doi:10.1093/hmg/ddh229

Exploiting human–fish genome comparisons for deciphering gene regulation

Nadav Ahituv[1,2], Edward M. Rubin[1,2] and Marcelo A. Nobrega[1,2,*]

[1]DOE Joint Genome Institute, Walnut Creek, CA 94598, USA, and [2]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Received June 14, 2004; Revised and Accepted July 19, 2004

Comparative genomics has served as an essential guide in the identification of functional coding and non-coding sequences in vertebrate genomes. Human–mouse pair-wise comparisons have limited utility for identifying functional conserved non-coding sequences, owing to the large number of sequences shared between these species. In searching for more stringent filters to uncover non-coding elements more likely to be of functional importance in the human genome, human–fish sequence comparisons have emerged as an important strategy, leading to the efficient identification of enhancer elements. These sequences are unevenly distributed in the genome, tending to cluster around genes involved in key developmental processes, with recent studies suggesting that they represent genomic segments in which sequence variation can result in morphological changes and innovation. These elements, conserved over long evolutionary time, emerge as primary candidates that are likely to harbor sequence variation contributing to susceptibility of human disease phenotypes.

An expedition to a foreign land is always an adventure. A map and travel guides can give us a flavor of the land, but the only true experience comes from the journey itself. Such is the current annotation status of the non-coding fraction of the human genome. We have the map, the human genome, while comparative genomics serves as a travel guide, allowing us to emphasize points of interest. The true expeditions are in the form of functional analyses, but owing to the complexity, the time-consuming and labor-intensive natures of these endeavors, these journeys must be carefully planned.

Unlike protein coding sequences, whose code was cracked decades ago, the code of gene regulatory elements still remains an unsolved riddle. Compounding this difficulty, these sequences are often scattered over long distances from the genes that they regulate, making the entire 98% of the genome that does not correspond to protein coding sequences fair game in the search for these elements. It is with this mindset that comparative genomics became one of the preferred strategies to identify functionally important sequences in the vast non-coding tracts of the human genome. The cardinal principle of this idea is that functionally important sequences are conserved across long evolutionary periods, whereas the remaining majority of sequences in the genome evolve neutrally and eventually diverge beyond recognition.

Given the early availability of both the human and mouse genomes, sequence comparisons between these genomes have been the most commonly employed approach in comparative genomics. This has proven to be a successful guide in the identification of both protein-coding genes as well as non-coding sequences (1,2). Nevertheless, it has become evident that pair-wise comparisons between mammals alone are not sufficient for the prioritization of which non-coding sequences are the most likely to be functional in the genome. Given the uneven rate of evolution across vertebrate genomes, sequence similarities can be extensive in certain genomic regions, not because those sequences have been conserved under evolutionary pressure, but because in many genomic regions the evolutionary divergence between mammals is not sufficient to discern neutrally evolving sequences from functionally constrained ones.

Several strategies have recently been devised and employed aiming at better filtering conserved sequences to identify those with a higher likelihood of being functional. Prominent among these are the use of multiple species of comparable evolutionary divergence instead of simple pair-wise comparisons (3–6) and the use of evolutionarily distant species for pair-wise comparisons. The rational behind both methods is straightforward: increasing the total phylogenetic tree branch length enables the removal of similarities between neutrally evolving sequences, so that only truly functional sequences with selective constraints will remain conserved among the species studied. In multiple species comparisons this increased

[*]To whom correspondence should be addressed. Tel: +1 5104952301; Fax: +1 5104864229; Email: manobrega@lbl.gov

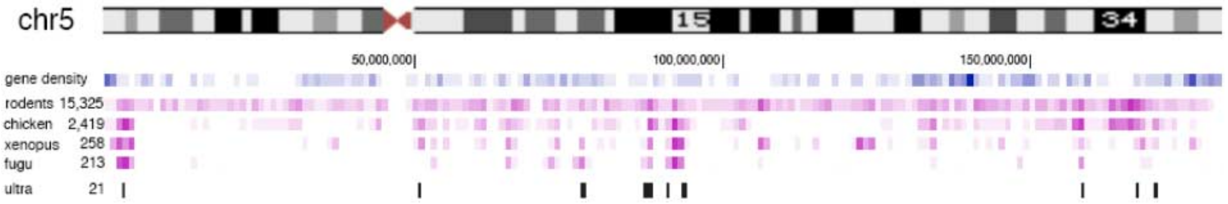

Figure 1. A schematic diagram of human chromosome 5 showing non-coding conservation density. The plot shows the normalized density and number of human/mouse/rat, human/mouse/chicken, human/mouse/frog and human/mouse/fugu non-coding elements, with the corresponding location and number of the ultraconserved non-coding elements on this chromosome (7).

branch length is obtained by adding species of similar evolutionary divergence, whereas evolutionarily distant species attain the increase in branch length by simple pair-wise comparisons of species at the extremes of vertebrate phylogeny, such as mammals and fish.

