Software

Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes

Pär G Engström*,†‡, David Fredman*,† and Boris Lenhard*‡

Addresses: *Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, Thormøhlensgate, N-5008 Bergen, Norway. †Sars Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway. ‡Programme for Genomics and Bioinformatics, Department of Cell and Molecular Biology, Karolinska Institutet, S-17177 Stockholm, Sweden.

Correspondence: Boris Lenhard. Email: boris.lenhard@bccs.uib.no

© 2008 Engström et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Metazoan genomes contain arrays of highly conserved noncoding elements (HCNEs) that span developmental regulatory genes and define regulatory domains. We describe Ancora http://ancora.genereg.net, a web resource that provides data and tools for exploring genomic organization of HCNEs for multiple genomes. Ancora includes a genome browser that shows HCNE locations and features novel HCNE density plots as a powerful tool to discover developmental regulatory genes and distinguish their regulatory elements and domains.

Rationale

Comparisons of metazoan genome sequences have revealed an abundance of genomic segments that are highly conserved across large evolutionary distances even though they do not encode proteins and do not tend to be near transcription start sites. For example, 256 non-exonic segments longer than 200 bp were found to be perfectly conserved between human, mouse and rat genomes; 140 of these were more than 10 kb away from any known gene [1]. Using less stringent criteria for length and sequence similarity, other investigators have found thousands of non-exonic segments in the human genome that are conserved in organisms as distant as fugu [2,3] and shark [4].

Several lines of evidence indicate that these highly conserved noncoding elements (HCNEs) play a fundamental role in regulating animal development and constraining genome evolution. In vertebrates, insects and worms, HCNEs tend to cluster in the vicinity of developmental regulatory genes [1-7]. Through experiments in transgenic animals in which cloned HCNEs are tested for the ability to drive transcription of a reporter gene, many HCNE sequences have shown the ability to induce part of the embryonic expression pattern of a developmental regulatory gene located in the genomic neighborhood of the endogenous HCNE [3,8-11]. These experiments have associated HCNEs and developmental genes separated by considerable genomic distances, up to 800 kb in human [8], suggesting that many HCNEs act as long-range regulatory elements. Hundreds of HCNEs have now been characterized as developmental enhancers in transgenic mice, frogs or zebrafish, and the list is growing rapidly [10,12-14].

The emerging model for explaining these observations is that an array of HCNEs defines a region of regulatory inputs of its target gene(s), and that the full complement of those inputs results in the expression pattern of the gene [3,8-11]. If this notion that HCNE arrays constitute regulatory domains is correct, chromosomal rearrangements within HCNE arrays...
should be selected against in evolution [15-17]. Accordingly, large HCNE arrays have been found to correspond to the largest and most deeply conserved blocks of synteny across vertebrates [18] and across insects [6]. In addition to HCNE arrays and their target genes, many of these synteny blocks contain unrelated (bystander) genes that do not appear to be regulated by the HCNEs, although they can be situated between HCNEs and target genes, as well as contain HCNEs in their introns. Kikuta et al. [18] termed these synteny blocks ‘genomic regulatory blocks’ (GRBs) and demonstrated that, for some GRBs, it is possible to distinguish bystander from target genes by comparing mammalian genome sequences with those of teleost fish (such as fugu and zebrafish). This is facilitated by a whole-genome duplication event that occurred in the teleost lineage [19] and caused each GRB to be present in two copies, thereby allowing some bystander genes to be disentangled from HCNE arrays during the subsequent rediploidization [18].

Despite a rising interest in HCNEs in the genomics and evolutionary community, there has been a lack of resources that provide information about HCNEs and allow researchers to explore the distribution of HCNEs along chromosomes. Here, we describe Ancora [20], a web resource consisting of: a genome browser where HCNE locations and HCNE density plots can be viewed over different genomes, with a number of adjustable parameters; data files that allow users to easily view HCNE locations and densities in the UCSC Genome Browser [21]; and a service that allows users to view HCNE data in the Ensembl browser [22] through the distributed annotation system (DAS) protocol for sharing sequence annotations [23]. We demonstrate how Ancora can be used to discover developmental regulatory genes and distinguish their chromosomal regulatory domains that correspond to the GRBs described above. The visualization of these regulatory domains is the most powerful and novel function of Ancora. We anticipate that Ancora will be particularly useful for assigning distal regulatory elements to their target genes, and for the discovery of hitherto unknown developmental regulatory genes, including noncoding RNAs.

