Development, Growth & Differentiation
The Japanese Society of Developmental Biologists
Develop. Growth Differ. (2018) **60**, 53–62 doi: 10.1111/dgd.12423

Review Article

Identifying co-opted transposable elements using comparative epigenomics

David Venuto[1] and Guillaume Bourque[1,2,3]*

[1]Department of Human Genetics, McGill University, Montréal, H3A 1B1, [2]Canadian Center for Computational Genomics, and [3]McGill University and Génome Québec Innovation Center, Montréal, H3A 0G1, Québec, Canada

The human genome gives rise to different epigenomic landscapes that define each cell type and can be deregulated in disease. Recent efforts by ENCODE, the NIH Roadmap and the International Human Epigenome Consortium (IHEC) have made significant advances towards assembling reference epigenomic maps of various tissues. Notably, these projects have found that approximately 80% of human DNA was biochemically active in at least one epigenomic assay while only approximately 10% of the sequence displayed signs of purifying selection. Given that transposable elements (TEs) make up at least 50% of the human genome and can be actively transcribed or act as regulatory elements either for their own purposes or be co-opted for the benefit of their host; we are interested in exploring their overall contribution to the "functional" genome. Traditional methods used to identify functional DNA have relied on comparative genomics, conservation analysis and low throughput validation assays. To discover co-opted TEs, and distinguish them from noisy genomic elements, we argue that comparative epigenomic methods will also be important.

Key words: comparative epigenomics, epigenetics, functional genome, genomics, transposable elements.

Most of the genome is biochemically active in at least one cell type

The human genome has been available for over 15 years (Lander *et al.* 2001) but most of its sequence remains vastly uncharacterized in terms of the functions it encodes. An estimated 1–2% of human DNA is protein coding but non-coding RNAs, alternative spliced sites and regulatory elements have also been shown to have important functions (Kung *et al.*, 2013). Recent efforts have been made to try to understand the role of non-coding DNA by examining the epigenomic states of the genome in various cell types. This includes a number of large-scale projects whose goal is the systematic mapping of regions of transcription, transcription factor binding sites, histone modifications and chromatin structure.

The first of these large-scale mapping efforts was the ENCODE project, which initially generated a catalogue of 1640 next generation sequencing data sets in 147 cell lines using various epigenomic mapping technologies (ENCODE Project Consortium, 2012). These include transcriptional assays using RNA-sequencing (RNA-seq), transcription factor binding and histone mark assays using Chromatin-Immunoprecipitation (ChIP)-seq, Bisulfite-seq to characterize DNA methylation, and DNAase-seq to identify regions of open chromatin. A major observation of ENCODE was that 80.4% of the genome was participating in at least one of the biochemical RNA or chromatin-associated event described above in at least one cell type. Actually, up to 99% of the genome was found to be within 1.7 kb of such biochemical events. Moreover, approximately half of the approximately 25 000 protein coding genes were measured to be expressed in a given cell type (ENCODE Project Consortium, 2012). Each cell type was also found to be regulated by a subset of perhaps 20 000–40 000 enhancers.

Similarly, the NIH Roadmap project sought to systematically characterize the epigenomic state of human primary cells (Kundaje *et al.* 2015). Part of the consortium mapped genome wide chromatin interactions in four human embryonic stem (ES)-cell derived lineages and found specific chromatin structure reorganization during ES-cell differentiation. These results highlighted that enhancer elements play an important role in shaping high order chromatin structure re-organization during differentiation (Dixon *et al.* 2015). In another project of the consortium, examining epigenetic states in different types of cancer, it was shown that chromatin

*Author to whom all correspondence should be addressed.
Email: guil.bourque@mcgill.ca
Received 15 November 2017; accepted 8 December 2017.
© 2018 Japanese Society of Developmental Biologists

accessibility and modification explain 86% of the variance in mutation rates among tumor genomes (Polak *et al.* 2015). That particular study suggested the local somatic mutation density to be directly linked to the epigenetic landscape. Additionally, in an analysis of 21 autoimmune diseases, approximately 90% of causal genetic variants were found to be in non-coding regions and approximately 60% mapped to immune cell specific enhancer regions as defined by epigenomic data, most of which gain histone acetylation marks and transcribe enhancer associated RNAs upon activation (Farh *et al.* 2015).

Finally, the International Human Epigenome Consortium (IHEC), has an objective to continue the efforts started with ENCODE and the NIH Roadmap and to produce human reference epigenomic maps in at least 1000 tissues (Bujold *et al.* 2016; Stunnenberg *et al.* 2016). In early studies from this consortium focusing on immune and neural cells, the majority of the genome was again found to be associated with at least one biochemical event (Schmitt *et al.* 2016). Projects such as these aim at building human reference epigenomic maps that can be used for comparison against diseased epigenomic landscapes.

How much of the genome is functional?