Both strategies are likely to simplify and optimize the task of sifting through large tracts of DNA to prioritize small regions for further experimental analysis. A wealth of recent functional studies using human–fish comparisons has already validated this approach as a powerful filter for the identification of functional non-coding sequences in the human genome. While several limitations, such as the small number of conserved sequences between these species, preclude the widespread use of this type of sequence comparison in most genomic regions, fascinating lessons have already been learned from the comparisons between humans and fish. Here, we review these findings, and argue for the use of deep evolutionarily conserved non-coding sequences as candidate intervals harboring sequence variations involved in morphological innovation and human disease.

CHARACTERISTICS OF HUMAN–FISH CONSERVED NON-CODING ELEMENTS

The initial observation of the compactness of the fugu genome, 7.5 times smaller than the human one, led to the suggestion that genes conserved between these species would represent the minimal set of genes required to construct a vertebrate organism (8). The sequencing and analyses of the fugu genome further supported this notion and human-fugu comparisons resulted in the immediate discovery of several novel human genes (9). The utility of this genome for also identifying functional non-coding sequences was first hinted before the full sequence of any vertebrate genome became available. In 1994 the mouse Hoxb-1 enhancer regulating expression in the neuroectoderm and mesoderm was identified in fugu and its sequence was shown to appropriately drive the expression of a reporter gene in an *in vivo* mouse transgenic assay (10). Ever since this pioneering study, a host of reports have supported the notion that human–fish comparisons efficiently identify functional non-coding sequences in the human genome (11–15). On the basis of these reports and others, a conventional threshold was informally created for the identification of these human–fish non-coding elements, requiring 70% identity over a minimum size of 100 bp.

An important observation emerging from these studies was that a significant portion of these human–fish conserved non-coding sequences are located in the vicinity of genes involved in early embryonic development, whose products frequently are DNA-binding proteins, suggesting their roles as transcription factors (16). Many of these transcription factors, thought to be regulated by these non-coding human–fish elements, are involved in various morphogenic processes during embryonic development that are, by and large, shared by most vertebrates. The molecular mechanisms underlying these similarities in morphogenesis have been gradually uncovered, and appear to be predicated precisely on the regionalized and/or dynamic expression throughout embryogenesis of these transcription factors, in whose chromosomal neighborhoods many human–fish conserved non-coding sequences reside. Thus, the observation that both the genes encoding the transcription factors and the genetic switches controlling their expression appear to have been conserved throughout hundreds of millions of years of evolution supports the notion that this set of sequences, literally 'fossil DNA' embedded in our genomes, constitutes the 'core genome' elements of vertebrates.

STRINGENT EVOLUTIONARY SEQUENCE CONSERVATION AS A FILTER

An alternative approach to identify conserved non-coding elements with a high probability of being functional is using high stringency of sequence conservation among mammals as a filter (7). Using a screen for sequences in the human genome that are at least 200 bp long and 100% identical in human, mouse and rat, Bejerano *et al.* (7) identified 481 such elements, termed ultraconserved sequences. Among these, 256 show no evidence of transcription and thus were dubbed 'non-exonic'. These non-exonic ultraconserved elements are frequently found in clusters near developmental genes encoding transcription factors. Two-thirds of these ultraconserved elements are conserved between human and fugu. This overlap between the mammalian ultraconserved and human–fish conserved elements suggests that they most likely represent a similar category of non-coding DNA. Although the catalog of ultraconserved elements is considerably smaller than that of conserved human–fugu non-coding elements, as exemplified in chromosome 5 (Fig. 1), one can increase their number by changing the stringency of the filter. For example, relaxing the stringency criteria of sequence identity to 100 bp and 100% identity captures more than 5000 highly conserved sequences, many of which are also conserved between human and fugu.