A comprehensive HCNE database
Ancora rests on a database of HCNEs conserved between various metazoan genomes (Figure 1). Building on our previously described strategies for detecting HCNEs [2,6,18] we have created a refined procedure that is not biased against a chosen base genome and better captures HCNEs duplicated in genome evolution. We identify HCNEs by scanning pairwise BLASTZ net whole-genome alignments (nets) [24] downloaded from the UCSC Genome Browser database [21] for regions with at least I identities over C alignment columns. Because different similarity criteria may be appropriate for different loci and investigations, we scan for conserved elements using at least two different window sizes (C = 30 and C = 50) and several different similarity thresholds (I/C) in the range 70–100% for each species pair. The algorithm that creates net alignments is designed to retain only the best alignment for each position in one of the genomes [24]. For each pairwise comparison, we therefore scan two sets of nets (one from the perspective of each genome) in order not to miss elements duplicated in either lineage. This is particularly important for comparisons between teleost fish and other vertebrates, because of the whole-genome duplication that occurred in the teleost lineage [19]. We subsequently merge highly conserved elements that overlap on both genomes, but not elements that coincide on only one of the genomes, so that duplicated elements remain distinct. After discarding elements whose genome coordinates overlap by one or more base-pairs with annotated exons, we remove repetitive sequences by considering overlap with known repeats and the number of high-identity alignments obtained by realignment of each sequence against the two respective genomes. We consider remaining elements as HCNEs. The exon and repeat annotations we use, and the realignment parameters we employ, are listed on the Ancora web site, where an up-to-date description of our HCNE detection procedure is maintained. To illustrate the effect of parameter changes on the number of HCNEs detected, Table 1 lists HCNE counts for some selected settings and genomes.

Exploring HCNEs and GRBs with the Ancora genome browser
Ancora contains a genome browser designed to explore the distribution of HCNEs on metazoan chromosomes (Figure 2a). The browser is currently set up to show the genomes of human, mouse, zebrafish and Drosophila melanogaster; we aim to expand this list in the future.

Basic usage
To put HCNEs in context, the browser also shows gene annotation from NCBI [25], Ensembl [22], the UCSC Genome Browser [21], Mouse Genome Informatics [26], the Zebrafish Information Network [27], FlyBase [28] and miRBase [29], as well as a selection of other annotation tracks from UCSC. The user can click on gene models to bring up detailed gene information pages from the original data sources. By default, the HCNEs are colored by the chromosome they align to in the other genome. This simplifies the identification of conserved HCNE arrays: a stretch of HCNEs in the same color suggests a conserved array. To visualize the tendency of HCNE arrays to correspond to large synteny blocks, we have included tracks showing human-zebrafish synteny blocks and Drosophila synteny blocks from recent analyses [6,18]. (The human-zebrafish synteny blocks should be interpreted with caution, however, because of artifacts in the underlying zebrafish genome assembly - in particular artificial segmental duplications, which may appear as overlapping synteny blocks on the human genome.) The user can move between the vertebrate genomes that the genome browser displays by clicking on HCNEs and synteny blocks, which link aligned
regions from the different genomes. Ancora also provides links that bring up the same region in other major genome browsers (Ensembl, UCSC and FlyBase) and the VISTA browser, which is useful for detailed examination of sequence conservation [30].

