during *in vitro* differentiation. Furthermore, only the deletion of one promoter-overlapping UCE resulted in a consistent expression difference of UCE associated genes. Apart from this special case,

no other UCE deletion caused an effect on expression or differentiation of marker genes, which is inconsistent with the idea that UCEs function as enhancers during early development.

Deletion of multiple UCEs has no adverse effects

Redundancy of UCEs, that is the presence of multiple elements with similar and overlapping functional roles, has been proposed as a possible explanation for the lack of phenotypes after deleting single UCEs. Consistent with enhancer redundancy being an important factor, previous studies found that deletions of single limb or eye enhancers did not lead to a detectable phenotype, while the deletion of pairs of enhancers resulted in measurable phenotypic defects (Antosova et al. 2016). Therefore, we tested the effect of deleting more than one UCE in the same cell. Since UCEs preserve synteny and distances between themselves and flanking genes (Simons et al. 2006; Sun et al. 2006; Li et al. 2018), we selected 32 pairs of UCEs located in the same genomic locus that are associated with the same genes (Figure 5A). We analyzed a total of 2454 clones with 76 clones per UCE pair on average (Figure 5B). For all 32 pairs of UCEs, we were able to obtain clones with a homozygous deletion of both UCEs, as confirmed by PCR and sequencing (Supplemental_Table_S4). The clones with double deletions were viable and stable under normal culture conditions. Although minor differences in the growth speed and morphology were observed, they did not exceed the degree of normal clonal variation. Next, we generated 27 triple and 9 quadruple deletions of UCEs located on the same chromosome, as well as clones with three different combinations of five homozygous deletions (Figure 5C). Again, the obtained cells had normal growth rates and morphology, indicating that multiple UCEs can be deleted in mESCs without obvious consequences.

Figure 5. Deletions of multiple UCEs are possible. A) Schematic representation of deleting multiple, potentially redundant UCEs that are associated with the same genes in a sequential fashion. B) Workflow of experiments. Double KO clones were generated by transfecting clones with one UCE deletion with sgRNAs that target the second UCE. After single cell sorting, PCR was used to identify double UCE deletion clones. This procedure was iterated to generate clones with multiple UCE deletions, as exemplified by the gel picture which shows a deletion of two additional UCEs in KO clone. White asterisks mark clones that carry a homozygous deletion of both UCEs. C) Gel picture with PCR analysis of wild-type (WT) cells and clones with deletions for 5 UCEs (KO). Note the reduced band size for the KO clones. For uce181 no DNA fragment could be amplified with the employed primers, probably reflecting a large deletion in this clone that deleted the primer binding sites.

In summary, we show that mESCs can tolerate deletions of multiple UCEs without detectable phenotypic changes or growth deviations. Thus, redundancy between UCEs does not seem to be a general factor that can explain the lack of observable phenotypes of single UCE deletions.

Additional copies of UCEs do not affect mESC growth and viability

UCEs have been shown to be significantly depleted in copy number variations, raising the possibility that the diploid ratio of the UCEs might be important for genome stability (McCole et al. 2014; McCole et al. 2018). Based on these observations, we tested whether UCE copy number and localization within the genome are important by inserting additional UCE copies into the genome of mESCs. If the extra elements are unfavorable for the cells, one would expect reduced cell proliferation, lower cell fitness or elevated cell death. We introduced four different UCEs or as a control a non-conserved non-exonic sequence and an empty vector into cells individually or combined using a piggyBac system (Figure 6A). Using a transposon-transposase ratio that generally results in stable integration of multiple copies (up to 15 per cell) (Wang et al. 2008; Saridey et al. 2009; Kahlig et al. 2010; Lu and Huang 2014), we generated six cell lines with different integrates. After allowing the transfected cells to grow for one week, we counted viable cells in each well. Strikingly, we observed no significant differences between the number of cells with additional copies of UCEs or control regions (Figure 6B). Thus, similar to the deletion of UCEs, insertion of additional UCE copies does not affect mESCs growth or viability.