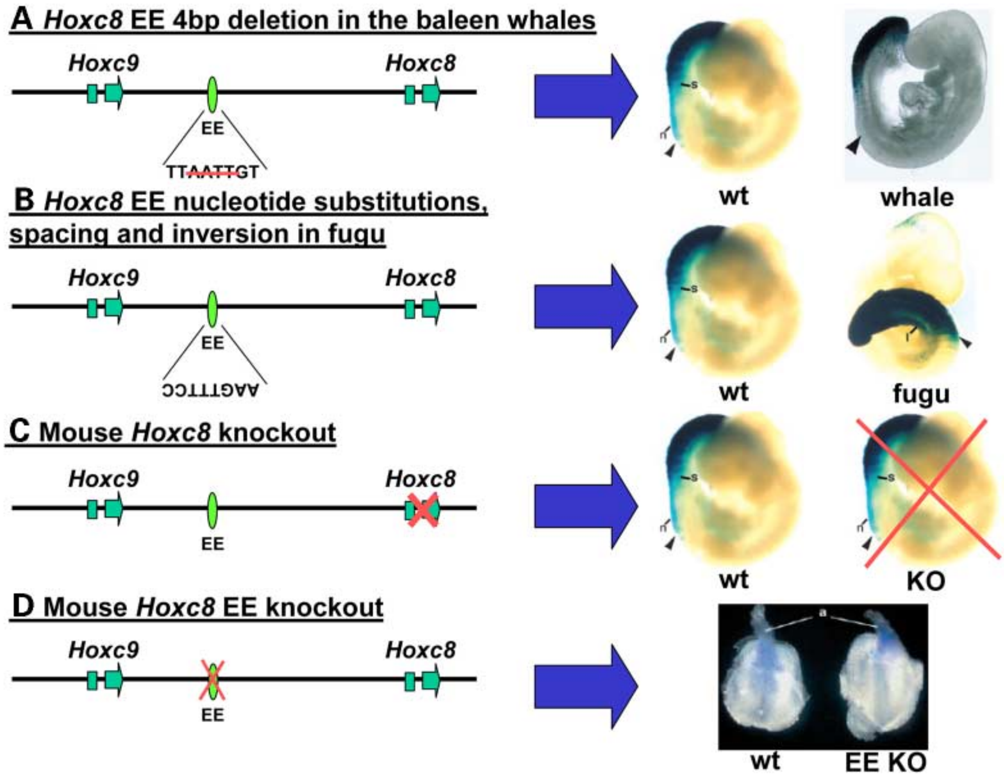
A Hoxc8 EE 4bp deletion in the baleen whales
Hoxc9
Hoxc8
EE
TTAATTGT
wt
whale
n
s
B Hoxc8 EE nucleotide substitutions, spacing and inversion in fugu
Hoxc9
Hoxc8
EE
AAGTTCC
wt
fugu
l
C Mouse Hoxc8 knockout
Hoxc9
Hoxc8
EE
wt
KO
D Mouse Hoxc8 EE knockout
Hoxc9
Hoxc8
EE
a
wt
EE KO

Figure 2. Nucleotide changes in the *Hoxc8* EE and their effects on expression. (**A**) A 4 bp deletion in the *Hoxc8* EE of the fin whale fails to show reporter gene expression at mouse embryonic day 9–9.5 in the posterior mesoderm, and exhibits staining at 4–5 somite levels posterior when compared with the wild-type (wt) expression pattern (adapted from 17 with the kind permission of National Academy of Sciences, USA, Copyright 1998). (**B**) Nucleotide changes and an inversion in the *Hoxc8* EE in fugu leads to expression of the reporter gene to more rostral levels at the 16th somite and more anterior levels in the lateral plate mesoderm when compared with the mouse enhancer in embryonic day 9–9.5 (adapted from 18 with the kind permission of National Academy of Sciences, USA, Copyright 2003). (**C**) Deletion of the *Hoxc8* gene leads to abolition of its expression. (**D**) Mouse whole-mount *in situ* hybridization at embryonic day 8 showing the lack of expression of *Hoxc8* in the anterior regions of the allantois in *Hoxc8* EE knockouts when compared with the wt (adapted from 39 with the kind permission of The Company of Biologists). Arrowhead indicates the position of the 14th somite. n, neural tube; s, somites; l, lateral plate mesoderm; a, allantois.