**GBrowse extensions in Ancora**
The Ancora genome browser was built using the GBrowse software [31], which is used by most model organism databases. The basic user interface should thus be familiar to most users. To visualize HCNE data in the most informative manner and to efficiently plot HCNE densities along entire chromosomes, we have extended GBrowse with a number of plugins and custom glyphs. The plugins that retrieve and render HCNE data can be configured by selecting a HCNE data set of interest from the 'Reports & Analysis' menu above the 'Scroll/Zoom' controls or by clicking on a HCNE density plot (Figure 2a). On the configuration page (Figure 3), the user can select which similarity thresholds to show HCNEs and HCNE densities for, and configure additional properties of the density plots (see below).

**Unique information revealed by HCNE density plots**
Plots of HCNE density along chromosomes highlight regions that harbor large HCNE arrays and, thus, are likely to contain key developmental regulatory genes and correspond to regulatory domains [2,6,18,32,33]. Unlike conservation profiles, which can be seen in several other genome browsers [21,22,30,34], HCNE density plots do not directly reflect conservation on the sequence/alignment level; instead, they show density distributions of HCNEs on a larger scale. The result is qualitatively different from a sequence based conservation plot such as the Conservation track in the UCSC browser.
Figure 2
A 1.7 Mb region around the human SHOX2 gene. (a) Ancora genome browser main view. SHOX2, a homeobox gene implicated in limb development [47], is embedded in an array of HCNEs detected by comparison with mouse and zebrafish genomes. Overlaid density plots show densities of HCNEs detected at similarity thresholds of 95% (yellow), 98% (orange) and 100% (red) in the mouse comparison and similarity thresholds of 70%, 80% and 90% in the zebrafish comparison, over a 50 column sliding window. Note that the density of the most strongly conserved HCNEs (red) peaks around SHOX2. Synteny blocks are based on net alignments with the zebrafish genome [18]; boxes indicate aligned segments, connecting lines indicate gaps and labels indicate alignment orientation and position in the zebrafish genome assembly. The centrally shown synteny block encompasses SHOX2, RSRC1 (a gene of unknown function) and the array of HCNEs conserved in zebrafish. (b) Conservation profiles for the same region in the UCSC Genome Browser [21]. Comparison between (a) and (b) demonstrates qualitatively different information provided by the HCNE density plots in (a).
At lower resolutions, the step size is at least one step per pixel base resolution, densities are computed for every base shown. chromosomal segment in steps of a size that is adapted to the will be combined into one plot labeled 'other' (Figure 5).

If the resulting number of (in the other genome) for which there are HCNEs in the displayed region, the browser will attempt to create one density plot for each chromosome to separate densities based on chromosomes in other genomes is used, and the plot for set three drawn on top of that for set two. If the option (Figure 2a), so that the plot for set two is drawn on top of that for set one, locations, or both. Density plots for the different sets will be overlaid.

For each selected set, the user can choose to see HCNE densities, HCNE track configuration page. Up to three HCNE sets from each pairwise comparison can be shown simultaneously. A set is selected by choosing a similarity threshold (for example, 70% identity over 50 alignment columns), and can be further restricted by an arbitrary threshold on HCNE size. Note that HCNEs may be larger than the window size (30 or 50 columns) used to identify them because the procedure that detects HCNEs merges overlapping conserved elements. For each selected set, the user can choose to see HCNE densities, HCNE locations, or both. Density plots for the different sets will be overlaid (Figure 2a), so that the plot for set two is drawn on top of that for set one, and the plot for set three drawn on top of that for set two. If the option to separate densities based on chromosomes in other genomes is used, the browser will attempt to create one density plot for each chromosome (in the other genome) for which there are HCNEs in the displayed region, or within half a sliding window to either side. If the resulting number of plots exceeds the number of plots requested on this configuration page, densities for the chromosomes with least HCNE sequence in this region will be combined into one plot labeled ‘other’ (Figure 5).