Figure 6. Random integration of additional UCE copies in the mESC genome does not have adverse effects on cellular viability. A) Scheme: uce5, 8 and 71 and the control region

that is devoid of genes and conserved elements were cloned into the piggyBac vector. Individual vectors or a mix of different vectors were then transfected into WT mESCs in arrayed format. B) Number of viable cells (log 10) with multiple integrations of UCEs. Mean and standard variation, overlaid by grey dots that represent the results of one transfection. The insert carried by the integrated piggyBac vectors are specified on the x-axis, with 'UCE mix' referring to the integration of vectors that carry *uce5*, *uce8*, *uce71* or *uce444*. Ordinary one-way ANOVA test was performed to analyze the significance of the mean differences between cells with additional copies of UCE and empty piggyBac-U6 vector. ns – not significant.

DISCUSSION

Here, we systematically investigated the influence of deleting UCEs on viability, fitness and developmental potential of embryonic stem cells. By targeting 93 UCEs representing 25% of all non-exonic UCEs, we show that clones with homozygous deletions of UCEs have a fitness comparable to wild-type cells, both under normal and stress culture conditions. UCE deletion clones were also indistinguishable from wild-type cells during early stages of *in vitro* differentiation. Finally, deletions of multiple UCEs or random integration of additional UCEs produces no obvious phenotype. Overall, we did not find a single UCE deletion that exerts a consistent and measurable phenotype on mESCs biology. It therefore remains mysterious why these extremely constrained non-exonic genomic regions can be deleted without apparent phenotypic consequences. However, several points should be considered that deserve future investigation.

First of all, technical limitations might have contributed to the inability to observe phenotypes. In order to identify knockout clones, we had to rely on single cell sorting. This process might have influenced the results, in case only very robust cells were able to survive this selection procedure. Second, while we found no effect of UCE deletion on proliferation, morphology and viability of mESCs, it is possible that deleting UCEs in other cell types results in measurable cellular phenotypes. Third, the previously found smaller size of animals after

deleting two UCEs near *Arx* (Nolte et al. 2014; Dickel et al. 2018) could be evidence for the involvement of UCEs in the crosstalk of different cells, as sensing and coordination between cells influences organism size (Gokhale and Shingleton 2015). Consequently, interactions between different cell types in tissues and organs might be of future interest. For such investigations, our set of UCE-KO ESC clones represents a valuable resource.

All in all, the most remarkable result of our work is that deletions of UCEs do not lead to any detectable effects in mESC, despite targeting a large set of 93 UCEs and performing comparisons under a number of different conditions. Moreover, our results do not support hypothesis that UCE copy number and localization might be crucial for mESCs and suggest that redundancy is not a general factor explaining why individual UCEs can be deleted without resulting in measurable mESC phenotypes. Our results are consistent with previous studies that observed no phenotype upon deleting few selected UCEs; however, the large sample size and extensive analyses justify more general conclusions. We believe that our results will fuel greater interest into this topic to better understand the connection between evolutionary sequence constraint and its biological function.

METHODS

UCE annotation. The list of UCEs was downloaded from the UCSC genome browser custom track `ct_Ultra_7128` of Human Build 35 (hg17) (access date: December 2015). Genomic UCE coordinates in the mouse mm9 and mm10 genome assemblies and the human hg18 and hg19 genome assemblies were obtained by aligning the UCE sequences using BLAT. We downloaded regulatory genomic annotations for human embryonic stem cells (H1-hESC), human endothelial cells (HUVEC), mouse brain (C57BL/6-strain, 8-week-old), and mouse embryonic stem cells (E14Tg2A). The human data was downloaded from UCSC Genome Browser tracks `wgEncodeAwgSegmentationCombinedH1hesc` (hg19 genome assembly, access date: January 2016) and `wgEncodeBroadHmmNhekHMM` (hg18 assembly, access date: January 2016). Mouse data was downloaded from the GitHub repository https://github.com/gireeshkbogu/chromatin_states_chromHMM_mm9 (access date: January

2016). Genomic regions bound by Oct4, Sox2 or Nanog were downloaded from (Gifford et al. 2013; Whyte et al. 2013; Galonska et al. 2015). We intersected these annotations with the UCE coordinates to select 93 UCEs. The control genomic region that is devoid of genes or conserved sequences is chr10:72191211-72191615 in the mouse mm10 assembly.

Cloning sgRNAs into a piggyBac vector. sgRNAs were designed for the regions spanning 300 bp up- and downstream of the chosen UCEs or control regions that is devoid of genes or conserved sequences (named BR10 and BR11) using the weight matrix from (Doench et al. 2016). All sgRNAs used in this study are listed in Supplemental_Table_S5.