An interesting finding revealed in this study (7) is that the ultraconserved sequences have a 20-fold decrease in the frequency of single nucleotide polymorphisms (SNPs). While one cannot exclude the possibility that an unknown biological phenomenon might account for the hypomutability of these elements, the most probable explanation for this low SNP frequency is that it reflects an unusually strong purifying selection, where minor sequence variations within these elements result in selective disadvantage. This finding supports the hypothesis that these sequences represent a particular subset of regulatory elements, with unique architectural constraints. Alternatively, this extreme conservation could reflect the presence of multiple functional sequences embedded within each ultraconserved element.

Table 1. Position effects causing human disease

Gene	Disease	Reference
FOXC2	Lymphedema-distichiasis	19
FOXL2	Blepharophimosis/ptosis/epicanthus inversus syndrome	20
FSHD	Facioscapulohumeral dystrophy	21
GLI3	Greig cephalopolysyndactyly	22
HBB complex	Γβ-Thalassemia	23,24
PAX6	Aniridia	25
PITX2	Rieger syndrome	26
PLP1	Pelizaeus–Merzbacher disease	27
POU3F4	X-linked deafness	28,29
SALL1	Townes–Brocks syndrome	30
SHH	Holoprosencephaly	31
SIX3	Holoprosencephaly	32
SOST	Van Buchem disease	33
SOX9	Campomelic displasia	34
SRY	Sex reversal	35,36
TWIST	Saethre–Chotzen syndrome	37,38

CONSERVED CIS-REGULATORY ELEMENTS: A SUBSTRATE FOR MORPHOLOGICAL INNOVATION

The extreme degree of conservation of these non-coding sequences over hundreds of millions of years of evolution implies that tampering with them may result in significant phenotypic consequences. A well-documented example of morphological variation, suspected to have arisen by modifications of non-coding sequences shared by mammals and non-mammalian vertebrates is *Hoxc8*, a gene whose boundaries of expression in the developing embryo are thought to