Discovering genes that encode developmental regulators
Since there is a strong association between HCNE arrays and developmental regulatory genes [1–7], it is likely that most regions of high HCNE-density contain at least one developmental regulatory gene, even in cases where no such gene has been annotated. Inspection of HCNE density can thus be used to formulate hypotheses about gene function and identify likely target genes of putative enhancer activity of HCNEs. In a study from 2004, Sandelin et al. [2] identified HCNEs conserved among human, mouse and fugu, and closely inspected the 50 most HCNE-rich regions for the presence of developmental regulatory genes. They found 41 of these regions to contain a gene known to be involved in embryonic development. Of the remaining nine regions, seven contained a gene known to be a transcription factor or predicted as such based on homology. In a recent study, one of these transcription factor genes (FLJ20321) was recognized as a homolog of the Drosophila gene castor and found to be upregulated in cell differentiation [35], confirming the prediction from HCNE density. Sandelin et al. focused on the 50 HCNE-densest regions they detected in the human genome. Inspection of other HCNE-dense regions has revealed that several coincide with microRNA gene loci [18], a class of regulators implicated in multiple aspects of development [36]. We predict that many additional HCNE-dense regions will be found to contain developmental regulators. By plotting HCNE densities along entire chromosomes, the Ancora genome browser...
makes it easy to survey genomes for HCNE-dense regions (Figure 4). HCNE density curves from multiple pairwise genome comparisons can be shown simultaneously, so that users can identify regions rich in HCNEs that are specific to a subset of species, or shared across many species, if so desired. By zooming in, the user can investigate these regions in detail by inspecting the genome annotation available in Ancora as well as annotation in the other genome browsers to which direct links are provided. As a demonstration of the immediate utility of Ancora, we identified 129 genomic regions in the human genome in which the density of human-zebrafish HCNEs (70% identity over 50 columns) surpassed 0.5% and, using the principles outlined here, identified putative target genes in 120 of these regions (Additional data file 1). The regions in which no target gene could be assigned are prime candidates for discovery of novel genes or non-coding RNA involved in developmental regulation.

**Detecting and interpreting duplicated GRBs**

As a result of whole-genome duplication in teleosts, many mammalian GRBs have two orthologous GRBs in teleost genomes. The Ancora genome browser makes it easy to locate such GRBs by coloring HCNEs according to the chromosome they align to in the other genome. For example, when viewing human-zebrafish HCNEs along human chromosomes, the hallmark of a GRB present in two copies in zebrafish is a HCNE-dense region where HCNEs occur mainly in two different colors. Such regions can also be discovered by activating an option that makes the genome browser separate HCNE density plots based on chromosomes in the other genome (Figure 3). Figure 5a shows an example: the GRB of PAX7, a transcription factor gene implicated in muscle development [37] and situated within an array of HCNEs. Most of the human-zebrafish HCNEs in this region are colored either gray or light green in the genome browser (Figure 5a) and align to orthologous loci on zebrafish chromosomes 23 and 11, respectively. Thus, this view quickly suggests that noncoding putative regulatory sequences have been preserved to a similar extent at both of the pax7 loci in zebrafish. In contrast, Figure 5b shows an example where duplicate GRBs have diverged to a large extent in zebrafish. Human LHX1, a LIM homeobox transcription factor gene implicated in head, neural and reproductive development [38] is within an array of HCNEs that extends into the neighboring genes AATF, which encodes a transcription factor involved in cell cycle control, and ACACA, which encodes a carboxylase involved in fatty acid synthesis. Most of the human-zebrafish HCNEs in this region are colored blue in the genome browser (Figure 5b).
and align to the region around \textit{lhx1a} on zebrafish chromosome 15. Thus, noncoding putative regulatory sequences appear to have been preserved to a much larger extent around \textit{lhx1a} than around \textit{lhx1b}. A detailed inspection of the zebrafish loci reveals that orthologs of \textit{AATF} and \textit{ACACA} have been retained at the \textit{lhx1b} locus, but lost from the \textit{lhx1a} locus, where there are more HCNEs (in Ancora, the zebrafish loci can be inspected by jumping from the displayed human locus to corresponding loci in the zebrafish genome by clicking on HCNEs or synteny blocks). Following the rationale in [18] this confirms that the HCNE array is unrelated to \textit{AATF} and \textit{ACACA}, and allows the classification of these two genes as bystanders.