PiggyBac and transposase vectors used in this study were purchased from System Bioscience. Selected sgRNAs for all target UCEs and control regions were cloned into the piggyBac-U6 vector (#PB510B-1, modified).

For cloning, two sgRNAs with U6/H1 promoters were used: Forward:

5'AAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAA
TAGCAAG3'

Reverse:

5'GCATAGCTCTTAAACXXXXXXXXXXXXXXXXXXXXGGATCCAAGGTGTCTCATAC3'

N = Denotes sgRNA1 sequence X= Denotes Reverse Complement of sgRNA2 sequence.

All sgRNA sequences used for primers constructions are listed in Supplemental_Table_S9.

PCR amplification was performed according to the PrecisionX Multiplex gRNA Cloning Kit (System Bioscience). BbsI-linearized piggyBac-U6 backbone was mixed with purified PCE product in a vector:insert 2:1 ratio and was directly used for transformation of One Shot MAX Efficiency DH5α-T1 Competent Cells (Invitrogen).

All the constructs were confirmed for the inserted sequence in both orientations using forward (5' GGCCTCCAAGGCCACTAG 3') and reverse (5' ATCCTTCGCGGCCGCAAA 3') primers.

ES cell culture. Mouse ES cells of cell line E14 were maintained in Dulbecco's modified Eagle medium (Invitrogen GmbH) supplemented with 15% ES specific serum (Pan-biotech), 0.1 mM nonessential amino acids (Invitrogen), 0.1 mM 2-mercaptoethanol (Invitrogen), 100 lg/ml penicillin/ streptomycin solution (Invitrogen), and 1000 U/ml leukemia inhibitory factor (LIF; produced in MPI-CBG). Cells were passaged every 2 days and maintained at 37°C and 5% CO₂. The medium was changed every day.

Transfection. E14 cells were transfected using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions in a "reverse" manner, where the transfection mix is added directly to cell suspension prior to seeding the cells. We used a piggyBac vector : transposase ratio = 2:1. After 24 h, cells were washed with PBS and the medium was changed to the selection medium. Selection of transfected clones was obtained with 400 µg/ml geneticin (Invitrogen) or 2 µg/ml puromycin (Invitrogen).

Single cell sorting. Cells were harvested and re-suspended in sorting buffer (1X PBS+1 mM EDTA+25 mM HEPES+1% BSA). Cells were sorted by FACS Aria II and III cell sorters (BD Bioscience). Cells were sorted on PI marker at 4°C and plated on 0.1 % gelatin coated 96-well plates filled with normal medium.

Identification of mono- and bi-allelic deletions. Genomic DNA extraction and PCR for deletion detection was performed using Phire Animal Tissue Direct PCR Kit (Thermo Scientific) according to manufacturer's instructions. All PCR primers are presented in Supplemental_Table _S6. PCR products were visualized by 2% agarose gel electrophoresis. PCR products with suspected bi-allelic deletions were sequenced directly with the PCR primers. Sequencing was carried by Eurofins Genomics.

Cell growth competition assay. The pool of GFP positive cells was mixed with the tested clones in equal proportions (2500 cells each) in one well of a 96-well plate. Cells were split every second day in a 1:10 ratio and the GFP percentage was monitored using MACS Quant for a total of 18 days. As a control, we used WT mESCs and cells with KO of region that is devoid of genes or conserved sequences (position in mm10 chr10:72191211-72191615). In addition to the mixed populations, all clones were seeded individually to monitor their normal growth. Results were analyzed with the FlowJo Analysis Software. Samples were removed from the analysis if GFP% in starting populations was >60 or <40, as the mixture was not equal anymore. To assess changes in GFP percentage over time, a linear regression analysis was performed in GraphPad Prism 6 (GraphPad Software), $p < 0.01$.

Stress induction experiments. Available clones with deletions and control wild-type cells were seeded in 96-well plates, and then divided into two groups with three biological replicates each. We tested multiple clones with the same UCE deletion. When only one clone out of multiple showed a significant effect, this effect was attributed to clonal variation or pipetting error. As a control, we used WT mESC, cells with Cas9 and clones transfected with sgRNAs for uce42 or uce291, but without deletions. Viable cells were counted based on staining with 10 μ g/ml PI (Molecular Probes, 1mg/ml) using FACS Calibur (BD Biosciences) in high-throughput mode. To assess whether there is a significant difference in cell number between standard culture conditions and stress conditions, we used a t-test (implemented in the GraphPad Prism 6 Software) and computed the false discovery rate (FDR) to correct for multiple testing.