With over 80% of the human genome participating in detectable epigenomic events but <2% annotated as genes, it leads to the question of what fraction represents a biologically meaningful sequence? (see Fig. 1). Actually, it has been suggested that much of the genome is non-functional (Doolittle & Sapienza 1980; Orgel & Crick 1980). Here, it is important to acknowledge that there is no agreement on what biologically meaningful, or functional, actually means. It has also been observed that species complexity does not correlate with genome size (even within the same genus) and that large organisms are not efficient at natural selection and have a large amount of neutral evolving DNA (Eddy 2012). This phenomenon is known as the *C*-value paradox, and many critiques opposing the synonymy of biochemical activity and functional activity have used it as evidence to counter ENCODE's loose definition of functionality (Eddy 2012; Doolittle 2013). It has also been argued that these functional regions as defined by ENCODE should show more sensitivity to deleterious mutations and that it is impossible to maintain biological function without evidence of selection (Graur *et al.* 2013). On the other side of the argument, it was suggested that our understanding of evolution dynamics is incomplete and that most transcription is non-random and is required for advanced organism development. Differential regulation of genes and their active splicing in cross-cell type comparisons were suggested as evidence of function. One can also argue that the *C*-value paradox is biased because polyploidy accounts for larger than expected genome sizes (Mattick & Dinger 2013). Taking all of this into account, the question of what fraction of the expressed non-coding elements or the ones acting as enhancer, promoters or transcription factor binding sites based on binding of specific histone marks or TFs, are functionally important, remains open for debate.

Stating that an element is functional ultimately relies on evidence such as the phenotypic results of mutation. Unfortunately, knockdown experiments, reporter assays and transfections in cell lines are limited by their modest throughput for validation. Loss of function tests can also be buffered by functional redundancy or can have no discernable phenotypic change (Kellis *et al.* 2014). In contrast, a general approach for the identification of functional regions in the genome uses phylogenetic methods to detect DNA sequence constraint. For protein coding genes, one can use non-synonymous sites, at which a newly arising mutation will alter the protein amino acid (AA) sequence, and synonymous sites, at which a newly arising mutation will change the codon, but not the protein AA sequence. If non-synonymous sites and synonymous sites were functionally equivalent, one would expect that the probability of a substitution at either class of sites would be the same. However, if changes at non-synonymous sites are considerably smaller than changes at synonymous sites on average, it is a sign of evolutionary constraint (Kimura 1991). This constraint is measured as the deficit in divergence at non-synonymous sites relative to expected value given by synonymous sites (Zhen & Andolfatto 2012). In the case of non-coding DNA synonymous sites, intronic DNA, or ancestral repeats are chosen as reference sites. Rates here are more challenging to estimate given the rapid evolution of noncoding DNA vs. protein coding DNA and the problems in selecting good unbiased reference sites (Gaffney & Keightley 2006; Pollard *et al.* 2006).

Using such comparative genomic approaches, a number of studies have estimated the proportion of the human genome that displays signs of purifying selection. For instance, using multi species pairwise alignments, this proportion was calculated to be approximately 10–15% (Ponting & Hardison 2011). More recently, also comparing genome sequences of different species, a study identified genomic regions that have evolved unexpectedly slowly (Rands *et al.* 2014). That study found that between 7.1 and 9.2% of the genome showed evidence of natural selection, a

© 2018 Japanese Society of Developmental Biologists

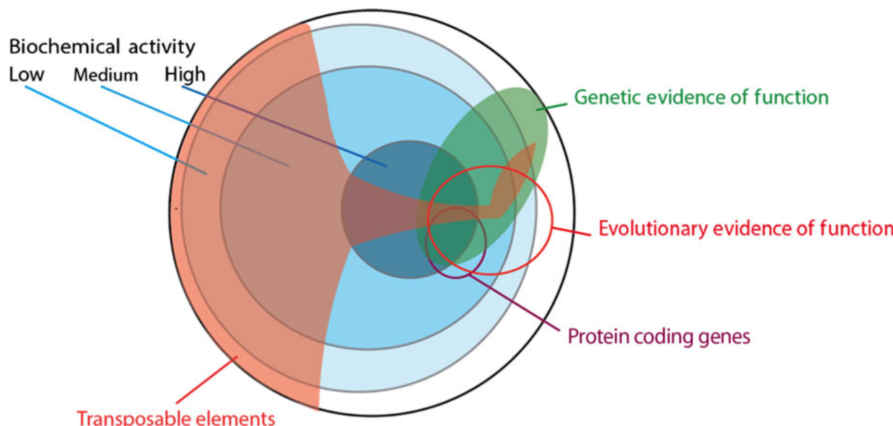

Fig. 1. Three levels of support for DNA being functional: biochemical, evolutionary and genetic evidence are shown. The outer circle represents all sequences in the human genome. The inner blue circles represent biochemical activity levels. The purple circle represents known coding genes. The green circle represents evidence of genetic selection (such as phenotype generation post mutation). The red circle represents mammalian conservation regions. The orange shape presents where transposable elements (TEs) interest each of these regions. Adapted and modified from (Kellis *et al.* 2014).

signature suggesting functional sequences. Unsurprisingly, these studies also revealed that the majority of mammalian-conserved regions consist of non-coding DNA (Waterston *et al.* 2002).