### Distinguishing chromosomal regulatory domains

By comparing HCNE arrays and synteny blocks, we have observed that the extent of a HCNE array often provides a good approximation of the extent of the corresponding GRB [6,18]. However, unless synteny conservation is taken into account, partitioning of HCNEs into separate arrays becomes arbitrary in regions with high noncoding conservation. In the Ancora genome browser, it is easy to visualize synteny conservation of HCNE arrays over large genomic segments by activating the option that separates HCNE density plots based on chromosome in the other genome. The result is an overview of how HCNE-dense regions have been partitioned over different chromosomes in evolution (Figure 6). Based on the assumption that fundamental regulatory domains have been maintained in evolution [6,15-18], the displayed separation of HCNE-dense regions across chromosomes should correspond to a separation of distinct regulatory domains. We expect the resolution of this approach to increase as more genomes are sequenced and assembled.

### Viewing HCNEs and density plots in other genome browsers

The Ancora genome browser provides the most flexible way to explore the HCNE data in Ancora. However, it is often useful to view these data in other browsers where it can be compared to other types of annotation. We aimed to make it as straightforward as possible to view HCNE data in the UCSC Genome Browser [21] and Ensembl [22] (Figure 7).

HCNE locations and precomputed density curves are available for download in the ‘bed’ and ‘wig’ formats used for UCSC Genome Browser custom tracks [39]. It is not necessary to download the .bed and .wig files to use them as custom tracks in the UCSC Genome Browser: the user can simply copy the URLs for track files of interest from the Ancora downloads section and paste them into the ‘add custom tracks’ form on the UCSC Genome Browser web site.

The Ensembl browser can display sequence annotations provided over the web through DAS, a method for data exchange [23]. Much of the Ancora data are available through DAS.

Ancora provides an interface where the user can add HCNE tracks to Ensembl ContigView. Tracks added in this way are stored as part of the user’s Ensembl preferences. Users who are familiar with DAS can also retrieve data directly from the DAS server. For example, the URL given in reference [40] provides a list of available tracks.

### Comparison to other tools

While the genome browsers at UCSC and Ensembl provide rich and diverse annotation sets including information about sequence conservation, they do not distinguish coding from noncoding conserved elements. To our knowledge, the Ancora genome browser is the first tool that makes it easy to visualize HCNE distributions on large genomic regions, up to whole chromosomes, and the browser is tailored to show data in a flexible manner at this level.

The ECR Browser [34] and VISTA Browser [30] allow detailed inspection of sequence conservation profiles across many genomes, highlight conserved elements in a user-customizable manner and distinguish noncoding from coding conservation. In the ECR Browser, one drawback is that thresholds for detection of conserved elements are uniform across all comparisons shown, irrespective of evolutionary distance. In contrast, Ancora and VISTA browsers can show results for multiple different thresholds simultaneously. A limitation of both the ECR and VISTA browsers is that they are not designed for visualizing the distribution of conserved elements on segments larger than a few megabases. The VISTA Browser can only display regions up to 5 Mb in size and the ECR Browser’s display of large regions is difficult to interpret because conserved elements are drawn close together. In contrast, the HCNE density plots in Ancora make it possible to view and intuitively interpret HCNE content at any scale. Ancora is therefore better suited for exploring conservation genome-wide and discovering regulatory domains at loci not known beforehand, while the ECR and VISTA browsers provide more functionality for close examination of sequence-level conservation profiles.