We tested the following stress conditions. For low temperature stress, the experimental group was placed in the incubator at 35°C, and cells from the control group were kept at 37°C. Viable cells were counted after 1-week incubation. For medium acidification, experimental plates with all available homozygous and heterozygous clones were kept without medium change for a week, while in the control plates medium was changed every day. For cytotoxic stress, absolute ethanol was added to ES medium to achieve a concentration of 1% and cells

were either seeded in this medium or normal medium with corresponding volume of water. Cell viability was accessed after 48 h. In the control group, the medium was also changed twice. For oxidative stress, cells were seeded directly in the medium with 12 μ M H₂O₂ and counted after 48 h.

EB formation. EB formation induced by the hanging drop method was carried out according to the protocol (Wang and Yang 2008) with modifications. To initiate EB formation, mESCs were dissociated into single cells with 0.1% trypsin-EDTA (Invitrogen) and cell suspension droplets consisting of 400-500 cells in 20 μ l of mESC medium without LIF were hung on the lids of 150 mm culture dishes (Corning). 8-12 embryoid bodies were formed from each clone. After two days, the EBs were transferred into a 96-well ultralow attachment plate (Corning) and left undisturbed for 3 days. For ESC three germ layer differentiation, EB cultured in suspension conditions were moved into 0.1% gelatin-coated 96-well tissue culture plates. mES medium without LIF supplement was changed every other day until day 12 when cells were harvested, trypsinized and used for further analysis.

NanoString gene expression analysis. A list of all tested genes is presented in Supplemental_Table_S5. Probes for gene expression assay were designed by the NanoString support team. Sample preparation was carried according to manufacturer's instructions from nCounter Vantage RNA Assays with Core TagSet + Extension TagSet protocol. For those UCEs where multiple deletion clones were available, we induced EB formation separately for the clones but mixed the lysates before the analysis. Probes were processed on an nCounter Prep Station and nCounter Digital Analyzer (NanoString Technologies).

For data analysis we first assessed the quality of the obtained data with the nSolver software (NanoString Technologies) and removed all samples with technical errors from the analysis. Reads counts were normalized using the same program based on the expression of

housekeeping genes. Second, outliers of the expression of every gene in every EB were identified using GraphPad Prism 6 at an FDR cutoff of 1%.

UCE associated genes were considered to be expressed based on NanoString results. Average maximum expression of negative controls for which no transcripts were present was used to set cut off at 30 counts.

UCE cloning into piggyBac-U6 vector. The genomic regions overlapping uce5, uce8, uce71, uce444, and the non-essential control region (position in mm10 chr9:12136109-12136360) were amplified from E14 wild-type gDNA with MyTaq DNA polymerase with the primers listed in Supplemental_Table_S3. Amplicons were then cloned into TA-vector and then re-cloned into an empty piggyBac-U6 cut with EcoRI. All the constructs were sequenced to confirm the inserted sequence in both orientations using PCR primers.

Viable cell counting. Cells cultivated in different conditions were harvested and stained with 10 µg/ml PI (Molecular Probes, 1mg/ml). Cells were analyzed in a MACS Quant analyzer (Miltenyi Biotec) in 96-well format with the same established settings. All the experiments were done in three replicates and results analyzed using FlowJo Analysis Software.

ACKNOWLEDGMENTS

The authors thank the members of the Buchholz and Hiller groups for helpful discussions and an ELBE PhD stipend for funding.

Figures were created with BioRender.

REFERENCES

Abascal F, Juan D, Jungreis I, Martinez L, Rigau M, Rodriguez JM, Vazquez J, Tress ML. 2018. Loose ends: almost one in five human genes still have unresolved coding status. *Nucleic Acids Res* **46**: 7070-7084.

Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol* **5**: e234.

Antosova B, Smolikova J, Klimova L, Lachova J, Bendova M, Kozmikova I, Machon O, Kozmik Z. 2016. The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. *PLoS Genet* **12**: e1006441.