Another important method to identify functional regions in the genome takes a population genetics approach and examines the distribution of polymorphic frequencies of a selected class of sites against putatively neutral sites. This is useful because purifying selection tends to decrease the frequencies of polymorphism at functional sites relative to neutral sites (Sawyer *et al.* 1987). Examining such frequencies has been used to prove negative selection on amino acid variants in a multiple animal species (Boyko *et al.* 2008). McDonald–Kreitman also proposed a statistical test to detect and measure the amount of adaptive evolution within a species by way of determining if adaptive evolution has occurred and the proportion of substitutions resulted from positive selection. In this context, the variation within a species is compared with divergence between species (nucleotide substitutions) at neutral and non-neutral sites. An excess of divergence relative to polymorphism at putatively selected sites is consistent with recurrent adaptive substitution (McDonald & Kreitman 1991). This test quantifies the intensity of selection.

It is important to note that all of these global methods provide a lower bound for the proportion of the genome that is functional since sequence constraint is not always detectable in functional DNA and selection cannot be measured instantaneously (see Fig. 1). Indeed, ENCODE has identified many seemingly functional elements that do not show signs of evolutionary constraint (Birney *et al.* 2007). For instance, the primate-specific genes DSCR9 and DSCR10 were found to be in the Down Syndrome Critical region and exhibit protein-coding expression, which appears to regulate disease state, clearly a functional role in a non-evolutionarily constrained element (Takamatsu *et al.* 2002).

Are TEs in the human genome functional?

Transposable elements are DNA elements that have the ability to move, copy and replicate throughout a genome via copy and paste or cut and paste mechanisms (Kidwell & Lisch 1997). Although very few elements in the human genome are still capable of transposition, they have made copies of themselves throughout the evolution of our species and now represent more than 50% of our DNA (Smit *et al.* 2013-2015). Following their discovery, TEs were presented as potential gene regulators (McClintock 1950; Britten & Davidson 1971). Later on, it was also suggested that these sequences could be simply parasitic, move due to selfish activities, and harbor no intrinsic function for their host (Comings 1973). Using the recent epigenomic maps, an interesting question is what fraction of TE-derived sequences is biochemically active (see Fig. 1). Moreover, what proportion has been co-opted to new functions in the human genome? Can they help explain the paradox between the approximately 80% biochemically active DNA versus the approximately 5–10% sequences that show evidence of purifying selection?

© 2018 Japanese Society of Developmental Biologists

TE-derived sequences have intrinsic cis and trans regulatory activity, which are sometimes co-opted. There are many examples in the human genome of TE-derived sequences that are functional to some level (Mills *et al.* 2007; Bourque 2009; Schmidt *et al.* 2012). Indeed, TEs have been found to have important regulatory roles, to overlap transcription factor binding motifs, to modulate expression of nearby genes via their promoters, and to give rise to non-coding of the sequences that are unusually constrained (Feschotte 2008). In a study using 75 DNAase-seq datasets obtained from various human cell types, it was reported that 44.1% of measured DNAase hypersensitivity sites overlapped TEs (Jacques *et al.* 2013). The TEs located in open chromatin regions were also found to be enriched in promoter or enhancer histone marks. In a different study, it was found that TEs contributed between 2 and 40% of binding sites depending on the transcription factor studied (Sundaram *et al.* 2014). Specific families and classes were enriched in different binding motifs indicating that some families have co-opted specific functions. The same study also showed that TEs have been important in the evolution between human and mouse, with only 1–2% of TE-derived binding peaks shared between human and mouse.

Notably, TEs also have the ability to rewire gene regulatory networks and to inflict change on a larger scale. For instance, the HERV-H repeat was found to be expressed as a long non-coding RNA that is essential in developing human embryonic stem cell identity (Lu *et al.* 2014; Wang *et al.* 2014). The long-terminal repeat (LTR) portion of HERV-H was also found to be enriched in OCT4 binding regions (an essential pluripotency transcription factor). More recently, it was shown that progesterone, a hormone for differentiation of endometrial stromal cells to form decidua, or the maternal component of the placenta, relies on activation of transcripts linked by MER20 elements (Lynch *et al.* 2011). These MER20 elements have been co-opted as binding sites for progesterone response molecules such as the components of the cAMP-signaling pathway. Another study investigated gene regulatory network evolution in various immune cell lines in addition to primary macrophages by stimulating via IFN-g and measuring STAT1 ChIP-seq peaks (Chuong *et al.* 2016). This study showed that MER41B was enriched in STAT1 binding post-stimulation in all cell types studied. Finally, another study showed that MER130 was highly enriched in active enhancer regions of the mouse cerebral wall. It was then experimentally validated that multiple MER130 instances were co-opted as enhancer elements in the mouse neocortex (Notwell *et al.* 2015).

Actually, the epigenomic state of TEs is often repressed, either via DNA methylation or histone modifications, to prevent potentially harmful de novo transposition events. In this context, the host genome has been shown to recruit silencing mechanisms to TEs, such as via the Piwi pathway in the germ line, and silence TE-derived regions in the genome. For instance, the X chromosome in many mammals, including humans, is rich in LINE elements. LINE elements are particularly enriched in the X-chromosome inactivation (XCI) center and are thought to be boosters in spreading effective silencing (Lyon 2006). LINEs are hypothesized to have an active role rather than a bystander role in XCI, as they are depleted in regions escaping X inactivation (Ross *et al.* 2005). A recent study has also shown that short tandem repeats with ancient homology to retro-transposons regulate XCI (Cohen *et al.* 2007). This is an example where the repression of TEs ends up having a more global impact on the host genome.