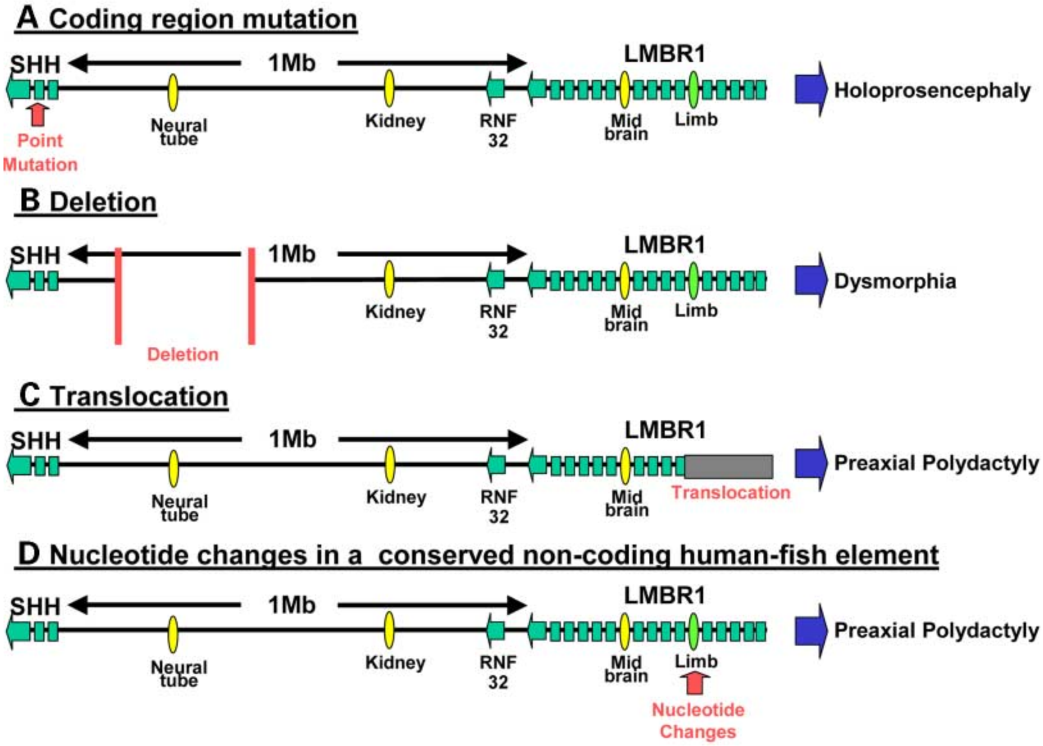

Figure 3. A schematic representation of the SHH region depicting the outcome of (**A**) point mutations in the SHH gene, (**B**) large scale deletion in the downstream region of SHH, (**C**) translocation (as described in 41) and (**D**) nucleotide changes (as described in 13). Exons are shown as light blue boxes, known functional regulatory elements are shown as green ovals and fabricated ones as yellow ovals.

influence vertebrae number (17,18). The *Hoxc8* early enhancer (EE) regulates expression of *Hoxc8* in the posterior regions of the neural tube and mesoderm (18). A detailed sequence analysis of this enhancer in 29 mammals found a 4 bp deletion segregating in the baleen whale lineage, a deletion that resulted in the failure of a reporter gene to be expressed in the posterior mesoderm in transgenic mice (Fig. 2A) (17). This deletion was suggested to underlie the variation in the number of thoracic vertebrae and alterations of axial structures and appendages in these whales as part of their adaptation to aquatic life (17). Other studies report that sequence variations in this enhancer observed in fugu and zebrafish, including nucleotide substitutions, spacing alterations and inversions also lead to different expression patterns in the neural tube and somites in transgenic mice (Fig. 2B) (18).

The *Hoxc8* EE was recently deleted in mice, also resulting in an axial skeletal phenotype (39), though different and milder than that resulting from the deletion of the *Hoxc8* gene (Fig. 2C and D). This example highlights a key concept about the impact of sequence variation within cis-regulatory elements regulating developmentally important genes. While several mouse models that knockout transcription factors involved in morphogenesis exhibit either lethality or severe phenotypes, mutations in individual cis-regulatory elements of these genes may be tolerated and lead to milder phenotypes by affecting only a subset of the gene functions. Thus, sequence variation in regulatory elements can clearly serve as the raw material for morphological, physiological and behavioral modifications, but can they also lead to human disease?

MUTATIONS IN CONSERVED ELEMENTS AS A CAUSE FOR HUMAN DISEASE

In the past, genetic mapping of various diseases was largely facilitated by the cytogenetic characterization of chromosomal deletions and translocations in patients affected by those diseases. In several instances this characterization revealed that these chromosomal aberrations lie near the causative gene, but do not interrupt its coding region, and were termed position effects (Table 1). For example, mutations in the coding region of the gene encoding for the developmental transcription factor SALL1 lead to autosomal dominant Townes–Brocks syndrome, while a thoroughly characterized translocation in one patient 180 kb telomeric to SALL1 also leads to a similar phenotype (30). One likely explanation for these position effects is that non-coding cis-regulatory sequences have been removed from the vicinity of the gene/s that they normally regulate. The observation that many diseases arise from the disruption in the linearity between cis-regulatory sequences and the genes that they regulate raises the possibility that sequence variation within cis-regulatory sequences might also result in disease processes.

Nevertheless, proving that sequence changes in a non-coding element cause a particular phenotype is a more complex problem. An example illustrating how sequence variation in a human–fish conserved non-coding element might lead to disease phenotypes is the Sonic Hedgehog (SHH) limb enhancer (Fig. 3). Early studies of chromosomal rearrangements in patients with holoprosencephaly led to the discovery that mutations in the coding region of SHH cause this disorder (40). During limb development, SHH is expressed in the zone of polarizing activity (ZPA) and is required for the anterior–posterior patterning of the digits. A conserved human–fugu element that is thought to regulate SHH expression in the ZPA was recently characterized, 1 Mb away from the SHH gene (13,41). SNPs within this element were found to segregate exclusively in patients with preaxial polydactyly (PPD) (13), suggesting that these sequence variations are likely, but not proven, to be causing PPD. Further characterization and understanding of the function of these highly conserved non-coding elements will hopefully enable to alter this circumstantial evidence to a more explicit verdict.