Arzumnyan A, Anni H, Rubin R, Rubin E. 2009. Effects of ethanol on mouse embryonic stem cells. *Alcohol Clin Exp Res* **33**: 2172-2179.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. 2004. Ultraconserved Elements in the Human Genome. *Science* **304**: 1321-1325.

Belinsky GS, Antic SD. 2013. Mild Hypothermia Inhibits Differentiation of Human Embryonic and Induced Pluripotent Stem Cells. *BioTechniques* **55**.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Dermitzakis E, Reymond A, Antonarakis SE. 2005. Conserved non-genic sequences — an unexpected feature of mammalian genomes. *Nat Rev Genet* **6**: 151-157.

Dickel DE, Ypsilanti AR, Pla R, Zhu Y, Barozzi I, Mannion BJ, Khin YS, Fukuda-Yuzawa Y, Plajzer-Frick I, Pickle CS et al. 2018. Ultraconserved Enhancers Are Required for Normal Development. *Cell* **172**: 491-499 e415.

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**: 184-191.

Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R. 1985. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *Embryol exp Morph* **87**: 27-45.

Drake JA, Bird C, Nemesh J, Thomas DJ, Newton-Cheh C, Reymond A, Excoffier L, Attar H, Antonarakis SE, Dermitzakis ET et al. 2005. Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nature Genetics* **38**: 223-227.

Eekels JJ, Pasternak AO, Schut AM, Geerts D, Jeeninga RE, Berkhout B. 2012. A competitive cell growth assay for the detection of subtle effects of gene transduction on cell proliferation. *Gene Ther* **19**: 1058-1064.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**: 817-825.

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**: 215-216.

Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**: 490-493.

Galonska C, Ziller MJ, Karnik R, Meissner A. 2015. Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. *Cell Stem Cell* **17**: 462-470.

Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R et al. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**: 1149-1163.

Gokhale RH, Shingleton AW. 2015. Size control: the developmental physiology of body and organ size regulation. *Wiley Interdiscip Rev Dev Biol* **4**: 335-356.

Guo Q, Mintier G, Ma-Edmonds M, Storton D, Wang X, Xiao X, Kienzle B, Zhao D, Feder JN. 2018. 'Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. *Sci Rep* **8**: 2080.

Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes JA, Birney E et al. 2013. Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Research* **41**: 827-841.

Jacobs K, Zambelli F, Mertzanidou A, Smolders I, Geens M, Nguyen Ha T, Barbé L, Sermon K, Spits C. 2016. Higher-Density Culture in Human Embryonic Stem Cells Results in DNA Damage and Genome Instability. *Stem Cell Reports* **6**: 330-341.

Juarez-Carreño S, Morante J, Dominguez M. 2018. Systemic signalling and local effectors in developmental stability, body symmetry, and size. *Cell Stress* **2**: 340-361.

Kahlig KM, Saridey SK, Kaja A, Daniels MA, George AL, Jr., Wilson MH. 2010. Multiplexed transposon-mediated stable gene transfer in human cells. *Proc Natl Acad Sci U S A* **107**: 1343-1348.

Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, Wilson RK, Salama SR, Haussler D. 2007. Human genome ultraconserved elements are ultraselected. *Science* **317**.

Lagha M, Bothma J, Levine M. 2012. Mechanisms of transcriptional precision in animal development. *Trends in Genetics* **28**: 409–416.

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. 2007. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**: 926-929.

Leahy A, Xiong J, Kuhnert F, Stuhlmann H. 1999. Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. *Journal of Experimental Zoology* **284**: 67-81.

Li L, Barth NKH, Hirth E, Taher L. 2018. Pairs of Adjacent Conserved Noncoding Elements Separated by Conserved Genomic Distances Act as Cis-Regulatory Units. *Genome Biol Evol* **10**: 2535-2550.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**: 476-482.

Lu X, Huang W. 2014. PiggyBac mediated multiplex gene transfer in mouse embryonic stem cell. *PLoS One* **9**: e115072.

McCole R, Fonseka C, Koren A, Wu C. 2014. Abnormal Dosage of Ultraconserved Elements Is Highly Disfavored in Healthy Cells but Not Cancer Cells. *PLoS Genet* **10**.