Finally, some TEs are themselves under purifying selection, such as the MER121 subfamily, which is also over-represented near developmental genes (Lowe *et al.* 2007). Other studies also pointed out that genetic methods examining TE sequence divergence rates can be useful in examining purifying selection in TE-derived sequences (Kamal *et al.* 2006). All of this suggests that TEs can have an impact on the regulatory landscape and that they have, in some cases, been co-opted by the host genome. At this point, only a small fraction of TEs have a demonstrated function but given the exciting examples that have been reported, it will be important to determine the proportion of TEs that has acquired such roles (see Fig. 1).

Comparative epigenomics to identify functional DNA

As mentioned above, genome-wide approaches used to identify functional non-coding DNA typically rely on evidence of sequence constraint (Meadows & Lindblad-Toh 2017). Such approaches study sequence composition, organization and function, both within and across species borders. Comparative epigenomic approaches use the same concepts but instead of examining sequence constraint between species, they examine constraint of the biochemical activity (see Fig. 2). If the epigenomic landscape is not conserved, regions could change state from being repressed to being a promoter or an enhancer, or vice-versa. Epigenomic marks are therefore useful in annotating loss and gain of biochemical activity, and variation from the neutral rates can be used to isolate putatively functional DNA (Chuong *et al.* 2017). If this activity is

© 2018 Japanese Society of Developmental Biologists

conserved or highly divergent between species, a list of candidate regions exhibiting these features could be prioritized for further targeted experiments.

Recently, a few studies have sought to study the evolution of the human regulatory landscape through comparative epigenomic studies. For instance, using human, mouse, and pig pluripotent stem cells, one group looked at the genomic distributions of DNA methylation, H2A.Z, H3K4me1/2/3, H3K9me3, H3K27me3, H3K27ac, H3K36me3, transcribed RNAs, and P300, TAF1, OCT4, and NANOG binding (Xiao *et al.* 2012). That study reported that epigenomic conservation is strong in both rapidly evolving and slowly evolving DNA sequences, but not in neutrally evolving sequences. Evolutionary changes of the epigenome and the transcriptome exhibited a linear correlation as predicted by DNA methylation. Also found via ChIP-seq was that NANOG's in vivo binding sites were surrounded by cell-type-specific epigenomic patterns. The evolutionary changes of epigenetic mark intensities were predictive of binding intensity changes of OCT4. In another study, it was found that gene expression differences across primate species were associated with epigenetic differences (Cain *et al.* 2011). That study used gene expression data and H3K4me3 histone modification data in lymphoblastoid cell lines from humans, chimpanzees, and rhesus macaques. H3K4me3 was found to be consistently enriched near annotated transcription start sites, and highly expressed genes were more likely than lowly expressed genes to have the histone modification near their TSS.

A comparative epigenomic approach was also applied to try to understand how tissue-specific epigenetic patterns were conserved between human, rat, and mouse (Zhou *et al.* 2017). Using whole methylome sequencing data in blood, brain and sperm tissues it was found that 11–37% of differentially methylated regions (DMRs) were conserved in the cross-species comparison. All categories of conserved DMRs were found to be enriched in known genetically conserved elements along with a significantly higher PhyloP score than non-conserved DMRs. Conserved DMRs were also found to be enriched in transcriptionally active histone marks and associated with specific conserved transcription factor binding motifs. These studies highlight how epigenetic data can explain evolutionary dynamics between species and that some regions are epigenetically conserved even though they are not genetically conserved (see Fig. 1).

Comparative epigenomics to identify co-opted transposable elements

Many TE families represent relatively new DNA in the human genome. Unfortunately, even if some of the

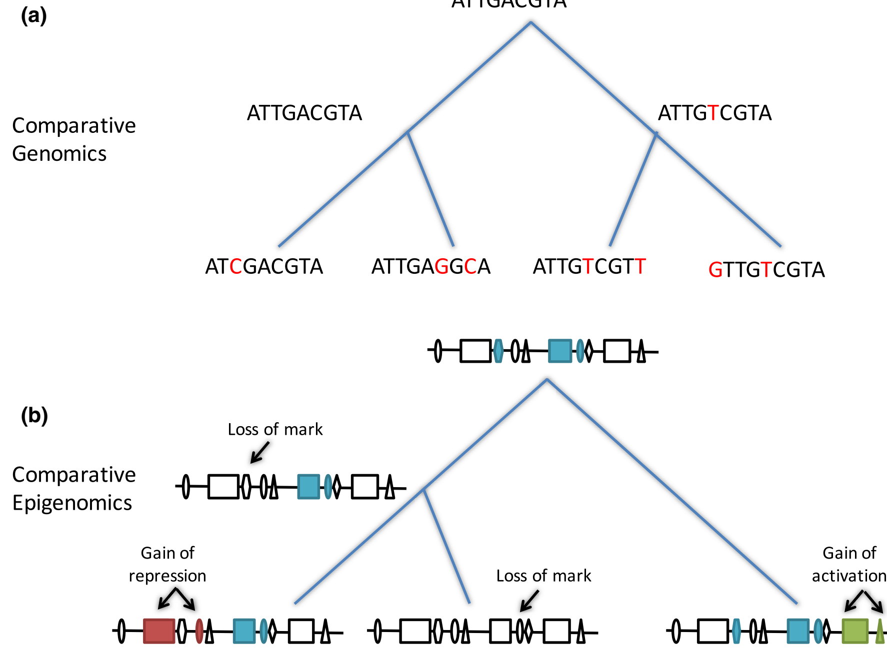

Fig. 2. (a) Comparative genomics ancestral genome reconstruction using phylogenetic tree (maximum parsimony method). (b) Ancestral epigenome reconstruction using parsimony method and counting loss, change in state, and adding of epigenomic marks as mutations.