The CONDOR database [14] holds information on about 6,800 HCNEs from about 120 blocks of conserved synteny between human and fugu and provides a graphical interface to view the distribution of HCNEs in those regions. While there are several similarities between Ancora and CONDOR, Ancora has the advantage of providing HCNE data for entire genomes. Another difference between the two resources is that the Ancora HCNE sets are not as stringently defined in terms of conservation as those in CONDOR, where HCNEs are required to be conserved among four diverged vertebrates. In Ancora, we have chosen to provide a range of HCNE data sets from different pairwise comparisons and with different similarity thresholds (Figure 1 and Table 1), so that users can choose to look at the data appropriate for their questions. A valuable section of CONDOR provides
Duplicated GRBs. Zebrafish HCNEs and their density distribution are shown for the human (a) PAX7 and (b) LHX1 loci. HCNEs are colored by the zebrafish chromosome they map to. In (a), most HCNEs are colored light green or gray and map to zebrafish chromosomes 11 or 23, respectively. In (b), most HCNEs are colored blue and map to zebrafish chromosome 15 (because this region contains many HCNEs, they are collapsed on a single row in this screenshot). Comparison of synteny blocks to exon locations indicate that orthologs of AATF and ACACA are present next to the LHX1 ortholog (lhx1b) on zebrafish chromosome 5, but not on zebrafish chromosome 15 where lhx1a is located; this can be confirmed by detailed inspection of the zebrafish loci.

Figure 5
Duplicated GRBs. Zebrafish HCNEs and their density distribution are shown for the human (a) PAX7 and (b) LHX1 loci. HCNEs are colored by the zebrafish chromosome they map to. In (a), most HCNEs are colored light green or gray and map to zebrafish chromosomes 11 or 23, respectively. In (b), most HCNEs are colored blue and map to zebrafish chromosome 15 (because this region contains many HCNEs, they are collapsed on a single row in this screenshot). Comparison of synteny blocks to exon locations indicate that orthologs of AATF and ACACA are present next to the LHX1 ortholog (lhx1b) on zebrafish chromosome 5, but not on zebrafish chromosome 15 where lhx1a is located; this can be confirmed by detailed inspection of the zebrafish loci.
Distinguishing regulatory domains. Screenshot from the Ancora genome browser showing 26 Mb of human chromosome 10 and HCNE densities from comparisons with mouse, zebrafish and medaka. HCNE density curves are separated based on chromosomes in these organisms (Zv7_NA53 is a contig that has not been assigned to a chromosome). To illustrate the use of this view for distinguishing chromosomal regulatory domains, rectangles have been manually added to the screenshot around density peaks indicating clusters of HCNEs in conserved synteny. Rectangles are labeled with regulatory genes annotated in the corresponding genomic regions.
developmental expression patterns for about 100 HCNEs that have been investigated by reporter assays in zebrafish. We are preparing to link similar data to Ancora.

**Summary**

Ancora is a new web resource that provides data and tools for exploring HCNEs and their association with developmental
regulatory genes. Built upon a database of HCNEs conserved between various metazoan genomes, Ancora provides a genome browser for visualizing the distribution of those elements on chromosomes in the context of other types of annotation integrated from different sources. One of the novel features of Ancora is the possibility to display highly customizable plots of HCNE density along chromosomes. HCNE density plots are qualitatively different from conservation profiles available in other genome browsers [21,22,30,34]: they clearly reveal regions of extensive noncoding conservation and highlight larger chromosomal regulatory domains (GRBs) that have been maintained in evolution. The GRBs typically coincide with loci of developmental regulatory genes, for which HCNEs appear to act as enhancers [3,8-12]. Consequently, we anticipate that Ancora will be highly useful for developing overlooked regulatory genes and their distal cis-regulatory elements. We have illustrated how Ancora can be used to define the chromosomal regulatory domains of those genes and distinguish genes that appear to be functionally associated with HCNEs from unrelated ‘bystander’ genes within the same GRB. The HCNE data in Ancora are also available for download and can easily be displayed in the popular general-purpose genome browsers at UCSC [21] and Ensembl [22].

Abbreviations
DAS, distributed annotation system; GRB, genomic regulatory block; HCNE, highly conserved noncoding element.

Authors’ contributions
BL conceived of and supervised the study. PGE and DF designed and implemented the pipeline for detecting HCNEs. PGE implemented the web resource. All authors participated in writing the paper and approved the final version.