McCole RB, Erceg J, Saylor W, Wu CT. 2018. Ultraconserved Elements Occupy Specific Arenas of Three-Dimensional Mammalian Genome Organization. *Cell Rep* **24**: 479-488.

Merino MM, Levayer R, Moreno E. 2016. Survival of the Fittest: Essential Roles of Cell Competition in Development, Aging, and Cancer. *Trends Cell Biol* **26**: 776-788.

Monteiro CB, Costa MF, Reguenga C, Lima D, Castro DS, Monteiro FA. 2014. Paired related homeobox protein-like 1 (Prrxl1) controls its own expression by a transcriptional autorepression mechanism. *FEBS Lett* **588**: 3475-3482.

Morata G, Ripoll P. 1974. Minutes: mutants of *Drosophila* autonomously affecting cell division rate. *Developmental biology* **42**: 211-221.

Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M, Jr. 2007. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev* **21**: 708-718.

Nobrega M, Ovcharenko I, Afzal V, Rubin E. 2003. Scanning Human Gene Deserts for Long-Range Enhancers. *Science* **302**: 413–413.

Nolte MJ, Wang Y, Deng JM, Swinton PG, Wei C, Guindani M, Schwartz RJ, Behringer RR. 2014. Functional analysis of limb transcriptional enhancers in the mouse. *Evol Dev* **16**: 207-223.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**: 499-502.

Penzo-Mendez A, Chen YJ, Li J, Witze E, Stanger B. 2015. Spontaneous Cell Competition in Immortalized Mammalian Cell Lines. *PLoS One* **10**.

Poh YC, Chen J, Hong Y, Yi H, Zhang S, Chen J, Wu DC, Wang L, Jia Q, Singh R et al. 2014. Generation of organized germ layers from a single mouse embryonic stem cell. *Nat Commun* **5**: 4000.

Rebelo S, Reguenga C, Osorio L, Pereira C, Lopes C, Lima D. 2007. DRG11 immunohistochemical expression during embryonic development in the mouse. *Dev Dyn* **236**: 2653-2660.

Saito K, Fukuda N, Hayashi N. 2013. Moderate Low Temperature Preserves the Stemness of Neural Stem Cells (Methods). doi:10.1007/978-94-007-6262-6_12: 137-145.

Sakuraba Y, Kimura T, Masuya H, Noguchi H, Sezutsu H, Takahasi K, Toyoda A, Fukumura R, Murata T, Sakaki Y et al. 2008. Identification and characterization of new long conserved noncoding sequences in vertebrates. *Mamm Genome* **19**: 703–712.

Sancho M, Di-Gregorio A, George N, Pozzi S, Sanchez JM, Pernaute B, Rodriguez TA. 2013. Competitive interactions eliminate unfit embryonic stem cells at the onset of differentiation. *Dev Cell* **26**: 19-30.

Saridey SK, Liu L, Doherty JE, Kaja A, Galvan DL, Fletcher BS, Wilson MH. 2009. PiggyBac Transposon-based Inducible Gene Expression In Vivo After Somatic Cell Gene Transfer. *Molecular Therapy* **17**: 2115-2120.

Simons C, Pheasant M, Makunin IV, Mattick JS. 2006. Transposon-free regions in mammalian genomes. *Genome Res* **16**: 164-172.

Sun H, Skogerbo G, Chen R. 2006. Conserved distances between vertebrate highly conserved elements. *Human Molecular Genetics* **15**: 2911-2922.

Tamori Y, Deng WM. 2011. Cell competition and its implications for development and cancer. *J Genet Genomics* **38**: 483-495.

The EPC Dunham I Kundaje A Aldred SF Collins PJ Davis CA Doyle F Epstein CB Fietze S Harrow J et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57.

Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet* **40**: 158-160.

Wang J, Wei Y, Zhao S, Zhou Y, He W, Zhang Y, Deng W. 2017. The analysis of viability for mammalian cells treated at different temperatures and its application in cell shipment. *PLoS One* **12**: e0176120.

Wang W, Lin C, Lu D, Ning Z, Cox T, Melvin D, Wang X, Bradley A, Liu P. 2008. Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **105**: 9290-9295.

Wang X, Yang P. 2008. In vitro differentiation of mouse embryonic stem (mES) cells using the hanging drop method. *J Vis Exp* doi:10.3791/825.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**: 307-319.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K et al. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* **3**: e7.