© 2018 Japanese Society of Developmental Biologists

newer TE insertions have been co-opted, they may not have had enough longevity to develop signs of sequence constraint. In this context, mapping the epigenome and looking for constraint of biochemical activity may provide some additional clues (Rivera & Ren 2013). Indeed, the availability of epigenomic data enables the development approaches to help isolate functional non-coding DNA that lacks evidence of genetic selection (Ludwig 2002). For TEs, it is important to remember that they should rapidly be decaying from an active, transposition-able, state to an inactive state (Hellen & Brookfield 2013). Comparative epigenomics allows the identification of old families that have unusually high epigenomic activity, which could be evidence of regulatory function or co-option. Such methods have already been extremely useful in finding TEs with important regulatory functions and also non-coding TE-derived RNAs (Kunarso *et al.* 2010; Lu *et al.* 2014; Chuong *et al.* 2016).

Notably, in an analysis of binding events for 26 pairs of orthologous transcription factors across two comparable human and mouse cell lines, only approximately 2% of more than 130 000 TE-derived peaks were found to be conserved between the two species (Sundaram *et al.* 2014). This study also identified outliers of old TE families that should have been repressed and decayed in activity but are still located in accessible regions. In Ramsay *et al.* (2017) a TE orthologue annotation pipeline was proposed for cross-species comparisons of transcription. Using RNA-seq data from three primate species in iPSC cell lines, TEs that were conserved in expression as well as in sequence were identified. In that study, many TE families exhibiting an unusually high number of orthologous instances with conserved expression were identified. TE families that were found to be expressed over a long and sustained evolutionary period of time could potentially be used as evidence of functionality in humans.

There are three important points to remember at this stage: (i) epigenomic activity frequently lacks evidence of genetic selection, (ii) TEs often represent relatively young DNA and (iii) many TEs are biochemically active. Those three points together suggest that comparative epigenomics can be a useful tool for functional DNA annotation, in particular for TEs.

Discussion

For decades, we have been in trying to characterize the role, if any, of all the DNA elements of the human genome. Evolutionary and genetic evidence of function examine constraint that is global and will not depend on the cell type being studied. This type of evidence will therefore be ineffective to identify cell type specific regulatory functions especially if they are relatively novel. Epigenomic data, on the other hand, provides evidence that can be used to identify the function of DNA elements in the specific cell types that they are active in. Such epigenomic maps, when available, enable the study of regulatory or expressed elements in different scenarios such as differentiation, development and disease.

TEs show a remarkably different evolutionary profile relative to the rest of the genome and so typical methods relying on sequence constraint to identity functional elements frequently will not apply. Indeed, genomic evidence of constraint does not appear instantaneously, and many insertions will not have been present long enough to acquire signatures of natural selection. For TEs, an important clue that may indicate regulatory function or co-option, is sustained biochemical activity in regions that should have otherwise been epigenetically silenced to prevent transposition. This supports the case of using cross-species biochemical evidence to identify potentially functional elements.

The use of comparative approaches to identify functional DNA elements in the genome has been gaining in popularity and we propose to apply and adapt these approaches specifically for TEs. We believe that these methods can be used to systematically examine the impact of TEs, especially new elements, on gene regulation and transcript innovation. The large faction of the genome that is derived from TEs is too vast to explore using only low-throughput functional assays. In this context, genome-wide comparative epigenomic maps can provide a tool for gaining evidence of function and prioritizing regions for further study.

Acknowledgements

This work was supported by a grant by a grant from the Canadian Institute for Health Research (CIHR-MOP-115090). GB is supported by the Fonds de Recherche Québec Santé (FRQS-25348) and by the Top Global University program at Kyoto University. We would also like to thank Patricia Goerner-Potvin for useful comments on the manuscript.

References

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N.,

© 2018 Japanese Society of Developmental Biologists

Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W. K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C. L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaöz, U., Siepel, A., Taylor, J., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya- Burgos, J. I., Löytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., Comparative Sequencing, NISC Program, College, Baylor of Genome, Medicine Human Center, Sequencing, Genome, Washington University Center, Sequencing, Institute, Broad, Oakland, Children's Hospital Institute, Research, Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N., Yu, Y., Ruan, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Armengol, L., Bird, C. P., De Bakker, P. I., Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrímsdóttir, I. B., Huppert, J., Zody, M. C., Abecasis, G. R., Estivill, X., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V., Maskeri, B., McDowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B., & De Jong, P. J. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816.

Bourque, G. 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr. Opin. Genet. Dev.* **19**, 607–612.