Additional data files
The following additional data are available with the online version of this paper. Human genomic regions in which the density (in a 300 kb sliding window) of human-zebrafish conserved elements act in vivo to regulate transcription. Nucleic Acids Res 2005, 33:5437-5445.

Visel A, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Rubin EM: VISTA Enhancer Browser - a database of tissue-specific human enhancers. Nucleic Acids Res 2007, 35:D88-92.

Woolfe A, Goode D, Cooke J, Callaway H, Smith S, Snell P, McEwen GK, Elgar G: CONDOR: a database resource of developmentally associated conserved non-coding elements. BMC Dev Biol 2007, 7:100.

Ahituv N, Prabhakar S, Poulin F, Rubin EM, Couronne O: Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. Hum Mol Genet 2005, 14:3057-3063.

Kleinjan DA, van Heyningen V: Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 2005, 76:8-32.

Becker TS, Lenhard B: The random versus fragile balance models of chromosome evolution: a matter of resolution. Mol Genet Genomics 2007, 278:487-491.

Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, Fredman D, Akalin A, Caccamo M, Seal Y, Howe K, Ghislain J, Pezeron G, Mourrain P, Ellingsen S, Oates AC, Thibes C, Choquet S, Foucher I, Adolf B, Geling A, Lenhard B, Becker TS: Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res 2007, 17:545-555.

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Boucher L, Fischer L, Ozouf-Coraz C, Berton A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dalsina C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthovery V, Vinbu C, Castelli V, Katinka M, Vacherie B, Bleton C, Skalli Z, Cattolico L, Poujol J, et al.: Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 2004, 431:946-957.

Ancora [http://ancora.genereg.net]
21. Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, Siepel A, Rosenbloom KR, Rhein B, Raney BJ, Pohl A, Peterson JS, Hsu F, Hinrichs AS, Harte RA, Diekhans M, Clawson H, Bejerano G, Barber GP, Baetsch R, Haussler D, Kent WJ, et al. The UCSC genome browser database: update 2007. Nucleic Acids Res 2007, 35:D688-673.

22. Heinbrand T, Aiken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Costes G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Houck K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, et al. Ensembl 2007. Nucleic Acids Res 2007, 35:D651-617.

23. Dowell RD, Jokerst RM, Day A, Eddy SR, Stein L. The distributed annotation system. BMC Bioinformatics 2001, 2:7.

24. Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: duplication, deletion, and rearrangement in the structure of the human genome. Proc Natl Acad Sci USA 2003, 100:11484-11489.

25. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeria E, Sherry ST, Sirotkin AV, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2007, 35:D35-12.

26. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE. The mouse genome database (MGD): new features facilitating a model system. Nucleic Acids Res 2007, 35:D630-637.

27. Sprague J, Bayraktaroglu L, Bradford Y, Conlin T, Dunn N, Fashena D, Frazer K, Haendel M, Howe DG, Knight J, Mani P, Moxon SA, Pich C, Ramachandran S, Schaper K, Segersdottir E, Shao X, Singer A, Song P, Sprunger B, Van Slyke CE, Wheeler DA. The zebrafish information network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. Nucleic Acids Res 2007, 35:D768-772.

28. Crosby MA, Goodman JL, Strelets VB, Zhang P, Gelbart WM. FlyBase: genomes by the dozen. Nucleic Acids Res 2007, 35:D486-491.

29. Griffiths-Jones S, Grocock RJ, von Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006, 34:D140-144.

30. Brudno M, Poliakov A, Minovitsky S, Arora M, Flicek P, Nataleschi DAS, Washietl S, Frazer K, Ovcharenko I, Maglott DR, Ostell J, Miller V, Pruitt KD, Bhardwaj L, Haussler D, Batalov S, Weinstock GM, Lamont TB, Lanfear D, Dessimoz C, Baertsch R, Gerstein M, Birney E, et al. The human protein reference database—2005 update. Nucleic Acids Res 2005, 33:242-244.