SUMMARY

Exploiting pair-wise sequence comparisons between evolutionarily distant species and stringent sequence identity filters have proven to be a powerful means to identify functional non-coding sequences with significant impact on the organism. Nevertheless, these strategies have their limitations. A very prominent one is the relatively small number of conserved sequences and genes thought to be regulated by these elements. One strategy that could be used to overcome this predicament is the use of species that are closer than fish for comparisons with humans, such as chicken and marsupials. Multiple-species sequence comparisons can be used as an alternative comparative genomic analysis to help address this limitation, although more evidence is still needed that this filter also represents a valid strategy for prioritizing mammalian sequences whose functions are testable using the current experimental settings.

It is likely that comparative analyses alone will not suffice to correctly predict all functional sequences in the genome. Other large scale technologies to identify regulatory elements such as chromatin immunoprecipitation (ChIP) (reviewed in (42) and genome-wide analysis of DNase hypersensitive sites (43) are beginning to fill that gap. Bioinformatics tools also add layers of information to be considered together with sequence conservation, such as transcription factor binding site identification and clustering as well as generating more refined alignments (44–46). Meanwhile, the extreme sequence conservation over long evolutionary periods of the non-coding elements discussed in this review makes them primary candidates for genetic screens seeking sequence variation associated with morphological innovation and disease. Nature has already given us an invaluable clue to the importance of these exquisitely conserved sequences by keeping them untouched for hundreds of millions of years. It is about time we pay attention to them.

ACKNOWLEDGEMENTS

We thank Claire Weer and other members of the Rubin laboratory for suggestions and criticisms, Olivier Couronne for help with the figures and Josephine S. Porch for help with handling of the manuscript. This work was funded by the NIH-NHLBI Programs for Genomic Application Grant HL66681 (E.M.R.) through the U.S. Department of Energy under contract DE-AC03-76SF00098.

REFERENCES

1. Pennacchio, L.A., Olivier, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M. and Rubin, E.M. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science*, **294**, 169–173.
2. Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M. and Frazer, K.A. (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, **288**, 136–140.
3. Dermitzakis, E.T., Reymond, A., Scamuffa, N., Ucla, C., Kirkness, E., Rossier, C. and Antonarakis, S.E. (2003) Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science*, **302**, 1033–1035.
4. Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., Margulies, E.H., Blanchette, M., Siepel, A.C., Thomas, P.J., McDowell, J.C. *et al.* (2003) Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, **424**, 788–793.
5. Frazer, K.A., Tao, H., Osoegawa, K., de Jong, P.J., Chen, X., Doherty, M.F. and Cox, D.R. (2004) Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. *Genome Res.*, **14**, 367–372.
6. Margulies, E.H., Blanchette, M., Haussler, D., Green, E.D. and NISC Comparative Sequencing Program (2003) Identification and characterization of multi-species conserved sequences. *Genome Res.*, **13**, 2507–2518.
7. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.
8. Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. and Aparicio, S. (1993) Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. *Nature*, **366**, 265–268.
9. Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A. *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, **297**, 1301–1310.
10. Marshall, H., Studer, M., Popperl, H., Aparicio, S., Kuroiwa, A., Brenner, S. and Krumlauf, R. (1994) A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*. *Nature*, **370**, 567–571.
11. Nobrega, M.A., Ovcharenko, I., Afzal, V. and Rubin, E.M. (2003) Scanning human gene deserts for long-range enhancers. *Science*, **302**, 413.
12. Santagati, F., Abe, K., Schmidt, V., Schmitt-John, T., Suzuki, M., Yamamura, K. and Imai, K. (2003) Identification of cis-regulatory elements in the mouse Pax9/Nkx2-9 genomic region: implication for evolutionary conserved synteny. *Genetics*, **165**, 235–242.
13. Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E. and de Graaff, E. (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.*, **12**, 1725–1735.
14. Bagheri-Fam, S., Ferraz, C., Demaille, J., Scherer, G. and Pfeifer, D. (2001) Comparative genomics of the SOX9 region in human and *Fugu rubripes*: conservation of short regulatory sequence elements within large intergenic regions. *Genomics*, **78**, 73–82.
15. Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R. and Brenner, S. (1995) Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl Acad. Sci. USA*, **92**, 1684–1688.
16. Boffelli, D., Nobrega, M.A. and Rubin, E.M. (2004) Comparative genomics at the vertebrate extremes. *Nat. Rev. Genet.*, **5**, 456–465.
17. Shashikant, C.S., Kim, C.B., Borbely, M.A., Wang, W.C. and Ruddle, F.H. (1998) Comparative studies on mammalian Hoxc8 early enhancer sequence reveal a baleen whale-specific deletion of a cis-acting element. *Proc. Natl Acad. Sci. USA*, **95**, 15446–15451.