Boyko, A. R., Williamson, S. H., Indap, A. R., Degenhardt, J. D., Hernandez, R. D., Lohmueller, K. E., Adams, M. D., Schmidt, S., Sninsky, J. J., Sunyaev, S. R., White, T. J., Nielsen, R., Clark, A. G., & Bustamante, C. D. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* **4**, e1000083.

Britten, R. J., & Davidson, E. H. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **46**, 111–138.

Bujold, D., Morais, D. A., Gauthier, C., Cote, C., Caron, M., Kwan, T., Chen, K. C., Laperle, J., Markovits, A. N., Pastinen, T., Caron, B., Veilleux, A., Jacques, P. E., & Bourque, G. 2016. The international human epigenome consortium data portal. *Cell Syst.* **3**, 496–499. e2.

Cain, C. E., Blekhman, R., Marioni, J. C., & Gilad, Y. 2011. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* **187**, 1225–1234.

Chuong, E. B., Elde, N. C., & Feschotte, C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083–1087.

Chuong, E. B., Elde, N. C., & Feschotte, C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86.

Cohen, D. E., Davidow, L. S., Erwin, J. A., Xu, N., Warshawsky, D., & Lee, J. T. 2007. The DXPas34 repeat regulates random and imprinted X inactivation. *Dev. Cell* **12**, 57–71.

Comings, D. E. 1973. Evolution of genetic systems. *Am. J. Human Genet.* **25**, 340–342.

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V. V., Ecker, J. R., Thomson, J., & Ren, B. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336.

Doolittle, W. F. 2013. Is junk DNA bunk? A critique of ENCODE. *Proc. Natl Acad. Sci. USA* **110**, 5294–5300.

Doolittle, W. F., & Sapienza, C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.

Eddy, S. R. 2012. The C-value paradox, junk DNA and ENCODE. *Curr. Biol.* **22**, R898–R899.

ENCODE Project Consortium 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

Farh, K. K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., Shoresh, N., Whitton, H., Ryan, R. J., Shishkin, A. A., Hatan, M., Carrasco-Alfonso, M. J., Mayer, D., Luckey, C. J., Patsopoulos, N. A., De Jager, P. L., Kuchroo, V. K., Epstein, C. B., Daly, M. J., Hafler, D. A., & Bernstein, B. E. 2015. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343.

Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405.

Gaffney, D. J., & Keightley, P. D. 2006. Genomic selective constraints in murid noncoding DNA. *PLoS Genet.* **2**, e204.

© 2018 Japanese Society of Developmental Biologists

Graur, D., Zheng, Y., Price, N., Azevedo, R. B., Zufall, R. A., & Elhaik, E. 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.* **5**, 578–590.

Hellen, E. H. B., & Brookfield, J. F. Y. 2013. Transposable element invasions. *Mob. Genet. Elements*, **3**, e23920.

Jacques, P. E., Jeyakani, J., & Bourque, G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9**, e1003504.

Kamal, M., Xie, X., & Lander, E. S. 2006. A large family of ancient repeat elements in the human genome is under strong selection. *Proc. Natl Acad. Sci. USA* **103**, 2740–2745.

Kellis, M., Wold, B., Snyder, M. P., Bernstein, B. E., Kundaje, A., Marinov, G. K., Ward, L. D., Birney, E., Crawford, G. E., Dekker, J., Dunham, I., Elnitski, L. L., Farnham, P. J., Feingold, E. A., Gerstein, M., Giddings, M. C., Gilbert, D. M., Gingeras, T. R., Green, E. D., Guigo, R., Hubbard, T., Kent, J., Lieb, J. D., Myers, R. M., Pazin, M. J., Ren, B., Stamatoyannopoulos, J. A., Weng, Z., White, K. P., & Hardison, R. C. 2014. Defining functional DNA elements in the human genome. *Proc. Natl Acad. Sci. USA* **111**, 6131–6138.

Kidwell, M. G., & Lisch, D. 1997. Transposable elements as sources of variation in animals and plants. *Proc. Natl Acad. Sci. USA* **94**, 7704–7711.

Kimura, M. 1991. The neutral theory of molecular evolution: a review of recent evidence. *Jpn. J. Genet.*, **66**, 367–386.

Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., & Bourque, G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y. C., Pfenning, A. R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R. A., Shoresh, N., Epstein, C. B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R. D., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore, R., Chuah, E., Tam, A., Canfield, T. K., Hansen, R. S., Kaul, R., Sabo, P. J., Bansal, M. S., Carles, A., Dixon, J. R., Farh, K. H., Feizi, S., Karlic, R., Kim, A. R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R. C., Siebenthall, K. T., Sinnott-Armstrong, N. A., Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Beaudet, A. E., Boyer, L. A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J., Li, W., Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T. D., Tsai, L. H., Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello, J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J. A., Wang, T., & Kellis, M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330.

Kung, J. T. Y., Colognori, D., & Lee, J. T. 2013. Long noncoding RNAs: past, present, and future. *Genetics*, **193**, 651–669.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., Mcewan, P., Mckernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., Mcmurray, A., Matthews, L., Mercer, S., Milne, S.., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., Mcpherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de La Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., De Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., Szustakowki, J. & International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Lowe, C. B., Bejerano, G., & Haussler, D. 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc. Natl Acad. Sci. USA* **104**, 8005–8010.