31. Stein LD, Mungall CJ, Shu S, Caudy M, Sawaya P, Mungall K, Mungall C, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S. The generic genome browser: a building block for a model organism system database. Genome Res 2002, 12:1599-1610.

32. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Lekstrom-Himes J, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostlander EA, Ponting CP, Gilbert R, Smith DR, Dejong P, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Groner S, et al. Genome sequence, comparative analysis and haplotype structure of the Fugu rubripes. Nature 2005, 432:803-819.

33. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SM, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidson L, Deskin JE, Fazakerley DJ, Glass J, Gu W, et al. Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature 2007, 447:167-177.

34. Ovcharenko I, Nobrega MA, Loos G, Stubbs L. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res 2004, 32:W280-286.

35. Liu Z, Yang X, Tan F, Cullion K, Thiele CJ. Molecular cloning and characterization of human Castor, a novel human gene underregulated during cell differentiation. Biochem Biophys Res Commun 2006, 344:834-844.

36. Zhao Y, Srivastava D. A developmental view of microRNA function. Trends Biochem Sci 2007, 32:189-197.

37. Buckingham M. Skeletal muscle progenitor cells and the role of Pax genes. C R Biol 2007, 330:530-533.

38. Hunter CS, Rhodes SJ. LIM-homeodomain genes in mammalian development and human disease. Mol Biol Rep 2005, 32:67-77.

39. UCSC Genome Bioinformatics - FAQ: Data File Formats [http://genome.ucsc.edu/FAQ/FAQformat].

40. Ancora DAS Tracks [http://ancora.genereg.net/das/dsn].

41. International Chicken Genome Sequencing Consortium: Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 2004, 432:695-716.

42. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehul P, Christofers A, Rash T, Hoos N, Smith A, Gelpke MD, Roach J, Oh T, Ho Y, Wong M, Detter C, Verhoeof F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, et al. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 2002, 297:1301-1310.

43. Kasahara M, Naronu K, Sasaki S, Nakatani Y, Qu VW, Ahsan B, Yamada M, Nakayasu Y, Doi K, Kasa C, jingi T, Kobayashi D, Shimida A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta M, et al. The medaka draft genome and insights into vertebrate evolution. Nature 2007, 477:714-719.

44. Celniker SE, Wheeler DA, Kronmiller B, Carlson JW, Halpern A, Patel S, Adams M, Champe M, Dugan SP, Friese E, Hodgson A, George RA, Hoskins RA, Laverty T, Muzy DN, Nelson CR, Pacibé JM, Park S, Pfeiffer BD, Richards S, Sodergren E, Sivarskas R, Tobor PE, Wan K, Stapleton M, Sutton GG, Venter C, Winstock G, Scherer SE, et al. Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster genome sequence. Genome Biol 2002, 3:research0079.

45. Richards S, Liu Y, Betancourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hibuzis M, Chen R, Meisel RP, Couronne O, Hua S, Smith MA, Zhang P, Liu J, Bussemaker HJ, van Batenburg MF, Howells SL, Scherer SE, Sodergren E, Matthews BB, Crosby MA, Schroeder AJ, Ortiz-Barrientos D, Rives CM, Metzker ML, Muzy DN, Scott G, Stef- fen D, Wheeler DA, et al. Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. Genome Res 2005, 15:1-18.

46. Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart WM, Iyer VN, Pollard DA, Sacconk TB, Larraquen AM, Singh ND, Abad JP, Abr D, Adryan B, Aguade M, Askali H, Anderson WW, Acedo CF, Ardel D, Arguello R, Antieri K, Barbash DA, Barker D, Barsanti P, Batterham P, Babzoglu S, et al. Evolution of genes and genomes on the Drosophila phylogeny. Nature 2007, 450:203-218.

47. Cobb J, Dierich A, Huss-Garcia Y, Duboule D. A mouse model for human short stature syndromes identifies Shox2 as an upstream regulator of Runx2 during long-bone development. Proc Natl Acad Sci USA 2006, 103:4511-4515.