18. Anand, S., Wang, W.C., Powell, D.R., Bolanowski, S.A., Zhang, J., Ledje, C., Pawashe, A.B., Amemiya, C.T. and Shashikant, C.S. (2003) Divergence of Hoxc8 early enhancer parallels diverged axial morphologies between mammals and fishes. *Proc. Natl Acad. Sci. USA*, **100**, 15666–15669.
19. Fang, J., Dagenais, S.L., Erickson, R.P., Arlt, M.F., Glynn, M.W., Gorski, J.L., Seaver, L.H. and Glover, T.W. (2000) Mutations in FOXC2 (MFH-1), a forkhead family transcription factor, are responsible for the hereditary lymphedema-distichiasis syndrome. *Am. J. Hum. Genet.*, **67**, 1382–1388.
20. Crisponi, L., Uda, M., Deiana, M., Loi, A., Nagaraja, R., Chiappe, F., Schlessinger, D., Cao, A. and Pilia, G. (2004) FOXL2 inactivation by a translocation 171 kb away: analysis of 500 kb of chromosome 3 for candidate long-range regulatory sequences. *Genomics*, **83**, 757–764.
21. van Deutekom, J.C., Lemmers, R.J., Grewal, P.K., van Geel, M., Romberg, S., Dauwerse, H.G., Wright, T.J., Padberg, G.W., Hofker, M.H., Hewitt, J.E. *et al.* (1996) Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum. Mol. Genet.*, **5**, 581–590.
22. Vortkamp, A., Gessler, M. and Grzeschik, K.H. (1991) GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature*, **352**, 539–540.
23. Kioussis, D., Vanin, E., deLange, T., Flavell, R.A. and Grosveld, F.G. (1983) Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature*, **306**, 662–666.
24. Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) Gamma delta beta-thalassemia due to a de novo mutation deleting the 5′ beta-globin gene activation-region hypersensitive sites. *Proc. Natl Acad. Sci. USA*, **86**, 7470–7474.
25. Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M. *et al.* (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum. Mol. Genet.*, **4**, 415–522.
26. Flomen, R.H., Vatcheva, R., Gorman, P.A., Baptista, P.R., Groet, J., Barisic, I., Ligutic, I. and Nizetic, D. (1998) Construction and analysis of a sequence-ready map in 4q25: Rieger syndrome can be caused by haploinsufficiency of RIEG, but also by chromosome breaks approximately 90 kb upstream of this gene. *Genomics*, **47**, 409–413.
27. Inoue, K., Osaka, H., Thurston, V.C., Clarke, J.T., Yoneyama, A., Rosenbarker, L., Bird, T.D., Hodes, M.E., Shaffer, L.G. and Lupski, J.R. (2002) Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. *Am. J. Hum. Genet.*, **71**, 838–853.
28. de Kok, Y.J., Merkx, G.F., van der Maarel, S.M., Huber, I., Malcolm, S., Ropers, H.H. and Cremers, F.P. (1995) A duplication/paracentric inversion associated with familial X-linked deafness (DFN3) suggests the presence of a regulatory element more than 400 kb upstream of the *POU3F4* gene. *Hum. Mol. Genet.*, **4**, 2145–2150.
29. de Kok, Y.J., Vossenaar, E.R., Cremers, C.W., Dahl, N., Laporte, J., Hu, L.J., Lacombe, D., Fischel-Ghodsian, N., Friedman, R.A., Parnes, L.S. *et al.* (1996) Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Hum. Mol. Genet.*, **5**, 1229–1235.
30. Marlin, S., Blanchard, S., Slim, R., Lacombe, D., Denoyelle, F., Alessandri, J.L., Calzolari, E., Drouin-Garraud, V., Ferraz, F.G., Fourmaintraux, A. *et al.* (1999) Townes–Brocks syndrome: detection of a SALL1 mutation hot spot and evidence for a position effect in one patient. *Hum. Mutat.*, **14**, 377–386.
31. Roessler, E., Ward, D.E., Gaudenz, K., Belloni, E., Scherer, S.W., Donnai, D., Siegel-Bartelt, J., Tsui, L.C. and Muenke, M. (1997) Cytogenetic rearrangements involving the loss of the *Sonic Hedgehog* gene at 7q36 cause holoprosencephaly. *Hum. Genet.*, **100**, 172–181.