Lu, X., Sachs, F., Ramsay, L., Jacques, P.-É., Göke, J., Bourque, G., & Ng, H.-H. 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **21**, 423–425.

Ludwig, M. Z. 2002. Functional evolution of noncoding DNA. *Curr. Opin. Genet. Dev.*, **12**, 634–639.

Lynch, V. J., Leclerc, R. D., May, G., & Wagner, G. P. 2011. Transposon-mediated rewiring of gene regulatory networks

© 2018 Japanese Society of Developmental Biologists

contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **43**, 1154–1159.

Lyon, M. F. 2006. Do LINEs have a role in X-chromosome inactivation? *J. Biomed. Biotechnol.* **2006**, 59746.

Mattick, J. S., & Dinger, M. E. 2013. The extent of functionality in the human genome. *HUGO J.* **7**, 2.

McClintock, B. 1950. The origin and behavior of mutable loci in maize. *Proc. Natl Acad. Sci. USA* **36**, 344–355.

McDonald, J. H., & Kreitman, M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 652–654.

Meadows, J. R. S., & Lindblad-Toh, K. 2017. Dissecting evolution and disease using comparative vertebrate genomics. *Nat. Rev. Genet.* **18**, 624–636.

Mills, R. E., Bennett, E. A., Iskow, R. C., & Devine, S. E. 2007. Which transposable elements are active in the human genome? *Trend Genet.* **23**, 183–191.

Notwell, J. H., Chung, T., Heavner, W., & Bejerano, G. 2015. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat. Commun.* **6**, 6644.

Orgel, L. E., & Crick, F. H. 1980. Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.

Polak, P., Karlic, R., Koren, A., Thurman, R., Sandstrom, R., Lawrence, M., Reynolds, A., Rynes, E., Vlahovicek, K., Stamatoyannopoulos, J. A., & Sunyaev, S. R. 2015. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**, 360–364.

Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J. S., Bejerano, G., Baertsch, R., Rosenbloom, K. R., Kent, J., & Haussler, D. 2006. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, e168.

Ponting, C. P., & Hardison, R. C. 2011. What fraction of the human genome is functional? *Genome Res.* **21**, 1769–1776.

Ramsay, L., Marchetto, M. C., Caron, M., Chen, S. H., Busche, S., Kwan, T., Pastinen, T., Gage, F. H., & Bourque, G. 2017. Conserved expression of transposon-derived noncoding transcripts in primate stem cells. *BMC Genom.* **18**, 214.

Rands, C. M., Meader, S., Ponting, C. P., & Lunter, G. 2014. 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet.* **10**, e1004525.

Rivera, C. M., & Ren, B. 2013. Mapping human epigenomes. *Cell* **155**, 39–55.

Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G. R., Burrows, C., Bird, C. P., Frankish, A., Lovell, F. L., Howe, K. L., Ashurst, J. L., Fulton, R. S., Sudbrak, R., Wen, G., Jones, M. C., Hurles, M. E., Andrews, T. D., Scott, C. E., Searle, S., Ramser, J., Whittaker, A., Deadman, R., Carter, N. P., Hunt, S. E., Chen, R., Cree, A., Gunaratne, P., Havlak, P., Hodgson, A., Metzker, M. L., Richards, S., Scott, G., Steffen, D., Sodergren, E., Wheeler, D. A., Worley, K. C., Ainscough, R., Ambrose, K. D., Ansari-Lari, M. A., Aradhya, S., Ashwell, R. I., Babbage, A. K., Bagguley, C. L., Ballabio, A., Banerjee, R., Barker, G. E., Barlow, K. F., Barrett, I. P., Bates, K. N., Beare, D. M., Beasley, H., Beasley, O., Beck, A., Bethel, G., Blechschmidt, K., Brady, N., Bray-Allen, S., Bridgeman, A. M., Brown, A. J., Brown, M. J., Bonnin, D., Bruford, E. A., Buhay, C., Burch, P., Burford, D., Burgess, J., Burrill, W., Burton, J., Bye, J. M., Carder, C., Carrel, L., Chako, J., Chapman, J. C., Chavez, D., Chen, E., Chen, G., Chen, Y., Chen, Z., Chinault, C., Ciccodicola, A., Clark, S. Y., Clarke, G., Clee, C. M., Clegg, S., Clerc-Blankenburg, K., Clifford, K., Cobley, V., Cole, C. G., Conquer, J. S., Corby, N., Connor, R. E., David, R., Davies, J., Davis, C., Davis, J., Delgado, O., Deshazo, D., Dhami, P., Ding, Y., Dinh, H., Dodsworth, S., Draper, H., Dugan-Rocha, S., Dunham, A., Dunn, M., Durbin, K. J., Dutta, I., Eades, T., Ellwood, M., Emery-Cohen, A., Errington, H., Evans, K. L., Faulkner, L., Francis, F., Frankland, J., Fraser, A. E., Galgoczy, P., Gilbert, J., Gill, R., Glöckner, G., Gregory, S. G., Gribble, S., Griffiths, C., Grocock, R., Gu, Y., Gwilliam, R., Hamilton, C., Hart, E. A., Hawes, A., Heath, P. D., Heitmann, K., Hennig, S., Hernandez, J., Hinzmann, B., Ho, S., Hoffs, M., Howden, P. J., Huckle, E. J., Hume, J., Hunt, P. J., Hunt, A. R., Isherwood, J., Jacob, L., Johnson, D., Jones, S., de Jong, P. J., Joseph, S. S., Keenan, S., Kelly, S., Kershaw, J. K., Khan, Z., Kioschis, P., Klages, S., Knights, A. J., Kosiura, A., Kovar-Smith, C., Laird, G. K., Langford, C., Lawlor, S., Leversha, M., Lewis, L., Liu, W., Lloyd, C., Lloyd, D. M., Loulseged, H., Loveland, J. E., Lovell, J. D., Lozado, R., Lu, J., Lyne, R., Ma, J., Maheshwari, M., Matthews, L. H., McDowall, J., McLaren, S., McMurray, A., Meidl, P., Meitinger, T., Milne, S., Miner, G., Mistry, S. L., Morgan, M., Morris, S., Müller, I., Mullikin, J. C., Nguyen, N., Nordsiek, G., Nyakatura, G., O'Dell, C. N., Okwuonu, G., Palmer, S., Pandian, R., Parker, D., Parrish, J., Pasternak, S., Patel, D., Pearce, A. V., Pearson, D. M., Pelan, S. E., Perez, L., Porter, K. M., Ramsey, Y., Reichwald, K., Rhodes, S., Ridler, K. A., Schlessinger, D., Schueler, M. G., Sehra, H. K., Shaw-Smith, C., Shen, H., Sheridan, E. M., Shownkeen, R., Skuce, C. D., Smith, M. L., Sotheran, E. C., Steingruber, H. E., Steward, C. A., Storey, R., Swann, R. M., Swarbreck, D., Tabor, P. E., Taudien, S., Taylor, T., Teague, B., Thomas, K., Thorpe, A., Timms, K., Tracey, A., Trevanion, S., Tromans, A. C., D'urso, M., Verduzco, D., Villasana, D., Waldron, L., Wall, M., Wang, Q., Warren, J., Warry, G. L., Wei, X., West, A., Whitehead, S. L., Whiteley, M. N., Wilkinson, J. E., Willey, D. L., Williams, G., Williams, L., Williamson, A., Williamson, H., Wilming, L., Woodmansey, R. L., Wray, P. W., Yen, J., Zhang, J., Zhou, J., Zoghbi, H., Zorilla, S., Buck, D., Reinhardt, R., Poustka, A., Rosenthal, A., Lehrach, H., Meindl, A., Minx, P. J., Hillier, L. W., Willard, H. F., Wilson, R. K., Waterston, R. H., Rice, C. M., Vaudin, M., Coulson, A., Nelson, D. L., Weinstock, G., Sulston, J. E., Durbin, R., Hubbard, T., Gibbs, R. A., Beck, S., Rogers, J., & Bentley, D. R. 2005. The DNA sequence of the human X chromosome. *Nature* **434**, 325–337.

Sawyer, S. A., Dykhuizen, D. E., & Hartl, D. L. 1987. Confidence interval for the number of selectively neutral amino acid polymorphisms. *Proc. Natl Acad. Sci. USA* **84**, 6225–6228.

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Goncalves, A., Kutter, C., Brown, G. D., Marshall, A., Flicek, P., & Odom, D. T. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335–348.

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., Li, Y., Lin, S., Lin, Y., Barr, C. L., & Ren, B. 2016. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059.

Smit, A., Hubley, R., & Green, P.. 2013-2015. *RepeatMasker Open-4.0.* [Online]. Available: http://www.repeatmasker.org.

Stunnenberg, H. G., The International Human Epigenome Consortium, & Hirst, M. 2016. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell*, **167**, 1145–1149.

© 2018 Japanese Society of Developmental Biologists

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M. P., & Wang, T. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976.

Takamatsu, K., Maekawa, K., Togashi, T., Choi, D. K., Suzuki, Y., Taylor, T. D., Toyoda, A., Sugano, S., Fujiyama, A., Hattori, M., Sakaki, Y., & Takeda, T. 2002. Identification of two novel primate-specific genes in DSCR. *DNA Res.* **9**, 89–97.

Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V., Schumann, G. G., Chen, W., Lorincz, M. C., Ivics, Z., Hurst, L. D., & Izsvak, Z. 2014. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S. P., Zdobnov, E. M., Zody, M. C., & Lander, E. S. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562.

Xiao, S., Xie, D., Cao, X., Yu, P., Xing, X., Chen, C. C., Musselman, M., Xie, M., West, F. D., Lewin, H. A., Wang, T., & Zhong, S. 2012. Comparative epigenomic annotation of regulatory DNA. *Cell* **149**, 1381–1392.

Zhen, Y., & Andolfatto, P. 2012. Methods to detect selection on noncoding DNA. *Method Mol. Biol.* **856**, 141–159.

Zhou, J., Sears, R. L., Xing, X., Zhang, B., Li, D., Rockweiler, N. B., Jang, H. S., Choudhary, M. N. K., Lee, H. J., Lowdon, R. F., Arand, J., Tabers, B., Gu, C. C., Cicero, T. J., & Wang, T. 2017. Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. *BMC Genom.* **18**, 724.

© 2018 Japanese Society of Developmental Biologists