32. Wallis, D.E., Roessler, E., Hehr, U., Nanni, L., Wiltshire, T., Richieri-Costa, A., Gillessen-Kaesbach, G., Zackai, E.H., Rommens, J. and Muenke, M. (1999) Mutations in the homeodomain of the human *SIX3* gene cause holoprosencephaly. *Nat. Genet.*, **22**, 196–198.
33. Balemans, W., Patel, N., Ebeling, M., Van Hul, E., Wuyts, W., Lacza, C., Dioszegi, M., Dikkers, F.G., Hildering, P., Willems, P.J. *et al.* (2002) Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. *J. Med. Genet.*, **39**, 91–97.
34. Wirth, J., Wagner, T., Meyer, J., Pfeiffer, R.A., Tietze, H.U., Schempp, W. and Scherer, G. (1996) Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from SOX9. *Hum. Genet.*, **97**, 186–193.
35. McElreavey, K., Vilain, E., Abbas, N., Costa, J.M., Souleyreau, N., Kucheria, K., Boucekkine, C., Thibaud, E., Brauner, R., Flamant, F. *et al.* (1992) XY sex reversal associated with a deletion 5′ to the SRY ‘HMG box’ in the testis-determining region. *Proc. Natl Acad. Sci. USA*, **89**, 11016–11020.
36. McElreavey, K., Vilain, E., Barbaux, S., Fuqua, J.S., Fechner, P.Y., Souleyreau, N., Doco-Fenzy, M., Gabriel, R., Quereux, C., Fellous, M. *et al.* (1996) Loss of sequences 3′ to the testis-determining gene, SRY, including the Y pseudoautosomal boundary associated with partial testicular determination. *Proc. Natl Acad. Sci. USA*, **93**, 8590–8594.
37. Krebs, I., Weis, I., Hudler, M., Rommens, J.M., Roth, H., Scherer, S.W., Tsui, L.C., Fuchtbauer, E.M., Grzeschik, K.H., Tsuji, K. *et al.* (1997) Translocation breakpoint maps 5 kb 3′ from TWIST in a patient affected with Saethre–Chotzen syndrome. *Hum. Mol. Genet.*, **6**, 1079–1086.
38. Rose, C.S., Patel, P., Reardon, W., Malcolm, S. and Winter, R.M. (1997) The TWIST gene, although not disrupted in Saethre–Chotzen patients with apparently balanced translocations of 7p21, is mutated in familial and sporadic cases. *Hum. Mol. Genet.*, **6**, 1369–1373.
39. Juan, A.H. and Ruddle, F.H. (2003) Enhancer timing of Hox gene expression: deletion of the endogenous Hoxc8 early enhancer. *Development*, **130**, 4823–4834.
40. Belloni, E., Muenke, M., Roessler, E., Traverso, G., Siegel-Bartelt, J., Frumkin, A., Mitchell, H.F., Donis-Keller, H., Helms, C., Hing, A.V. *et al.* (1996) Identification of *Sonic hedgehog* as a candidate gene responsible for holoprosencephaly. *Nat. Genet.*, **14**, 353–356.
41. Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N. *et al.* (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl Acad. Sci. USA*, **99**, 7548–7553.
42. Buck, M.J. and Lieb, J.D. (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*, **83**, 349–360.
43. Crawford, G.E., Holt, I.E., Mullikin, J.C., Tai, D., Blakesley, R., Bouffard, G., Young, A., Masiello, C., Green, E.D., Wolfsberg, T.G. *et al.* (2004) Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc. Natl Acad. Sci. USA*, **101**, 992–997.
44. Ovcharenko, I., Boffelli, D. and Loots, G.G. (2004) eShadow: a tool for comparing closely related sequences. *Genome Res.*, **14**, 1191–1198.
45. Ovcharenko, I., Loots, G.G., Hardison, R.C., Miller, W. and Stubbs, L. (2004) zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res.*, **14**, 472–477.
46. Loots, G.G., Ovcharenko, I., Pachter, L., Dubchak, I. and Rubin, E.M. (2002) rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res.*, **12**, 832–839.